UNIVERSIDAD PERUANA DE CIENCIAS APLICADAS

ESCUELA DE POSTGRADO

**PROGRAMA DE MAESTRÍA EN DIRECCIÓN DE SISTEMAS Y
TECNOLOGÍAS DE LA INFORMACIÓN**

**Propuesta de analytics a los patrones de comportamiento en el proceso de
clasificación socioeconómica en el MIDIS**

TRABAJO DE INVESTIGACIÓN

**Para optar el grado académico de Maestro en Dirección de Sistemas y Tecnologías de la
Información**

AUTORES

Atalaya Tello, Karina (0000-0002-9131-6638)

Flores Aedo, Nancy (0000-0002-2158-0839)

Flores Alvarado, Ángela (0000-0002-9164-3612)

ASESOR

Medina La Plata Edison Humberto (0000-0001-5305-6516)

Lima, 21 de Enero de 2019

DEDICATORIA

A mi querida madre Elvira, por el apoyo incondicional en cada etapa de mi vida, por ser un ejemplo de fortaleza ante la adversidad y mantener ese espíritu de apoyo hacia los que la rodean.

A mi colegio y a mis padres y hermanas, por los valores que me inculcaron y que fueron el pilar para realizar este trabajo, buscando contribuir con el bienestar de los demás.

A mi hijo Cristhian, el motor que me impulsa a continuar y ser mejor cada día.

Karina Atalaya

Quiero agradecer primeramente a Dios Jehová, mis padres y a mis hermanos y en especial a Walter y Maribel, por ser las personas que me guiaron siempre por el bien, haciendo de mí una persona perseverante y con valores.

Nancy Flores

A mi esposo José, a mi hija Vanessa y a mi hermana Ana, por su paciencia, comprensión, aliento y soporte que me permitió culminar este gran reto.

Angela María Flores Alvarado

AGRADECIMIENTOS

A Dios, quien ha guiado siempre mis pasos y me ha sostenido firmemente en cada momento de mi vida. A mis compañeras, por el esfuerzo realizado para alcanzar nuestra meta y a los profesores con quienes compartimos horas de estudio y esfuerzo que nos trajeron hasta este momento.

Karina Atalaya

A Dios porque me dio la vida. A mis padres porque permitieron que venga a este mundo maravilloso y desde el cielo guían mi camino. A mis hermanos por su apoyo emocional y en especial a Walter y Maribel por su apoyo incondicional a lo largo de mis estudios. A mis profesores por haberme transmitido sus conocimientos.

Nancy Flores

En primer lugar, a Dios, por haberme dado la oportunidad de desarrollar mi capacidad y conseguir mis metas. A mi familia, por haberme brindado su soporte y alentarme a lo largo de la duración de la maestría para alcanzar mis metas. A mis compañeros y profesores de la UPC, especialmente a nuestro profesor y asesor Edison Medina por sus aportes para mejorar y culminar nuestro trabajo.

Angela María Flores Alvarado

RESUMEN

El proceso de clasificación socioeconómica, realizado por el Ministerio de Desarrollo e Inclusión Social (MIDIS), permite asignar el nivel socioeconómico a los hogares y sus integrantes, principalmente, a través de mecanismos de evaluación del hogar en base a sus características socioeconómicas, demográficas y de condición de vida. Esta evaluación permite identificar a hogares y personas en situación de pobreza, vulnerabilidad o exclusión, como potenciales beneficiarios de los programas y subsidios del Estado.

La aplicación de Analytics y de la minería de datos, a través de técnicas como la clusterización, para la búsqueda de patrones mediante la identificación de agrupaciones con características similares, y la clasificación, para catalogar nuevos registros, permitió determinar la clasificación socioeconómica de los hogares y de sus integrantes; así como establecer la clasificación socioeconómica de nuevos registros de hogares. Para la aplicación de estas técnicas de minería de datos se siguieron las fases establecidas en la metodología CRISP-DM. Se recopiló información desde los formatos de recojo de información de las características de los hogares, procediéndose a considerar las variables más relevantes para la construcción del modelo o patrón

El presente trabajo tiene por finalidad hallar un patrón de comportamiento para el proceso de clasificación socioeconómica de la población, a través de la aplicación de Analytics, buscando contribuir con la reducción del porcentaje de filtración de hogares que reciben apoyo de los programas sociales sin corresponderles, para así a tener una política de apoyo social efectiva.

Palabras clave: Analytics; patrón; minería de datos; clasificación socioeconómica.

Proposal of Analytics to behavior patterns in the socioeconomic classification process in
the MIDIS

ABSTRACT

The socioeconomic classification process made by Ministry of development and social inclusion (MIDIS), allows socioeconomic level assignment to households and their members, mainly through household assessment mechanisms base on their socioeconomic, demographic and social characteristics. This evaluation allows it possible to identify households and people in situations of poverty, vulnerability or exclusion, as potential beneficiaries of State programs and subsidies.

The application of analytics and data mining, through techniques such as clustering, for the search of patterns by identifying clusters with similar characteristics, and classification, to catalog new records, allowed determining the socioeconomic classification of households; as well as establishing the socioeconomic classification of new household registers. For the application of these data mining techniques, the CRISP-DM methodology was used. Used Information was collected from the information collection formats of the characteristics of the households and the most relevant variables were considered for the model construction.

The purpose of this paper is to find a pattern of behavior for the socioeconomic classification process, through the application of Analytics, seeking to contribute to the reduction of the percentage of households that receive support from social programs without corresponding them, in order to have an effective social support policy.

Keywords: Analytics; pattern; Data mining; Socioeconomic Cluster

TABLA DE CONTENIDOS

1	INTRODUCCIÓN.....	15
2	MARCO TEÓRICO.....	16
2.1	LA POBREZA	16
2.1.1	MÉTODOS DE MEDICIÓN DE LA POBREZA	17
2.2	SISTEMA DE FOCALIZACIÓN	17
2.2.1	FOCALIZACIÓN.....	17
2.2.2	MECANISMOS DE FOCALIZACIÓN	18
2.2.3	FOCALIZACIÓN EN AMÉRICA LATINA	18
2.2.4	ERRORES EN LA FOCALIZACIÓN	19
2.2.4.1	ERRORES DE FOCALIZACIÓN EN EL DISEÑO	20
2.2.4.2	ERRORES DE FOCALIZACIÓN EN LA IMPLEMENTACIÓN ..	20
2.3	PROGRAMAS SOCIALES Y SUBSIDIOS DEL ESTADO	21
2.4	BUSINESS INTELLIGENCE	21
2.4.1	ALMACENAMIENTO DE DATOS.....	22
2.4.1.1	DATA WAREHOUSE	22
2.4.1.2	DATAMART.....	22
2.4.2	EXTRACTION, TRANSFORMATION AND LOAD (ETL)	23
2.4.3	ONLINE ANALYTICAL PROCESSING (OLAP).....	23
2.4.3.1	RELATIONAL ONLINE ANALYTICAL PROCESSING.....	24
	(ROLAP)	24
2.4.3.2	MULTIDIMENSIONAL ONLINE ANALYTICAL PROCESSING	
	(MOLAP).....	25
2.4.3.3	HYBRID ONLINE ANALYTICAL PROCESSING (HOLAP).....	26
2.4.4	DIFERENCIAS ENTRE ROLAP Y MOLAP	27
2.4.5	VENTAJAS Y DESVENTAJAS DE ROLAP.....	27
2.4.6	VENTAJAS Y DESVENTAJAS DE MOLAP	28
2.5	BUSINESS ANALYTICS	29
2.5.1	BUSINESS ANALYTICS EN LAS ORGANIZACIONES.....	30
2.5.2	BIG DATA	30
2.5.2.1	VOLUMEN	31

2.5.2.2	VARIEDAD.....	31
2.5.2.3	VELOCIDAD.....	32
2.5.3	TECNOLOGÍAS DE BIG DATA	32
2.5.4	BASE DE DATOS NOSQL	32
2.5.5	IMPLEMENTACIÓN DE BIG DATA	33
2.5.6	MINERÍA DE DATOS.....	33
2.5.6.1	PROCESO KDD	34
2.5.6.2	TÉCNICAS DE MINERÍA DE DATOS	35
2.5.6.3	TIPOS DE ANÁLISIS DE DATOS	35
2.5.7	ANÁLISIS PREDICTIVO.....	36
2.5.7.1	CONOCIMIENTO DEL NEGOCIO	36
2.5.7.2	DATA.....	36
2.5.7.3	MÉTODOS.....	37
2.5.7.4	MODELOS Y ALGORITMOS DEL ANÁLISIS PREDICTIVO	37
2.5.7.4.1	CLASIFICACIÓN	37
2.5.7.4.2	CLUSTERING (APRENDIZAJE NO SUPERVISADO):	37
2.5.7.4.3	REGLAS DE ASOCIACIÓN	38
2.5.7.4.4	REGRESIÓN	38
2.5.7.4.5	MATRIZ DE CONFUSIÓN	38
2.5.7.5	PRINCIPALES HERRAMIENTAS DE PREDICCIÓN.....	39
2.5.7.5.1	HERRAMIENTAS DE CÓDIGO ABIERTO.....	39
2.5.7.5.2	HERRAMIENTAS COMERCIALES DISPONIBLES	39
2.5.8	BENCHMARKING DE TECNOLOGÍAS BUSINESS ANALYTICS	42
2.5.9	METODOLOGÍA DE IMPLEMENTACIÓN DE BUSINESS ANALYTICS.....	43
2.5.9.1	BUSINESS ACTIVITY MONITORING (BAM)	43
2.5.9.2	CROSS INDUSTRY STANDARD PROCESS FOR DATA MINING (CRISP-DM)	44
2.5.9.3	SAMPLE, EXPLORE, MODIFY, MODEL AND ASSESS	47
	(SEMMA)	47
3	SITUACIÓN PROBLEMÁTICA	48
3.1	OBJETO DE ESTUDIO: LA INSTITUCIÓN	48
3.1.1	MISIÓN	49

3.1.2	VISIÓN	49
3.1.3	OBJETIVOS ESTRATÉGICOS.....	49
3.1.4	ANÁLISIS FODA.....	50
3.2	MARCO REGULATORIO DEL PROCESO DE FOCALIZACIÓN.....	53
3.3	SITUACIÓN DE LA POBREZA EN EL PERÚ.....	53
3.3.1	SITUACIÓN DE LA POBREZA EN AMÉRICA LATINA	53
3.3.2	SITUACIÓN DE PERÚ Y OTROS PAÍSES DE AMÉRICA LATINA ..	54
3.3.3	VARIACIÓN DE LA POBREZA EN EL PERÚ	55
3.3.3.1	DISTRIBUCIÓN GEOGRÁFICA DE LA POBREZA EN PERÚ ...	57
3.3.3.2	PERFIL DE POBLACIÓN EN POBREZA.....	58
3.3.3.3	COBERTURA EN SALUD	59
3.4	PROCESO DE DETERMINACIÓN DE LA CLASIFICACIÓN SOCIOECONÓMICA (CSE) ACTUAL.....	60
3.4.1	RECOPIACIÓN DE INFORMACIÓN	61
3.4.1.1	FUENTES MANUALES Y/O ELECTRÓNICOS	61
3.4.1.2	FUENTES EXTERNAS	66
3.4.2	PROCESO DE CLASIFICACIÓN SOCIOECONÓMICA	68
3.4.2.1	EVALUACIÓN DE LA CLASIFICACIÓN SOCIOECONÓMICA	68
3.4.2.2	REEVALUACIÓN DE LA CLASIFICACIÓN SOCIOECONÓMICA	69
3.4.3	PADRÓN GENERAL DE HOGARES.....	70
3.4.3.1	ACCESO AL RESULTADO DE LA CLASIFICACIÓN SOCIOECONÓMICA	71
3.5	INTERVENCIONES PÚBLICAS FOCALIZADAS.....	74
3.5.1	PROGRAMAS SOCIALES Y SUBSIDIOS DEL ESTADO.....	76
3.5.2	PRESUPUESTO ASIGNADO	80
3.6	DESCRIPCIÓN DE LA SITUACIÓN PROBLEMÁTICA	81
3.6.1	IMPACTO DE LA CSE EN LA POBREZA	81
3.6.1.1	PRESUPUESTO	81
3.6.1.2	FILTRACIÓN.....	82
3.6.2	FACTORES DE IMPACTO A LA CLASIFICACIÓN SOCIOECONÓMICA	85

3.6.2.1	PERSONA	85
3.6.2.2	INFORMACIÓN	86
3.6.2.3	TECNOLOGÍA.....	87
4	SOLUCIÓN PROPUESTA.....	88
4.1	ANTECEDENTES	88
4.2	SELECCIÓN DE LA METODOLOGIA.....	90
4.3	APLICACIÓN DE LA METODOLOGIA	91
4.3.1	COMPRENSIÓN DEL NEGOCIO	91
4.3.1.1	DETERMINAR LOS OBJETIVOS DE NEGOCIO.....	91
4.3.1.2	EVALUACIÓN DE LA SITUACIÓN ACTUAL	92
4.3.1.3	DEFINIR LOS OBJETIVOS DE MINERÍA DE DATOS	94
4.3.1.4	PLAN DE PROYECTO	94
4.3.2	COMPRENSIÓN DE LOS DATOS	95
4.3.2.1	RECOPIACIÓN INICIAL DE DATOS	95
4.3.2.2	PROPORCIONAR UNA DESCRIPCIÓN DE LOS DATOS	101
4.3.2.3	EXPLORACIÓN O ANÁLISIS DE LOS DATOS.....	110
4.3.2.4	VERIFICACIÓN DE CALIDAD DE LOS DATOS	118
4.3.3	PREPARACIÓN DE LOS DATOS	118
4.3.3.1	SELECCIÓN DE LOS DATOS RELEVANTES	118
4.3.3.2	DEPURACIÓN DE LOS DATOS NO RELEVANTES.....	120
4.3.3.3	ESTRUCTURACIÓN DE LOS DATOS Y SUS ATRIBUTOS	122
4.3.3.4	INTEGRACIÓN DE LOS DATOS ESTRUCTURADOS Y NO ESTRUCTURADOS	127
4.3.3.5	FORMATEO DE DATOS GENERANDO UN INFORME DE CALIDAD	127
4.3.4	MODELADO	129
4.3.4.1	SELECCIÓN DEL MODELO Y SUPUESTOS	129
4.3.4.2	ELABORAR EL PLAN DE PRUEBAS	130
4.3.4.2.1	ALGORITMO K-MEANS	130
4.3.4.2.2	ALGORITMO NAIVE BAYES	130

4.3.4.3	CONSTRUCCIÓN DEL MODELO CONSIDERANDO LOS PARÁMETROS ESCOGIDOS, MODELO, DESCRIPCIÓN DE LOS MODELOS.....	131
4.3.4.3.1	USO ALGORITMO NAIVE BAYES	143
4.3.4.4	EVALUACIÓN DEL MODELO	153
4.3.4.5	EVALUACION.....	163
4.3.4.5.1	EVALUACIÓN DE RESULTADOS.....	164
4.3.4.5.2	REVISIÓN DEL PROCESO	164
4.3.4.5.3	DETERMINACIÓN DE LOS PASOS O ACCIONES A SEGUIR	165
4.4	EVALUACIÓN DE BENEFICIOS	165
4.4.1	EVALUACIÓN ECONÓMICA	165
4.4.1.1	AHORRO ESPERADO	165
4.4.1.2	PRESUPUESTO DEL PROYECTO	166
4.4.1.3	FLUJO DE CAJA DEL PROYECTO	168
4.4.2	BENEFICIOS SOCIALES	172
5	CONCLUSIONES	174
6	RECOMENDACIONES	175
7	REFERENCIAS	176
8	ANEXOS	180

ÍNDICE DE TABLAS

Tabla 1 <i>Diferencias entre ROLAP y MOLAP</i>	27
Tabla 2 <i>Ventajas y desventajas de ROLAP</i>	27
Tabla 3 <i>Ventajas y desventajas de MOLAP</i>	28
Tabla 4 <i>Principales Herramientas Predictivas</i>	43
Tabla 5 <i>Análisis FODA MIDIS- FORTALEZAS</i>	51
Tabla 6 <i>Análisis FODA MIDIS- OPORTUNIDADES</i>	51
Tabla 7 <i>Análisis FODA MIDIS- DEBILIDADES</i>	52
Tabla 8 <i>Análisis FODA MIDIS- AMENAZAS</i>	52
Tabla 9 <i>% de Personas en situación de pobreza y pobreza extrema 2012 – 2016</i>	55
Tabla 10 <i>Información que almacena el PGH</i>	70
Tabla 11 <i>Índice de filtración en los Programas Sociales (Porcentaje del PBI)</i>	83
Tabla 12 <i>Filtración en los Programas Sociales</i>	84
Tabla 13 <i>Plan de Proyecto</i>	95
Tabla 14 <i>Fuentes de Datos</i>	95
Tabla 15 <i>Datos contenidos en el formato S100</i>	102
Tabla 16 <i>Datos contenidos en el format FSU</i>	105
Tabla 17 <i>Depuración datos no relevantes</i>	119
Tabla 18 <i>Datos que no serán utilizados para el análisis</i>	120
Tabla 19 <i>Datos para normalización</i>	122
Tabla 20 <i>Ejemplo de campo normalizado:</i>	123
Tabla 21 <i>Normalización de datos</i>	124
Tabla 22 <i>Creación de un nuevo campo ingreso promedio por hogar</i>	125
Tabla 23 <i>Datos a consolidar</i>	126
Tabla 24 <i>Campos calculados</i>	126
Tabla 25 <i>Campos formateados</i>	128
Tabla 26 <i>Pruebas fuente de datos de entrenamiento – k-means</i>	130
Tabla 27 <i>Pruebas fuente de datos de entrenamiento – k-means</i>	130
Tabla 28 <i>Lista de variables para el proceso</i>	131
Tabla 29 <i>Detalle Resultados K-means</i>	135
Tabla 30 <i>Detalle resultados Data Prueba K-means</i>	140
Tabla 31 <i>Detalle Resultado Naive Bayes</i>	146

Tabla 32 <i>Costo asociado a la Filtración Actual</i>	165
Tabla 33 <i>Costo asociado a la Filtración Propuesta</i>	166
Tabla 34 <i>Partidas presupuestales</i>	167
Tabla 35 <i>Presupuesto de Proyecto</i>	168
Tabla 36 <i>Porcentaje Depreciación</i>	169
Tabla 37 <i>Inversión Inicial del Proyecto</i>	170
Tabla 38 <i>Cálculo Depreciación</i>	170
Tabla 39 <i>Estado de Resultados</i>	171
Tabla 40 <i>Flujo de Caja Libre</i>	172
Tabla 41 <i>Indicadores Financieros y sociales del Proyecto</i>	172
Tabla 42 <i>Localización territorial de pobreza</i>	182
Tabla 43 <i>Evolución del analfabetismo según condición de pobreza, 2007-2017</i>	188

ÍNDICE DE FIGURAS

<i>Figura 1.</i> Relational Online Analytical Processing.....	25
<i>Figura 2.</i> Multidimensional online Analytical Processing.....	26
<i>Figura 3.</i> Hybrid Online Analytical Process	26
<i>Figura 4.</i> Arquitectura Business Intelligence.....	29
<i>Figura 5.</i> Secuencia pasos en el proceso KDD	35
<i>Figura 6.</i> Familias de algoritmos utilizados en el modelado predictivo	38
<i>Figura 7.</i> Fases de la Metodología CRISP-DM	47
<i>Figura 8.</i> Fases de la Metodología SEMMA	48
<i>Figura 9.</i> Mapa de objetivos estratégicos del MIDIS	50
<i>Figura 10.</i> Porcentaje de pobreza y pobreza extrema	54
<i>Figura 11.</i> Evolución de la distribución de la población (pobreza extrema, pobreza).....	56
<i>Figura 12.</i> Evolución de la distribución de la población (pobreza extrema, pobreza)	56
<i>Figura 13.</i> Distribución de la pobreza a nivel nacional	57
<i>Figura 14.</i> Evolución de población con seguro de Salud.....	60
<i>Figura 15.</i> Proceso de Determinación de Clasificación Socioeconómica.....	60
<i>Figura 16.</i> Formato S100	64
<i>Figura 17.</i> Formato FSU	66
<i>Figura 18.</i> Pantalla acceso al PGH.....	72
<i>Figura 19.</i> Pantalla consulta PGH.....	73
<i>Figura 20.</i> Resultado de consulta al PGH	73
<i>Figura 21.</i> Proceso de Clasificación Socioeconómica Actual	74
<i>Figura 22.</i> Presupuesto de los Programas Sociales 2007-2017.....	81
<i>Figura 23.</i> Tendencia presupuesto Programas Sociales vs Índice de Pobreza y pobreza extrema	82
<i>Figura 24.</i> Filtración en Cajamarca.....	84
<i>Figura 25.</i> Formato FSU – pagina1	96
<i>Figura 26.</i> Formato FSU – pagina2	97
<i>Figura 27.</i> Formato FSU – pagina3	98
<i>Figura 28.</i> Formato FSU – pagina4	99
<i>Figura 29.</i> Formato S100 – pagina1.....	100
<i>Figura 30.</i> Formato S100 – pagina2.....	101

<i>Figura 31. Cantidad de Viviendas por Ubicación Geográfica</i>	110
<i>Figura 32. Cantidad de Hogares por Tipo de Vivienda.....</i>	111
<i>Figura 33. Cantidad de Hogares por Tenencia de Propiedad</i>	111
<i>Figura 34. Cantidad de Hogares por Tipo de Material de Pared</i>	112
<i>Figura 35. Cantidad de Hogares por Tipo de Material de Techo.....</i>	112
<i>Figura 36. Cantidad de Hogares por Tipo de Material de Techo</i>	113
<i>Figura 37. Cantidad de Hogares por Tipo de Alumbrado.....</i>	113
<i>Figura 38. Cantidad de Hogares por Tipo de Abastecimiento de Agua</i>	114
<i>Figura 39. Cantidad de Hogares por Tipo de Conexión de Servicios Higiénicos</i>	114
<i>Figura 40. Cantidad de Hogares por Tipo de Combustible usado para cocinar</i>	115
<i>Figura 41. Cantidad de Hogares por Tipo de Artefactos</i>	115
<i>Figura 42. Cantidad de integrantes del hogar por idioma de su niñez</i>	116
<i>Figura 43. Porcentaje de Integrantes de Hogares que cuenta con Seguro de Salud</i>	116
<i>Figura 44. Porcentaje de Integrantes de Hogares que cuenta con Seguro de Salud</i>	117
<i>Figura 45. Porcentaje de Integrantes de Hogares beneficiarios de programas sociales....</i>	117
<i>Figura 46. Cantidad de Integrantes de Hogares por Sector en el que laboran}</i>	118
<i>Figura 47. Estructuración de datos</i>	122
<i>Figura 48. Selección de archivo para procesamiento</i>	132
<i>Figura 49. Variables para procesamiento</i>	133
<i>Figura 50. Selección de algoritmo</i>	133
<i>Figura 51. Parametrización 3 clústeres</i>	134
<i>Figura 52. Resultados 1 Clustering – Kmeans.....</i>	134
<i>Figura 53. Resultados2 clustering – Kmeans.....</i>	135
<i>Figura 54. Distribución de Clúster.....</i>	138
<i>Figura 55. Selección de Data de Prueba K-means.....</i>	139
<i>Figura 56. Resultado1 Data Prueba K-means</i>	139
<i>Figura 57. Resultado2 Data Prueba K-means</i>	140
<i>Figura 58. Distribución data Prueba K-means.....</i>	143
<i>Figura 59. Selección archivo Naive Bayes.....</i>	144
<i>Figura 60. Variables Data Prueba Naive Bayes.....</i>	144
<i>Figura 61. Resultado2 Naive Bayes</i>	145
<i>Figura 62. Evaluación de Atributos CorrelationAttributeEval – Mayor relevancia</i>	153

<i>Figura 63. Evaluación de Atributos CorrelationAttributeEval – Menor relevancia</i>	154
<i>Figura 64. Evaluación de Atributos GainRatioAttributeEval – Mayor relevancia</i>	154
<i>Figura 65. Evaluación de Atributos GainRatioAttributeEval – Menor relevancia</i>	155
<i>Figura 66. Evaluación de Atributos InfoGainAttributeEval – Mayor relevancia</i>	155
<i>Figura 67. Evaluación de Atributos InfoGainAttributeEval – Menor relevancia</i>	156
<i>Figura 68. Análisis Ingreso promedio familiar</i>	157
<i>Figura 69. Análisis acceso al agua del rio</i>	158
<i>Figura 70. Análisis Material de la pared tipo adobe</i>	158
<i>Figura 71. Análisis cantidad de propiedades</i>	159
<i>Figura 72. Resultado Ejecución Data Prueba</i>	160
<i>Figura 73. Resultados de Ejecución de Prueba en formato ARTF de Weka</i>	160
<i>Figura 74. Extracto de resultados de Ejecución Data de Prueba en CSV</i>	161
<i>Figura 75. Información anómala según ingreso promedio familiar</i>	161
<i>Figura 76. Posible filtración de información</i>	162
<i>Figura 77. Resultado SSE</i>	163
<i>Figura 78. Resultados Precisión y Exhaustividad</i>	163
<i>Figura 79. Pobreza por grupos de edades</i>	180
<i>Figura 80. Pobreza según lengua materna</i>	181
<i>Figura 81. Evolución de la pobreza 2016-2017 por Región Natural</i>	183
<i>Figura 82. Miembros de hogar en condición de pobreza, 2017</i>	183
<i>Figura 83. Material predominante en paredes de viviendas en condición de pobreza.</i>	184
<i>Figura 84. Material predominante en piso de viviendas en condición de pobreza, 2017.</i>	185
<i>Figura 85. Abastecimiento de agua en viviendas en condición de pobreza, 2017</i>	185
<i>Figura 86. Evolución de la pobreza monetaria, según área de residencia 2007-2017</i>	186
<i>Figura 87. Evolución de la pobreza extrema, según área de residencia 2007-2017</i>	187

1 INTRODUCCIÓN

El presente trabajo se enmarca en el nuevo paradigma del análisis de los datos, donde la información y el conocimiento son el pilar para la toma de decisiones.

Conocer el pasado y establecer los patrones en la información, permitirán adelantarse al futuro y hará posible que una organización genere valor, alcanzando los objetivos estratégicos planteados y generando nuevos retos.

El presente trabajo de investigación es una propuesta para el uso de Analytics en el marco del proceso de Clasificación Socioeconómica en el MIDIS, buscando determinar patrones de comportamiento que contribuyan a tener una política efectiva en los programas sociales con el fin de mejorar la calidad de vida de las personas en situación vulnerable.

El capítulo 1 corresponde a la parte teórica, donde se desarrollan conceptos básicos como pobreza, programas sociales, sistemas de focalización y errores en América Latina. Se describen también conceptos de almacenamiento y procesamiento de datos como Big data, Business Intelligence, Business Analytics, herramientas y algoritmos de análisis predictivo y algunas metodologías de implementación de Business Analytics.

El capítulo 2 busca mostrar la situación problemática que deseamos resolver, describiendo al MIDIS y sus objetivos estratégicos. Analizaremos la distribución de la pobreza en el Perú en los últimos años, describiendo el proceso de clasificación socioeconómico actual, el presupuesto asignado a los programas sociales y los conceptos de filtración y subcobertura.

El capítulo 3 describe la puesta en práctica de la solución de la propuesta mediante la aplicación de las distintas etapas de la metodología seleccionada, utilizando los datos disponibles para su posterior evaluación de resultados con respecto a los objetivos de la institución. Se describe también la evaluación de la presente propuesta en el marco de los beneficios sociales que puede generar.

Finalmente, el **capítulo 4** corresponde a las conclusiones y recomendaciones de la propuesta.

2 MARCO TEÓRICO

2.1 LA POBREZA

Algunos autores conceptualizan la pobreza desde el punto de vista de las diferentes escuelas económicas: escuela clásica, escuela keynesiana, escuela neoclásica y la escuela marxista (Davis & Sanchez-Martinez, 2015).

- Para la escuela clásica la pobreza es el resultado de una mala decisión tomada por los individuos, cuya productividad se refleja a través de los salarios, no influye el funcionamiento de los mercados ya que estos son eficientes.
- Para la escuela Keynesiana la pobreza es un resultado coyuntural y se asume como el estado en el cual las personas no tienen la capacidad de encontrar un trabajo estable, viéndose afectados por el funcionamiento del mercado, por ejemplo, cuando hay recesión la pobreza se incrementa.
- La escuela Neoclásica considera a la pobreza bajo el enfoque de escasez de activos y servicios tales como vivienda, salud, educación, pensiones, seguros de vida, etc. Este enfoque señala que los hogares dueños de un adecuado nivel de activos son menos afectados por las fluctuaciones en sus ingresos que los hogares sin un adecuado nivel de activos, a los que considera como hogares pobres.
- La escuela marxista clasifica la pobreza a través de un enfoque estructural de ingresos y considera que las causas de la pobreza se deben a cómo está estructurado el sistema capitalista y la forma de aliviarlo es a través de políticas regulatorias como el salario mínimo.

La forma de medir la pobreza puede definirse como (Martínez Álvarez, 2003):

- Pobreza objetiva. La pobreza es medida a través de indicadores de criterios objetivos (hechos) de los hogares o de las personas, tales como el nivel de ingresos y gastos, consumo de ciertos bienes y servicios, entre otros. La medida está basada en un nivel de vida mínimo que responde a criterios establecidos a priori.
- Pobreza subjetiva. La pobreza es medida a través de la percepción de la persona, quien no está satisfecha con su situación, que se siente excluido de lo que considera como una condición de vida normal, considerándose como pobre, independientemente de sus posibilidades económicas.

2.1.1 MÉTODOS DE MEDICIÓN DE LA POBREZA

(Perez & Rodriguez, 2015), clasifica los métodos de medición de la pobreza:

- Método de pobreza subjetiva: Para su cálculo, se realiza un método de encuesta de opinión donde se le pregunta al jefe de hogar si se considera pobre. La limitación de este enfoque es que las percepciones de las personas no son las mismas, algunos pueden sobreestimar sus necesidades mínimas, lo que sesga el resultado.
- Método de líneas de pobreza monetaria. Para su cálculo, se estima el costo de una canasta básica de bienes de consumo y se identifica como pobre al individuo u hogar cuyos ingresos no satisface lo necesario para adquirir dicha canasta. Este método sólo está basado en los ingresos que perciben los individuos u hogares.
- Línea de pobreza multidimensional. No sólo considera indicadores de ingresos o gastos del individuo u hogar, sino que considera privaciones o carencias que padecen los individuos u hogares pobres como los servicios de salud, educación, acceso a servicios como electricidad, agua y desagüe, y otros aspectos relacionados a calidad de vida.
- Necesidades Básicas Insatisfechas (NBI). Define un conjunto de necesidades básicas y considera como pobres a aquellos hogares que carecen de uno o más de esas necesidades. Los indicadores más usuales empleados como NBI son: hacinamiento, vivienda inadecuada, falta de servicios como agua y desagüe, educación primaria incompleta de los hijos y nivel educativo del jefe de hogar. Su limitación está en la subjetividad en la definición de los indicadores de NBI y en que les otorga el mismo peso.

2.2 SISTEMA DE FOCALIZACIÓN

2.2.1 FOCALIZACIÓN

Debido a la presión política para limitar los gastos estatales, se limitó la capacidad de los gobiernos para financiar programas universales. Ante esto, surge la focalización como una política para identificar a la población pobre con el fin de transferirle recursos limitados de manera igualitaria y excluir a los que no lo necesitan (Peyre Dutrey, 2007).

La focalización describe los mecanismos para identificar a los hogares o individuos que se definen como elegibles para la transferencia de recursos y al mismo tiempo excluir a aquellos que se consideran no elegibles (Sabates-Wheeler, Hurrell, & Devereux, 2015).

De acuerdo con lo mencionado, entendemos que la focalización permite clasificar a la población y que los gobiernos dirijan sus políticas, programas de protección social y limitados presupuestos a la población en condición de pobreza y vulnerabilidad.

2.2.2 MECANISMOS DE FOCALIZACIÓN

(Sabates-Wheeler et al., 2015) señala los diferentes mecanismos de focalización:

- Focalización geográfica: cobertura total de una comunidad o localidad, es rápida y barata porque no requiere una evaluación individual y no discrimina entre los residentes de la comunidad o localidad. Puede darse errores de inclusión, ya que puede beneficiar a personas no pobres.
- Focalización usando pruebas de medios (Proxy means tests): se le considera como el mecanismo más preciso para identificar a la población pobre, pero es costoso al requerir una evaluación individual donde se obtiene información detallada y sensible de la población, la que debe ser obtenida, verificada y reevaluada regularmente, ya que su condición puede cambiar en el tiempo.
- Focalización de enfoques categóricos: se centra en criterios para beneficiar a la población; por ejemplo, el criterio edad (infantes, niños o personas mayores), condiciones como mujeres embarazadas o personas con discapacidad. Puede darse errores de inclusión al destinar la ayuda a personas no pobres. También puede darse errores de exclusión al excluir; por ejemplo, a niños o personas mayores que no tienen documento de identificación y; por lo tanto, no pueden demostrar su identidad y edad.
- Focalización basada en la comunidad: este tipo de focalización es realizada por las propias comunidades que identifican a sus miembros más pobres. Las comunidades seleccionan a los hogares que consideran requieren la ayuda de los programas de protección social. Puede ser propensa a prejuicios de los líderes comunitarios.

2.2.3 FOCALIZACIÓN EN AMÉRICA LATINA

(Castañeda et al., 2005) presenta una breve descripción de la focalización aplicada en algunos países de América:

- **Brasil:** Cuenta con Ministerio de Desarrollo Social y un Registro Único que es una herramienta que permite identificar las carencias de los hogares pobres, con el fin de ofrecerles acciones, servicios y programas. Se basa en una recolección de datos por barrido de encuestas.
- **Colombia:** Usan el SISBEN, un sistema de información que identifica clasifica y selecciona a las personas y familias de acuerdo a su condición socioeconómica como potenciales beneficiarios de subsidios dentro de los programas sociales. Realiza una recopilación de datos por barrido de encuesta.
- **Costa Rica:** Cuenta con una Estrategia Nacional para la Reducción de la Pobreza Puente al Desarrollo. Utiliza una combinación de procesos de recopilación de datos como el uso de un mapa de pobreza rural que visualiza información de las condiciones de la población y sus necesidades básicas; así como el uso de información censal de la Encuesta Nacional de Hogares del Instituto Nacional de Estadísticas y Censos (INEC) que es plasmada en el mapa de pobreza.
- **Chile:** Cuenta con un Ministerio de Desarrollo Social que diseña y aplica políticas, planes y programas de desarrollo social para erradicar la pobreza y brindar protección a personas vulnerables. Desde los años 80, Chile aplica un instrumento denominado Ficha de Estratificación Social conocida como ficha CAS, a través del cual recopila datos, evalúa la condición socioeconómica del hogar y determina el acceso a beneficios que otorgan las prestaciones sociales.
- **México:** Utiliza un mecanismo de focalización basado en la identificación de zonas geográficas de mayor concentración de pobreza usando información del Censo General de población y Vivienda. Además, usa recolección de datos socioeconómica de los hogares.

2.2.4 ERRORES EN LA FOCALIZACIÓN

La traducción de (Brown, Ravallion, & Van de Walle, 2016) señala que la medición del rendimiento de la focalización se basa en:

- *“La Tasa de Error de Inclusión (IER), también conocida como filtración, está dada por la proporción de población identificada como pobre que en realidad no lo es. Los errores de inclusión implican un gasto presupuestal sin ningún impacto directo sobre la tasa de pobreza.*

- *La Tasa de Error de Exclusión (EER), también llamada subcobertura, es definida como la proporción de la población pobre que no está identificada como tal”.*

De lo señalado, la IER hace que los esfuerzos realizados por los países para reducir la pobreza y vulnerabilidad no se dirija a la población que realmente requiere de la ayuda de los programas de protección social. Asimismo, este gasto presupuestal dirigido a quienes no debe ir hace que los programas de protección social con sus escasos recursos presupuestales no puedan dar cobertura a toda la población identificada como pobre.

Si el objetivo es la reducción de la pobreza, los responsables políticos con un presupuesto determinado deberían estar más preocupados por los errores de exclusión que por los errores de inclusión (Brown et al., 2016). Sin embargo; si se reduce los errores de inclusión, puede generar que los programas de protección social tengan mayor presupuesto para incluir a las personas pobres.

(Sabates-Wheeler et al., 2015) indica que los errores de inclusión y exclusión pueden surgir en la etapa de diseño y/o durante la implementación de la focalización.

2.2.4.1 ERRORES DE FOCALIZACIÓN EN EL DISEÑO

Los errores de diseño ocurren cuando se utilizan medidas indirectas para identificar a los hogares pobres o cuando se seleccionen criterios de focalización sin vinculación explícita con las medidas de pobreza objetivamente cuantificables.

Los errores de focalización generan un malgasto presupuestal y no reduce la pobreza.

2.2.4.2 ERRORES DE FOCALIZACIÓN EN LA IMPLEMENTACIÓN

Los errores de inclusión en la implementación pueden ocurrir por numerosas razones como el dado por los beneficios que prestan los programas de protección social para las personas, que hace que estas tergiversen sus datos con el fin de salir elegibles y beneficiarse de los programas de protección social o cuando las comunidades excluyen deliberadamente a algunos miembros.

2.3 PROGRAMAS SOCIALES Y SUBSIDIOS DEL ESTADO

Según (World Bank Group, s. f.), el Banco Mundial afirma que el 36 % de las personas más pobres que recibió beneficios de los programas de protección social salió de la pobreza extrema. También afirma que a través de los programas de transferencia monetaria se redujo la desigualdad y la brecha de pobreza en alrededor del 45%, lo que demuestra que estos programas tienen un impacto sustancial en la lucha mundial contra la pobreza. El impacto de las redes de protección social en la pobreza se mide sobre la base de los datos disponibles de los hogares de 79 países, comparando el bienestar de los beneficiarios de estas redes con aquel de que gozaría si no hubieran recibido dicho apoyo.

Pese a la creciente aprobación de programas de protección social en los países en los últimos años, la cobertura mundial de la población pobre y vulnerable continúa siendo insuficiente. Alrededor de 2500 millones de personas en todo el mundo están cubiertas por alguna red de protección social, 650 millones de las cuales pertenecen al 20% más pobre de la población. Sin embargo, tan sólo una de cada cinco personas que viven en un país de ingreso bajo está cubierta por una red de protección social. Además, los países con un alto riesgo de sufrir desastres naturales suelen tener menos cobertura de redes de protección social.

Del informe presentado podemos ver cómo los programas sociales contribuyen a disminuir la brecha de pobreza, por tal motivo la necesidad de mejorar los procesos para que la ayuda llegue a la población que lo necesita.

2.4 BUSINESS INTELLIGENCE

Según (Gómez & Bautista, 2010) señala que business intelligence: *“Se define como la habilidad corporativa para tomar decisiones. Esto se logra mediante el uso de metodologías, aplicaciones y tecnologías que permiten reunir, depurar, transformar datos, y aplicar en ellos técnicas analíticas de extracción de conocimiento, los datos pueden ser estructurados para que indiquen las características de un área de interés, generando el conocimiento sobre los problemas y oportunidades del negocio para que puedan ser corregidos y aprovechados respectivamente”*.

Podemos concluir que Business Intelligence (BI) es un conjunto de estrategias para la creación de conocimiento a partir del análisis de datos recopilados de una organización. El proceso de BI se apoya de metodologías, herramientas y tecnologías.

Según (Tamayo & Moreno, 2006) señalan que BI consta de tres componentes básicos:

- Almacenamiento de datos: Data Warehouse (DW) y data mart (DM)
- Extracción de datos: Extract, Transform and Load (ETL)
- Herramienta de análisis: Llamada On Line Analytical Processing (OLAP)

2.4.1 ALMACENAMIENTO DE DATOS

2.4.1.1 DATA WAREHOUSE

Según (Gómez & Bautista, 2010) define el Data Warehouse como: *“El proceso de extraer datos de distintas aplicaciones (internas y externas), para que una vez depurados y especialmente estructurados sean almacenados en un depósito de datos consolidado para el análisis del negocio. Requiere una combinación de metodologías, técnicas, hardware y los componentes de software que proporcionan en conjunto la infraestructura para soportar el proceso de información. La estructura que se defina debe reflejar las necesidades y características del negocio, sus departamentos, equipos de trabajo y directivos, esto permitirá responder a interrogantes generados al tratar de tomar las decisiones y con el tiempo se va convirtiendo en la memoria corporativa; describiendo el pasado y el presente de la empresa. Data Warehouse desglosa, resume, ordena y compara, pero no descubre, ni predice”*.

De lo señalado por (Gómez & Bautista, 2010) podemos concluir que un Data Warehouse (DW) es un almacén de grandes cantidades de datos, los mismos que son recopilados e integrados desde múltiples fuentes y aunque no descubre ni predice genera información relevante para ayudar a las organizaciones a la toma de decisiones.

2.4.1.2 DATAMART

Según (Vicente, 2008) señala que: *“Un data mart es un subconjunto del DW relativo a un departamento o área de negocio concreto. El data mart puede funcionar de manera autónoma o estar enlazado al DW corporativo central”*.

Bajo el concepto definido en el párrafo anterior, entendemos que un datamart es un subconjunto de almacén de datos especializado en un área de negocio específico. Por ejemplo, los datos de ventas, finanzas, etc.

2.4.2 EXTRACTION, TRANSFORMATION AND LOAD (ETL)

Según (Bustamante Martínez, Galvis Lista, & Gómez Flórez, 2013) señalan que *“El proceso de extracción, transformación y carga – ETL (Extraction, Transformation and Load) es una de las actividades técnicas más críticas en el desarrollo de soluciones de inteligencia de negocios – BI”*.

Por lo mencionado en el párrafo anterior el proceso de ETL permite extraer, transformar y limpiar datos generando información relevante, la cual será almacenada en un DATAMART o DW. Dicha información podrá ser utilizada para la definición de estrategias que permitan generar valor en la organización.

Según (Bustamante Martínez et al., 2013) señalan que el proceso de ETL cuenta con 3 fases o secuencias que se detallan a continuación:

- **Extracción**

Se extrae información de diferentes sistemas y/o fuentes de información (data estructurada).

- **Transformación**

La información que fue extraída de diferentes fuentes, en esta fase se realiza la validación, limpieza y transformación de los datos.

- **Carga**

Almacena los datos en los DATAMART o DW, que fueron validados en la fase de transformación, para que finalmente puedan ser explotados por diferentes herramientas de BI.

2.4.3 ONLINE ANALYTICAL PROCESSING (OLAP)

Según (Tamayo & Moreno, 2006) señala que *“La tecnología OLAP facilita el análisis de datos en línea en un DW, proporcionando respuestas rápidas a consultas analíticas complejas. OLAP es utilizado generalmente para ayudar en la toma de decisiones y*

presenta los datos a los usuarios a través de un modelo de datos intuitivo y natural. Con este estilo de presentación los usuarios finales pueden ver y entender con mayor facilidad la información de sus BD, lo que permite a las organizaciones reconocer el valor de sus datos”.

Según el artículo anterior OLAP es otro componente de BI, que facilita la explotación de los datos en línea del DW, permitiendo que el usuario final pueda visualizar la información de manera intuitiva.

Existen 3 tipos de sistemas OLAP:

2.4.3.1 RELATIONAL ONLINE ANALYTICAL PROCESSING

(ROLAP)

Según (Tamayo & Moreno, 2006) señala que “*ROLAP utiliza una arquitectura de tres niveles. La base de datos relacional maneja el almacenamiento de datos, el motor OLAP proporciona la funcionalidad analítica, y alguna herramienta especializada es empleada para el nivel de presentación. El nivel de aplicación es el motor OLAP, que ejecuta las consultas de los usuarios. El motor OLAP se integra con el nivel de presentación a través del cual los usuarios realizan los análisis OLAP*”. Ver figura de título Relational Online Analytical Processing (ROLAP).

De lo indicado en el artículo anterior, podemos concluir que esta tecnología consta de 3 componentes, los cuales son:

- Base de datos: Corresponde a la base de datos relacional, la cual permite el manejo, acceso y obtención de datos.
- Motor OLAP: El cual se encarga de la ejecución de las consultas del usuario a través de sentencias SQL.
- Interfaz OLAP: Corresponde a la visualización de datos, en los cuales los resultados de las consultas son devueltas a los usuarios para su explotación mediante herramientas front end.

Figura 1. Relational Online Analytical Processing

Fuente: Elaboración propia

2.4.3.2 MULTIDIMENSIONAL ONLINE ANALYTICAL PROCESSING (MOLAP)

Según (Tamayo & Moreno, 2006), un sistema MOLAP usa una base de datos multidimensional que almacena la información multidimensionalmente y se encarga del manejo, acceso y obtención de los datos. En MOLAP, el nivel de aplicación es el responsable de la ejecución de las consultas OLAP y el nivel de presentación brinda una interfaz a los usuarios finales para que visualicen los análisis OLAP.

De lo indicado en el artículo anterior, podemos concluir que esta tecnología consta de 2 componentes, los cuales son:

- **Base de datos Multidimensional (BDMD):** Comprende la capa de base de datos y la capa de aplicación. La BDMD es la encargada del manejo, acceso y obtención de los datos y de la ejecución de las consultas OLAP.
- **Interface OLAP:** Corresponde a la visualización de datos.

Ver figura de título Multidimensional Online Analytical Processing (MOLAP).

Figura 2. Multidimensional online Analytical Processing

Fuente: Elaboración propia

2.4.3.3 HYBRID ONLINE ANALYTICAL PROCESSING (HOLAP)

Según (Tamayo & Moreno, 2006) una base de datos con volúmenes muy grandes se debe mantener en un almacén ROLAP; mientras que los agregados se deben almacenar en un almacén MOLAP.

Ver figura de título Hybrid Online Analytical Processing (ROLAP).

Figura 3. Hybrid Online Analytical Process

Fuente: Fuente propia

2.4.4 DIFERENCIAS ENTRE ROLAP Y MOLAP

Tabla 1

Diferencias entre ROLAP y MOLAP

	MULTIDIMENSIONAL	RELACIONAL
DATOS	Detalle y pre calculados (Agregados)	Detalle y agregados
ESTRUCTURA	Matrices comprimidas	Tablas dimensionales
ADMINISTRACIÓN	Especialista en BDMD	Administrador de BD
ACCESO	Lenguaje especializado	SQL

Fuente: Marysol Tamayo, Francisco Javier Moreno, 2006

2.4.5 VENTAJAS Y DESVENTAJAS DE ROLAP

Tabla 2

Ventajas y desventajas de ROLAP

VENTAJAS	DESVENTAJAS
<ul style="list-style-type: none">● Se puede aprovechar la tecnología relacional, ya que se facilita aprovechar las inversiones realizadas en hardware y en SGBD relacionales.● Uso de la seguridad e integridad de los SGBD relacionales.● Capaz de manejar conjunto de datos muy grandes por encima de un terabyte.● Pueden utilizarse SGBD relacionales gratuitos● Soportan un gran número de dimensiones.● Escalable (adición de dimensiones a un esquema existente)	<ul style="list-style-type: none">● Es menor en rendimiento frente a MOLAP en BD pequeñas.● Limitación para consultas complejas.● Utilización de mucho almacenamiento en disco.

Fuente: (Tamayo & Moreno, 2006)

2.4.6 VENTAJAS Y DESVENTAJAS DE MOLAP

Tabla 3

Ventajas y desventajas de MOLAP

VENTAJAS	DESVENTAJAS
<ul style="list-style-type: none">● Ofrece un buen rendimiento cuando se trabaja sobre datos agregados, totales, subtotales, series temporales y diversos grados de detalles de los datos.● Facilita el estudio de alto nivel de los datos, al ofrecer una mayor flexibilidad y rapidez de acceso para el análisis de los mismos.● Almacenamiento de datos y consultas bastantes eficientes.● La complejidad de los datos de oculta a los usuarios.● El análisis se hace sobre datos agregados y métricas o indicadores pre calculados.● Mayor rendimiento frente a ROLAP en el procesamiento de consulta de BD pequeñas.● Almacena agregados para facilitar un acceso rápido.	<ul style="list-style-type: none">● La asimilación de los conceptos multidimensionales en especial, cuando se tienen hipercubos de muchas dimensiones.● La construcción y poblado de las estructuras multidimensionales pueden demandar mucho tiempo.● Están limitadas a tener diez o más dimensiones debido a la complejidad del manejo de las mismas.● No se puede acceder a datos que no están en el cubo (Ibarra 2005).● Existen pocas herramientas gratuitas que lo soporten.

Fuente: (Tamayo & Moreno, 2006)

Según (Affeldt & da Silva Júnior, 2013) explican que “*Los datos albergados en el Data Warehouse o en cada Datamart se explotan utilizando herramientas comerciales de análisis, reporting, alertas, etc. En estas herramientas se basa también la construcción de productos BI más completos, como los sistemas de soporte a la decisión (DSS), los sistemas de información ejecutiva (EIS) y los cuadros de mando (CMI) o Balanced Scorecard*”.

La arquitectura de BI se representa de acuerdo a la siguiente figura:

Figura 4. Arquitectura Business Intelligence

Fuente: (Affeldt & da Silva Júnior, 2013)

2.5 BUSINESS ANALYTICS

Para la definición de Business Analytics, se ha tomado como base la tesis de (Rosemberg, 2017) en la cual se toman 3 conceptos importantes:

- Según (Năstase & Stoica, 2011), Business Analytics utiliza el análisis de los datos para entender el negocio y su integración con el entorno. Cumple con la finalidad de generar información confiable y oportuna para la toma de decisiones.
- De acuerdo con lo definido por (Laureano, da Silva Laureano, & Grencho, 2016), Business Analytics es un proceso que utiliza 3 tipos de análisis para obtener resultados de los datos: el análisis de los datos actuales y las causas que los generan (análisis descriptivo); la predicción de comportamientos futuros (análisis predictivo) y la detección de patrones y tendencias (análisis de prescripción).
- Adicionalmente, según (Schniederjans, Schniederjans, & Starkey, 2014), business analytics es un proceso basado en técnicas estadísticas, sistemas de información y metodologías de investigación, los cuales permiten transformar los datos para identificar patrones de comportamiento y generar información útil.

De estos 3 conceptos podemos concluir que Business Analytics tiene como objetivo principal, generar información confiable para la toma de decisiones dentro de una

organización, basándose en el uso de técnicas para la recolección de datos, tratamiento, predicción y obtención de resultados. Entender el negocio no solo con datos actuales, sino en base a patrones de comportamiento le permitirá a la organización ser más competitiva dentro de su entorno.

2.5.1 BUSINESS ANALYTICS EN LAS ORGANIZACIONES

Según el artículo de (Sharma, Mithas, & Kankanhalli, 2014) en referencia a artículos como (Chen, Chiang, & Storey, 2012) p. 1166–1168, muchas organizaciones están tomando importancia al análisis de los datos y cómo la tecnología relacionada al Business Analytics ayuda a entender mejor su negocio.

Referencia también a (LaValle, Lesser, Shockley, Hopkins, & Kruschwitz, 2011), p. 22, que menciona cómo las organizaciones están tomando sus decisiones basándose cada vez más en un riguroso análisis de datos, y que la analítica está siendo utilizada no solo para la gestión estratégica sino también para las operaciones diarias.

Rajeev Sharma argumenta que, si bien hay evidencia de que el análisis de datos puede generar valor, según su punto de vista, lo primero que se debe definir es el proceso para la toma de decisiones, en donde el rol de la organización se centra en definir los procesos y la asignación de recursos, así como la orquestación de los procesos.

Según Tarek Soukieh, en su disertación del año 2015 sobre el efecto de la interacción de los usuarios con Business Analytics, menciona que la mayoría de las organizaciones se encuentran utilizando BA para mejorar su eficiencia, reducir sus costos al encontrar al cliente correcto y generar ingresos al reducir la rotación. Pero también, el análisis de datos se está utilizando para crear ideas innovadoras para mejorar el proceso de toma de decisiones.

2.5.2 BIG DATA

Según (Hernández-Leal, Duque-Méndez, & Moreno-Cadavid, 2017), el término Big Data describe un gran volumen de datos, tanto estructurados y no estructurados. Responde a la necesidad de contar con modelos y métodos para el almacenamiento y tratamiento de los datos. No solo se basa en el almacenamiento de un gran volumen de datos, sino a la velocidad de acceso y procesamiento de este. Se pueden encontrar 4 fases de agrupación

para las tecnologías de soporte: generación, adquisición, almacenamiento y análisis de datos.

Fase 1 – generación: En esta fase se genera gran cantidad de información, con distinta estructura y con características de gran valor.

Fase 2 – adquisición: recolección de los datos del proceso diario, evitando duplicidad o redundancia para no incrementar el espacio de almacenamiento innecesariamente.

Fase 3 – almacenamiento: en esta fase se hacen uso de las nuevas tecnologías que permiten almacenar los diferentes tipos de datos de manera masiva

Fase 4 – análisis de datos: Es la fase que permite obtener información relevante para la toma de decisiones. Requiere el uso de tecnología en tiempo real.

Según (Camargo-Vega, Camargo-Ortega, & Joyanes-Aguilar, 2015), Big Data considera 3 dimensiones importantes: Volumen, Velocidad y Variedad.

2.5.2.1 VOLUMEN

En los últimos años, la generación de datos ha tenido un crecimiento exponencial, llegando a cifras en el orden de terabytes, petabytes y exabytes. Según (Camargo-Vega et al., 2015) *“En el año 2000 se generaron 800.000 petabytes (PB), de datos almacenados y se espera que esta cifra alcance los 35 zettabytes (ZB) en el 2020. Las redes sociales también generan datos, es el caso de Twitter, que por sí sola genera más de 7 terabytes (TB) diariamente, y de Facebook, 10 TB de datos cada día. Algunas empresas generan terabytes de datos cada hora de cada día del año, es decir, las empresas están inundadas de datos”*.

2.5.2.2 VARIEDAD

Actualmente los datos han pasado de ser solo datos de tipo texto a ser representados como datos estructurados y no estructurados. Los datos no estructurados son los que se generan principalmente en las páginas web, redes sociales, foros, etc. La tecnología ha permitido el incremento de este tipo de datos (imágenes, sonido, etc.).

2.5.2.3 VELOCIDAD

Los datos se crean con mayor velocidad debido al incremento de los productos de desarrollo de software. Estas 3 características se relacionan entre sí, generando una respuesta adecuada a la necesidad del procesamiento de la información.

2.5.3 TECNOLOGÍAS DE BIG DATA

Según (Camargo-Vega et al., 2015), *“Para el manejo de datos es necesario tener dos componentes básicos, tanto el hardware como el software; respecto al primero, se tienen tecnologías tales como arquitecturas de Procesamiento Paralelo Masivo (MPP), que ayudan de forma rápida a su procesamiento. Para el manejo de datos no estructurados o semiestructurados es necesario acudir a otras tecnologías; es aquí donde aparecen nuevas técnicas y tecnologías, como MapReduce o Hadoop, diseñado para el manejo de información estructurada, no estructurada o semiestructurada”*.

2.5.4 BASE DE DATOS NOSQL

Según (Camargo-Vega et al., 2015), *“el término NoSQL significa no solo SQL. El nombre fue creado por Carlos Strozzi para denominar su base de datos que no ofrecía SQL”*.

Estas bases de datos difieren de las bases de datos relacionales ya que no usan SQL ni almacenan los datos en tablas con filas y columnas, tampoco hacen uso de sentencia tipo join y utilizan la memoria principal del computador.

Dependiendo de la forma en que se almacene la información, se pueden establecer varias bases de datos NoSQL, según (Camargo-Vega et al., 2015) presentamos una lista de algunas de dichas bases:

- **DynamoDB:** Desarrollada en Amazon, guarda grandes cantidades de información en unidades de estado sólido que le permiten una mayor velocidad para encontrar información.
- **Cassandra:** Iniciado por Facebook, de código abierto. Base de datos distribuida, que almacena los datos en forma de clave-valor, descentralizada y que puede atender muchas solicitudes a la vez sin mostrar degradación.
- **Voldemort:** Creada por linkedin para resolver la escalabilidad de las bases de datos relacionales, ambiente distribuido, código bajo licencia Apache.

- **Google BigTable:** Creado por Google el 2014, El sistema de archivos usado por BigTable es GFS (Google File System).
- **Hbase:** De código abierto, con almacenamiento de datos en formato clave-valor. Puede almacenar datos de tipos estructurados, semiestructurados y no estructurados. Necesita varios servidores y trabaja con un volumen de datos no tan grande.

2.5.5 IMPLEMENTACIÓN DE BIG DATA

Según (Camargo-Vega et al., 2015), se debe seguir un procedimiento para implementar big data en la organización, tales como:

- Entender el negocio y los datos. Este primer paso pide un análisis detallado con las personas que hoy laboran y entienden los procesos y los datos que la empresa maneja.
- El segundo paso consiste en determinar los problemas y cómo los datos pueden ayudar.
- Al momento de conocer los procesos es muy posible que se encuentren los problemas de la empresa o del negocio.
- Establecer expectativas razonables, es decir, definir metas alcanzables; esto se puede lograr si al implementar la solución de un problema éste no presenta alguna mejora, y se debe buscar otra solución.
- Existe una recomendación especial, y es que cuando se inicia un proyecto de Big Data es necesario trabajar en paralelo con el sistema que hoy está funcionando.
- Al tratar de implementar un proyecto de Big Data se debe ser flexible con la metodología y las herramientas; esto se debe a que las dos anteriores son recientes y pueden llegar a presentar problemas al implementarlas. Esto se puede solucionar realizando investigación e inversión en este tipo de tecnología.
- Es importante mantener el objetivo de Big Data en mente; esto porque el proceso es pesado y porque no es tedioso, máxime cuando los métodos y herramientas que usan Big Data para el análisis de datos aún pueden presentar problemas, y la idea es que se mantenga en mente la meta final del proyecto sin desanimarse pronto.

2.5.6 MINERÍA DE DATOS

Según (Riquelme Santos, Ruiz, & Gilbert, 2006): *“Todo apunta a la necesidad de metodologías de análisis inteligente de datos, las cuales puedan descubrir conocimiento*

útil de los datos. El término KDD (iniciales de Knowledge Discovery in Databases), acuñado en 1989 se refiere a todo el proceso de extracción de conocimiento a partir de una base de datos y marca un cambio de paradigma en el que lo importante es el conocimiento útil que seamos capaces de descubrir a partir de los datos.

En realidad, los términos MD y KDD son a menudo confundidos como sinónimos. En general se acepta que la MD es un paso particular en el proceso consistiendo en la aplicación de algoritmos específicos para extraer patrones (modelos) de los datos. Otros pasos en el proceso KDD, son la preparación de los datos, la selección y limpieza de estos, la incorporación de conocimiento previo, y la propia interpretación de los resultados de minería. Estos pasos aplicados de una manera iterativa e interactiva aseguran que un conocimiento útil se extraiga de los datos”.

Como indica (Riquelme Santos et al., 2006), en la era digital, existen numerosos tipos de datos que pueden ser procesados, almacenados y distribuidos. Con el progreso en la informática y las tecnologías se hace posible contar con una gran base de distintos tipos de datos como imágenes, textos, video y otros.

La minería de datos intenta aprovechar esta explosión de información para descubrir conocimiento, creando un proceso que busca descubrir patrones y tendencias.

Las técnicas de la minería de datos se basan en modelos matemáticos, la inteligencia artificial y de la estadística. Estas técnicas se aplican en un conjunto de datos para obtener unos resultados.

Según (Lara, Zepahua, Rodríguez-Mazahua, Alor-Hernández, & Contreras, 2016), se define a la minería de datos como una herramienta de generación de conocimiento basada en grandes cantidades de datos. Así mismo, se incluyen diferentes fuentes de información con formatos o tipos de datos distintos como texto, XML, hojas de cálculo, entre otros.

2.5.6.1 PROCESO KDD

De acuerdo a (Lara et al., 2016), *“El proceso KDD (Knowledge Discovery from Data, Descubrimiento de Conocimiento a partir de los Datos), es el proceso de extracción no trivial de información, previamente desconocida y potencialmente útil, a partir de los datos. El proceso KDD implica una secuencia iterativa de una serie de pasos”.*

Figura 5. Secuencia pasos en el proceso KDD

Fuente: (Lara et al., 2016)

2.5.6.2 TÉCNICAS DE MINERÍA DE DATOS

De acuerdo con el artículo de (Riquelme Santos et al., 2006), se pueden identificar dos pasos importantes para la Minería de Datos: la elección del modelo y el ajuste a los datos.

Para elegir el modelo adecuado es necesario verificar el tipo de dato y el objetivo a conseguir. Existen diversos modelos de acuerdo con los objetivos establecidos, tales como árboles de decisión, redes neuronales, clustering, entre otros.

2.5.6.3 TIPOS DE ANÁLISIS DE DATOS

(Mahmoud, 2017) menciona también que el análisis de datos se divide en 4 tipos, los cuales son:

- Análisis Descriptivo: El cual describe la situación actual y responde a la pregunta: ¿Qué está pasando ahora?
- Análisis Diagnóstico: El cual describe o responde a la pregunta ¿Por qué está pasando?
- Análisis Predictivo: El cual responde a la pregunta ¿Qué podrá pasar en el futuro?
- Análisis Prescriptivo: El cual responde a la pregunta ¿Cuál es la correcta decisión o solución?

2.5.7 ANÁLISIS PREDICTIVO

De acuerdo con el artículo de (Mahmoud, 2017), el análisis predictivo es utilizado generalmente para predecir el futuro mediante el análisis de la información del pasado y del estudio de la data histórica para descubrir las relaciones y los patrones en dicha información. El beneficio de este análisis predictivo difiere de empresa a empresa, pero se puede mencionar entre los principales beneficios la prevención del riesgo, reducir tiempos y costos, así como la mejor administración de los recursos, todo esto permitirá que la organización esté en la posibilidad de tomar mejores decisiones basadas en hechos y no solo en la intuición.

Una de las principales tareas dentro del análisis predictivo es la correcta elección del modelo a utilizar, la selección incorrecta de las variables y algoritmos pueden afectar a los resultados finales.

- El proceso del análisis predictivo se divide en cinco fases:
- La identificación del problema
- La recolección y preparación de la data
- El análisis de la data y el desarrollo del modelo
- El desarrollo o ejecución
- La observación y el control del modelo predictivo.

De acuerdo a (Rantzer, 2016), el proceso de análisis predictivo generalmente requiere la ejecución de 3 fases para tener éxito:

2.5.7.1 CONOCIMIENTO DEL NEGOCIO

Siendo la parte más importante dentro del proyecto del análisis predictivo, se puede utilizar para el preprocesamiento y la selección de las características a considerar, puede inclusive decidir si el proyecto es viable o no.

El conocimiento del negocio permite determinar la forma de crear valor para el negocio a fin de satisfacer alguna necesidad de esta.

2.5.7.2 DATA

Los datos son vitales para el proceso de aprendizaje automático, que es la base del análisis predictivo. Sin embargo, será necesario considerar algunas características antes del inicio del proyecto, como son:

- Las fuentes de datos se distribuyen o provienen de múltiples fuentes, siendo difíciles de recopilar.
- Los datos pueden obtenerse en cooperación con otros actores, lo cual puede no garantizar el acceso a los mismos.
- La calidad de los datos es muy importante, por lo que la misma puede comprometer severamente el desempeño del proyecto.

2.5.7.3 MÉTODOS

Se necesita la habilidad y tecnología para implementar los algoritmos requeridos y poder crear valor a partir de los datos.

De lo descrito anteriormente, podemos concluir que, el análisis predictivo será de gran beneficio en una empresa para poder identificar los patrones de comportamiento y utilizar esta información para tomar acciones que contribuyan a los objetivos de la organización. Lo más importante es conocer lo que se quiere obtener y no solo obtener unos datos y tomar decisiones en función a ellos. Es la evolución del análisis de los datos para el negocio, enfocado u orientado a los objetivos estratégicos.

2.5.7.4 MODELOS Y ALGORITMOS DEL ANÁLISIS PREDICTIVO

(Mahmoud, 2017) indica que la selección de los modelos predictivos permite descubrir los patrones entre las variables dependientes y las variables explicativas para predecir un resultado.

Para los modelos se pueden utilizar varios algoritmos, entre los cuales se pueden mencionar a:

2.5.7.4.1 CLASIFICACIÓN

Basado en una o varias variables decisivas o numéricas llamadas atributos o predictores, los cuales permiten predecir una variable final que representa a la clase u objetivo definidos

2.5.7.4.2 CLUSTERING (APRENDIZAJE NO SUPERVISADO):

Basado en agrupaciones, cada agrupación contiene datos similares. Este proceso ayuda a identificar relaciones que no eran conocidas dentro de los datos procesados.

2.5.7.4.3 REGLAS DE ASOCIACIÓN

Permite encontrar asociaciones importantes en los datos, identificando los conjuntos de elementos con mayor afinidad. Utilizado por ejemplo para identificar el cliente que, al comprar determinado producto, es más menos probable que compre otro producto o conjunto de productos.

2.5.7.4.4 REGRESIÓN

Predice el valor del objetivo numérico mediante la construcción de un modelo basado en uno o más variables predictivas.

(Mahmoud, 2017) grafica también los modelos predictivos agrupándolos en familias de algoritmos y las diferentes formas de medir los errores, tal como se muestra en la figura siguiente:

Figura 6. Familias de algoritmos utilizados en el modelado predictivo

Fuente: (Mahmoud, 2017)

2.5.7.4.5 MATRIZ DE CONFUSIÓN

Según (Ariza-López, Rodríguez-Avi, & Alba-Fernández, 2018): “Una matriz de confusión o matriz de error, es una tabla de contingencia que sirve como herramienta estadística para el análisis de observaciones emparejadas.

El contenido de una matriz de confusión es un conjunto de valores que contabilizan el grado de semejanza entre observaciones emparejadas: un conjunto de datos bajo control (CDC) y un conjunto de datos de referencia (CDR), para los que se ha establecido una

clasificación. Usualmente el CDR es la verdad terreno, es decir, la realidad, y suele conocerse por medio de un muestreo”.

2.5.7.5 PRINCIPALES HERRAMIENTAS DE PREDICCIÓN

De acuerdo con el artículo de (Mohmmad, Butt, & Zaman, 2017), las herramientas de análisis predictivo se dividen en 2 grupos y describe también las principales herramientas utilizadas en la actualidad

2.5.7.5.1 HERRAMIENTAS DE CÓDIGO ABIERTO

Entre los que se pueden mencionar los siguientes:

- Apache Mahout
- GNU Octave
- KNIME
- OpenNN
- Orange
- R
- Scikit-learn
- WEKA
- RapidMiner
- Tanagra
- Data Science Studio (DSS)
- H2O
- Lavastorm Public Edition
- LIBLINER

2.5.7.5.2 HERRAMIENTAS COMERCIALES DISPONIBLES

Entre los que se pueden mencionar los siguientes:

- Alpine Data Labs
- Alteryx
- Angoss Knowledge STUDIO
- BIRT Analytics
- MATLAB
- IBM SPSS Statistics and IBM SPSS Modeler

- KXEN Modeler
- Mathematica
- Minitab
- Neural Designer
- Oracle Data Mining (ODM)
- Pervasive
- Predixion Software
- RCASE
- Revolution Analytics
- SAS and SAS Enterprise Miner
- STATA
- Statgraphics
- STATISTICA
- TeleRetail
- TIBCO

Entre las principales herramientas actualmente utilizadas se pueden mencionar a:

- **R:** Herramienta de código abierto utilizada especialmente por estadistas y especialistas en minería de datos para el desarrollo de software estadístico. Proporciona una amplia gama de métodos y técnicas para medios estadísticos, modelos predictivos, análisis de datos y visualización. Muchos de sus métodos se pueden implementar fácilmente en C, C++ y Fortran, los cuales se conectarán luego a R. Es software libre licenciado bajo GNU general. En general, R es un paquete de estadísticas bien soportado, de código abierto, dirigido por la línea de comandos, utilizado por los analistas estadísticos y predictivos.
- **WEKA:** (Waikato Environment for Knowledge Analysis) es un software gratuito con licencia bajo la Licencia Pública General de GNU, de aprendizaje automático popular, desarrollado en la universidad de Waikato Nueva Zelanda. Es una herramienta de análisis predictivo de código abierto basada en Java, conformado por muchos algoritmos de extracción de datos y aprendizaje automático, se considera también el preprocesamiento de datos, la clasificación, agrupación y la extracción de reglas de asociación. Esta herramienta provee tres interfaces gráficas para el usuario:

- Explorador, para el análisis de datos exploratorios que permiten soportar el preprocesamiento, selección de atributos, aprendizaje y visualización
- Experimentador, para un entorno experimental que permite probar y evaluar algoritmos de aprendizaje automático
- Flujo de conocimiento, para nuevos procesos
- **RAPIDMINER:** Es una herramienta de minería de datos basada en GUI de código abierto. Proporciona un entorno integrado para los procesos de minería de datos, minería de texto, análisis de negocio, análisis predictivo y aprendizaje automático. Este software no depende de la plataforma, lo que contribuye a que se pueda instalar en cualquier sistema operativo. La licencia es propiedad de AGPL y se puede descargar desde www.rapidminer.com. Esta herramienta contiene más de 100 esquemas de aprendizaje para la clasificación de regresión y análisis de agrupamiento y soporta alrededor de veintidós formatos de archivos. Implementa también diversos procedimientos para la creación y evaluación de modelos. Trabaja bien con diferentes bases de datos y archivos Excel, lo cual lo hace popular entre los analistas de datos, mineros de datos y estadísticos.
- **KNIME:** Konstanz Information Miner es una herramienta de código abierto escrita en java y basada en plataforma Eclipse. Está licenciado por la licencia pública general de GNU y es compatible con varios sistemas operativos como Linux, OS X y Windows. Esta herramienta se utiliza especialmente para investigación farmacéutica, informes empresariales, inteligencia empresarial, minería de datos y análisis predictivo.

Cuenta con una plataforma modular que le permite al usuario crear flujos de manera visual, ejecutar tareas de análisis y validar los resultados obtenidos.

Una diferencia con respecto a las otras herramientas es que puede interactuar con programas de visualización.
- **ORANGE:** Herramienta de código abierto y aprendizaje automático para principiantes y expertos. Software basado en GUI multiplataforma, especializado para la visualización de datos y minería de datos. Licenciado por GNU (General Public License), bajo código C/ C++ y Python.

Este software permite el procesamiento de datos, así como la creación de modelos que utilizan diversas técnicas de exploración. Permite una programación visual de

front-end para el análisis y visualización de datos y una secuencia de comandos para el back-end.

2.5.8 BENCHMARKING DE TECNOLOGÍAS BUSINESS ANALYTICS

Según (Lara et al., 2016), se realizan las siguientes evaluaciones para Minería de datos, presentamos a continuación un cuadro resumen de las principales herramientas y conceptos denominado Evaluación de Tecnologías BA

- **Herramientas:** entre las principales herramientas de acceso público se mencionan a KNIME (Konstanz Information Miner), Orange, RapidMiner y WEKA (Entorno para Análisis del Conocimiento de la Universidad de Waikato).

De la comparación realizada con el fin de determinar la mejor herramienta y técnica para el análisis de los datos se determinó que no existe una herramienta que sea mejor que otra y que se aplique a cualquier tipo de clasificación. De manera individual, la herramienta Weka obtuvo mejores resultados, así como la técnica de árboles de decisión.

- **Técnicas de clasificación:** Se propuso una técnica de clasificación denominada Neuro-fuzzy para Minería de Datos, la cual, según las pruebas realizadas obtuvo mejores resultados que métodos de clasificación convencionales como Función de Base Radial de Redes Neuronales (RBFNN) y Sistema de Inferencia Adaptativa Neuro-difusa (ANFIS).

Para la evaluación se tomaron en cuenta la precisión, falsos positivos, falsos negativos, error medio y estadística kappa.

- **Modelos de Clasificación:** Se propuso la evaluación de dos modelos de clasificación más utilizados por la minería de datos. Los algoritmos evaluados utilizando la herramienta WEKA fueron MLPNN (Multilayer Perception Neural Network) y NB (Naïve Bayes). Los resultados obtenidos determinaron que el algoritmo MLPNN obtuvo la mejor tasa de precisión.
- **Estimación de costos:** Se realizó una evaluación de técnicas de análisis de datos para determinar el costo de estas. las técnicas evaluadas fueron LR (Linear Regression), ANN (Artificial Neural Network), SRV (Support Regression Vector) y KNN (K-Nearest Neighbors). Los resultados demostraron que los de mayor costo

tenían mejores resultados en velocidad y precisión, con lo cual se establece la importancia de la técnica a utilizar para alcanzar el objetivo deseado.

Tabla 4
Principales Herramientas Predictivas

Evaluación	Opciones	Conclusiones
Herramientas	- KNIME - Orange - RapidMiner - WEKA	Se deduce que no hay una sola herramienta o técnica mejor que otra.
Algoritmos de clasificación	- RBFNN (Función de Base Radial de Redes Neuronales) - NFS (Sistema de Inferencia Adaptativa Neuro-difusa)	Se evaluó la precisión, el error, entre otros, siendo que NFS obtuvo mejores resultados.
Modelos de Clasificación	- MLPNN (Multilayer Perception) - NB (Naive Bayes)	Se evaluaron utilizando la herramienta WEKA, siendo que MLPNN obtuvo mejores resultados.
Técnicas de minería vs costo	- LR (Linear Regression) - ANN (Artificial Neural Network) - SRV (Support Regression Vector) - KNN (K-Nearest Neighbors)	El costo del software tuvo mayor velocidad y precisión.

Fuente: (Lara et al., 2016)

2.5.9 METODOLOGÍA DE IMPLEMENTACIÓN DE BUSINESS ANALYTICS

Existen diversas metodologías para aplicar Business Analytics. A continuación, detallaremos 3 de las más conocidas en los últimos años.

2.5.9.1 BUSINESS ACTIVITY MONITORING (BAM)

Según (Hindle & Vidgen, 2018), BAM es una metodología de analítica de negocios cuyo propósito en las organizaciones es agregar valor en el negocio. Está compuesta por cuatro fases: Situación problemática, Mapeo de asignación de negocio, Análisis e implantación de analítica. A continuación, describiremos cada fase:

Fase01: Estructuración de la situación problemática

En esta fase se identifica la problemática del negocio, se estudia cómo funciona el modelo de negocio de manera integral.

Fase02: Modelo de asignación de negocio

El objetivo de esta fase es describir el modelo de negocio actual, para lo cual podemos utilizar herramientas estratégicas como Business Model Canvas (BMC), el mismo que está compuesto por nueve componentes básicos:

- Las necesidades de segmentos de clientes.

- Propuestas de valor de una organización.
- Entrega de propuesta de valor a través de canales.
- Relaciones con los clientes.
- Actividades claves.
- Recursos claves.
- Socios claves.
- Estructura de costos.
- Estructura de ingresos.

Fase03: Análisis

Luego de haber realizado el modelo de negocio con la situación actual, podemos identificar puntos de influencia para el análisis del negocio que son más propensos a conducir la creación de valor y el mejor uso de los recursos.

Fase04: Implantación de la analítica

Para la implementación de la analítica se debe considerar 3 procesos:

- Recopilación de datos

Los datos existentes se recogen, revisan y se evalúan, verificando la calidad de los datos.

- Construcción

Los datos internos son mejorados y combinados con fuentes de datos externos.

- Despliegue.

En esta fase se realiza el despliegue de los modelos predictivos que son construidos para apoyar la toma de decisiones.

2.5.9.2 CROSS INDUSTRY STANDARD PROCESS FOR DATA MINING (CRISP-DM)

Según (Shafique & Qaiser, 2014), la metodología CRISP-DM nos proporciona un marco de trabajo estándar y lineamientos para proyectos de minería de datos.

De acuerdo a (Parra, Galeano, & Rodriguez, 2010), el ciclo de vida de un proyecto se realiza en 6 fases, los cuales son:

Fase01: Comprensión del negocio

Dentro de esta fase se busca entender el negocio, los requerimientos y objetivos principales para la elaboración del proyecto, considerando las siguientes categorías:

- Definir los objetivos y criterios de éxito del negocio.
- Evaluación de la situación actual considerando recursos, requisitos, supuestos, riesgos, contingencias, costos y beneficios.
- Definición de los objetivos a nivel de minería de datos.
- Elaboración del plan del proyecto en el que detallaremos las fases, actividades y tareas para el cumplimiento de los objetivos planteados.

Fase02: Comprensión de los datos

En esta fase trabajaremos la calidad de los datos, para lo cual debemos tener en cuenta las siguientes categorías:

- Recopilación inicial de datos.
- Proporcionar una descripción de los datos.
- Exploración o análisis de los datos.
- Verificación de calidad de los datos.

Fase03: Preparación de los datos

El objetivo de esta fase es analizar los datos que se consideran importantes en el proceso de selección, depuración y transformación. Las categorías para esta son:

- Selección de los datos relevantes.
- Depuración de los datos no relevantes.
- Estructuración de los datos y sus atributos.
- Integración de los datos estructurados y no estructurados.
- Formateo de datos generando un informe de calidad.

Fase04: Modelado

El objetivo de esta fase será el de disponer un modelo que nos permita alcanzar los objetivos del proyecto y del negocio, definidos en la primera fase.

Las categorías para esta fase son:

- Selección del modelo y supuestos.
- Elaborar el plan de pruebas.
- Construcción del modelo considerando los parámetros escogidos, modelos, descripción de los modelos.
- Evaluación del modelo

Fase05: Evaluación

Esta fase se centra en la evaluación de los modelos obtenidos y de ser necesario realizaremos un feedback por cada fase. Las categorías para esta fase son:

- Evaluación de resultados
- Revisión del proceso.
- Determinación de los pasos o acciones a seguir.

Fase06: Despliegue

En esta fase se implantará el modelo seleccionado, los mismos que serán presentados a los usuarios. Las categorías para esta fase son:

- Plan de implementación.
- Plan de monitoreo y mantenimiento.
- Presentación del informe final.
- Revisión del proyecto y registro de lecciones aprendidas.

En la figura 7 podemos visualizar las fases de la metodología.

Figura 7. Fases de la Metodología CRISP-DM

Fuente: (Parra et al., 2010)

2.5.9.3 SAMPLE, EXPLORE, MODIFY, MODEL AND ASSESS (SEMMA)

Según (Shafique & Qaiser, 2014) señala que SEMMA es una metodología para minería de datos el cual fue desarrollado por SAS Institute. Permite la comprensión, organización, desarrollo y mantenimiento de los proyectos de minería de datos. El ciclo de vida se define en 5 fases, las cuales son:

Fase 01: Muestreo

En esta fase se considera la entrada de datos. El proceso se inicia con la extracción de la población muestral sobre la que se va a aplicar el análisis para extraer información significativa.

Fase 02: Exploración

El objetivo de esta fase es explorar los datos para comprender e identificar tendencias y anomalías.

Fase 03: Manipulación

Esta fase se centra en la modificación de los datos mediante la creación, selección y transformación de las variables.

Fase 04: Modelado

En esta fase de modelado podemos hacer uso de técnicas para el modelado de datos.

Fase 05: Valoración

Finalmente, en esta fase podemos realizar actividades como evaluación, medidas y reportes.

En la figura 8 podemos visualizar las fases de la metodología SEMMA.

Figura 8. Fases de la Metodología SEMMA

Fuente: Elaboración propia

3 SITUACIÓN PROBLEMÁTICA

3.1 OBJETO DE ESTUDIO: LA INSTITUCIÓN

La Ley N° 29792, Ley de creación, organización y funciones del Ministerio de Desarrollo e Inclusión Social (MIDIS), determina su naturaleza jurídica, como un organismo del poder ejecutivo, cuyo objetivo principal es mejorar la calidad de vida de la población que se encuentra en situación de pobreza y vulnerabilidad, promoviendo el acceso a servicios públicos y a oportunidades y al desarrollo de sus propias capacidades.

El MIDIS es el organismo rector de las políticas nacionales que promueven el desarrollo y la inclusión social, con énfasis en la reducción de la desnutrición crónica infantil y la promoción del desarrollo infantil temprano. Coordina y articula con las entidades del sector público de los tres niveles de gobierno; así como con organizaciones del sector privado y de la sociedad civil, vinculados con el cumplimiento de las políticas nacionales en materia de promoción del desarrollo e inclusión social, con el fin de concertar acciones conjuntas, hacer seguimiento al cumplimiento de los acuerdos, evaluar los resultados e impactos que se generan y provocar aprendizajes conjuntos de manera de marcar el rumbo del Estado hacia una lucha eficaz contra la pobreza y quebrar la transmisión intergeneracional de la pobreza.

La Ley N° 29951, Ley de Presupuesto del Sector Público 2013, dispone que el MIDIS tiene a cargo la operación del Sistema de Focalización de Hogares, correspondiéndole generar y administrar la información del Padrón General de Hogares bajo estándares de calidad, seguridad y confidencialidad, así como certificar la clasificación socioeconómica de los potenciales usuarios de los programas sociales y de subsidios del Estado que se ejecuten bajo criterios de focalización individual.

El MIDIS tiene a su cargo siete programas sociales:

- Programa de Cooperación para el Desarrollo Social (Foncodes).
- Programa Nacional de Apoyo Directo a los más Pobres (Juntos).
- Programa Nacional de Asistencia Solidaria (Pensión 65).
- Programa Nacional de Alimentación Escolar (Qali Warma).
- Programa Nacional Cuna Más (Cuna Más).
- Programa Nacional Plataformas de Acción para la Inclusión Social (País).
- Programa Nacional de Entrega de Pensión no Contributiva a Personas con Discapacidad Severa en Situación de Pobreza (Contigo).

3.1.1 MISIÓN

Garantizar que las políticas y programas sociales de los diferentes sectores y niveles de gobierno actúen de manera coordinada y articulada para cerrar las brechas de acceso a servicios públicos universales de calidad y de acceso a las oportunidades que abre el crecimiento económico.

3.1.2 VISIÓN

El Perú ha erradicado la pobreza extrema y la desnutrición crónica infantil, y las niñas y niños de todo el territorio nacional reciben servicios universales de salud y educación de calidad.

3.1.3 OBJETIVOS ESTRATÉGICOS

De acuerdo con el plan estratégico institucional vigente, los objetivos estratégicos generales del MIDIS son:

- Mejorar las condiciones de vida de la población en proceso de inclusión.
- Reducir la prevalencia de la desnutrición crónica infantil en niños menores de 3 años.

- Incrementar los niveles de desarrollo cognitivo, social, físico y emocional en la primera infancia.
- Incrementar competencias para el desarrollo personal educativo y ocupacional de niños, niñas y adolescentes de acuerdo a la edad.
- Incrementar el ingreso autónomo de los hogares.
- Incrementar la protección y bienestar del adulto mayor.
- Ejecutar los programas del MIDIS con efectividad, articuladamente y con calidad.
- Ejercer la rectoría de la política nacional y sectorial de desarrollo e inclusión social.
- Mejorar la capacidad de gestión para la implementación de la política y la ejecución de los programas MIDIS.

Los objetivos del 2 al 6 están vinculados a las funciones de los programas del MIDIS y contribuyen a lograr el objetivo 1. Los objetivos 7 al 9 son la base para lograr los demás objetivos. La relación de estos objetivos es mostrada en la figura 9.

Figura 9. Mapa de objetivos estratégicos del MIDIS

Fuente: Página web del MIDIS, 2013

3.1.4 ANALISIS FODA

De acuerdo a la Resolución Ministerial N° 260-2015-MIDIS que aprueba el Plan Operativo Institucional 2015, el análisis FODA es el siguiente:

Tabla 5
Análisis FODA MIDIS- FORTALEZAS

Nro	FORTALEZAS DEL MIDIS
1	Alta Dirección del MIDIS con alto nivel de preparación y liderazgo
2	MIDIS posee la rectoría de las políticas de desarrollo e inclusión social y el rol articulador en esa materia, según se le ha asignado en el marco legal vigente
3	Empoderamiento del rol articulador del MIDIS
4	Personal competente y comprometido con la institución
5	Los programas del MIDIS están presentes a nivel nacional y generan alto impacto
6	Enfoque de diseño y gestión de políticas basado en evidencia brinda mayor solidez a las intervenciones
7	Los programas del MIDIS cuentan con capacidad de articulación intersectorial e intergubernamental
8	Los programas del MIDIS cuentan con instrumentos técnicos de gestión que ayudan al logro de resultados orientados al cierre de brechas
9	Existe conocimiento y aprendizajes acumulados en la implementación de los programas del MIDIS

Fuente: Página web del MIDIS, 2018

Tabla 6
Análisis FODA MIDIS- OPORTUNIDADES

Nro	OPORTUNIDADES DEL MIDIS
1	Alineamiento de las Prioridades Nacionales de la Política Social 2016-2021 con la Agenda 2030 de Desarrollo Sostenible
2	Política General de Gobierno contiene un eje de inversión social que prevé el fortalecimiento de los programas del MIDIS
3	Impulso a la simplificación administrativa en la gestión desde el ente rector en materia de modernización
4	Red social conformada por organizaciones sociales de base, dirigencias vecinales, comunidades campesinas, grupos étnicos y otros grupos sociales con interés en participar en acciones que favorezcan procesos de inclusión social
5	La tecnología de la información y comunicación aplicada a la gestión social gubernamental facilita los procesos de inclusión al posibilitar la interconexión de los sistemas de información del Estado
6	Proceso de articulación ya iniciado en los sectores y en los tres niveles de gobierno
7	Alta incidencia de la pobreza en personas de todos los rangos etarios que habitan en zonas rurales, hablan una lengua nativa o han estudiado como máximo hasta el nivel primario demanda soluciones urgentes y especializadas
8	Algunas empresas que operan en zonas de pobreza, disponen de recursos y muestran disposición para contribuir con el Estado en materia de Desarrollo e Inclusión Social
9	El fortalecimiento de las relaciones bilaterales y la existencia de programas de cooperación internacional en temas de desarrollo e inclusión social brindan la posibilidad de trabajar conjuntamente y compartir experiencias y aprendizajes
10	Los gobiernos regionales y locales tienen un marco normativo que les asigna la función de ejercer liderazgo en temas de desarrollo e inclusión social en su ámbito de influencia territorial
11	Algunas organizaciones sin fines de lucro han desarrollado capacidades de gestión de proyectos sociales y han sistematizado lecciones aprendidas

Fuente: Página web del MIDIS, 2018

Tabla 7
Análisis FODA MIDIS- DEBILIDADES

Nro	DEBILIDADES DEL MIDIS
1	Limitaciones en el proceso de focalización generan filtraciones y subcoberturas
2	Ausencia de un sistema de información integrado de los programas sociales afecta su ejecución, monitoreo y evaluación
3	Ausencia de una estrategia de comunicación en medios externos
4	El Reglamento del Sistema Nacional de Desarrollo e Inclusión Social (SINADIS) aún no está implementado
5	Las estructuras de los programas del MIDIS a nivel regional y local no se han integrado
6	Bajo nivel de alineamiento y coordinación entre los programas del MIDIS en temas como articulación territorial, información y comunicación con usuarios y otros
7	Incertidumbre en la continuidad laboral limita el desarrollo y compromiso de los trabajadores con el MIDIS

Fuente: Página web del MIDIS, 2018

Tabla 8
Análisis FODA MIDIS- AMENAZAS

Nro	AMENAZAS DEL MIDIS
1	Vulnerabilidad del sector de Desarrollo e Inclusión Social frente a factores políticos podría afectar o discontinuar los avances de la actual política
2	La corrupción es un problema latente en el país que afecta el desarrollo de los programas sociales y de las intervenciones en materia de Desarrollo e Inclusión Social en todos los sectores y niveles de gobierno
3	Los desastres naturales podrían incidir en la continuidad operativa de los programas del MIDIS
4	Impacto del medio ambiente en zonas de influencia podría afectar la ejecución de los programas del MIDIS
5	Balanza fiscal deficitaria podría implicar una menor asignación de recursos a programas del MIDIS
6	Desconocimiento de los criterios de focalización y priorización de programas del MIDIS por parte de la población
7	Presión mediática respecto de los avances en materia de desarrollo e inclusión social
8	Potenciales conflictos de autoridad con otros sectores rectores de temas relacionados y resistencia al cambio en algunos gobiernos regionales y locales podrían afectar el trabajo articulado
9	Los gobiernos regionales y locales tienen déficit de competencias para la priorización y gestión de proyectos de desarrollo e inclusión social, especialmente en zonas alejadas o de alto riesgo
10	La estrategia social y de gestión de conflictos y la estructura orgánica de algunos gobiernos y locales están desalineadas y requieren fortalecerse para asumir su rol de liderazgo en el desarrollo sostenible de sus territorios bajo los lineamientos del MIDIS
11	La administración pública no ha incorporado la innovación como fuente de mejoras radicales que faciliten la adopción del enfoque territorial en la gestión de los programas y servicios sociales, contemplando la diversidad socio-económica y cultural de los usuarios
12	Algunas intervenciones tienen un enfoque que promueve, en algunos segmentos de la población objetivo, la idea de que es un derecho recibir programas asistencialistas
13	Falta de cobertura total por parte del RENIEC en las zonas rurales alejadas, dificulta el desarrollo de algunos programas
14	El sistema estadístico nacional no cuenta con niveles de desagregación de la información suficiente y de calidad para las zonas rurales priorizadas para el sector Desarrollo e Inclusión Social
15	Asignación presupuestal inercial y desconectada de una efectiva gestión por resultados, de metas de mediano y largo plazo y de prioridades territoriales que se agrava con remesas presupuestales que llegan a las regiones a destiempo
16	El sistema de planeamiento gubernamental es aún incipiente y no cuenta con un proceso de seguimiento y evaluación de los planes en los diversos niveles de gobierno para promover mejoras
17	Los sectores sociales, en los tres niveles de gobierno, y el sistema de gestión de recursos humanos no han establecido estrategias exitosas para asegurar la cobertura de los puestos y la idoneidad del personal (médicos, maestros, especialistas en proyectos de desarrollo e inclusión social, etc.) para atender las necesidades diferenciadas de las zonas priorizadas
18	La mayoría de los distritos rurales en situación de pobreza presentan dispersión poblacional, geografía accidentada y bajo acceso a servicios públicos de calidad

Fuente: Página web del MIDIS, 2018

3.2 MARCO REGULATORIO DEL PROCESO DE FOCALIZACIÓN

- Ley N° 30435 - Ley que crea el Sistema Nacional de Focalización (SINAFO)
Ley que crea el Sistema Nacional de Focalización (SINAFO) y establece los principios, alcance, organización, ámbito, procesos, procedimientos, infracciones y sanciones que lo regulan; y especifica los integrantes del sistema y sus funciones en concordancia con la política social y en articulación con el Sistema Nacional de Desarrollo e Inclusión Social (SINADIS).
- Directiva N° 004-2017-MIDIS, establece lineamientos, estructura, contenidos, adecuación y fases del proceso de focalización de las intervenciones públicas focalizadas
Establece las disposiciones que regulan la elaboración y/o actualización de la lista de intervenciones públicas focalizadas, con el fin de contribuir a la asignación eficiente de los recursos públicos asociados a la Política Social y contribuir al cierre de brechas relativas a los problemas y carencias que las intervenciones buscan resolver.
- Directiva N° 006-2017-MIDIS, regula la operatividad del Sistema de Focalización de Hogares - SISFOH
Establecer las disposiciones para la operatividad del SINAFO, con el fin de que los hogares cuenten con Clasificación Socioeconómica (CSE) en el Padrón General de Hogares (PGH) y se certifique ante las entidades a cargo de las Intervenciones Públicas Focalizadas.
- Decreto Legislativo N° 1246, Decreto Legislativo que aprueba diversas medidas de simplificación administrativa
En el Artículo 3 refiere a la implementación progresiva de la interoperabilidad en beneficio del ciudadano. Esto significa que las entidades de la Administración Pública que posean y administren información, deben ponerla a disposición de manera gratuita y permanente a las entidades del Poder Ejecutivo para la interoperabilidad.

3.3 SITUACIÓN DE LA POBREZA EN EL PERÚ

3.3.1 SITUACIÓN DE LA POBREZA EN AMÉRICA LATINA

Según la Comisión Económica para América Latina y el Caribe (CEPAL): *“De acuerdo con la información más reciente, en 2016 el número de personas pobres en América Latina llegó a 186 millones, es decir, el 30,7% de la población, mientras que la pobreza*

extrema afectó al 10% de la población, cifra equivalente a 61 millones de personas”. A continuación, podemos observar en la figura 10, las estadísticas de los últimos años, donde se evidencia que desde el año 2002, el porcentaje de pobreza y pobreza extrema fueron disminuyendo hasta el año 2015 donde inicia una nueva tendencia de incremento de estos valores.

Figura 10. Porcentaje de pobreza y pobreza extrema

Fuente: CEPAL 2002-2017

De acuerdo con la figura anterior, podemos observar que, en el año 2017, el porcentaje de pobreza se encuentra en el rango del 30.7%, lo cual representa que 3 personas de cada 10, no tienen lo suficiente para cubrir sus necesidades básicas.

3.3.2 SITUACIÓN DE PERÚ Y OTROS PAÍSES DE AMÉRICA LATINA

Las cifras de la situación de pobreza correspondientes a los últimos años por cada país de América Latina se visualizan en la tabla 9.

Tabla 9

% de Personas en situación de pobreza y pobreza extrema 2012 – 2016

País	Pobreza				Pobreza extrema			
	2012	2014	2015	2016	2012	2014	2015	2016
Argentina	...	...	...	30,3	...	...	...	6,1
Bolivia (Estado Plurinacional de)	43,3	39,2	38,6	...	21,6	17,2	16,8	...
Brasil ^b	15,9	13,3	...	...	5,3	4,2	...	...
Chile	22,2	14,4	11,7	...	8,1	4,5	3,5	...
Colombia	32,7	28,5	27,8	28,0	10,4	8,1	7,9	8,5
Costa Rica ^c	20,6	22,4	21,7	20,5	6,3	6,7	7,2	6,3
Ecuador	27,3	22,5	23,3	22,9	11,2	7,7	8,5	8,7
El Salvador ^c	34,5	31,8	34,9	32,7	8,9	7,6	8,1	7,9
Guatemala	...	59,3	...	...	...	23,4	...	...
Honduras	71,1	68,2	68,7	65,7	50,9	44,6	44,7	42,5
México ^d	51,6	53,2	...	50,6	20,0	20,6	...	17,5
Nicaragua	...	29,6	...	...	...	8,3	...	...
Panamá	26,5	25,8	23,0	22,1	11,1	11,0	10,3	9,9
Paraguay	31,4	27,2	26,6	28,9	7,4	5,5	5,4	5,7
Perú	25,8	22,7	21,8	20,7	6,0	4,3	4,1	3,8
República Dominicana	42,2	36,4	31,5	30,0	11,1	8,1	6,8	6,1
Uruguay	12,4	9,7	9,7	9,4	0,5	0,3	0,3	0,3
Venezuela (República Bolivariana de) ^e	21,2	32,6	...	...	6,0	9,5	...	...

Fuente: CEPAL 2012 - 2016

De las cifras mostradas concluimos que Perú es uno de los países con menor índice de pobreza después de Uruguay y Costa Rica.

3.3.3 VARIACIÓN DE LA POBREZA EN EL PERÚ

En los últimos años, según cifras oficiales del INEI, la pobreza del Perú ha presentado una primera etapa de reducción, desde el año 2006 al año 2016, en la cual el índice de pobreza se redujo de 35.4% a 17% mientras que la pobreza extrema tuvo una disminución de 13.8% a 3.8%. A continuación, se puede observar esta tendencia en la figura 11.

Figura 11. Evolución de la distribución de la población (pobreza extrema, pobreza)

Fuente: Comisión Consultiva para la estimación de la pobreza del INEI 2004 - 2016

Sin embargo, en el último año de análisis (2017), el índice de pobreza tuvo un incremento en promedio del 1%. Como se puede observar en la figura 12.

Este 1% de incremento es equivalente a 375 mil personas que durante el año 2017 pasaron a condición de pobreza.

Figura 12. Evolución de la distribución de la población (pobreza extrema, pobreza)

Fuente: INEI 2007-2017

Observamos que, según las cifras recientes, a pesar de la reducción de la pobreza en la última década, 2 de cada 10 peruanos no tiene lo suficiente para cubrir sus necesidades básicas.

3.3.3.1 DISTRIBUCIÓN GEOGRÁFICA DE LA POBREZA EN PERÚ

Analizando cómo se encuentra distribuida la pobreza dentro del territorio peruano, se puede observar también que existen departamentos donde se concentra un mayor porcentaje de pobreza y pobreza extrema. Esto quiere decir, que hay condiciones especiales para los grupos de población que están afectando su condición de pobreza.

A continuación, se presenta la figura 13, donde el índice de pobreza supera el 30% en las zonas rurales y en determinados departamentos como Cajamarca y Huancavelica, el índice supera el 40%.

Figura 13. Distribución de la pobreza a nivel nacional

Fuente: INEI - Encuesta Nacional de Hogares 2016-2017

3.3.3.2 PERFIL DE POBLACIÓN EN POBREZA

A continuación, se presenta el perfil de las personas clasificadas como pobres en relación a una serie de variables o características. Tomaremos una muestra de estas variables para analizar las tendencias. El detalle de las cifras se encontrará en el anexo del documento (informe técnico INEI 2017).

- **Edad:**

La mayor cantidad de pobreza agrupa a la población infantil y adolescente. Se atribuye a este factor las situaciones de riesgo como desnutrición, abandono, poco acceso a servicios médicos, entre otros. Estas condiciones reducen las oportunidades de la población infantil y adolescente al ser difícilmente rebatibles. El INEI señala: *“En el año 2017, la pobreza afectó al 32,7% de niñas y niños menores de cinco años de edad, al 32,5% de los que tienen de 5 a 9 años, al 30,0% de 10 a 14 años y al 22,8% de los adolescentes de 15 a 19 años de edad”*.

- **Lengua Materna:**

El mayor porcentaje de pobreza se concentra en la población cuya lengua materna ha sido una lengua nativa como el quechua, aimara u otra.

Según el INEI: *“la pobreza afectó al 33,0% (6,9% pobre extremo y 26,1% pobre no extremo) de las personas que mencionaron tener como lengua materna una lengua nativa, siendo casi el doble de la incidencia en la población con lengua materna castellano, 18,6% (2,9% pobre extremo y 15,7% pobre no extremo)”*.

- **Región natural:**

La población en condición de pobreza se encuentra distribuida de la siguiente manera: en la sierra se concentra el 46.2%, en la Costa 36.9% y en la Selva el 16.9%. Mientras que la distribución de la pobreza extrema: en la sierra está el 66.7%, en la selva el 21.1% y en la costa el 12.2%.

- **Número de miembros por familia:**

Las familias en situación de pobreza tienen mayor número de niños y niñas en comparación con las familias no pobres, alcanzando un promedio de 4.6 niños para la condición de pobreza y 4.8 niños para la condición de extrema pobreza.

- **Material predominante en paredes exteriores:**

El material predominante de las viviendas de las familias en pobreza es el adobe/quincha con un 42.4%, aunque se puede observar que este porcentaje se ha reducido en 3.6% en el último año.

- **Piso de la vivienda:**

El 58.8% de la población en situación de pobreza presenta como material del piso la tierra. Con respecto del año anterior 2016, este porcentaje se redujo en 3.8%

- **Formas de abastecimiento de agua para consumo humano:**

Solo el 69.9% de la población en situación de pobreza se abastecía de la red pública dentro de la vivienda para el consumo de agua, mientras que un número importante que representa el 11.7% utiliza otras fuentes como ríos, acequias o manantiales.

- **Área de residencia (rural/urbana):**

Se pudo determinar también que la población en situación de pobreza tiene mayor concentración en el área rural, alcanzando el 44.4% de su población.

Se observa también que el último año, la pobreza en áreas urbanas ha tenido un incremento de 1.2%, mientras que en el área rural el incremento fue de 0.6%.

- **Analfabetismo:**

La población analfabeta alcanzó el 12.9% de la población pobre, afectando en mayor medida a la población en extrema pobreza, alcanzando el 19.4%.

Comparando el porcentaje de analfabetismo según el área de residencia, se observa que el 18.4% corresponden al área rural y el 8.5% al área urbana.

3.3.3.3 COBERTURA EN SALUD

Podemos también analizar el impacto que ha tenido el estado en algunos de los servicios básicos de la población.

Como se puede observar en la figura 14, la cobertura de salud por parte del estado se ha ido incrementando en la última década, pasando de 36.2 % a 82.3% para los pobres.

Figura 14. Evolución de población con seguro de Salud

Fuente: INEI 2007 - 2017

En el año 2017, el 82,3% de la población pobre tenía algún tipo de seguro de salud (mayoritariamente SIS), mientras que el 74,8% de los no pobres accedieron a este servicio, siendo 7,5 puntos porcentuales menos que los pobres.

3.4 PROCESO DE DETERMINACIÓN DE LA CLASIFICACIÓN SOCIOECONÓMICA (CSE) ACTUAL

El MIDIS mediante la Dirección General de Focalización tiene a cargo el proceso de la clasificación socioeconómica, que permite identificar la población en situación de pobreza y proporciona esa información a las intervenciones públicas focalizadas para que estas las consideren en su proceso de afiliación. En la figura 15 se muestra el proceso para la determinación de la clasificación socioeconómica.

Figura 15. Proceso de Determinación de Clasificación Socioeconómica

Fuente: Elaboración propia

Como apoyo para el proceso de clasificación socioeconómica, el MIDIS cuenta con el Sistema Integrado de Gestión de Operaciones de Focalización (SIGOF) que contiene módulos para el recojo de información en las municipalidades y el procesamiento para la determinación de la clasificación socioeconómica de hogares. En las municipalidades se cuenta con un módulo cliente/servidor dado que no todas las municipalidades cuentan con internet.

3.4.1 RECOPIACIÓN DE INFORMACIÓN

El proceso de recopilación de información tiene como objetivo obtener los datos provenientes de los instrumentos de recojo de información y fuentes externas.

Este proceso consiste en:

- El ciudadano (nacional o extranjero con residencia en el Perú) se dirige a la Unidad Local de Empadronamiento (ULE) de la municipalidad a la que pertenece la vivienda donde reside y presenta la solicitud de clasificación socioeconómica (S100). El ciudadano deberá mostrar su documento de identidad y recibo de luz y agua de la vivienda.
- La ULE realiza la digitación del S100 en el SIGOF. Este sistema genera un archivo que es enviado por correo electrónico a la Dirección de Operaciones de Focalización (DOF) del MIDIS.
- A solicitud del MIDIS, el personal de la ULE visita la vivienda y completa la Ficha Socioeconómica Única (FSU). Luego, digita la información del FSU en el SIGOF, este sistema genera un archivo que es enviado por correo electrónico a la DOF.
- De ser el caso, la ULE realiza la corrección de los formatos de recojo de información.
- La municipalidad envía de manera formal los formatos de recojo de información.

3.4.1.1 FUENTES MANUALES Y/O ELECTRÓNICOS

Esta información contiene variables que permiten el recojo de información del ciudadano que solicita la clasificación socioeconómica.

- **FORMATO S100:** Los datos que forman parte del formato son las siguientes:

Datos de la solicitud

- Numero de Solicitud
- Ubigeo
- Número de DNI de jefe de hogar
- Fecha de la Solicitud

Datos de la persona solicitante

- Apellidos y nombres
- Tipo de solicitante (Representante de hogar o servidor de IPF)
- Correo electrónico
- Teléfono (Celular)

Datos de los declarantes de hogar

- Tipo de documento de identidad (DNI, CNV, AN, CE)
- Número de DNI
- Apellidos y nombres
- Fecha de nacimiento
- Sexo
- Parentesco con el jefe de hogar
- Indicador de residente habitual (Si/No)

Datos de acceso a servicios públicos

- Número de suministro de luz
- Nombre de la empresa de luz
- Número de Suministro de agua
- Nombre de la empresa de agua
- Indicador de servicio de luz compartido (Si/No)

Datos de la vivienda

- Tipo de vía
- Nombre de la vía
- Referencia de ubicación de la vía

Datos de la ubicación geográfica

- Zona de la vivienda (Rural/Urbano)
- Código Centro Poblado
- Nombre del Centro Poblado

Datos de declaración y autorización del solicitante

- Número de propiedades
- Número de vehículos motorizados
- Ingresos en moneda nacional

Datos de la ubicación censal

- Número de conglomerado
- Número de zona
- Número de manzana
- Número de vivienda
- Número de hogares que habitan
- Número de hogares en la vivienda

Datos de residencia en un pueblo indígena

- Nombre del pueblo indígena

Datos del servidor público que orienta al llenado de la solicitud

- Número de documento de identidad
- Apellidos y nombres
- Tipo de entidad (IPF, ULE, DOF)

PERÚ Ministerio de Desarrollo e Inclusión Social	FORMATO S100 SOLICITUD DE LA CLASIFICACIÓN SOCIOECONÓMICA (CSE) O DE VERIFICACIÓN DE LA CSE	Dirección General de Focalización (DGFO) Sistema de Focalización de Hogares (SISFOH)
SOLICITUD: 		UBIGEO: 	DOCUMENTO DE IDENTIDAD DEL SOLICITANTE:
		FECHA: 	DD - MM - AA
1. DATOS DE LA PERSONA SOLICITANTE <small>(Complete con letra impresa)</small>			
1.1 APELLIDO PATERNO 			
1.2 APELLIDO MATERNO 			
1.3 NOMBRES 			
1.4 TIPO DE SOLICITANTE 1 REPRESENTANTE DE HOGAR ☐ 2 SERVIDOR/A DE UNA INTERVENCIÓN PÚBLICA FOCALIZADA U OTRA INSTITUCIÓN ☐			
1.5 CORREO ELECTRÓNICO 1.6 TELÉFONO: <small>(Celular)</small>			
2. ¿EN EL HOGAR SE PRESENTÓ ALGUNA DE LAS SIGUIENTES SITUACIONES?			
2.1 NO TIENE CSE ☐		2.5 NO ESTOY DE ACUERDO CON EL RESULTADO DE CSE (REEVALUACIÓN) ☐	
2.2 INCORPORACIÓN DE MENORES DE EDAD ☐		2.6 SE REQUIERE VERIFICAR INFORMACIÓN DEL HOGAR ☐	
2.3 INCORPORACIÓN O SALIDA DEL HOGAR DE UNO O MÁS INTEGRANTES ☐		2.7 OTRAS CIRCUNSTANCIAS ☐	
2.4 CAMBIO DE DOMICILIO DE TODOS/AS LOS Y LAS INTEGRANTES DEL HOGAR ☐			
3. REGISTRO DE PADRE, MADRE O TUTOR <small>Consiguar el documento de identidad de la madre, padre o tutor cuando el solicitante haya marcado la situación "Incorporación de menores de edad".</small>			
3.1 PADRE ☐	3.2 MADRE ☐	3.4 DOCUMENTO DE IDENTIDAD 	Si se trata de tutor, consiguar adicionalmente los siguientes datos:
3.3 TUTOR ☐			3.3.1 RESOLUCIÓN JUDICIAL N°: 3.3.2 EXPEDIENTE:

Figura 16. Formato S100

Fuente: Página web del SISFOH – MIDIS, 2017

- **FICHA SOCIOECONÓMICA ÚNICA (FSU):** Los datos de este formato son:

Datos de la Ubicación geográfica

- Departamento
- Provincia
- Distrito
- Ubicación de vivienda (Rural/Urbano)

Datos de la ubicación censal

- Número de conglomerado
- Número de zona
- Número de manzana
- Número de vivienda
- Número de hogares que habitan
- Apellidos y nombres del informante

Datos del empadronador

- Apellidos y nombres
- Número de DNI

Datos de las características de la vivienda

- Tipo de vivienda (Casa independiente, vivienda en quinta, etc.)
- Estado de la propiedad en donde vive (Alquilada, propia, etc.)
- Tipo de material de las paredes
- Tipo de material de los techos
- Tipo de material de los pisos
- Tipo de alumbrado
- Tipo de abastecimiento de agua
- Tipo de servicio higiénico

Datos del hogar

- Número de habitaciones
- Tipo de combustible que emplea para cocinar
- Tipo de Artefactos domésticos
- Número de suministro de luz
- Número de suministro de agua
- Número de personas que viven en el hogar (mujeres y varones)

Datos de las características de la población

- Tipo de relación de parentesco
- Estado civil
- Tipo de seguro de salud
- Tipo de idioma aprendido en la niñez
- Nivel educativo
- Grado aprobado
- Tipo de ocupación (aplicable a partir de los 6 años)
- Sector en el que trabaja
- Tipo de discapacidad
- Tipo de beneficiario de IPF

SisFoh
Sistema de Focalización de Hogares

13. Ficha Socioeconómica Única - FSU

MINISTERIO DE ECONOMÍA Y FINANZAS
Dirección Nacional de Asesoría Económica y Social

PADRÓN GENERAL DE HOGARES
FICHA SOCIOECONÓMICA ÚNICA

SisFoh
Sistema de Focalización de Hogares

I. LOCALIZACIÓN DE LA VIVIENDA

UBICACIÓN GEOGRÁFICA

1. DEPARTAMENTO: 2. PROVINCIA: 3. DISTRITO:

4. CENTRO POBLADO: CATEGORÍA:

5. MONTO URBANIZABLE APLICABLE PARA DISTRITOS POBLADOS URBANOS: CATEGORÍA:

UBICACIÓN CENSAL

6. ZONA: 7. MANZANA: 8. N° FRENTE DE MANZANA: 9. VIVIENDA:

10. CUANTOS HOGARES HABITA EN ESTA VIVIENDA: 11. HOGAR:

12. APELLIDOS Y NOMBRES DEL INFORMANTE: N° ORDEN:

13. UBICACIÓN DE LA VIVIENDA

TIPO DE VÍA: 1. AVENIDA 2. JIRÓN 3. CALLE 4. PASAJE 5. CARRETERA 6. OTRO

NOMBRE DE LA VÍA: N° DE LA PUERTA:

BLINDA: PISO: INTERIOR: MANIZANA: LOTE: ROL: TELÉFONO DEL DUEÑO:

PERSONAL RESPONSABLE

APELLIDOS Y NOMBRES DEL EMPADRONADOR(A): CARGO DEL EMPADRONADOR(A):

APELLIDOS Y NOMBRES DEL JEFE(A) DE BRIGADA: CARGO DEL JEFE(A) DE BRIGADA:

APELLIDOS Y NOMBRES DEL REVISOR(A): CARGO DEL REVISOR(A):

APELLIDOS Y NOMBRES DEL SUPERVISOR(A): CARGO DEL SUPERVISOR(A):

II. ENTREVISTA Y SUPERVISIÓN

VISITA

FECHA (DD-MM-AA): RESULTADO VISITA (1):

PRIMERA: SEGUNDA: TERCERA:

JEFE(A) DE BRIGADA

FECHA (DD-MM-AA): RESULTADO VISITA (2):

PRIMERA: SEGUNDA: TERCERA:

REVISOR(A)

FECHA (DD-MM-AA): RESULTADO REVISIÓN (3):

PRIMERA: SEGUNDA: TERCERA:

RESULTADO FINAL DEL EMPADRONAMIENTO

FECHA (DD-MM-AA): RESULTADO (4):

(5) CÓDIGOS DE RESULTADO

1. Consultado 2. Revisado 3. Revisado 4. Revisado 5. No se inició la entrevista 6. Vivienda deshabitada 7. Otro (Especificar: para el resultado final)

Figura 17. Formato FSU

Fuente: Página web del SISFOH – MIDIS, 2017

3.4.1.2 FUENTES EXTERNAS

El MIDIS utiliza datos proporcionados por diferentes Instituciones Públicas para la validación y cruce de información, las cuales se detallan a continuación:

- RENIEC**

Organismo que proporciona información sobre la identidad de las personas que tienen documento de identidad nacional. Los datos que proporciona son:

- Tipo de documento de identidad
- Apellidos y nombres
- Sexo
- Fecha de nacimiento
- Fecha de fallecimiento
- Ubigeo
- Dirección

- **SUPERINTENDENCIA NACIONAL DE MIGRACIONES**

Organismo que proporciona información sobre la identidad de las personas extranjeras.

Los datos que proporciona son:

- Tipo de documento de identidad
- Apellidos y nombres
- Sexo
- Fecha de nacimiento
- Ubigeo
- Dirección

- **SUNAT**

Facilita la información de la planilla de las personas para verificar los ingresos declarados por estas. Proporciona los siguientes datos:

- Tipo de documento de identidad
- Apellidos y nombres
- Sector en el que trabaja
- Servicio que realiza
- Ingresos en moneda nacional
- Fecha último ingreso

- **OSINERGMIN**

Órgano Supervisor que centraliza la información del consumo de energía eléctrica de las viviendas que han sido registradas bajo un contrato con los ciudadanos propietarios cuyos consumos son registrados en recibos que son expedidos en forma mensual por las empresas distribuidoras de energía eléctrica. Los datos son:

- Tipo de documento de identidad del abonado
- Apellidos y nombres
- Número de Suministro de luz
- Monto
- Fecha alta de suministro
- Fecha de baja del suministro

3.4.2 PROCESO DE CLASIFICACIÓN SOCIOECONÓMICA

El proceso de clasificación socioeconómica tiene por objetivo la determinación de la clasificación socioeconómica de los hogares.

3.4.2.1 EVALUACIÓN DE LA CLASIFICACIÓN SOCIOECONÓMICA

El procedimiento para la evaluación y reevaluación de la clasificación socioeconómica de un hogar es el siguiente:

- Se obtiene, a través del SIGOF, los correos electrónicos enviados por las municipalidades y extrae el archivo adjunto (formato S100).
- Se procede a validar la consistencia y a realizar el cotejo de la información del formato S100 con las bases de datos de las fuentes externas:
 - Nivel de ingresos de los integrantes del hogar a través de las planillas de SUNAT.
 - Consumo de energía eléctrica a través de OSINERGMIN.

Sin validación del formato FSU

No se requiere utilizar el formato FSU cuando el hogar reporta información en las bases de datos de fuentes externas que permiten identificarlo como “No pobre”.

Con validación del formato FSU

Se requiere utilizar el formato FSU cuando se cumple uno de los siguientes criterios:

- El hogar no reporta información en las bases de datos de fuentes externas o,
- El hogar reporta información en las bases de datos de fuentes externas que permiten identificarlo como “Pobre”.
- El resultado de la consistencia y/o evaluación es enviado a la ULE.
- Si se requiere el formato FSU, se realiza la revisión del sustento manualmente. El resultado de esta revisión es registrado en el SIGOF.
- El SIGOF procesa la información y se realiza el cálculo de la clasificación socioeconómica.
- Se remite formalmente la respuesta a las municipalidades mediante oficio.

3.4.2.2 REEVALUACIÓN DE LA CLASIFICACIÓN SOCIOECONÓMICA

A solicitud del ciudadano, se puede realizar una reevaluación de la clasificación socioeconómica. El procedimiento sigue los mismos pasos descritos en la evaluación de la clasificación socioeconómica.

Las intervenciones públicas focalizadas también pueden indicar a los ciudadanos que forman parte de su grupo beneficiario, que soliciten una reevaluación de la clasificación socioeconómica en caso detecten cambio de domicilio.

- **Metodología de evaluación de la clasificación socioeconómica**

La Resolución Ministerial N° 151-2016-MIDIS aprueba la “Metodología para la Determinación de la Clasificación Socioeconómica” y dispone que sea de aplicación únicamente para las nuevas solicitudes de clasificación socioeconómica y para la actualización y reevaluación de la clasificación socioeconómica.

La metodología señala los siguientes pasos para la determinación de la clasificación socioeconómica:

Paso 1: Evaluación de afiliación a seguros de salud privados.

Se evalúa si el jefe del hogar o cónyuge tiene afiliación a un seguro privado de salud.

Paso 2: Evaluación de tenencia de vehículos

Se evalúa si el jefe del hogar o cónyuge posee un vehículo a su nombre.

Paso 3: Evaluación de ingresos del hogar

La evaluación de ingresos se realiza a través de la información declarada y registrada en SUNAT de acuerdo a los tipos de rentas correspondientes.

Paso 4: Evaluación del cálculo del índice de focalización de hogares.

Se comparan los ingresos per cápita del hogar de acuerdo a su ubicación (Lima Metropolitana, resto urbano y rural).

Paso 5: Evaluación del gasto en consumo de electricidad.

La evaluación de gastos en consumo de electricidad se realiza a través de la información registrada en OSINERGMIN.

A partir de la información recolectada, se determina la clasificación socioeconómica del hogar en función de la pobreza monetaria, basada en la influencia que ejercen las características de un hogar sobre la probabilidad que este sea pobre, usando el concepto de Proxy Means Test (PMT).

La metodología de clasificación socioeconómica hace referencia a herramientas estadísticas como el modelo de regresión para el cálculo de la probabilidad de que un hogar sea clasificado como pobre, pobre extremo o no pobre.

3.4.3 PADRÓN GENERAL DE HOGARES

El Sistema de Focalización de Hogares (SISFOH) administra la información del Padrón General de Hogares (PGH). Este padrón contiene información socioeconómica de los ciudadanos para que las intervenciones públicas focalizadas del Estado identifiquen qué personas pueden ser parte de los grupos poblacionales priorizados y de esta manera accedan a beneficios estatales.

Para obtener el PGH se considera lo siguiente:

- El resultado de la determinación de la clasificación socioeconómica es incorporado en la base de datos del Padrón General de Hogares (PGH).
- Los recién nacidos, niños, niñas y adolescentes menores de dieciocho (18) años son incorporados al PGH, siempre que el hogar cuente con clasificación socioeconómica vigente y que se verifique la relación de parentesco de la madre o del padre con el menor de edad o tutor a quien se le asigna legalmente el cuidado del menor.
- Son incorporados en el PGH y se les otorga la clasificación socioeconómica de pobre extremo, a las personas que forman parte de la base de datos oficial de pueblos indígenas listados por el Ministerio de Cultura.

La información que conforma el PGH es la siguiente:

Tabla 10
Información que almacena el PGH

Código de hogar	Código que identifica de manera única a cada hogar
Código de la CSE	Código que identifica de manera única la clasificación socioeconómica del hogar

Nivel de la CSE	Resultado aproximado de la medición de las condiciones socioeconómicas del hogar, y que se basa en la metodología para la determinación de la clasificación socioeconómica. El resultado tiene tres niveles: (i) pobre extremo, (ii) pobre y (iii) no pobre
Fecha de inicio de la vigencia de la CSE	Fecha de registro de la clasificación socioeconómica en el PGH
Fecha fin de la vigencia de la CSE	Fecha en la cual la clasificación socioeconómica cumple con el periodo de vigencia asignado
Estado de la vigencia de la CSE	Indica si la clasificación socioeconómica se encuentra: (i) vigente, (ii) caduca (iii) cancelada
Características del estado “vigente”	Subvariables de la variable vigente del estado de la clasificación socioeconómica: i) Pedido de verificación, ii) Indicios de presunto fraude y iii) Cambio en la composición del hogar
Norma que aprueba el método de cálculo de la CSE	Norma con la que se aprueba la Metodología de determinación de la clasificación socioeconómica

Fuente: Página web del MIDIS - 2017

3.4.3.1 ACCESO AL RESULTADO DE LA CLASIFICACIÓN SOCIOECONÓMICA

Las diversas Intervenciones Públicas Focalizadas pueden acceder a la información del “Padrón General de Hogares” a través de diversos mecanismos:

- **Servicio de consulta: “Web Service”**

Permite verificar la información socioeconómica de las personas y mostrarla en los sistemas de afiliación de las Intervenciones Públicas Focalizadas.

Las Intervenciones Públicas Focalizadas que necesitan acceder a dicha información deben contactarse con la Dirección de Operaciones de Focalización.

- **Aplicativo de Consulta al Padrón General de Hogares (PGH) del SISFOH**

La Dirección de Operaciones de Focalización (DOF) ha implementado un servicio de consulta en línea que permite conocer de manera rápida la clasificación socioeconómica de los hogares.

Para acceder al servicio de consulta se debe ingresar a la página web: <http://www.midis.gob.pe/padron> luego el ciudadano o ciudadana deberá seguir una secuencia sencilla de pasos para completar la consulta de clasificación socioeconómica (CSE) de su hogar. Ver figura 18,19 y 20.

PERÚ Ministerio de Desarrollo e Inclusión Social

CONSULTA DEL PADRÓN GENERAL DE HOGARES (PGH) DEL SISFOH PARA EL CIUDADANO

Ingrese el número de su documento:
DNI/ Carné de extranjería
o el DNI / Carné de extranjería de algún integrante de su hogar a Consultar

Tipo de Documento: ☒ DNI ☐ Carné de extranjería

41672348

Ingrese el número de verificación que aparece en el recibo

458 A58

Refrescar imagen ↻

Consultar

© Ministerio de Desarrollo e Inclusión Social - SISFOH

Figura 18. Pantalla acceso al PGH

Fuente: Página web del MIDIS, 2018

CONSULTA DEL PADRÓN GENERAL DE HOGARES (PGH) DEL SISFOH PARA EL CIUDADANO

Sr./Sra. [REDACTED] su hogar está constituido por [REDACTED] integrante(s) y se encuentra registrado en el PGH.

Para continuar con la consulta de Clasificación Socioeconómica (CSE) responda las siguientes preguntas que corresponden a la última información declarada por su hogar:

Seleccione la fecha de nacimiento de su HIJO/A:

☐ 01/11/1960 ☐ 12/03/2000 ☐ 30/01/2010

Seleccione el nombre de su JEFE:

☒ ESTHER LUCIA ☐ MIRLA LUCIA ☐ MARIA ELISABETH ☐ TERESA

Digite el número de UBIGEO DE NACIMIENTO que se encuentra en su DNI: :

Digite Ubigeo de Nacimiento según DNI

Digite la FECHA DE EMISIÓN de su DNI: :

Digite fecha de emisión del DNI

> Continuar

Figura 19. Pantalla consulta PGH

Fuente: Página web del MIDIS, 2018

RESULTADO DE CONSULTA AL PADRÓN GENERAL DE HOGARES (PGH)

Estimado ciudadano(a) [REDACTED] a continuación se muestra el resultado de la Clasificación Socioeconómica (CSE) de su hogar.

Fecha y Hora de la consulta: 20/10/2018 a las 18:17:22

- Código de Hogar: [REDACTED]
- Código Clasificación Socioeconómica: [REDACTED]
- Clasificación Socioeconómica: NO POBRE
- Vigencia de Clasificación Socioeconómica: Desde 14/01/2014 Hasta 14/01/2019
- Departamento: [REDACTED]
- Provincia: [REDACTED]
- Distrito: [REDACTED]
- Centro Poblado: [REDACTED]
- Núcleo Urbano: No hay información
- Dirección de la Vivienda LOTE: [REDACTED]
- Estado de la CSE: VIGENTE
- Integrantes del Hogar:

Tipo Documento	Nº Documento	Nombres	Apellido Paterno	Apellido Materno
DNI	[REDACTED]	[REDACTED]	[REDACTED]	[REDACTED]
DNI	[REDACTED]	[REDACTED]	[REDACTED]	[REDACTED]
DNI	[REDACTED]	[REDACTED]	[REDACTED]	[REDACTED]

☐ Enviar resultado por Correo: [REDACTED] **Enviar** **PDF**

Figura 20. Resultado de consulta al PGH

Fuente: Página web del MIDIS, 2018

Figura 21. Proceso de Clasificación Socioeconómica Actual

Fuente: Elaboración propia

3.5 INTERVENCIONES PÚBLICAS FOCALIZADAS

Se considera intervención pública a todo bien o servicio que brinda el Estado destinado a lograr un propósito social específico sobre personas, hogares, viviendas, centros poblados, comunidades, grupos poblacionales o jurisdicciones geográficas en situación de pobreza, vulnerabilidad o exclusión. Los programas sociales y subsidios del Estados son un tipo de intervención pública.

La Intervención Pública Focalizada (IPF) requiere definir y evaluar criterios de elegibilidad para que una determinada población acceda a los bienes y/o servicios que esta provee. En ese sentido, desde el año 2011, mediante la Ley 29911, se estableció que la selección de usuarios y usuarias de los programas sociales o subsidios del Estado debe tener en cuenta la evaluación del SISFOH, que se determina a través de la clasificación socioeconómica.

Según el Ministerio de Economía y Finanzas (MEF), un programa social está referido a la parte del gasto social que puede ser calificada según:

Su naturaleza universal o focalizada:

- Programa universal: Que no discrimina al beneficiario. En el Perú están definidos como tal: Educación Inicial, Educación Primaria, Educación Secundaria y la Provisión Básica de Salud.

Geográficamente se ubican en todo el territorio nacional mediante más de 42 mil locales educativos y más de 7 mil establecimientos de salud.

- Programa focalizado: Que atiende a determinado grupo, región o problema y se implementa empleando focalización geográfica y/o individual, como es el caso del Seguro Integral de Salud, el Programa del Vaso de Leche, el Programa de Comedores Populares, los programas de alimentación y nutrición del PRONAA, el Programa JUNTOS, los Programas de Inversión Social y Productiva, Provias Descentralizado, FONCODES, AGRORURAL, Electrificación Rural, etc.

Geográficamente se ubican de manera diversa; por ejemplo:

- A nivel nacional, como el Programa del Vaso de Leche que se distribuye a los 1834 municipios del país (pero focaliza sus beneficiarios entre niños menores de 14 años, madres gestantes y en periodo de lactancia, personas mayores de 65 años y afectados por TBC), o como el Seguro Integral de Salud.
- En determinados distritos, como el Programa JUNTOS atiende a 638 comunas (pero focaliza su atención en familias pobres con niños menores de 14 años y madres gestantes).
- Pueden dirigirse a determinado grupo objetivo, como el programa Wawa-Wasi (que con 23 millones en el 2007 focaliza su atención en niños menores de 3 años).

Si otorgan beneficios individuales o colectivos:

- Beneficios individuales: Donde la prestación se entrega directamente al beneficiario: Por ejemplo, todo lo que es educación y salud en general, Programa del Vaso de Leche, Programa JUNTOS, Wawa-Wasi, entre otros.

- Beneficios colectivos: Donde la prestación benéfica a la comunidad o región en general: Por ejemplo, lo que hacen FONCODES, PROVIAS Descentralizado o Electrificación Rural: carreteras rurales, puentes, estaciones eléctricas.”

3.5.1 PROGRAMAS SOCIALES Y SUBSIDIOS DEL ESTADO

Según información de la plataforma digital única del Estado Peruano para orientación a la ciudadanía, se identifican las siguientes Intervenciones Públicas Focalizadas.

- **Seguro Integral de Salud (SIS)**

Es un seguro de salud dirigido para todos los ciudadanos peruanos que no cuenten con otro seguro de salud vigente. El

SIS cubre los medicamentos, procedimientos, operaciones, insumos, bonos de sepelio y traslados, pero la cobertura varía según el tipo de SIS. Cubre más de 1400 enfermedades, incluidos varios tipos de cáncer.

Existen 4 tipos de seguros provistos por el SIS:

- SIS Gratuito: Para personas en pobreza extrema, pobreza, madres gestantes, niños, bomberos, entre otros.
- SIS Independiente: Para cualquier ciudadano que pueda pagarlo.
- SIS Microempresas: Para los dueños de una microempresa que quieran inscribir a sus trabajadores.
- SIS Emprendedor: Para trabajadores independientes que no tengan empleados a su cargo (como peluqueros, modistas, gasfiteros, etc.).

- **Pensión 65 - Programa Nacional de Asistencia Solidaria**

El Programa Nacional de Asistencia Solidaria Pensión 65 protege a los adultos mayores de 65 años a más que se encuentran en condición de pobreza extrema, brindándoles una subvención económica de 250 soles de manera bimestral. Con esto, se ayuda a que sus necesidades básicas sean atendidas, que sean revalorados por su familia y comunidad, y se contribuye a dinamizar pequeños mercados y ferias locales.

Este programa del Ministerio de Desarrollo e Inclusión Social (MIDIS) se crea el 19 de octubre del 2011 (Decreto Supremo N° 081–2011-PCM) como respuesta del Estado ante la necesidad de brindar protección a un sector especialmente vulnerable.

- **JUNTOS - Programa Nacional de Apoyo Directo a los más Pobres**

El Programa Nacional de Apoyo Directo a los más Pobres – Juntos entrega incentivos monetarios a hogares de distritos que presentan índices de 40% a más de pobreza. Con esto, buscan contribuir a la reducción de la pobreza y evitar que los niños, niñas, adolescentes de hasta 19 años y gestantes sigan viviendo en la pobreza que afectó a sus padres y generar capital humano dentro de los hogares.

Este programa del Ministerio de Desarrollo e Inclusión Social (MIDIS) promueve la corresponsabilidad, ya que sus incentivos monetarios están condicionados al cumplimiento de compromisos que garantizan el acceso y uso de servicios de salud, nutrición y educación.

Requisitos:

- Vivir en distrito de intervención hace al menos 6 meses.
- Tener mínimo un miembro objetivo entre sus integrantes: gestante, niña, niño, adolescente o joven de máximo 19 años.
- Ser considerado como pobre según el Sistema de Focalización de Hogares.

- **CONTIGO - Programa Nacional de entrega de la pensión no contributiva a personas con discapacidad severa en situación de pobreza**

El programa está a cargo del MIDIS y fue creado con el objetivo de mejorar la calidad de vida de las personas con discapacidad severa en situación de pobreza, el programa entrega una pensión de 300 nuevos soles bimestrales (cada 2 meses).

Requisitos para acceder al programa CONTIGO:

- Contar con un certificado de discapacidad con calificación de severa, emitido por los establecimientos de salud autorizados por el Ministerio de Salud.

- No percibir ingreso o pensión que provenga del ámbito público o privado, incluyendo prestaciones económicas que otorguen el Seguro Social de Salud – ESSALUD.
- Encontrarse en situación de pobreza bajo los criterios del Sistema de Focalización de Hogares – SISFOH.

- **FISE - Fondo de Inclusión Social Energético**

El Fondo de Inclusión Social Energético (FISE) se crea con el propósito de llevar energía menos contaminante a poblaciones más vulnerables en todo el país, a través de tres fines: la masificación del gas natural para viviendas y vehículos, la ampliación de la frontera energética utilizando energías renovables, la promoción para el acceso al GLP (balones de gas doméstico) en los sectores vulnerables urbanos y rurales.

Requisitos para acceder al FISE:

1. El beneficio del FISE consiste en la entrega mensual de un vale de descuento para la compra de un balón de GLP de 10 kilogramos en un agente autorizado, y la entrega por única vez de un Kit de cocina para los beneficiarios que no cuenten con cocina a GLP.
2. Potenciales beneficiarios aquellos que se encuentren en una situación de pobreza o pobreza extrema, que cuenten con un consumo de electricidad inferior a 30 kw/h por mes y que cuenten con el Documento Nacional de Identidad (DNI).

- **PRONABEC - Programa Nacional de Becas y Crédito Educativo**

PRONABEC es un programa dependiente del Ministerio de Educación (MINEDU), tiene el objetivo de mejorar la igualdad en el acceso a la educación superior, garantizando la permanencia y culminación de los estudios de los beneficiarios.

El componente BECA 18 de PRONABEC financia los estudios de pregrado en universidades públicas o privadas nacionales y extranjeras, así como estudios técnicos en institutos de educación superior tecnológicos públicos o privados nacionales, con énfasis en carreras vinculadas al desarrollo científico y tecnológico del país y que posibiliten una adecuada inserción laboral de los graduados.

Requisitos para acceder a BECA 18:

- Ser peruano.
- Tener hasta 22 años cumplidos, como máximo, al 31 de diciembre

Excepto:

- REPARED (víctimas del proceso de violencia ocurrida entre 1980 – 2000): sin límite de edad.
- FF.AA.: hasta 25 años cumplidos, como máximo, al 31 de diciembre.
- Haber terminado la educación secundaria entre los años solicitados por MINEDU (excepto FF.AA. y REPARED).
- Haber estudiado toda la educación secundaria en una institución educativa en el territorio nacional.
- Acreditar alto rendimiento académico.
- Haber ingresado a una IES elegible en una carrera elegible.
- Encontrarse en situación de pobreza, según el Sistema de Focalización de Hogares (SISFOH), para los postulantes a las modalidades Ordinaria y EIB.

- **JOVENES PRODUCTIVOS - Programa Nacional de Empleo Juvenil**

Es un programa del Ministerio de Trabajo y Promoción del Empleo que facilita el acceso de los jóvenes al mercado laboral formal a través de capacitación laboral, asistencia técnica para el emprendimiento e intermediación laboral.

Requisitos para acceder al programa JOVENES PRODUCTIVOS:

- Contar con DNI
- Ser elegible en Sistema Integral de Focalización de Hogares – SISFOH.
- Los jóvenes que hayan cumplido los 18 años hasta los 29 años.

- **TRABAJA PERU - Programa Nacional para la Generación de Empleo Social Inclusivo**

Es un programa del MTPE, creado con el objetivo de generar empleo, y promover el empleo sostenido y de calidad en la población desempleada y subempleada de las áreas urbanas y rurales, en condición de pobreza y extrema pobreza.

Requisitos para acceder al programa TRABAJA PERU:

- Fotocopia de DNI Vigente.
- Ser elegible en Sistema Integral de Focalización de Hogares – SISFOH.
- Fotocopia de Seguro Integral de Salud - SIS vigente.
- Declaración jurada simple de persona desempleada.

- **MI VIVIENDA - Bono de protección de viviendas vulnerables a los riesgos sísmicos**

El Bono Protección de Viviendas Vulnerables a los Riesgos Sísmicos (BPVVRS) otorga a los hogares en riesgo un bono para la reducción de la vulnerabilidad de las viviendas frente a los efectos de los riesgos sísmicos. A la fecha el bono interviene en 12 distritos priorizados de Lima. El bono es entregado por el fondo Mivivienda.

Requisitos para acceder al Bono Mivivienda:

- Tener un nivel socioeconómico de pobre o pobre extremo.
- Ser propietario del predio inscrito en Registros Públicos o COFOPRI.
- La vivienda debe ser evaluada por el Ministerio de Vivienda, Construcción y Saneamiento (MVCS).
- DNI del Propietario.

3.5.2 PRESUPUESTO ASIGNADO

En base a la información obtenida de la consulta amigable del MEF, el Estado Peruano tiene como uno de sus principales propósitos la disminución de la pobreza, para lo cual designa un presupuesto anual para los programas sociales.

Como podemos observar en la figura 22, el presupuesto anual ha venido incrementándose considerablemente desde el año 2007 hasta el año 2017.

Este crecimiento de la última década representa un aumento de más del 200% del presupuesto, variando desde los 3,387 millones de soles hasta los 11,208 millones de soles en el último año.

Figura 22. Presupuesto de los Programas Sociales 2007-2017

Fuente: Se tomó como fuente de información PROGR-SOC para la Elaboración: Propia

3.6 DESCRIPCIÓN DE LA SITUACIÓN PROBLEMÁTICA

Según la información del año 2017, se observa que la pobreza en el Perú alcanzó el índice de 21,7%, lo cual representa 6 millones 906 mil personas en situación de pobreza.

El Estado Peruano, a través del MIDIS, tiene por objetivo principal contribuir con la mejora en la calidad de vida de la población vulnerable, y por ello se encuentra trabajando para reducir el índice de pobreza. Es en este sentido que las intervenciones públicas focalizadas deben llegar a las personas que más lo necesitan.

3.6.1 IMPACTO DE LA CSE EN LA POBREZA

Tomando como base el concepto de focalización y cómo éste es utilizado para identificar a los hogares o individuos que se definen como elegibles para las transferencias de recursos y al mismo tiempo excluir a aquellos que se consideran no elegibles, se analizarán dos aspectos que se encuentran impactados por el proceso de Clasificación Socioeconómico actual.

3.6.1.1 PRESUPUESTO

Al realizar un análisis sobre las tendencias del presupuesto asignado para los programas Sociales del Estado y el índice de pobreza, notamos que, si bien la pobreza ha estado disminuyendo, el presupuesto del Estado continúa incrementándose año a año. La figura 23 nos muestra la evolución de ambas tendencias a nivel comparativo.

Figura 23. Tendencia presupuesto Programas Sociales vs Índice de Pobreza y pobreza extrema

Fuente: Elaboración propia.

3.6.1.2 FILTRACIÓN

La filtración permite identificar si una intervención pública focalizada está otorgando beneficios a personas no pobres; es decir a población que no es su público objetivo.

Para poder entender por qué el Estado tiene que asignar mayor presupuesto a los programas sociales si el índice de pobreza ha disminuido más de 50%, analizamos otro hecho importante, según la tabla 11, existe un alto porcentaje de filtraciones en los programas sociales en los países de América Latina.

Tabla 11

Índice de filtración en los Programas Sociales (Porcentaje del PBI)

Indicadores de filtraciones en los programas sociales (porcentaje del PBI)				
País	En programas de transferencias monetarias	En otros programas	Total filtraciones	Filtraciones como porcentaje de los gastos de los programas sociales
Argentina	1.9	0.1	2.0	56.2
Paraguay	0.2	0.9	1.1	56.2
Uruguay	0.7	0.4	1.1	62.4
Bolivia	0.7	0.1	0.8	50.9
Chile	0.6	0.1	0.7	65.2
Brasil	0.4	0.2	0.6	51.6
Costa Rica	0.3	0.3	0.6	42.1
Colombia	0.1	0.4	0.5	50.6
Belice	-	0.4	0.4	39.0
El Salvador	0.1	0.4	0.5	46.3
Jamaica	0.2	0.2	0.4	66.6
República Dominicana	0.1	0.2	0.3	44.3
Panamá	0.2	0.1	0.3	48.0
México	0.1	0.1	0.2	28.9
Nicaragua	-	0.2	0.2	31.7
Guatemala	0.1	0.1	0.2	24.3
Perú	0.1	-	0.1	35.1
Honduras	0.1	-	0.1	6.2
Promedio	0.3	0.2	0.5	45.0

Fuente: FIEL y BID sobre la base de datos de encuestas de hogares y datos de presupuestos nacionales

- **Filtración en el Perú**

En el caso peruano, este porcentaje de filtración se encuentra en un promedio del 35.1%, esto quiere decir que, si analizamos la región más vulnerable del Perú (Cajamarca), donde el índice de pobreza y extrema pobreza supera el 40%, tendremos al menos un 14% de filtración.

Esto representa que, de cada 100 peruanos de Cajamarca, 40 no pueden cubrir sus necesidades básicas y, si 35% son infiltrados, quiere decir, que no debieron ser considerados pobres (debido a que no lo son) entonces al menos 14 personas que debieron ser cubiertas por los IPF no podrán contar con este beneficio, debido a que el presupuesto fue asignado a personas no pobres.

A continuación, en la figura 24, podemos observar el porcentaje de filtración en la población de Cajamarca, que representa más de 230,000 personas que estarían recibiendo ayuda de las intervenciones públicas sin corresponderle.

Figura 24. Filtración en Cajamarca

Fuente: Elaboración propia

• Costo Financiero de la Filtración por IPF

En un documento de discusión elaborado por encargo de la Universidad del Pacífico en 2013, determina que los principales programas sociales han presentado altas tasas de filtración al año 2012: Vaso de Leche (50.9%), Comedores Populares (44%), Desayuno Escolar (37%), Sistema Integral de Salud (SIS) (39.7%), Cuna Más (32.4%), Juntos (17.8%) y Pensión 65 (22.1%).

Tabla 12

Filtración en los Programas Sociales

Programa Social	Presupuesto	Beneficiarios	Filtraciones (como % de beneficiarios)	Filtraciones (absoluto)	Costo promedio por beneficiario	Costo de filtraciones	Costo de filtraciones como % PBI
Vaso de Leche	363,000,000	1,714,851	50.9%	872,203	212	184,627,998	0.03%
Comedor Popular	128,037,445	434,092	44.0%	190,863	295	56,296,020	0.01%
Desayuno Escolar	317,300,889	1,363,411	37.0%	504,475	233	117,404,367	0.02%
SIS	647,420,454	9,503,409	39.7%	3,775,277	71	267,916,923	0.05%
Cuna Más	186,487,089	57,817	32.4%	18,732	3225	60,419,533	0.01%
Juntos	852,844,592	518,339	17.8%	92,346	1645	151,940,426	0.03%
Pensión 65	264,357,731	133,099	22.1%	29,106	1986	58,404,859	0.01%
Total						897,010,126	0.17%

Fuente: ENAHO 2012 - INEI

Considerando los 3 programas sociales que utilizan la información del Padrón General de Hogares (SIS, Juntos, Pensión 65), observamos que el costo financiero de la filtración representa el 0.09% del BPI, equivalente a S/ 478,262,208.

- **Medidas para reducir el nivel de filtración**

Aunque no existen estudios recientes sobre el nivel de filtración en las intervenciones públicas focalizadas, el Ministerio de Desarrollo e Inclusión Social, en el 2017, ha establecido medidas para prevenir el fraude en las intervenciones públicas focalizadas ejecutadas por los diversos programas sociales del Estado en el marco del Sistema Nacional de Focalización (SINAFO).

Las medidas adoptadas son:

- Las IPFs podrán aplicar multas que van desde media (0,5) unidad impositiva tributaria (UIT) hasta 2 UIT, y de manera complementaria, se puede declarar a una persona inelegible para una Intervención Pública Focalizada hasta por cinco años.
- Supervisión por parte de las IPFs a sus procesos de afiliación con el objetivo de detectar los hechos que puedan afectar la integridad, exactitud, veracidad y actualización de la información que manejan.
- Inspección del MIDIS a los hogares de los solicitantes de clasificación socioeconómica y a los hogares que cuentan con clasificación socioeconómica vigente para verificar la veracidad de la información con la que se determinó la respectiva clasificación.

3.6.2 FACTORES DE IMPACTO A LA CLASIFICACIÓN SOCIOECONÓMICA

Se analizará tres aspectos que impactan al proceso de Clasificación Socioeconómico actual.

3.6.2.1 PERSONA

En base a lo indicado por (Lindert, Linder, Hobbs, & De la Brière, 2007), que señala que todos los programas de transferencia resultan en un grado de fraude y errores y que el desafío es desarrollar sistemas efectivos para detectar, remediar y minimizar la ocurrencia de estas irregularidades, tomamos los 2 siguientes puntos como referencia:

- **Error:** No todas las irregularidades que ocurren son resultado de un fraude intencional. Muchas implican errores involuntarios de medición o de registro de información (al hacer las encuestas). Dentro del error, se hace una distinción entre "error del cliente", donde se proporciona información incorrecta sin intenciones

fraudulentas, y "error oficial" que surge de un error o por un acto u omisión de un funcionario, que no fue ni causado ni conocido por el funcionario ni por el cliente en el momento en que se cometió.

- **Fraude:** El fraude intencional también existe en todos los programas de transferencia y puede surgir en diferentes formas por diferentes actores:
 - Por los políticos, registrando posibles beneficiarios (amigos, parientes, miembros del partido) para propósitos clientelistas, para obtener apoyo político a cambio de beneficios del programa;
 - Por los propios beneficiarios, proporcionando información falsa sobre el ingreso o composición de la familia para sesgar las decisiones de elegibilidad.

En el proceso de clasificación socioeconómica, podemos detallar estos puntos de la siguiente manera:

- **Errores en el registro:**

En el proceso de recopilación de información, donde se realiza el recojo de información mediante el formato S100 y/o FSU, el solicitante puede incurrir en errores al completar los datos. Asimismo, el empadronador (ULE) puede inducir al ciudadano a cometer un error por falta de capacitación. El error también puede darse al momento de digitar la información en el SIGOF.

- **Fraude:**

En el proceso de recopilación de información, donde se realiza el recojo de información mediante el formato S100 y/o FSU, se ha dado casos donde el poblador solicitante puede indicar datos falsos como el domicilio o no señalar todos los ingresos económicos. Asimismo, durante los procesos electorales, algunas autoridades o candidatos pueden ofrecer y favorecer a la población que los apoya o aporta a sus campañas, incurriendo al ingreso de datos falsos.

3.6.2.2 INFORMACIÓN

De acuerdo con la metodología planteada por el MIDIS, las siguientes bases de datos externas deberían ser utilizadas para cotejar la información del ciudadano:

- SEGUROS DE SALUD PRIVADOS
- SUNARP (tenencia de vehículos)

- OSINERGMIN (consumo de energía eléctrica)
- SUNAT (ingresos del hogar).

El Estado Peruano con el Decreto Supremo Nro. 1246 aprobó las diversas medidas de simplificación administrativa, y que consiste en la interconexión gratuita entre entidades públicas para poner a disposición, acceder o suministrar la información de las bases de datos actualizadas de las entidades del Estado, oportunidad que el MIDIS podría aprovechar para utilizar las bases de datos tales como:

- SEGUROS DE SALUD PRIVADOS (para verificar los seguros de salud privados de los ciudadanos).
- SUNARP (para verificar la tenencia de inmuebles o vehículos a nombre de los ciudadanos).

3.6.2.3 TECNOLOGÍA

El MIDIS utiliza un sistema de focalización de hogares (SIGOF), el cual tiene principalmente 2 módulos, uno para el recojo de la información y otro para el procesamiento de la información recogida y para la determinación del CSE:

- Módulo del sistema SIGOF para el recojo de la información: Se basa en el software CSPRO, el cual está adecuado a las necesidades de la entidad para el recojo de información. Es utilizado por las municipalidades, como plataforma para el registro de información del Formato de Solicitud S100 y registro de la Ficha Socioeconómica Única – FSU.

El CSPRO es un software orientado a censos y encuestas que permite digitar información en formularios predeterminados según los requerimientos de información establecidos, desarrollado por U.S. Census Bureau y ICF International Serpro S.A.

- Módulo del sistema SIGOF para la determinación de la clasificación socioeconómica: Módulo que permite validar y calificar en base a puntuaciones de las respuestas brindadas por los ciudadanos en los formatos de recojo de información, determinando la CSE. Por otro lado, este aplicativo web está diseñado para que las municipalidades puedan realizar seguimiento del estado del trámite de su FSU enviada.

La metodología de clasificación socioeconómica definida por el MIDIS, la cual hace referencia a modelos y métodos estadísticos basados en algoritmos como el bayesiano y regresión logística; sin embargo, la misma no han sido aplicadas al momento de la construcción del aplicativo SIGOF.

Existen en el mercado, herramientas que cuentan con modelos matemáticos y estadísticos incorporados que podrían utilizarse para el proceso de determinación de la CSE.

4 SOLUCIÓN PROPUESTA

4.1 ANTECEDENTES

Para el presente trabajo, mencionaremos 2 artículos importantes en el análisis de la propuesta que realizamos:

- **Proceso de clasificación de una población para la inmunización infantil en la India**

Según el artículo de (A. Sharma, Mansotra, & Shastri, 2015), en India, un país con gran cantidad de población y cuyos índices de enfermedades y pobreza son muy elevados, se buscaba contar con un proceso que le permita inmunizar a la población infantil que más lo requería.

Por lo tanto, se realizó un proceso de investigación para averiguar los distritos de los Estados de Jammu y Cachemira donde se requería mayor atención en términos de inmunización infantil mediante el desarrollo de un modelo basado en minería de datos, utilizando la clasificación de Naive Bayes.

Además de la construcción del modelo, se utilizaron varias medidas para la evaluación de los resultados.

Según este mismo artículo, la minería de datos es una herramienta poderosa con potencial para influir en la salud, medicina y el ambiente. Además, proporciona varias técnicas y algoritmos que permiten descubrir patrones en datos históricos.

La minería de datos es un paso necesario en el proceso de descubrimiento de conocimiento KDD, este proceso involucra también otros pasos como la limpieza de datos, la colección o

unión de diferentes fuentes de datos, la selección de data relevante, los procesos de transformación, la interpretación y evaluación de patrones.

Según el tipo de datos que se extraen, se encuentran dos categorías para la minería de datos que son descriptiva y predictiva.

La primera, que corresponde a tipo de aprendizaje no supervisado, hace referencia al descubrimiento de patrones basados en data real e identifica las relaciones entre los atributos, agrupando los datos según sus similitudes en grupos conocidos como clúster.

La minería predictiva corresponde al tipo de aprendizaje supervisado, donde se utiliza data histórica para predecir un resultado y se le indica el modelo a seguir. Existen varias técnicas como clasificación, regresión, entre otras.

En este artículo se utilizó Java NetBeans 6.1 y el algoritmo de clasificación Naive Bayes. El objetivo fue determinar dos clases diferentes, los distritos de prioridad y los de no prioridad, y se siguieron dos pasos:

- La carga de datos para leer los datos de entrada.
- Aplicación del algoritmo y la clasificación de los datos de prueba en base a los datos de entrenamiento.

Se realizó la medición del performance utilizando la matriz de confusión para determinar valores correctamente clasificados y los que no. Se midió la sensibilidad, el rango de error, la especificidad y la precisión.

Según los resultados obtenidos, de los 22 distritos analizados, 21 de ellos se clasificaron correctamente y solo 1 se clasificó de manera incorrecta, alcanzando una precisión del 95%. Esto significa que, con el modelo definido, se podrían clasificar nuevos distritos como prioritarios y no prioritarios fácilmente.

- **Identificación de familias en pobreza en Indonesia.**

Según el artículo de (Redjeki, Guntara, & Anggoro, 2015), la pobreza en Indonesia es una de las prioridades del gobierno. La determinación de la clasificación del estado de pobreza requiere mucho esfuerzo y para el artículo actual se utilizó el algoritmo Naive Bayes.

Se utilizaron 11 variables para el análisis y cada uno con determinado valor. Estas variables o indicadores fueron: alimentos, ropa, vivienda, salud, ingresos, educación, riqueza, propiedad, agua, electricidad y número de miembros en la familia. La clasificación utilizada fue pobre, muy pobre, vulnerable.

La data utilizada fue de las familias pobres de Kabupaten Bantul. 219 datos divididos en dos partes: 175 datos (80%) que se utilizaron para el entrenamiento y 44 datos (20%) para pruebas.

Los resultados de la fase de prueba se utilizaron para calcular la probabilidad de cada clasificación, utilizando la probabilidad obtenida en la fase de entrenamiento.

La implementación del algoritmo fue desarrollada en Java y los resultados fueron utilizados para mapear a las familias pobres.

Los resultados de este estudio fueron:

- El método Naive Bayes pudo realizar la clasificación de familias pobres con una precisión del 93%.
- Se realizaron solo dos grupos de clasificación, donde 25 familias eran pobres y 19 propensas a pobres. El modelo pudo identificar correctamente a 41 de ellas y los otros 3 no pudieron identificarse.

4.2 SELECCIÓN DE LA METODOLOGIA

Como consecuencia de la evolución y explosión de los datos y de las tecnologías que permiten su captura y almacenamiento, el KDD ha cobrado mayor relevancia para el procesamiento de éstos dentro de las organizaciones.

La toma de decisiones se basa más en el análisis de datos que en la intuición, por lo que disponer de la información en el momento oportuno es de vital importancia. Por tal motivo, los informes consolidados o detallados deben ser contruidos mediante la implementación de procesos adecuados.

Para llevar a cabo la gestión adecuada de dichos procesos, es necesario el uso de una metodología confiable que permita que los proyectos tengan los cimientos necesarios para la obtención de buenos resultados sobre la puesta en marcha.

Para el presente trabajo, con la finalidad de que el MIDIS pueda contar con información que le permita tomar decisiones acerca del proceso de clasificación socioeconómica, se utilizará la metodología CRISP-DM ya que, según (Parra et al., 2010), esta metodología es una de las más utilizadas actualmente, basada en estándares internacionales que le permiten contar con fases estructuradas y confiables para los usuarios. Las fases que componen la metodología se relacionan entre sí, lo que le permite generar procesos iterativos.

Una de las ventajas más importantes de esta metodología es que no restringe la herramienta a utilizar para la construcción del Data Warehouse, lo cual nos permitirá poder realizar diferentes pruebas con la data recolectada.

4.3 APLICACIÓN DE LA METODOLOGIA

4.3.1 COMPRENSIÓN DEL NEGOCIO

Siguiendo la metodología seleccionada, en este punto se determinará los objetivos del negocio, se realizará la evaluación de la situación actual, se definirán los objetivos de la minería de datos y se planteará el plan de proyectos.

4.3.1.1 DETERMINAR LOS OBJETIVOS DE NEGOCIO

Para determinar los objetivos del negocio se explicará el contexto del proceso de la clasificación socioeconómica. Asimismo, se determinarán los criterios de éxito para alcanzar los objetivos planteados.

- **Contexto**

El proceso de determinación de la clasificación socioeconómica se enmarca en la Ley del Sistema Nacional de Focalización – SINAFO y el Sistema de Focalización de Hogares - SISFOH. La entidad cuenta con normatividad relacionada a regular la operatividad del SISFOH, la metodología de la clasificación socioeconómica y la simplificación del proceso de clasificación socioeconómica.

La entidad realiza la clasificación socioeconómica de los hogares, a través de la información proporcionada por la población, la información recabada por el empadronador de la municipalidad y el cruce de esta información con las bases de datos de otras entidades. Esta información es almacenada en una base de datos que contiene las solicitudes de clasificación socioeconómica (S100) y las fichas únicas socioeconómicas

(FSU). Sobre ambos formatos se aplican reglas de consistencia a fin de obtener información confiable para determinar la clasificación socioeconómica. También se cuenta con una base de datos resultante del proceso de clasificación socioeconómica conocida como Padrón General de Hogares.

- **Objetivos del negocio**

Del objetivo estratégico específico definido por la institución en su Plan Estratégico Institucional vigente; así como el objetivo del proceso de clasificación socioeconómica se define como objetivos del negocio:

- Dirigir y supervisar acciones para la determinación de la clasificación socioeconómica de los potenciales usuarios de los programas y subsidios del Estado que se ejecutan bajo criterios de focalización individual (por hogar).
- Fortalecer la focalización y la gestión de usuarios contribuyendo al logro de objetivos de los programas sociales, con efectividad, articuladamente y con calidad (PEI vigente)

- **Criterios de éxito del negocio**

Desde el punto de vista del negocio se establece como criterio de éxito contar con la información que permita determinar la clasificación socioeconómica de la población.

Otro criterio de éxito del negocio es contar con otras fuentes de información que permitan cotejar la información del hogar para una correcta determinación de la clasificación socioeconómica.

4.3.1.2 EVALUACIÓN DE LA SITUACIÓN ACTUAL

En este capítulo, mostramos la situación actual de los recursos con los que cuenta el MIDIS, enfocado en el proceso de clasificación socioeconómica:

- **Inventario de Recursos**

- **Software:**

El MIDIS dispone de un sistema denominado SIGOF, el mismo que contiene información de la Clasificación Socioeconómica de Hogares (CSE). El SIGOF es un sistema que recopila información de los formatos S100, FSU y son cotejados con

otras bases de datos administrativas del Estado, que son proporcionado por RENIEC, Superintendencia Nacional de Migraciones, SUNAT, y OSINERGMIN.

Actualmente el MIDIS no cuenta con aplicativos que permitan analizar los datos haciendo uso de la minería de datos.

- **Personal:**

El MIDIS cuenta con personal de administración de bases de datos, analistas funcionales, analistas programadores, estadísticos, economistas, abogados, empadronadores, y personal especializado en focalización individual y geográfica.

No se cuenta con personal directamente vinculado al proceso de clasificación socioeconómica con conocimientos en herramientas de minería de datos.

- **Datos:**

Se cuenta con información de los ciudadanos contenidos en los formatos FSU y S100, así como información que provee el RENIEC sobre la identidad de las personas.

- **Requisitos**

Se requiere contar con información de hogares pertenecientes a diferentes regiones del Perú y que correspondan a un periodo de empadronamiento reciente para el análisis de la información.

- **Restricciones**

Debido a que los datos de la población para el proceso de clasificación socioeconómica es información sensible, para el presente trabajo no se cuenta con el acceso directo a la base de datos de MIDIS.

Para el presente estudio, se cuenta con algunos documentos físicos de los formatos S100 y FSU.

- **Riesgos y contingencias**

No contar con información fiable ni los mecanismos que permitan para verificar los datos.

4.3.1.3 DEFINIR LOS OBJETIVOS DE MINERÍA DE DATOS

Dentro de la presente propuesta, se determinan los siguientes objetivos de minería de datos:

Objetivo1:

En el análisis descriptivo, se busca identificar los patrones de los distintos grupos de clasificación socioeconómica, para lo cual se identificarán las variables que tienen mayor impacto en el proceso de clasificación socioeconómica de los hogares.

Objetivo2:

En el análisis predictivo, se busca que, en base al modelo definido, se pueda determinar el clúster de clasificación socioeconómico al que pertenece un hogar. Es decir, se pueda predecir en base al comportamiento de una familia, cuál debería ser el grupo socioeconómico al que pertenece.

Objetivo3:

Se busca poder identificar a los hogares que presenten alguna anomalía o que no puedan ser fácilmente agrupados a fin de determinar si corresponde a fraude o error.

Criterios de éxito de minería de datos

Entre los criterios de éxito para la minería de datos del presente trabajo se definen:

- Identificar al menos 3 variables que no tienen impacto en el proceso de clasificación socioeconómico: Este criterio tiene la finalidad de mostrar cómo el Analytics puede contribuir a identificar variables que no requieren análisis en el proceso, de tal manera que la institución pueda enfocarse en aquellas que tienen mayor relevancia.
- Establecer un modelo con al menos 80% de precisión en el proceso de clasificación socioeconómico: El mismo se realizará validando la información que pudo ser identificada correctamente.

4.3.1.4 PLAN DE PROYECTO

Como plan de proyecto se propone el siguiente esquema basado en las fases de la metodología seleccionada.

Tabla 13
Plan de Proyecto

Fase	Tiempo	Recursos	Riesgos
Comprensión de negocio	2 mes	Analistas	Desestimación de Proyecto
Comprensión de datos	2 meses	Analistas	Problemas de datos, problemas tecnológicos
Preparación de datos	2 meses	Experto en Base de Datos	Problemas para encontrar los datos adecuados
Modelado	3 meses	Experto en minería de datos	Incapacidad para encontrar un modelo adecuado
Evaluación	2 meses	Experto en minería de datos	Desestimación del Proyecto, incapacidad para evaluar resultados

Fuente: Elaboración propia

4.3.2 COMPRENSIÓN DE LOS DATOS

4.3.2.1 RECOPIACIÓN INICIAL DE DATOS

Para el presente trabajo se recolectó información de los formatos físicos S100 y FSU, correspondientes al periodo 2017/2018.

No se cuenta con información que permita la validación de la información y completar los datos de los hogares, por lo que se considerará que los datos son válidos para el presente trabajo. La información encontrada en los formatos físicos fue digitada en un único archivo en Excel, con el fin de tener la data simplificada para el proceso de minería de datos.

- **Fuente de datos:**

Tabla 14
Fuentes de Datos

ITEM	FUENTE DE DATOS	UBICACIÓN DE ORIGEN	UBICACIÓN DESTINO
1	S100	Formato físico	Excel
2	FSU	Formato físico	Excel

Fuente: Elaboración propia

Formato FSU

El formato FSU, es un mecanismo de recojo de información de la vivienda, el hogar y las personas residentes habituales, con el objeto de determinar la clasificación socioeconómica de las personas y/o sus hogares.

En la figura 23 mostramos el formato FSU, el cual se encuentra dividido en 4 hojas y contiene los datos que fueron digitados para el presente trabajo.

PADRÓN GENERAL DE HOGARES
FICHA SOCIOECONÓMICA ÚNICA

SisFoh
Sistema de Sistematización de Hogares

I. LOCALIZACIÓN DE LA VIVIENDA

UBICACIÓN GEOGRÁFICA

1. DEPARTAMENTO 2. PROVINCIA 3. DISTRITO

4. CENTRO POBLADO CÓDIGO CATEGORÍA

5. NÚCLEO URBANO (NO APLICABLE PARA CENTROS POBLADOS RURALES) CATEGORÍA

UBICACIÓN CENSAL

6. CONDOMINIO 7. ZONA 8. MANZANA 9. N.º FRENTES DE MANZANA 10. VIVIENDA N.º 11. ¿CUÁNTOS HOGARES HABITAN EN ESTA VIVIENDA? 12. HOGAR N.º

Los ítems N.º 6, 7, 8 y 9 no son aplicables para centros poblados rurales. Si (a): HOGAR es un conjunto de personas sean parientes o no, que ocupan en su totalidad o en parte una vivienda, comparten el mismo (a) comida principal y atienden en común otros necesidades básicas, con cargo a un presupuesto común.

13. APELLIDOS Y NOMBRES DEL INFORMANTE N.º ORDEN

14. UBICACIÓN DE LA VIVIENDA

TIPO DE VÍA: 1. AVENIDA 2. JIRÓN 3. CALLE 4. PASAJE 5. CARRETERA 6. OTRO

NOMBRE DE LA VÍA (EN LOS CENTROS POBLADOS RURALES, NOCAR ALGUNA REFERENCIA DE UBICACIÓN) N.º DE PUERTA

BLOCK PISO INTERIOR MANZANA LOTE KM. TELÉFONO DEL DOMICILIO

PERSONAL RESPONSABLE

APELLIDOS Y NOMBRES DEL EMPADRONADOR (A) DNI DEL EMPADRONADOR (A)

APELLIDOS Y NOMBRES DEL JEFE (A) DE BRIGADA DNI DEL JEFE (A) DE BRIGADA

APELLIDOS Y NOMBRES DEL REVISOR (A) DNI DEL REVISOR (A)

APELLIDOS Y NOMBRES DEL SUPERVISOR (A) DNI DEL SUPERVISOR (A)

II. ENTREVISTA Y SUPERVISIÓN

VISITA EMPADRONADOR(A)

FECHA (dd/mm/aa) RESULTADO VISITA (*)

PRIMERA SEGUNDA TERCERA

JEFE(A) DE BRIGADA

FECHA (dd/mm/aa) RESULTADO VISITA (*)

PRIMERA SEGUNDA TERCERA

REVISOR(A)

FECHA (dd/mm/aa) RESULTADO REVISIÓN (*)

PRIMERA SEGUNDA TERCERA

RESULTADO FINAL DEL EMPADRONAMIENTO

FECHA (dd/mm/aa) RESULTADO (*)

(*) CÓDIGOS DE RESULTADO

1. Completa 2. Incompleta 3. Rechazo 4. Ausente 5. No se inició la entrevista 6. Vivienda desocupada 7. Otro (Especificar) (Solo para el resultado final)

3. ¿Se obtuvo la firma del informante?

1. Si 2. No

a. Informante menor de edad b. Se negó a firmar c. No sabe firmar

00000001

CONSTANCIA DE EMPADRONAMIENTO

INFORMANTE

Estimado(a) Señor(a): Debido a la importancia y los objetivos del "Empadronamiento Distrital de Población y Vivienda", es necesario contar con la conformidad que su hogar ha sido empadronado y que la información proporcionada y los datos registrados en la Ficha Socioeconómica Única se ajustan a la verdad. Para dar fe de lo antes mencionado, se solicita su firma y huella digital por duplicado. Asimismo, yo, como responsable de haber realizado la entrevista en su hogar, firmo el documento.

1. ¿Vive Ud. permanentemente en la vivienda empadronada? 1. Si 2. No

2. Observe el sector de la vivienda empadronada y anote las características que se solicitan:

CARACTERÍSTICAS DE LA VIVIENDA

a. Número de "pisos" o niveles de la vivienda

b. Color del frente de la vivienda

1. No está pintado

Firma del Empadronador Firma del Informante Huella digital (informante)

3. ¿Se obtuvo la firma del informante?

1. Si 2. No

a. Informante menor de edad b. Se negó a firmar c. No sabe firmar

00000001

Figura 25. Formato FSU – página 1

Fuente: Página web del SISFOH – MIDIS, 2017

Pag. 02

III. CARACTERÍSTICAS DE LA VIVIENDA

ESCRIBA CON LETRA MAYÚSCULA IMPRINTA, NO ACENTUE LAS PALABRAS, CORRIJA CON CUIDADO, MARQUE A LA DERECHA LAS PALABRAS QUE CORRESPONDAN

1. TIPO DE VIVIENDA <small>(Observe la vivienda y marque solo el óvalo que corresponde)</small> <ol style="list-style-type: none"> 1 ☐ Casa independiente 2 ☐ Departamento en edificio 3 ☐ Vivienda en quinta 4 ☐ Vivienda en casa de vecindad (Callejón, solar o corredón) 5 ☐ Chozo o cabaña 6 ☐ Vivienda improvisada 7 ☐ Local no destinado para habitación humana 8 ☐ Otro (Especifique) _____ 	4. EL MATERIAL PREDOMINANTE EN LOS TECHOS ES: <small>(Lee cada alternativa y marque solo el óvalo que corresponde)</small> <ol style="list-style-type: none"> 1 ☐ Concreto armado? 2 ☐ Madera? 3 ☐ Tejas? 4 ☐ Planchas de calamina, fibra de cemento o similares? 5 ☐ Calia o estera con torta de barro? 6 ☐ Estera? 7 ☐ Paja, hojas de palmera? 8 ☐ Otro (Especifique) _____ 	7. EL ABASTECIMIENTO DE AGUA EN SU VIVIENDA PROCEDE DE: <small>(Lee cada alternativa y marque solo el óvalo que corresponde)</small> <ol style="list-style-type: none"> 1 ☐ Red pública dentro de la vivienda? 2 ☐ Red pública fuera de la vivienda, pero dentro del edificio? 3 ☐ Pilón de uso público? 4 ☐ Camión - cisterna u otro similar? 5 ☐ Pozo? 6 ☐ Río, acequia, manantial o similar? 7 ☐ Otro (Especifique) _____
2. SU VIVIENDA ES: <small>(Lee cada alternativa y marque solo el óvalo que corresponde)</small> <ol style="list-style-type: none"> 1 ☐ Alquilada? 2 ☐ Propia, pagándola a plazos? 3 ☐ Propia totalmente pagada? 4 ☐ Propia por inversión? 5 ☐ Cedió por el centro de trabajo? 6 ☐ Cedió por otro hogar o institución? 7 ☐ Otro (Especifique) _____ 	5. EL MATERIAL PREDOMINANTE EN LOS PISOS ES: <small>(Lee cada alternativa y marque solo el óvalo que corresponde)</small> <ol style="list-style-type: none"> 1 ☐ Parquet o madera pulida? 2 ☐ Láminas azulejadas, vinílicos o similares? 3 ☐ Losetas, terrazo o similares? 4 ☐ Madera (Entablado)? 5 ☐ Cemento? 6 ☐ Tierra? 7 ☐ Otro (Especifique) _____ 	3. ¿SU HOGAR TIENE: <small>(Lee cada alternativa y marque solo los óvalos que correspondan)</small> <ol style="list-style-type: none"> 1 ☐ Equipo de sonido? 2 ☐ Televisor a color? 3 ☐ DVD? 4 ☐ Licuadora? 5 ☐ Refrigeradora / congeladora? 6 ☐ Cocina a gas? 7 ☐ Teléfono fijo? 8 ☐ Plancha eléctrica? 9 ☐ Lavadora? 10 ☐ Computadora? 11 ☐ Horno microondas? 12 ☐ Internet? 13 ☐ Celular? 14 ☐ Otro? 15 ☐ No tiene ninguno
3. EL MATERIAL PREDOMINANTE EN LAS PAREDES EXTERIORES ES: <small>(Lee cada alternativa y marque solo el óvalo que corresponde)</small> <ol style="list-style-type: none"> 1 ☐ Ladrillo o bloque de cemento? 2 ☐ Piedra o sillar con cal o cemento? 3 ☐ Adobe o tapia? 4 ☐ Quincha (Caja con barro)? 5 ☐ Piedra con barro? 6 ☐ Madera? 7 ☐ Estera? 8 ☐ Otro (Especifique) _____ 	6. ¿CUÁL ES EL TIPO DE ALUMBRADO QUE TIENE SU VIVIENDA? <small>(Lee cada alternativa y marque solo el óvalo que corresponde)</small> <ol style="list-style-type: none"> 1 ☐ Electricidad? 2 ☐ Keroseno (Mechero/Lámpara)? 3 ☐ Petróleo / Gas (Lámpara)? 4 ☐ Vela? 5 ☐ Otro (Especifique) _____ 6 ☐ NO TIENE 	8. EL SERVICIO HIGIÉNICO (Water, latrina, etc.) QUE TIENE SU VIVIENDA ESTÁ CONECTADO A: <small>(Lee cada alternativa y marque solo el óvalo que corresponde)</small> <ol style="list-style-type: none"> 1 ☐ Red pública dentro de la vivienda? 2 ☐ Red pública fuera de la vivienda, pero dentro del edificio? 3 ☐ Pozo séptico? 4 ☐ Pozo ciego o negro / latrina? 5 ☐ Río, acequia o canal? 6 ☐ NO TIENE
IV. DATOS DEL HOGAR		
1. ¿CÓMO CONTAR BAÑO, COCINA, PASADIZOS, NI GARAJE? ¿CUÁNTAS HABITACIONES OCUPA ESTE HOGAR? Anote la respuesta en los recuadros: 		
2. ¿CUÁL ES EL COMBUSTIBLE QUE MÁS SE UTILIZA EN EL HOGAR PARA COCINAR? <small>(Lee cada alternativa y marque solo el óvalo que corresponde)</small> <ol style="list-style-type: none"> 1 ☐ Electricidad? 2 ☐ Gas? 3 ☐ Keroseno? 4 ☐ Carbón? 5 ☐ Leña? 6 ☐ Bestia o estiércol? 7 ☐ Otro (Especifique) _____ 8 ☐ NO COCINA 		

4. ¿CUÁL ES EL NÚMERO DE SU SUMINISTRO DE LUZ (1) O AGUA (2)?
 1 ☐ Luz 2 ☐ Agua 3 ☐ NO TIENE

5. ¿CUÁNTAS PERSONAS VIVEN PERMANENTEMENTE EN ESTE HOGAR?

TOTAL	HOMBRES	MUJERES
<div style="border: 1px solid black; width: 30px; height: 30px; display: flex; align-items: center; justify-content: center;"> <div style="width: 10px; height: 10px; border: 1px solid black;"></div> <div style="width: 10px; height: 10px; border: 1px solid black;"></div> <div style="width: 10px; height: 10px; border: 1px solid black;"></div> </div>	<div style="border: 1px solid black; width: 30px; height: 30px; display: flex; align-items: center; justify-content: center;"> <div style="width: 10px; height: 10px; border: 1px solid black;"></div> <div style="width: 10px; height: 10px; border: 1px solid black;"></div> <div style="width: 10px; height: 10px; border: 1px solid black;"></div> </div>	<div style="border: 1px solid black; width: 30px; height: 30px; display: flex; align-items: center; justify-content: center;"> <div style="width: 10px; height: 10px; border: 1px solid black;"></div> <div style="width: 10px; height: 10px; border: 1px solid black;"></div> <div style="width: 10px; height: 10px; border: 1px solid black;"></div> </div>

Incluya:

- A las personas que viven permanentemente en este hogar,
- A las personas que viven permanentemente en este hogar pero se encuentran ausentes temporalmente por trabajo, negocios, vacaciones, internados en un establecimiento de salud u otros motivos por un período igual o menor a seis (6) meses.

No incluya:

- A las personas que viven permanentemente en otro hogar,
- A las personas que viven permanentemente en instituciones tales como hospitales psiquiátricos, instituciones de protección a menores, asilos, cárceles, conventos, etc.
- A las personas que están ausentes del hogar por un período mayor a seis (6) meses.

Figura 26. Formato FSU – pagina2

Fuente: Página web del SISFOH – MIDIS, 2017

Pag. 03

V. CARACTERÍSTICAS DE LA POBLACIÓN										PARENTESCO										NIVEL EDUCATIVO										NIVEL OCUPACIONAL										NIVEL DE DISCAPACIDAD																																																																																																			
<p>Apellido (En el primer renglón registre el apellido paterno y en el segundo el apellido materno, según DNI)</p> <p>Nombre (En el tercer renglón registre los nombres separados por un espacio, según DNI)</p> <p>Fecha de Nacimiento (separada) / Edad (Años completos. Marque meses sólo para menores de 01 año)</p> <p>Tipo de documento: Seleccione una alternativa / Número de documento (escribir el número de documento. En el caso de la Partida de Nacimiento, poner sólo el número de CUI, de lo contrario no completar)</p> <p>(1) / (2) / (3) / (4) / (5) / (6) / (7)</p>										<p>Parentesco con el jefe de hogar</p> <p>(8)</p>										<p>Nº de núcleo familiar (si no es al núcleo)</p> <p>(9)</p>										<p>Sexo (Si es M indicar al ser masculino)</p> <p>(10)</p>										<p>Estado Civil (1) Soltero (2) Casado (3) Viudo (4) Divorciado</p> <p>(11)</p>										<p>¿Qué tipo de seguro de salud tiene?</p> <p>(12)</p>										<p>¿Qué tipo de seguro de salud tiene?</p> <p>(13)</p>										<p>¿Qué tipo de seguro de salud tiene?</p> <p>(14)</p>										<p>¿Qué tipo de seguro de salud tiene?</p> <p>(15)</p>										<p>¿Qué tipo de seguro de salud tiene?</p> <p>(16)</p>										<p>¿Qué tipo de seguro de salud tiene?</p> <p>(17)</p>										<p>¿Qué tipo de seguro de salud tiene?</p> <p>(18)</p>										<p>¿Qué tipo de seguro de salud tiene?</p> <p>(19)</p>										<p>¿Qué tipo de seguro de salud tiene?</p> <p>(20)</p>									
<p>01</p> <p>APELLIDO PATERNO</p> <p>APELLIDO MATERNO</p> <p>NOMBRES</p> <p>Fecha de Nacimiento: / /</p> <p>Edad (años):</p> <p>Meses:</p> <p>Tipo de Documento: (1) DNI (2) Cui (3) Cui Ex (4) No tiene doc.</p> <p>Número de Documento:</p>										<p>(8)</p>										<p>(9)</p>										<p>(10)</p>										<p>(11)</p>										<p>(12)</p>										<p>(13)</p>										<p>(14)</p>										<p>(15)</p>										<p>(16)</p>										<p>(17)</p>										<p>(18)</p>										<p>(19)</p>										<p>(20)</p>									
<p>02</p> <p>APELLIDO PATERNO</p> <p>APELLIDO MATERNO</p> <p>NOMBRES</p> <p>Fecha de Nacimiento: / /</p> <p>Edad (años):</p> <p>Meses:</p> <p>Tipo de Documento: (1) DNI (2) Cui (3) Cui Ex (4) No tiene doc.</p> <p>Número de Documento:</p>										<p>(8)</p>										<p>(9)</p>										<p>(10)</p>										<p>(11)</p>										<p>(12)</p>										<p>(13)</p>										<p>(14)</p>										<p>(15)</p>										<p>(16)</p>										<p>(17)</p>										<p>(18)</p>										<p>(19)</p>										<p>(20)</p>									
<p>03</p> <p>APELLIDO PATERNO</p> <p>APELLIDO MATERNO</p> <p>NOMBRES</p> <p>Fecha de Nacimiento: / /</p> <p>Edad (años):</p> <p>Meses:</p> <p>Tipo de Documento: (1) DNI (2) Cui (3) Cui Ex (4) No tiene doc.</p> <p>Número de Documento:</p>										<p>(8)</p>										<p>(9)</p>										<p>(10)</p>										<p>(11)</p>										<p>(12)</p>										<p>(13)</p>										<p>(14)</p>										<p>(15)</p>										<p>(16)</p>										<p>(17)</p>										<p>(18)</p>										<p>(19)</p>										<p>(20)</p>									
<p>04</p> <p>APELLIDO PATERNO</p> <p>APELLIDO MATERNO</p> <p>NOMBRES</p> <p>Fecha de Nacimiento: / /</p> <p>Edad (años):</p> <p>Meses:</p> <p>Tipo de Documento: (1) DNI (2) Cui (3) Cui Ex (4) No tiene doc.</p> <p>Número de Documento:</p>										<p>(8)</p>										<p>(9)</p>										<p>(10)</p>										<p>(11)</p>										<p>(12)</p>										<p>(13)</p>										<p>(14)</p>										<p>(15)</p>										<p>(16)</p>										<p>(17)</p>										<p>(18)</p>										<p>(19)</p>										<p>(20)</p>									
<p>05</p> <p>APELLIDO PATERNO</p> <p>APELLIDO MATERNO</p> <p>NOMBRES</p> <p>Fecha de Nacimiento: / /</p> <p>Edad (años):</p> <p>Meses:</p> <p>Tipo de Documento: (1) DNI (2) Cui (3) Cui Ex (4) No tiene doc.</p> <p>Número de Documento:</p>										<p>(8)</p>										<p>(9)</p>										<p>(10)</p>										<p>(11)</p>										<p>(12)</p>										<p>(13)</p>										<p>(14)</p>										<p>(15)</p>										<p>(16)</p>										<p>(17)</p>										<p>(18)</p>										<p>(19)</p>										<p>(20)</p>									

(8) Relación de Parentesco

1. Jefe

2. Cónyuge

3. Hija

4. Yerno

5. Nieto

6. Padre

7. Hermano

8. Tío

9. Primos

10. Otros

(12) Estado Civil

1. Soltero

2. Casado

3. Viudo

4. Divorciado

5. Separado

6. Otro

(13) Seguro de Salud

1. Seguro Social

2. Seguro Privado

3. Seguro de Salud

4. Seguro de Vida

5. Seguro de Accidentes

6. Seguro de Faltas

7. Seguro de Retiro

8. Seguro de Pensiones

9. Seguro de Indemnización

10. Seguro de Seguro

(14) Idioma o Lengua

1. Quechua

2. Aymara

3. Guaraní

4. Mestizo

5. Otro

(15) Nivel Educativo

1. Ninguno

2. Primaria

3. Secundaria

4. Superior

5. Superior Universitario

6. Superior Técnico

7. Superior Tecnológico

8. Superior Tecnológico

9. Superior Tecnológico

10. Superior Tecnológico

(16) Último año o Grado aprobado

1. Ninguno

2. Primaria

3. Secundaria

4. Superior

5. Superior Universitario

6. Superior Técnico

7. Superior Tecnológico

8. Superior Tecnológico

9. Superior Tecnológico

10. Superior Tecnológico

(17) Ocupación

1. Trabajador dependiente

2. Trabajador independiente

3. Empleador

4. Trabajador del hogar

5. Trabajador no remunerado

6. Desempleado

7. Desempleado a la espera de un empleo

8. Desempleado

9. Desempleado

10. Desempleado

(18) Sector

1. Agrícola

2. Pecuaria

3. Forestal

4. Pesca

5. Minería

6. Industria

7. Comercio

8. Servicios

9. Otros

10. Otros

(19) Tipo de Discapacidad

1. Visual

2. Auditiva

3. Física

4. Psíquica

5. Mental

6. Sensorial

7. Intelectual

8. Comunicacional

9. Social

10. Otros

(20) Beneficiario de:

1. Seguro Social

2. Seguro Privado

3. Seguro de Salud

4. Seguro de Vida

5. Seguro de Accidentes

6. Seguro de Faltas

7. Seguro de Retiro

8. Seguro de Pensiones

9. Seguro de Indemnización

10. Seguro de Seguro

Figura 27. Formato FSU – pagina3

Fuente: Página web del SISFOH – MIDIS, 2017

V. CARACTERÍSTICAS DE LA POBLACIÓN										Pag. 04																			
<p>Apellidos (En el primer renglón registre el apellido paterno y en el segundo el apellido materno, según DNI) Nombres (En el tercer renglón registre los nombres separados por un espacio, según DNI) Fecha de Nacimiento (dd-mm-aaaa) / Edad (Años cumplidos. Marque meses solo para menores de 01 año) Tipo de documento: Seleccionar sólo una alternativa / Número de documento (seleccionar el número de documento. En el caso de la Partida de Nacimiento, poner sólo el número de CUI, de lo contrario no completar)</p>										Parentesco con jefe de hogar	N° de niños menores de 18 años	Sexo	Estado Civil	¿Cuál es su nivel de escolaridad?	¿Cuál es su nivel de alfabetización?	¿Cuál es su nivel de educación superior?	¿Cuál es su nivel de educación superior?	¿Cuál es su nivel de educación superior?	¿Cuál es su nivel de educación superior?	¿Cuál es su nivel de educación superior?	¿Cuál es su nivel de educación superior?	¿Cuál es su nivel de educación superior?	¿Cuál es su nivel de educación superior?	¿Cuál es su nivel de educación superior?	¿Cuál es su nivel de educación superior?	¿Cuál es su nivel de educación superior?	¿Cuál es su nivel de educación superior?	¿Cuál es su nivel de educación superior?	¿Cuál es su nivel de educación superior?
										(8)	(9)	(10)	(11)	(12)	(13)	(14)	(15)	(16)	(17)	(18)	(19)	(20)							
06										1	2	3	4	5	6	7	8	9	10	11	12	13	14						
APELLIDO PATERNO										1	2	3	4	5	6	7	8	9	10	11	12	13	14						
APELLIDO MATERNO										1	2	3	4	5	6	7	8	9	10	11	12	13	14						
NOMBRES										1	2	3	4	5	6	7	8	9	10	11	12	13	14						
Fecha de Nacimiento										1	2	3	4	5	6	7	8	9	10	11	12	13	14						
Edad (años)										1	2	3	4	5	6	7	8	9	10	11	12	13	14						
Meses										1	2	3	4	5	6	7	8	9	10	11	12	13	14						
Tipo de Documento										1	2	3	4	5	6	7	8	9	10	11	12	13	14						
Número de Documento										1	2	3	4	5	6	7	8	9	10	11	12	13	14						
07										1	2	3	4	5	6	7	8	9	10	11	12	13	14						
APELLIDO PATERNO										1	2	3	4	5	6	7	8	9	10	11	12	13	14						
APELLIDO MATERNO										1	2	3	4	5	6	7	8	9	10	11	12	13	14						
NOMBRES										1	2	3	4	5	6	7	8	9	10	11	12	13	14						
Fecha de Nacimiento										1	2	3	4	5	6	7	8	9	10	11	12	13	14						
Edad (años)										1	2	3	4	5	6	7	8	9	10	11	12	13	14						
Meses										1	2	3	4	5	6	7	8	9	10	11	12	13	14						
Tipo de Documento										1	2	3	4	5	6	7	8	9	10	11	12	13	14						
Número de Documento										1	2	3	4	5	6	7	8	9	10	11	12	13	14						
08										1	2	3	4	5	6	7	8	9	10	11	12	13	14						
APELLIDO PATERNO										1	2	3	4	5	6	7	8	9	10	11	12	13	14						
APELLIDO MATERNO										1	2	3	4	5	6	7	8	9	10	11	12	13	14						
NOMBRES										1	2	3	4	5	6	7	8	9	10	11	12	13	14						
Fecha de Nacimiento										1	2	3	4	5	6	7	8	9	10	11	12	13	14						
Edad (años)										1	2	3	4	5	6	7	8	9	10	11	12	13	14						
Meses										1	2	3	4	5	6	7	8	9	10	11	12	13	14						
Tipo de Documento										1	2	3	4	5	6	7	8	9	10	11	12	13	14						
Número de Documento										1	2	3	4	5	6	7	8	9	10	11	12	13	14						
09										1	2	3	4	5	6	7	8	9	10	11	12	13	14						
APELLIDO PATERNO										1	2	3	4	5	6	7	8	9	10	11	12	13	14						
APELLIDO MATERNO										1	2	3	4	5	6	7	8	9	10	11	12	13	14						
NOMBRES										1	2	3	4	5	6	7	8	9	10	11	12	13	14						
Fecha de Nacimiento										1	2	3	4	5	6	7	8	9	10	11	12	13	14						
Edad (años)										1	2	3	4	5	6	7	8	9	10	11	12	13	14						
Meses										1	2	3	4	5	6	7	8	9	10	11	12	13	14						
Tipo de Documento										1	2	3	4	5	6	7	8	9	10	11	12	13	14						
Número de Documento										1	2	3	4	5	6	7	8	9	10	11	12	13	14						
10										1	2	3	4	5	6	7	8	9	10	11	12	13	14						
APELLIDO PATERNO										1	2	3	4	5	6	7	8	9	10	11	12	13	14						
APELLIDO MATERNO										1	2	3	4	5	6	7	8	9	10	11	12	13	14						
NOMBRES										1	2	3	4	5	6	7	8	9	10	11	12	13	14						
Fecha de Nacimiento										1	2	3	4	5	6	7	8	9	10	11	12	13	14						
Edad (años)										1	2	3	4	5	6	7	8	9	10	11	12	13	14						
Meses										1	2	3	4	5	6	7	8	9	10	11	12	13	14						
Tipo de Documento										1	2	3	4	5	6	7	8	9	10	11	12	13	14						
Número de Documento										1	2	3	4	5	6	7	8	9	10	11	12	13	14						

(8) Reducción de Parentesco

1. Jefe
2. Cónyuge
3. Hijos
4. Hermanos
5. Nietos
6. Padres/Vuagos
7. Hermanos
8. Trabajador del hogar
9. Pensiónista
10. Otros parientes

(12) Estado Civil

1. Soltero/a
2. Casado/a
3. Viudo/a
4. Separado/a
5. Divorciado/a
6. Viudo/a

(13) Seguro de Salud

1. Seguro Social
2. Seguro privado
3. Seguro médico
4. Seguro dental
5. Seguro de vida
6. Seguro de invalidez
7. Seguro de discapacidad
8. Seguro de vejez
9. Seguro de maternidad
10. Seguro de paternidad
11. Seguro de maternidad
12. Seguro de paternidad
13. Seguro de maternidad
14. Seguro de paternidad
15. Seguro de maternidad
16. Seguro de paternidad
17. Seguro de maternidad
18. Seguro de paternidad
19. Seguro de maternidad
20. Seguro de paternidad

(14) Idioma o lengua aprendida en la niñez

1. Guaraní
2. Mestizo
3. Mestizo
4. Mestizo
5. Mestizo
6. Mestizo
7. Mestizo
8. Mestizo
9. Mestizo
10. Mestizo
11. Mestizo
12. Mestizo
13. Mestizo
14. Mestizo
15. Mestizo
16. Mestizo
17. Mestizo
18. Mestizo
19. Mestizo
20. Mestizo

(15) Nivel Educativo

1. Ninguno
2. Primaria
3. Secundaria
4. Tercera
5. Superior no universitaria
6. Superior universitaria
7. Posgrado o doctorado
8. Otro

(16) Último año o Grado aprobado

1. Trabajador dependiente
2. Trabajador independiente
3. Empleado
4. Trabajador del hogar
5. Trabajador temporal
6. Desempleado
7. Desempleado a la espera de un trabajo
8. Jubilado
9. Otro
10. Sin actividad

(17) Ocupación (según el CIEC)

1. Agrícola
2. Pecuaria
3. Industrial
4. Comercio
5. Servicios
6. Construcción
7. Transportación
8. Comunicaciones
9. Otros
10. Estado (Cód.)

(18) Tipo de Discapacidad

1. Visual
2. Auditiva
3. Física
4. Psíquica
5. Intelectual
6. Sensorial
7. Mixta
8. Otra
9. No tiene discapacidad
10. No tiene discapacidad
11. No tiene discapacidad
12. No tiene discapacidad
13. No tiene discapacidad
14. No tiene discapacidad
15. No tiene discapacidad
16. No tiene discapacidad
17. No tiene discapacidad
18. No tiene discapacidad
19. No tiene discapacidad
20. No tiene discapacidad

(19) Beneficiario de:

1. Voto de Nulidad
2. Control popular
3. Desempleado a la espera de un trabajo
4. Seguro de maternidad
5. Seguro de paternidad
6. Seguro de maternidad
7. Seguro de paternidad
8. Seguro de maternidad
9. Seguro de paternidad
10. Seguro de maternidad
11. Seguro de paternidad
12. Seguro de maternidad
13. Seguro de paternidad
14. Seguro de maternidad
15. Seguro de paternidad
16. Seguro de maternidad
17. Seguro de paternidad
18. Seguro de maternidad
19. Seguro de paternidad
20. Seguro de maternidad

Figura 28. Formato FSU – pagina4

Fuente: Página web del SISFOH – MIDIS, 2017

- Formato S100

El formato S100, es un instrumento de recojo de datos de los hogares que permite a las personas solicitar su clasificación socioeconómica mediante las Unidades Locales de Empadronamiento.

En la figura 29 mostramos el formato S100, el cual se encuentra dividido en 2 hojas y contiene los datos que fueron digitados para el presente trabajo.

5. ACCESO A SERVICIOS PÚBLICOS

5.1 INDIQUE SI CUENTA CON EL RECIBO DE SERVICIOS PÚBLICOS DE:

1 LUZ ☐ 2 AGUA ☐ 3 LUZ Y AGUA ☐ 4 NINGUNO ☐

En caso marcar algún servicio detallar el número de suministro y el nombre de la empresa proveedora. Luego continuar con la pregunta 6.2. En caso de marcar ninguno continuar con la sección 6.

5.2 ¿EL SERVICIO DE LUZ ES COMPARTIDO, COMUNAL Y/O COOPERATIVO? 1 SI ☐ 2 NO ☐

NÚMERO DE SUMINISTRO DE LUZ:

NOMBRE DE LA EMPRESA DE LUZ:

NÚMERO DE SUMINISTRO DE AGUA:

NOMBRE DE LA EMPRESA DE AGUA:

6. DIRECCIÓN DE LA VIVIENDA U HOGAR DONDE RESIDE EL HOGAR O PERSONA

6.1 TIPO DE VÍA (Marcar una opción): 1 AVENIDA ☐ 2 JUNCIÓN ☐ 3 CALLE ☐ 4 PASAJE ☐ 5 CARRETERA ☐ 6 OTRO ☐

6.2 NOMBRE DE LA VÍA:

6.2.1 N° DE PUERTA:

6.2.2 BLOCK 6.2.3 PISO 6.2.4 INTERIOR 6.2.5 MANZANA 6.2.6 LOTE 6.2.7 KM 6.2.8 NÚCLEO URBANO

6.3 REFERENCIA DE UBICACIÓN DE LA VIVIENDA O LUGAR:

7. UBICACIÓN GEOGRÁFICA

7.1 LA VIVIENDA O LUGAR SE ENCUENTRA UBICADA EN UN: 1 URBANO ☐ 2 RURAL ☐

7.2 CÓDIGO CENTRO POBLADO:

7.3 NOMBRE DEL CENTRO POBLADO:

8. AUTORIZACIÓN Y DECLARACIÓN DE LA PERSONA SOLICITANTE

1- Expreso mi consentimiento para que el SISFOH use, en el marco de sus funciones, mis datos personales. Considerando que la atención de mi solicitud involucra el tratamiento de datos personales, que en algunos casos constituyen datos sensibles, por lo que doy mi libre y expreso consentimiento para: (i) el acceso a bases de datos personales, ya sean públicas o privadas, incluyendo datos sensibles, que contenga información personal del solicitante o de los integrantes de mi hogar; y (ii) Poner a disposición de las Intervenciones Públicas Focalizadas el resultado de la CSE de mi hogar, en el marco de sus funciones. (Artículo 14, del Decreto Supremo N° 33-2012-JUS Reglamento Ley N° 20123).

2- Autorizo que el resultado de la presente solicitud, o en su defecto las observaciones que se formulen, sean notificadas a la dirección de correo electrónico (e-mail) o al número de teléfono - celular señalado en la sección 1 "Datos de la Persona Solicitante" (según numeral 30.1 y numeral 30.4 del artículo 20 de la Ley N° 27444) vía SMS.

3- Declaro bajo juramento que quien suscribe y todos los integrantes de mi hogar (incluyendo al solicitante), residentes de forma permanente y, a la fecha, ningún miembro se encuentra ausente más de seis meses de este hogar. Comprensivamente incluyo en mi hogar a uno o varios jóvenes menores de 25 años de edad que se encuentran cursando estudios y que dependen económicamente de mi hogar o las personas que se encuentran prestando servicio militar.

4- Que, además, declaro que la suma de los ingresos provenientes de negocios e inversiones de todos los miembros de mi hogar, en los últimos doce (12) meses, expresados en nuevos soles, asciende a:

5- Que declaro contar con propiedades (Incluir todas las propiedades inmuebles registradas ante Registros Públicos - SUNARP así como aquellas que no han sido registradas). INDICAR NÚMERO DE PROPIEDADES: INDICAR NÚMERO DE VEHÍCULOS MOTORIZADOS:

6- En caso se compruebe que las autorizaciones y declaraciones proporcionadas en la presente solicitud no corresponden a la realidad o sean falsas, me someto a las responsabilidades administrativas, civiles y penales correspondientes (Artículo 19, Numeral 1.7 de Principios de Procedimiento Administrativo y Artículo 30, numeral 32.3 de la Ley N° 27444).

Firma de esta solicitud: Huella digital de esta solicitud:

9. FIRMA DEL SERVIDOR/A PÚBLICO/A DE LA INTERVENCIÓN PÚBLICA FOCALIZADA, U OTRA INSTITUCIÓN QUE SOLICITA LA VERIFICACIÓN

Firma del servidor público: Huella digital del servidor público:

10. UBICACIÓN CENSAL (Se llenará por la Municipalidad, Intervención Pública Focalizada, u otra institución encargada)

10.1 CONDOMINIO N°: 10.2 ZONA N°: 10.3 MANZANA N°: 10.4 N° PTE. DE MZA.: 10.5 VIVIENDA N°: 10.6 ¿CUÁNTOS HOGARES HABITAN EN ESTA VIVIENDA?: 10.7 HOGAR N°:

11. EL/LA SOLICITANTE DECLARA QUE EL HOGAR RESIDE EN EL ÁMBITO DE UN CC.PP. UBICADO EN EL LISTADO OFICIAL VIGENTE DEL MINISTERIO DE CULTURA

Según listado oficial vigente del Ministerio de Cultura.

1 SI ☐ 2 NO ☐ De marcar "SI" deberá indicar el Pueblo Indígena y la Comunidad nativa a la que pertenece. COMUNIDAD NATIVA: PUEBLO INDÍGENA:

DECLARACIÓN JURADA DE RESIDENCIA EN CC.PP. UBICADO EN EL ÁMBITO DEL LISTADO OFICIAL VIGENTE DEL MINISTERIO DE CULTURA REGISTRADO POR EL/LA ALCALDE/SA

Esta sección debe ser llenada en caso que la dirección registrada en el DNI de alguno de los integrantes del hogar no coincide con el lugar de residencia actual.

N° DEL DOCUMENTO DE IDENTIDAD DEL/DE LA ALCALDE/SA:

Firma del/la Alcalde/a: Huella digital del/la Alcalde/a:

12. DATOS DE RECEPCIÓN DE LA SOLICITUD

N° DEL DOCUMENTO DE IDENTIDAD DEL/DE LA SERVIDOR/A PÚBLICA QUE RECIBE/RECIBIENTA LA SOLICITUD: NOMBRES Y APELLIDOS DEL/DE LA SERVIDOR/A PÚBLICA QUE RECIBE/RECIBIENTA LA SOLICITUD:

INSTITUCIÓN QUE RECIBE LA SOLICITUD:

1 ULE ☐ 2 DOP ☐ 3 PENSIÓN 60 ☐ 4 JUNTOS ☐ 5 PROMABIC ☐ 6 SIS ☐

7 OTRO ☐ Detallar:

Firma del/la registrador/a: Huella digital del/la registrador/a:

Figura 30. Formato S100 – pagina2

Fuente: Página web del SISFOH – MIDIS, 2017

4.3.2.2 PROPORCIONAR UNA DESCRIPCIÓN DE LOS DATOS

En la tabla 14 y 15 se describen los datos de los formatos del S100 y FSU:

- **Formato S100**

Tabla 15

Datos contenidos en el formato S100

Item	Datos del formato S100	Descripción
Datos de generales de la Solicitud		
1	Número de Solicitud	Este número está impreso en el formato, es correlativo y asignado por la DOF
2	Ubigeo	Es el código de ubicación geográfica que identifica a un distrito. Se consigna el dato que aparece en el listado de centros poblados proporcionado por la DOF
3	Documento de Identidad	Se consigna el número del Documento Nacional de Identidad (DNI) de la persona que realiza la solicitud. De ser el caso también puede consignar el número de Carnet de Extranjería (CE).
4	Fecha	Fecha de presentación de la solicitud ante la ULE, o IPF bajo el siguiente formato DÍA / MES / AÑO.
Datos de la persona solicitante		
5	Apellido Paterno	Se consigna el apellido paterno conforme al documento de identidad
6	Apellido Materno	Se consigna el apellido materno conforme al documento de identidad
7	Nombres	Se consiga nombre (s) conforme al documento de identidad
8	Tipo de Solicitante	Persona que realiza la solicitud
9	Correo Electrónico	Correo del solicitante
10	Teléfono (Celular)	Número de solicitud
Registro de la situación del hogar respecto a su solicitud		
11	Situación del hogar	Se debe seleccionar una opción: <ul style="list-style-type: none"> • No tiene CSE. • Incorporación de menores de edad • Incorporación o salida del hogar de uno o más integrantes • Cambio de domicilio de todos/as los y las integrantes del hogar.

		<ul style="list-style-type: none"> • No estoy de acuerdo con el resultado de la CSE (Reevaluación) • Se requiere verificar información del hogar • Otras circunstancias
Datos de los integrantes del hogar		
12	Tipo de Documento	Marcar tipo de documento del integrante del hogar: <ul style="list-style-type: none"> • DNI • Certificado nacido vivo manual • Certificado nacido vivo electrónico • Acta de nacimiento • RENIEC CUI • Acta de Nacimiento Manual y Acta Nacimiento Electrónica • Carné de Extranjería • No tiene
13	Apellido Paterno	Se consigna el apellido paterno conforme al documento de identidad
14	Apellido Materno	Se consigna el apellido materno conforme al documento de identidad
15	Nombres	Se consiga nombre (s) conforme al documento de identidad
16	Fecha de Nacimiento	Se consiga la fecha de nacimiento conforme al documento de identidad
17	Sexo	Se consiga el sexo conforme al documento de identidad
18	Tipo de parentesco	Se consigna el tipo de parientes según la lista siguiente: <ul style="list-style-type: none"> • Jefe de Hogar • Cónyuge/Concubina • Hijo/a • Yerno/Nuera • Nieto/a • Padres/suegros • Hermano/a • Trabajador/a del Hogar

		<ul style="list-style-type: none"> • Pensionista • Otros parientes • Otros no parientes
19	Residente habitual	Se consigna en el recuadro el número 1 si la persona es residente habitual del hogar, o el número 2 si no es residente habitual.
Acceso a servicios público		
20	Cuenta con servicios público	Debe seleccionar una de las siguientes opciones: <ul style="list-style-type: none"> • Luz • Agua • Luz y agua • Ninguno
21	El servicio público es compartido	De tener el servicio de energía eléctrica compartido registre el número 1 en el recuadro de la opción “SI”, caso contrario registre el número 2 en el recuadro de la opción “NO”.
22	Numero de suministro de luz	Se debe consignar el número de suministro de luz
23	Nombre de la empresa de luz	Se debe consignar el nombre de la empresa de la luz
24	Número de suministro de agua	Se debe consignar el número de suministro de agua
25	Nombre de la empresa de agua	Se debe consignar el nombre de la empresa de agua
Datos de la vivienda		
26	Tipo de vía	Debe seleccionar una de las siguientes opciones: <ul style="list-style-type: none"> • Avenida • Jirón • Calle • Pasaje • Carretera • Otro
27	Nombre de la vía	Se consigna el nombre de la vía
Datos de la ubicación geográfica		
28	Nombre del centro poblado	Consignar el centro poblado del INEI
29	Código de centro poblado	Consignar el código centro poblado del INEI

Declaración de ingresos		
30	Monto de ingresos	Suma de ingresos por hogar
Datos de la ubicación censal		
31	Conglomerado	Se deben dejar en blanco
32	Zona	Se deben quedar en blanco
33	Manzana	Se deben quedar en blanco
34	Número de vivienda	Se deben quedar en blanco
35	Cantidad de hogares	Número de hogares que habitan en la vivienda
36	Número de hogar	Número que corresponde al hogar de la persona solicitante
Datos de recepción de la solicitud de la entidad		
37	Número de documento del servidor	Consignar número de documento del servidor
38	Nombres y apellido del servidor	Consignar apellidos y nombres del servidor
39	Institución que recibe la solicitud	Las instituciones pueden ser: ULE, DOF, PENSION 65, JUNTOS, PRONABEC, SIS, OTROS

- **Formato SFU**

Tabla 16
Datos contenidos en el formato FSU

Item	Datos del FSU	Descripción
1	Nro FSU	Número de la ficha de clasificación socioeconómica
Ubicación geográfica		
2	Cod.Dpto	Código del departamento
3	Dpto	Descripción del Departamento
4	Cod.Prov	Código de provincia
5	Provincia	Descripción de la provincia
6	Cod.Dist	Código del distrito
7	Distrito	Descripción del distrito

8	Centro poblado	Descripción del centro poblado
9	Cod.CCPP	Código del centro poblado
10	Categoría	Categoría
11	Núcleo urbano	Núcleo urbano
Ubicación censal		
12	Cong.Nº	Conglomerado
13	Zona Nº	Zona
14	Mz Nº	Manzana
15	Nº frente a Mz	Número de frente de manzana
16	Vivienda Nº	Número de vivienda
17	Cant.hogares	Cantidad de hogares
18	Hogar Nº	Número de hogar
19	Apellido pat	Apellido paterno
20	Apellido mat	Apellido materno
21	Nombres	Nombres
22	Nro. Orden	Número de orden
23	Tipo de vía	Tipo de vía
24	Nombre de vía	Nombre de la vía
25	Nº puerta	Número de la puerta
26	Block	Block a donde pertenece
27	Piso	Número de piso
28	Int	Interior
29	Mz	Manzana
30	Lt	Lote
31	Km	Kilómetro
33	Teléfono	Teléfono
34	Fecha empadronamiento.	Fecha de empadronamiento
35	Nº pisos vivienda	Número de pisos de vivienda
36	Color frontis	Color de frontis
Datos del hogar		

37	Tipo vivienda	1=Casa independiente 2=Departamento en edificio 3=Vivienda en quinta 4=Vivienda en casa de vecindad (Callejón solar o corralón) 5=Choza o cabaña 6=Vivienda improvisada 7=Local no destinado para habitación humana 8=Otro (Especifique)
38	Propiedad	1=Alquilada 2=Propia, pagándola a plazos 3=Propia totalmente pagada 4=Propia por invasión 5=Cedida por el centro de trabajo 6=Cedida por otro hogar o institución 7=Otro (Especifique)
39	Material pared	1=Ladrillo o bloque de cemento 2=Piedra o sillar con cal o cemento 3=Adobe o tapia 4=Quincha (Caña con barro) 5=Piedra con barro 6=Madera 7=Estera 8=Otro (Especifique)
40	Material techo	1=Concreto armado 2=Madera 3=Tejas 4=Planchas de calamina, fibra de cemento o similares 5=Caña o estera con torta de barro 6=Estera 7=Paja, Hojas de palmera

		8=Otro (Especifique)
41	Material piso	1=Parquet o madera pulida 2=Láminas, vinílicos o similares 3=Losetas, terrazos o similares 4=Madera (Entablados) 5=Cemento 6=Tierra 7=Otro (Especifique)
42	Alumbrado	1=Electricidad 2=Kerosene (Mechero/Lamparin) 3=Petróleo/Gas (Lampara) 4=Vela 5=Otro 6=No tiene
43	Agua	1=Red pública dentro de la vivienda 2=Red pública fuera de la vivienda y dentro del edificio 3=Pilón de uso público 4=Camión - Cisterna u otro similar 5=Pozo 6=Rio, acequia, manantial o similar
44	Servicio higiénico	1=Red pública dentro de la vivienda 2=Red pública fuera de la vivienda y dentro del edificio 3=Pozo séptico 4=Pozo ciego o negro/letrina 5=Rio, acequia o canal 6=No tiene
45	Horas llegar a vivienda	Cantidad de horas que se demoran desde su vivienda hasta llegar a la capital: 0=1 a 24 Hs 1=Más de 24 horas
46	Nº habitac. del hogar	Cantidad de habitaciones que ocupa el hogar

47	Combustible para cocinar	Combustible que más utiliza para cocinar
48	Aparatos	Tipo de electrodomésticos que tiene el hogar
49	Nº suministro	Numero de suministro
50	Tipo de suministro	Los tipos de suministro son: (1 – Luz; 2 – Agua; 3 – No tiene)
Acceso a servicios públicos		
51	Nro. de varones que viven en hogar	Número de hombres que viven en el hogar
52	Nro. De mujeres viven en hogar	Número de mujeres que viven en el hogar
Características de la población		
53	nro.doc	Número de documento de los integrantes
54	apellido pat	Apellido paterno
55	apellido mat	Apellido materno
56	Nombre	Nombres
57	Fecha nac	Fecha de Nacimiento
58	Tipo doc	Tipo de document
59	Sexo	Sexo
60	residente habitual	Persona que vive habitualmente
61	Parentesco c/jefe hogar	Parentesco que es jefe de hogar
62	Sexo	Sexo de los integrantes
63	Gestante	Esta variable indica si está en gestación
64	Estado civil	Estado civil de los integrantes
65	Tipo de seguro	Tipo de seguro que tienen
66	Idioma o lengua	Idiomas que hablan los integrantes
67	Lee y escribe	Leen o escriben los integrantes
68	Nivel educativo	Nivel educativo de cada integrante
69	Último año de estudios	Ultimo año de estudios
70	Ocupación último mes	Ocupación del último mes
71	Sector	Sector al que pertenece
72	Beneficiario de PPSS	Acceso al programa social

4.3.2.3 EXPLORACIÓN O ANÁLISIS DE LOS DATOS

Se realiza un análisis de los datos que serán considerados en la selección de datos relevantes, mediante consultas a la fuente de datos obtenida, se obtienen diversos cuadros estadísticos.

La figura 31 representa la cantidad de registros por departamento, provincia y distrito donde se encuentra ubicada la vivienda y la figura 32 representa la cantidad de hogares por tipo de vivienda.

En la figura 31 podemos observar que tenemos información de hogares que se encuentran distribuidos en 14 diferentes ubicaciones geográficas, siendo 8 ubicaciones las que cuentan con la mayor concentración de familias.

Figura 31. Cantidad de Viviendas por Ubicación Geográfica

Fuente: Elaboración Propia

La figura 32 muestra que prácticamente el total de hogares tiene como tipo de vivienda: casa independiente.

Figura 32. Cantidad de Hogares por Tipo de Vivienda

Fuente: Elaboración Propia

La figura 33 representa la cantidad de hogares por tenencia de la vivienda y la figura 34, la cantidad de hogares por tipo de material predominante en las paredes exteriores.

Figura 33. Cantidad de Hogares por Tenencia de Propiedad

Fuente: Elaboración Propia

En la figura 33 podemos observar que el 74% de los hogares tienen vivienda propia totalmente pagada.

Figura 34. Cantidad de Hogares por Tipo de Material de Pared

Fuente: Elaboración Propia

En la figura 35 podemos observar que el 67% de los hogares presentan el material de adobe o tapia en las paredes de su vivienda.

La figura 35 representa la cantidad de hogares por tipo de material predominante en el techo y la figura 36 la cantidad de hogares por tipo de material predominante en el piso.

Figura 35. Cantidad de Hogares por Tipo de Material de Techo

Fuente: Elaboración Propia

La figura 35 muestra que el 58% de los hogares presentan el material planchas de calamina en el techo de su vivienda, mientras que el 24% presenta tejas y esteras.

Figura 36. Cantidad de Hogares por Tipo de Material de Techo

Fuente: Elaboración propia

La figura 36 muestra que el 76% de los hogares presentan en el piso de su vivienda el material de tierra, mientras que el 8% presenta el material de cemento.

La figura 37 representa la cantidad de viviendas que cuentan con un tipo de alumbrado y la figura 38, la cantidad de viviendas que cuentan con un tipo de abastecimiento de agua.

Figura 37. Cantidad de Hogares por Tipo de Alumbrado

Fuente: Elaboración propia

La figura 37 muestra que el 68% de los hogares cuenta con electricidad en su vivienda y el 28% usa la vela para alumbrar su vivienda.

Figura 38. Cantidad de Hogares por Tipo de Abastecimiento de Agua

Fuente: Elaboración propia

La figura 38 muestra que el 73% de los hogares usa agua de río, acequia o manantial, y solo el 12% cuenta con red pública dentro de su vivienda.

La figura 39 representa la cantidad de hogares por tipo de conexión de servicios higiénicos y la figura 40 la cantidad de hogares por tipo de combustible para cocinar.

Figura 39. Cantidad de Hogares por Tipo de Conexión de Servicios Higiénicos

Fuente: Elaboración propia

La figura 39 muestra que el 59% de los hogares usa un pozo ciego o letrina como servicio higiénico, el 15% usa un pozo séptico y el 12% una red pública dentro de su vivienda.

Figura 40. Cantidad de Hogares por Tipo de Combustible usado para cocinar

Fuente: Elaboración propia

La figura 41 representa la cantidad de hogares que cuentan o no con un artefacto y la figura 36 muestra el idioma de niñez que es hablado por los integrantes de los hogares.

Figura 41. Cantidad de Hogares por Tipo de Artefactos

Fuente: Elaboración propia

La figura 41 muestra que el 45% de los hogares no cuenta con artefactos y el 44% cuenta con al menos un celular.

Figura 42. Cantidad de integrantes del hogar por idioma de su niñez

Fuente: Elaboración propia

La figura 42 muestra que el 51% de integrantes de los hogares habla el idioma castellano y el 48% habla el idioma quechua.

La figura 43 representa los integrantes de los hogares que cuenta o no con un seguro de salud, y la figura 44, los integrantes de los hogares que presentan o no alguna discapacidad.

Figura 43. Porcentaje de Integrantes de Hogares que cuenta con Seguro de Salud

Fuente: Elaboración propia

La figura 43 indica que el 87% de integrantes de los hogares cuentan con seguro de salud.

Figura 44. Porcentaje de Integrantes de Hogares que cuenta con Seguro de Salud

Fuente: Elaboración propia

La figura 44 muestra que el 97% de integrantes de los hogares no presentan discapacidad.

La figura 45 representa los integrantes de los hogares que son beneficiarios de programas sociales y la figura 46, los integrantes de los hogares por sector en el que trabajan.

Figura 45. Porcentaje de Integrantes de Hogares beneficiarios de programas sociales

Fuente: Elaboración propia

La figura 45 muestra que el 59% de integrantes de los hogares son beneficiarios de los programas sociales.

Figura 46. Cantidad de Integrantes de Hogares por Sector en el que laboran}

Fuente: Elaboración propia

La figura 46 muestra que el 57% de integrantes de los hogares laboran en el sector agrícola.

4.3.2.4 VERIFICACIÓN DE CALIDAD DE LOS DATOS

Debido a que los datos han sido recolectados desde documentos escaneados, muchas veces la información registrada no es visible ni clara. Pudiendo generar mala calidad en los datos. Adicionalmente a esto se han identificado las siguientes observaciones:

- Datos incompletos o en blanco
- Formatos mal escaneados con imagen borrosa
- Errores en el registro

4.3.3 PREPARACIÓN DE LOS DATOS

4.3.3.1 SELECCIÓN DE LOS DATOS RELEVANTES

De toda la información que se encuentran en los formatos S100 y FSU, se identificaron inicialmente cincuenta y nueve (59) campos relevantes para el análisis de los hogares:

Tabla 17
Depuración datos no relevantes

ITEM	CAMPO	MOTIVO DE EXCLUSIÓN
1	ID_FSU	Considerado irrelevante para el estudio
2	Cod_Dpto	Considerado irrelevante para el estudio
3	Dpto	Considerado irrelevante para el estudio
4	Cod_Provincia	Considerado irrelevante para el estudio
5	Provincia	Considerado irrelevante para el estudio
6	Cod_Distrito	Considerado irrelevante para el estudio
7	Distrito	Considerado irrelevante para el estudio
8	Centro_poblado	Dato alfanumérico incompleto
9	Cod_CCPP	Dato incompleto
10	Categoria	Dato alfanumérico incompleto
11	Nucleo_urbano	Dato alfanumérico incompleto
12	Cant_hogares	Considerado irrelevante para el estudio
13	Nro_hogares	Considerado irrelevante para el estudio
14	Fecha_Empadronamiento	Considerado irrelevante para el estudio
16	Color_Frontis	Considerado irrelevante para el estudio
17	Horas_llegar	Considerado irrelevante para el estudio
18	Vive_capital	Considerado irrelevante para el estudio
19	Nro_habitaciones	Considerado irrelevante para el estudio
20	Suministro	Considerado irrelevante para el estudio
21	Nro_suministro	Considerado irrelevante para el estudio
22	Nro_hombres	No será considerado para el estudio
23	Nro_mujeres	No será considerado para el estudio
24	Servicio_compartido	No es necesario por ser un dato alfanumérico y contar con el código correspondiente en otro campo
25	Nro_doc	No será considerado para el estudio
26	Apellido_Pat	No será considerado para el estudio
27	Apellido_Mat	No será considerado para el estudio

28	Nombres	No será considerado para el estudio
29	Fecha_nacimiento	No será considerado para el estudio
30	Tipo_Doc	No será considerado para el estudio
31	sexo	No será considerado para el estudio
32	Residente_habitual	No será considerado para el estudio
33	Parentesco	No será considerado para el estudio
34	Gestante	No será considerado para el estudio
35	Estado_civil	No será considerado para el estudio
36	Nivel_educativo	No será considerado para el estudio
37	Ult_estudio	No será considerado para el estudio
38	Ocupacion	No será considerado para el estudio

Fuente: Elaboración propia

4.3.3.2 DEPURACIÓN DE LOS DATOS NO RELEVANTES

De la validación de la información se identificaron campos con valores faltantes (no registrados por el empadronador), los cuales por ser tipo alfanumérico no serán considerados en el proceso.

A continuación, se detalla los campos que no serán usados por ser de tipo alfanumérico y otros por no estar completos en todos los registros.

Tabla 18

Datos que no serán utilizados para el análisis

ITEM	CAMPO	MOTIVO DE EXCLUSIÓN
1	Categoría	Dato alfanumérico incompleto
2	Centro poblado	Dato alfanumérico incompleto
3	Nucleo_urbano	Dato alfanumérico incompleto
4	Nro_hogares	Considerado irrelevante para el estudio
5	Fecha_empadronamiento	La información recolectada corresponde al mismo periodo
6	Nro_suministro	Considerado irrelevante para el estudio
7	Tipo_Doc	Considerado irrelevante para el estudio (mismo valor para toda la base)

8	Dpto	No es necesario por ser un dato alfanumérico y contar con el código correspondiente en otro campo
9	Provincia	No es necesario por ser un dato alfanumérico y contar con el código correspondiente en otro campo
10	Distrito	No es necesario por ser un dato alfanumérico y contar con el código correspondiente en otro campo
11	Apellido_Pat	No se utiliza porque la evaluación no es por integrante sin por hogar
12	Apellido_Mat	No se utiliza porque la evaluación no es por integrante sin por hogar
13	Nombres	No se utiliza porque la evaluación no es por integrante sin por hogar
14	Fecha_nacimiento	No se utiliza porque la evaluación no es por integrante sin por hogar
15	Parentesco	Considerado irrelevante para el estudio
16	Estado civil	Considerado irrelevante para el estudio
17	Ultimo año de estudio	Considerado irrelevante para el estudio
18	Sexo	Considerado irrelevante para el estudio
19	Número de Documento de Identidad	Considerado irrelevante para el estudio

Fuente: Elaboración propia

Adicionalmente, considerando los resultados del análisis de datos realizado en la fase de exploración de datos, se decide excluir también el dato de tipo de vivienda ya que el 96% de los hogares tienen el mismo valor (casa independiente), por lo que no sería relevante para el presente análisis.

Al finalizar este paso, quedan veinte (20) campos para continuar la estructuración de los datos.

4.3.3.3 ESTRUCTURACIÓN DE LOS DATOS Y SUS ATRIBUTOS

La estructuración de los datos consiste en realizar la normalización de los datos del hogar y el agrupamiento de los datos de los integrantes de un hogar.

Figura 47. Estructuración de datos

Fuente Elaboración propia

- **Normalización**

Para poder ejecutar la minería de datos, se opta por normalizar el archivo, para lo cual se dividen los siguientes campos en varias columnas, según la cantidad de valores posibles. En algunos hogares la selección puede ser de más de un valor:

Tabla 19
Datos para normalización

ITEM	CAMPO	TIPO	DESCRIPCIÓN
1	Propiedad	Numérico	Indica el grado de pertenencia de la vivienda
2	Material_pared	Numérico	Material con que se encuentra construida la pared de la casa
3	Material_techo	Numérico	Material con que se encuentra construido el techo de la casa

4	Material_piso	Numérico	Material con que se encuentra construido el piso de la casa
5	Alumbrado	Numérico	Identifica el tipo de alumbrado con que cuenta el hogar.
6	Agua	Numérico	Identifica la forma de abastecimiento de agua al hogar
7	SSHH	Numérico	Identifica el tipo de servicios higiénicos con que cuenta el hogar
8	Combustible_cocina	Numérico	Indica el tipo de combustible para la realización de la preparación de alimento.
9	Artefacto	Numérico	Indica los artefactos electrónicos con que cuenta la familia.

Fuente: Elaboración propia

Tabla 20
Ejemplo de campo normalizado:

Nº	CAMPO	TIPO	DESCRIPCIÓN
1	Tipo_propiedad_alquilada	Numérico	1=Alquilada
2	Tipo_propiedad_propiapagandola	Numérico	2=Propia pagándola a plazos
3	Tipo_propiedad_propiapagada	Numérico	3=Propia totalmente pagada
4	Tipo_propiedad_propiainvasion	Numérico	4=Propia por invasión
5	Tipo_propiedad_cedidatrabajo	Numérico	5=Cedida por el centro de trabajo

6	Tipo_propiedad _cedidahogarinstitution	Numérico	6=Cedida por otro hogar o institución
7	Tipo_propiedad_7	Numérico	7=Otros

Fuente: Elaboración propia

El resultado de la normalización es la siguiente:

Tabla 21
Normalización de datos

ITEM	CAMPO	TIPO	DESCRIPCIÓN
1	Propiedad_PropiaPagada	Numérico	Propiedad propia totalmente pagada
2	Propiedad_otros	Numérico	Otro tipo de propiedad
3	Material_pared_adobe	Numérico	Pared de adobe de la vivienda
4	Material_pared_otro	Numérico	Otro material en la pared de la vivienda
5	Material_techo_calamina	Numérico	Techo de calamina de la vivienda
6	Material_techo_otro	Numérico	Otro material en el techo de la vivienda
7	Material_piso_tierra	Numérico	Piso de tierra de la vivienda
8	Material_piso_otro	Numérico	Otro material en el piso de la vivienda
9	Alumbrado_electricidad	Numérico	Alumbrado por medio de electricidad
10	Alumbrado_vela	Numérico	Alumbrado por medio de vela
11	Alumbrado_otro	Numérico	Otro tipo de alumbrado
12	Agua_rio	Numérico	Agua por medio de un rio
13	Agua_otro	Numérico	Otro tipo de abastecimiento de agua
14	SSHH_pozo_ciego	Numérico	Servicios higiénicos mediante

			un pozo ciego
15	SSHH_otros	Numérico	Otro tipo de servicios higiénicos
16	Combustible_cocina_leña	Numérico	Combustible para la cocina usando leña
17	Combustible_cocina_otros	Numérico	Otro combustible para la cocina
18	Artefacto_celular	Numérico	Indica si el hogar cuenta con un celular
19	Artefacto_no_tiene	Numérico	Indica que el hogar no cuenta con artefactos
20	Artefacto_otros	Numérico	Indica que el hogar tiene otros tipo de artefacto (no celular)

Fuente: Elaboración propia

Adicionalmente se crean el siguiente campo a nivel de hogar, debido a que se analiza su relevancia en el proceso:

Tabla 22

Creación de un nuevo campo ingreso promedio por hogar

Nº	CAMPO	TIPO	DESCRIPCIÓN
1	Ingreso_promedio	Moneda	Promedio de ingreso familiar por hogar

Fuente: Elaboración propia

- **Agrupamiento**

Luego de identificar los datos relevantes a considerar en el proceso, se identifica también que la información registrada se encuentra con un nivel de detalle por integrante de un hogar; sin embargo, el proceso de clasificación socioeconómica se realiza en base a hogares. Por tal motivo, se procede a estructurar el archivo original obtenido a fin de realizar una consolidación de la información por hogar.

Listamos a continuación los datos que contienen información del detalle de los integrantes de un hogar y que serán consolidados para el proceso:

Tabla 23
Datos a consolidar

ITEM	CAMPO	TIPO	DESCRIPCIÓN
1	Tipo_seguro	Numérico	Tipo Seguro de Salud
2	Idioma	Numérico	Idioma o lengua aprendido en la niñez
3	Lee_escribe	Numérico	Indica si el integrante sabe leer y escribir
4	Sector	Numérico	Sector
5	Discapacidad	Numérico	Tipo de discapacidad
6	Beneficiario	Numérico	Indica si es beneficiario de algún programa social

Fuente: Elaboración propia

Estos son los campos que se crean como campos calculados a partir de la información de los integrantes del hogar:

Tabla 24
Campos calculados

ITEM	CAMPO	TIPO	DESCRIPCIÓN
1	Lee_escribe	numérico	Suma los integrantes que saben leer y escribir
2	Discapacitados	numérico	Suma de las discapacidades tipo 1 al 5 de los integrantes
3	No Discapacitados	numérico	Suma del valor campo discapacidad tipo 6 de los integrantes
4	Beneficiarios	numérico	Suma de los integrantes que acceden a los programas sociales (campo programas sociales tipo 1 al tipo 10)
5	No Beneficiarios	numérico	Suma de los integrantes que no acceden a los programas sociales (campo programas sociales tipo 11)
6	Suma Con Seguro	numérico	Cantidad de integrantes que tienen

			seguro
7	Suma sin Seguro	numérico	Cantidad de integrantes que no tienen seguro
8	Sector_agricola	Numérico	Indica si en el hogar hay trabajadores del sector agrícola
9	Sector_otros	Numérico	Indica si en el hogar hay trabajadores de otros sectores (no agrícola)
10	Idioma_quechua	numérico	Cantidad de integrantes que hablan el idioma quechua
11	Idioma_castellano	numérico	Cantidad de integrantes que hablan el idioma castellano
12	Idioma_otro	numérico	Cantidad de integrantes que hablan otros idiomas

Fuente: Elaboración propia

El resultado final de la normalización y agrupamiento de campos dio como resultado treinta y nueve (39) campos para continuar la fase de preparación de datos.

4.3.3.4 INTEGRACIÓN DE LOS DATOS ESTRUCTURADOS Y NO ESTRUCTURADOS

Para el presente trabajo no se consideran datos no estructurados debido a que no se cuenta con los mismos.

4.3.3.5 FORMATEO DE DATOS GENERANDO UN INFORME DE CALIDAD

La información considerada relevante según lo descrito anteriormente, correspondiente a las respuestas de las encuestas de los formatos FSU y S100 son numéricos.

Algunos campos que se formatearon fueron los siguientes:

Tabla 25
Campos formateados

ITEM	CAMPO	TIPO	VALORES
1	Propiedad	Numérico	Los datos vacíos se consideraron con el valor 7: Otro, resultando en Propiedad_otros
2	Material_pared	Numérico	Los datos vacíos se consideraron con el valor 8: Otro, resultando en Material_pared_otro
3	Material_techo	Numérico	Los datos vacíos se consideraron con el valor 8: Otro, resultando en Material_techo_otro
4	Material_piso	Numérico	Los datos vacíos se consideraron con el valor 7: Otro
5	Alumbrado	Numérico	Los datos vacíos se consideraron con el valor 6: No tiene, resultando en Alumbrado_otro
6	Agua	Numérico	Los datos vacíos se consideraron con el valor 7: Otro, resultando en Agua_otro
7	SSHH	Numérico	Los datos vacíos se consideraron con el valor 6: No tiene, resultando en SSHH_otro
8	Combustible_cocina	Numérico	Los datos vacíos se consideraron con el valor 7: Otro, resultando en Combustible_cocina_otro
9	Artefacto	Numérico	Los datos vacíos se consideraron con el valor 15: No tiene, resultando en Artefacto_otro
10	Idioma	Numérico	Los datos vacíos se consideraron con el valor 7: Otro, resultando en Idioma_otro
12	Discapacidad	Numérico	Los datos vacíos se consideraron con el valor 6: No tiene discapacidad,

			considerándose en No_Discapacitados
13	Beneficiario	Numérico	Los datos vacíos se consideraron con el valor 11: Ninguno, considerándose en No_Beneficiarios

Fuente: Elaboración propia

4.3.4 MODELADO

En esta fase de la metodología es necesario considerar que existen diversos modelos y algoritmos predictivos, los cuales deberán ser analizados de acuerdo a los objetivos que se deseen alcanzar.

En esta etapa se tomará en cuenta la información analizada en los puntos anteriores, elaborando un plan de pruebas para la ejecución del modelo seleccionado y realizando posteriormente el análisis de los resultados obtenidos.

4.3.4.1 SELECCIÓN DEL MODELO Y SUPUESTOS

Considerando que el proceso de clasificación socioeconómico intenta identificar si un hogar debe ser catalogado como pobre extremo, pobre o no pobre, y tomando en cuenta el marco teórico previamente descrito acerca de los modelos y algoritmos existentes, se ha considerado en el siguiente estudio que entre los modelos más adecuados para alcanzar este objetivo se encuentran los modelos de clasificación y clusterización.

Se tomaron como base también los artículos científicos descritos en la sección de antecedentes, donde se pudo evidenciar el uso y los resultados de la aplicación de la clasificación Naive Bayes para la identificación de distritos priorizados y no priorizados, así como para la identificación de familias pobres y propensas a pobres.

Cabe recalcar que para la obtención de óptimos resultados se recomienda la exploración de diferentes modelos predictivos.

Con la información que se ha logrado recopilar se intentará mostrar cómo se identifican patrones en los datos, que nos lleven a poder predecir con determinado nivel de certeza la clasificación de un hogar. Sin embargo, es importante también indicar que en el presente estudio se consideran solo los datos descritos en los pasos anteriores y que nuevos datos relacionados al hogar podrían cambiar los resultados.

4.3.4.2 ELABORAR EL PLAN DE PRUEBAS

En esta sección se presenta el plan de pruebas para medir la calidad de los modelos de minería de datos generados a partir de las técnicas seleccionadas.

Para ambas técnicas seleccionadas (clusterización y clasificación) se separarán los datos en dos conjuntos, uno de entrenamiento y otro de prueba, que permitirán construir el modelo basado en el conjunto de entrenamiento, para luego medir la calidad del modelo generado con el conjunto de prueba.

4.3.4.2.1 ALGORITMO K-MEANS

Tabla 26

Pruebas fuente de datos de entrenamiento – k-means

Fuente de datos de entrenamiento	DataEntrenamiento.csv
Conjunto de entrenamiento	99 registros (34.4%)
Fuente de datos de testeo	DataPrueba.csv
Conjunto de testeo	189 registros (65.6 %)
Medir la bondad	Porcentaje de clasificación incorrecta: 25%

Fuente: Elaboración propia

Se espera que el porcentaje de error en la clasificación de datos que alcance esta técnica tenga un valor máximo de 25%.

4.3.4.2.2 ALGORITMO NAIVE BAYES

Tabla 27

Pruebas fuente de datos de entrenamiento – k-means

Fuente de datos de entrenamiento	DataEntrenamiento.csv
Conjunto de entrenamiento	99 registros (34.4%)
Fuente de datos de testeo	DataPrueba.csv
Conjunto de testeo	189 registros (65.6 %)
Medir la bondad	Precisión: 80%

Fuente: Elaboración propia

La precisión indica qué proporción acierta un modelo respecto a una clase. Se espera que el modelo alcance una precisión de 80% al determinar cada clase.

4.3.4.3 CONSTRUCCIÓN DEL MODELO CONSIDERANDO LOS PARÁMETROS ESCOGIDOS, MODELO, DESCRIPCIÓN DE LOS MODELOS

Realizaremos la construcción del modelo tomando en cuenta los objetivos definidos para el estudio, para lo cual seleccionamos dos algoritmos:

- SimpleKmeans
- Naivebayes

Data de entrada:

Se considera dividir la información recolectada en 2 grupos: Data preparatoria y data de prueba.

La data de entrenamiento servirá para entrenar al modelo el cual nos permitirá evaluar los patrones en los datos, utilizaremos la información de 99 hogares.

La data de prueba servirá para probar el modelo entrenado, utilizaremos la información de 189 hogares sumados con la data de entrenamiento son un total de 288 hogares que serán analizados.

La cantidad de columnas o variables a analizar en la información original de los 288 hogares luego de la normalización y el agrupamiento es de 39 variables. Con fines prácticos, la información de estas variables ha sido consolidada de tal manera que nos permita demostrar los resultados que esperamos según los criterios de éxito definidos. A continuación, se listan las variables utilizadas para el estudio correspondiente:

Tabla 28
Lista de variables para el proceso

ITEM	VARIABLES	ITEM	VARIABLES
1	Nro_pisos_vivienda	21	Artefacto_Otro
2	Propiedad_PropiaPagada	22	Nro_integrantes
3	Propiedad_otros	23	Servicios
4	Material_pared_adobe	24	Ingreso_familiar
5	Material_pared_otro	25	ingreso_promedio_familiar
6	Material_techo_calamina	26	Cant_propiedades
7	Material_techo_otro	27	Cant_vehiculos
8	Material_piso_tierra	28	Lee_escribe
9	Material_piso_otro	29	Discapacitados

10	Alumbrado_electricidad	30	No_Discapitados
11	Alumbrado_Vela	31	Beneficiarios
12	Alumbrado_Otro	32	No_Beneficiarios
13	Agua_rio	33	Con_Seguro
14	Agua_otro	34	Sin_Seguro
15	SSHH_pozo_ciego	35	Sector_agricola
16	SSHH_otro	36	Sector_Otros
17	Combustible_cocina_leña	37	Idioma_quechua
18	Combustible_cocina_otro	38	Idioma_castellano
19	Artefacto_Celular	39	Idioma_Otro
20	Artefacto_No_Tiene		

Fuente: Elaboración propia

De la herramienta

Utilizaremos la herramienta gratuita Weka, la cual nos permitirá ejecutar el modelo, utilizando data en formato csv.

Procesamiento de la data de entrenamiento

- Data de entrenamiento: Como podemos visualizar en la figura 42, se selecciona el archivo con la data de entrenamiento en formato csv.

Figura 48. Selección de archivo para procesamiento

Fuente: Elaboración propia

Se puede observar que la herramienta carga los datos con las 39 variables que se quieren analizar para identificar los patrones de comportamiento.

Figura 49. Variables para procesamiento

Fuente: Elaboración propia

- **Selección del algoritmo:** Para la ejecución inicial de la data de entrenamiento, seleccionamos el algoritmo SimpleKMeans. Este algoritmo nos permitirá establecer la clasificación inicial de la data para que el modelo pueda ser entrenado.

Figura 50. Selección de algoritmo

Fuente: Elaboración propia

- **Para la clasificación:** Con la finalidad de identificar a los hogares según la siguiente clasificación: No pobre, pobre y extremo pobre, parametrizamos la aplicación para la distribución en tres clústeres.

Figura 51. Parametrización 3 clústeres

Fuente: Elaboración propia

- **Ejecución:** A continuación, se puede observar en la figura 46 el resultado de la ejecución del modelo SimpleKMeans, el cual realiza la distribución de los 99 hogares en 3 conjuntos o clúster.

Figura 52. Resultados 1 Clustering – Kmeans

Fuente: Elaboración propia

Figura 53. Resultados2 clustering – Kmeans

Fuente: Elaboración propia

Tabla 29
Detalle Resultados K-means

=== Run information ===	
Scheme:	weka.clusterers.SimpleKMeans -init 0 -max-candidates 100 -periodic-pruning 10000 -min-density 2.0 -t1 -1.25 -t2 -1.0 -N 3 -A "weka.core.EuclideanDistance -R first-last" -I 500 -num-slots 1 -S 10
Relation:	DataEntrenamiento_09012019
Instances:	99
Attributes:	41
	Nro_pisos_vivienda
	Propiedad_PropiaPagada
	Propiedad_otros
	Material_pared_adobe
	Material_pared_otro
	Material_techo_calamina
	Material_techo_otro
	Material_piso_tierra
	Material_piso_otro
	Alumbrado_electricidad
	Alumbrado_Vela
	Alumbrado_Otro
	Agua_rio
	Agua_otro
	SSHH_pozo_ciego
	SSHH_otro


```

Combustible_cocina_le?
Combustible_cocina_otro
Artefacto_Celular
Artefacto_No_Tiene
Artefacto_Otro
Nro_integrantes
Servicios
Ingreso_familiar
ingreso_promedio_familiar
Cant_propiedades
Cant_vehiculos
Lee_escribe
Discapacitados
No_Discapacitados
Beneficiarios
No_Beneficiarios
Con_Seguro
Sin_Seguro
Sector_agricola
Sector_Otros
Idioma_quechua
Idioma_castellano
Idioma_Otro

Ignored:

    ID_FSU
    Clasificacion

Test mode:    Classes to clusters evaluation on training data

=== Clustering model (full training set) ===

kMeans
=====
Number of iterations: 15
Within cluster sum of squared errors: 355.4950163692034

Initial starting points (random):

Cluster 0: 1,1,0,1,0,1,0,1,0,1,0,0,1,0,1,0,1,0,0,2,1,1000,500,1,0,1,0,2,0,2,2,0,1,1,2,0,0
Cluster 1: 1,1,0,1,0,1,0,1,0,1,0,0,1,0,1,0,0,1,0,4,4,150,37.5,0,0,3,0,4,2,2,4,0,1,3,0,4,0
Cluster 2: 1,1,0,1,0,1,0,1,0,1,0,0,1,0,0,0,1,0,0,1,0,4,4,800,200,0,0,1,0,4,2,2,4,0,1,3,4,0,0

Missing values globally replaced with mean/mode

Final cluster centroids:

Attribute                                Cluster#
Full Data                                0          1          2
(99.0)    (30.0)    (30.0)    (39.0)
=====
Nro_pisos_vivienda                      1.1414      1.1      1.0333      1.2564
Propiedad_PropiaPagada                   0.7374      0.6667      0.9      0.6667
Propiedad_otros                          0.2626      0.3333      0.1      0.3333
Material_pared_adobe                     0.6566      0.2333      0.7667      0.8974
Material_pared_otro                      0.3434      0.7667      0.2333      0.1026
Material_techo_calamina                  0.596       0.2       0.9      0.6667
Material_techo_otro                      0.404       0.8       0.1      0.3333
Material_piso_tierra                     0.8182      0.7667      1      0.7179

```

Material_piso_otro	0.1818	0.2333	0	0.2821
Alumbrado_electricidad	0.7172	0.8333	0.2667	0.9744
Alumbrado_Vela	0.2323	0.0667	0.6667	0.0256
Alumbrado_Otro	0.0505	0.1	0.0667	0
Agua_rio	0.6768	0	1	0.9487
Agua_otro	0.3131	0.9667	0	0.0513
SSHH_pozo_ciego	0.5657	0.5	0.8333	0.4103
SSHH_otro	0.3535	0.3667	0.0667	0.5641
Combustible_cocina_le?	0.7677	0.4667	0.9333	0.8718
Combustible_cocina_otro	0.2323	0.4667	0.0667	0.1795
Artefacto_Celular	0.4545	0.4333	0.0667	0.7692
Artefacto_No_Tiene	0.4343	0.3	0.9	0.1795
Artefacto_Otro	0.5758	1.1333	0.0667	0.5385
Nro_integrantes	3.6667	3.1	3.3333	4.359
Servicios	2.5455	2.3333	3.8	1.7436
Ingreso_familiar	1509.3939	2754	504.6667	1324.8718
ingreso_promedio_familiar	543.4627	1196.7778	174.7206	324.5604
Cant_propiedades	0.3737	0.0667	0.2	0.7436
Cant_vehiculos	0.0303	0	0	0.0769
Lee_escribe	2.7677	2.3667	2.3667	3.3846
Discapacitados	0.1212	0.2667	0.0667	0.0513
No_Discapacitados	3.5455	2.8333	3.2667	4.3077
Beneficiarios	1.5051	0.7	1.8	1.8974
No_Beneficiarios	2.1616	2.4	1.5333	2.4615
Con_Seguro	3.2525	2.3667	3.1333	4.0256
Sin_Seguro	0.4141	0.7333	0.2	0.3333
Sector_agricola	1.0606	0.8	1.1333	1.2051
Sector_Otros	2.6061	2.3	2.2	3.1538
Idioma_quechua	1.9091	0.9	1.7667	2.7949
Idioma_castellano	1.6869	1.9667	1.5667	1.5641
Idioma_Otro	0.0808	0.2667	0	0

Time taken to build model (full training data) : 0.07 seconds

=== Model and evaluation on training set ===

Clustered Instances

0 30 (30%)

1 30 (30%)

2 39 (39%)

Class attribute: Clasificacion

Classes to Clusters:

0 1 2 <-- assigned to cluster

2 4 28 | Pobre

26 0 1 | No pobre

2 26 10 | Pobre Extremo

Cluster 0 <-- No pobre

Cluster 1 <-- Pobre Extremo

Cluster 2 <-- Pobre

Incorrectly clustered instances: 19.0 19.1919 %

Fuente: Elaboración propia

Podemos observar los resultados obtenidos para la clasificación de los hogares como se detalla a continuación:

Cluster 0: No pobre

Cluster 1: Pobre Extremo

Cluster 2: Pobre

Distribución de los hogares procesados:

- 30 hogares son clasificados como No pobres
- 29 hogares son clasificados como Pobres
- 30 hogares son clasificados como Pobres Extremos

Luego de la ejecución del modelo, podemos observar en la figura 48 la distribución de los clústeres según centroide del modelo.

Figura 54. Distribución de Clúster

Fuente: Elaboración propia.

- **Procesamiento de la data de prueba**

Como podemos visualizar en la figura 49, se selecciona el archivo con la data de prueba en formato csv.

Figura 55. Selección de Data de Prueba K-means

Fuente: Elaboración propia.

Para ejecutar la data de prueba, seleccionamos en la herramienta la opción Supplied test set como se observa en la figura 50, en el cual se procesa la data de prueba de los 189 hogares, con las 39 las variables que serán analizadas y clasificadas mediante el algoritmo de SimpleKmeans.

Figura 56. Resultado1 Data Prueba K-means

Fuente: Elaboración propia

Figura 57. Resultado2 Data Prueba K-means

Fuente: Elaboración propia

Tabla 30

Detalle resultados Data Prueba K-means

=== Run information ===

```

Scheme:          weka.clusterers.SimpleKMeans -init 0 -max-candidates 100 -periodic-pruning
10000 -min-density 2.0 -t1 -1.25 -t2 -1.0 -N 3 -A "weka.core.EuclideanDistance -R first-
last" -I 500 -num-slots 1 -S 10
Relation:        DataPrueba_09012019
Instances:       189
Attributes:      40
                  Nro_pisos_vivienda
                  Propiedad_PropiaPagada
                  Propiedad_otros
                  Material_pared_adobe
                  Material_pared_otro
                  Material_techo_calamina
                  Material_techo_otro
                  Material_piso_tierra
                  Material_piso_otro
                  Alumbrado_electricidad
                  Alumbrado_Vela
                  Alumbrado_Otro
                  Agua_rio
                  Agua_otro
                  SSHH_pozo_ciego
                  SSHH_otro
                  Combustible_cocina_leña

```

```

Combustible_cocina_otro
Artefacto_Celular
Artefacto_No_Tiene
Artefacto_Otro
Nro_integrantes
Servicios
Ingreso_familiar
ingreso_promedio_familiar
Cant_propiedades
Cant_vehiculos
Lee_escribe
Discapacitados
No_Discapacitados
Beneficiarios
No_Beneficiarios
Con_Seguro
Sin_Seguro
Sector_agricola
Sector_Otros
Idioma_quechua
Idioma_castellano
Idioma_Otro

Ignored:

ID_FSU

Test mode:    user supplied test set: 189 instances

=== Clustering model (full training set) ===

kMeans
=====
Number of iterations: 13
Within cluster sum of squared errors: 632.8479869197822

Initial starting points (random):

Cluster
2,0,1,1,0,0,1,0,1,1,0,0,1,0,0,1,0,1,1,0,2,2,4,100,50,0,0,1,0,2,1,1,2,0,0,2,1,1,0
Cluster
1,1,0,0,1,0,1,1,0,1,0,0,0,1,0,0,1,0,0,1,0,4,2,400,100,1,0,3,2,2,0,4,1,3,1,3,1,3,0
Cluster
2,1,0,1,0,0,1,1,0,1,0,0,1,0,0,1,1,0,1,0,0,2,1,100,50,1,0,2,0,2,0,2,1,1,1,1,2,0,0

Missing values globally replaced with mean/mode

Final cluster centroids:

Attribute          Full Data          Cluster#
                    (189.0)    (71.0)    (36.0)    (82.0)
=====

```

Nro_pisos_vivienda	1.127	1.169	1.2222	1.0488
Propiedad_PropiaPagada	0.7407	0.7606	0.5278	0.8171
Propiedad_otros	0.2593	0.2394	0.4722	0.1829
Material_pared_adobe	0.6825	0.8028	0.1667	0.8049
Material_pared_otro	0.3175	0.1972	0.8333	0.1951
Material_techo_calamina	0.5767	0.7606	0.1389	0.6098
Material_techo_otro	0.4233	0.2394	0.8611	0.3902
Material_piso_tierra	0.7354	0.4507	0.7222	0.9878
Material_piso_otro	0.2646	0.5493	0.2778	0.0122
Alumbrado_electricidad	0.6561	0.9296	0.9167	0.3049
Alumbrado_Vela	0.3069	0.0704	0.0278	0.6341
Alumbrado_Otro	0.037	0	0.0556	0.061
Agua_rio	0.7566	0.9014	0.1111	0.9146
Agua_otro	0.2222	0.0986	0.7778	0.0854
SSHH_pozo_ciego	0.5979	0.4507	0.5278	0.7561
SSHH_otro	0.3175	0.5352	0.4167	0.0854
Combustible_cocina_leña	0.7672	0.8732	0.25	0.9024
Combustible_cocina_otro	0.2328	0.1549	0.6944	0.0976
Artefacto_Celular	0.4392	0.8732	0.5	0.0366
Artefacto_No_Tiene	0.4603	0.0704	0.1667	0.9268
Artefacto_Otro	0.5926	0.6901	1.5556	0.0854
Nro_integrantes	3.3757	3.6479	3.3889	3.1341
Servicios	2.672	1.9437	2.0556	3.5732
Ingreso_familiar	1193.4921	989.2958	2716.6667	701.5854
ingreso_promedio_familiar	419.6689	309.4333	998.7235	260.8977
Cant_propiedades	0.3968	0.7746	0.0278	0.2317
Cant_vehiculos	0.0159	0.0141	0	0.0244
Lee_escribe	2.3439	2.6479	2.3333	2.0854
Discapacitados	0.0952	0.0423	0.1667	0.1098
No_Discapacitados	3.291	3.6056	3.2222	3.0488
Beneficiarios	1.1534	1.4789	0.6389	1.0976
No_Beneficiarios	2.2222	2.169	2.75	2.0366
Con_Seguro	2.9101	3.1831	2.0833	3.0366
Sin_Seguro	0.4656	0.4648	1.3056	0.0976
Sector_agricola	0.9206	0.9718	0.6944	0.9756
Sector_Otros	2.455	2.6761	2.6944	2.1585
Idioma_quechua	1.5079	1.5352	0.6944	1.8415
Idioma_castellano	1.8201	2.0704	2.5278	1.2927
Idioma_Otro	0.0529	0.0423	0.1944	0
Time taken to build model (full training data) : 0.13 seconds				
=== Evaluation on test set ===				
Clustered Instances				
0	71 (38%)			
1	36 (19%)			
2	82 (43%)			

Fuente: Elaboración propia

Se observa el resultado obtenido de esta ejecución, en la cual los hogares fueron distribuidos de la siguiente manera:

- 71 hogares son clasificados como No pobres
- 36 hogares son clasificados como Pobres Extremos
- 82 hogares son clasificados como Pobres

En la figura 52 se puede observar la distribución de los 189 hogares en los clústeres según centroide del modelo de la data de prueba.

Figura 58. Distribución data Prueba K-means

Fuente: Elaboración propia

4.3.4.3.1 USO ALGORITMO NAIVE BAYES

Tomando en cuenta el artículo de (Redjeki et al., 2015), que señala que el modelo Naive Bayes fue uno de los algoritmos utilizados para la clasificación pobre, muy pobre y vulnerable, se ejecutará a continuación, una corrida en la herramienta WEKA utilizando este algoritmo:

- En la figura 53 podemos observar la selección del archivo de entrenamiento

Figura 59. Selección archivo Naive Bayes

Fuente: Elaboración propia

- Se observa en la figura 54 la carga de las 39 variables de la data de entrenamiento

Figura 60. Variables Data Prueba Naive Bayes

Fuente: Elaboración propia

- En la figura 55 se muestra el resultado de la ejecución con este algoritmo.

Figura 55. Resultado1 Naive Bayes

Fuente: Elaboración propia

Figura 61. Resultado2 Naive Bayes

Fuente: Elaboración propia

Tabla 31
Detalle Resultado Naive Bayes

```

=== Run information ===

Scheme:      weka.classifiers.bayes.NaiveBayes
Relation:    DataEntrenamiento_09012019
Instances:   99
Attributes:  41
             ID_FSU
             Nro_pisos_vivienda
             Propiedad_PropiaPagada
             Propiedad_otros
             Material_pared_adobe
             Material_pared_otro
             Material_techo_calamina
             Material_techo_otro
             Material_piso_tierra
             Material_piso_otro
             Alumbrado_electricidad
             Alumbrado_Vela
             Alumbrado_Otro
             Agua_rio
             Agua_otro
             SSHH_pozo_ciego
             SSHH_otro
             Combustible_cocina_leña
             Combustible_cocina_otro
             Artefacto_Celular
             Artefacto_No_Tiene
             Artefacto_Otro
             Nro_integrantes
             Servicios
             Ingreso_familiar
             ingreso_promedio_familiar
             Cant_propiedades
             Cant_vehiculos
             Lee_escribe
             Discapacitados
             No_Discapacitados
             Beneficiarios
             No_Beneficiarios
             Con_Seguro
             Sin_Seguro
             Sector_agricola
             Sector_Otros
             Idioma_quechua
             Idioma_castellano
             Idioma_Otro

```

Clasificacion			
Test mode: evaluate on training data			
=== Classifier model (full training set) ===			
Naive Bayes Classifier			
Attribute	Class		
	Pobre (0.34)	No pobre (0.27)	Pobre Extremo (0.38)
=====			
ID_FSU			
mean	141.3193	54.4913	180.5618
std. dev.	62.3455	44.3832	79.9647
weight sum	34	27	38
precision	2.7347	2.7347	2.7347
Nro_pisos_vivienda			
mean	1.1471	1.1111	1.1579
std. dev.	0.3542	0.3143	0.3646
weight sum	34	27	38
precision	1	1	1
Propiedad_PropiaPagada			
mean	0.7059	0.6296	0.8421
std. dev.	0.4556	0.4829	0.3646
weight sum	34	27	38
precision	1	1	1
Propiedad_otros			
mean	0.2941	0.3704	0.1579
std. dev.	0.4556	0.4829	0.3646
weight sum	34	27	38
precision	1	1	1
Material_pared_adobe			
mean	0.9118	0.1111	0.8158
std. dev.	0.2836	0.3143	0.3877
weight sum	34	27	38
precision	1	1	1
Material_pared_otro			
mean	0.0882	0.8889	0.1842
std. dev.	0.2836	0.3143	0.3877
weight sum	34	27	38
precision	1	1	1
Material_techo_calamina			
mean	0.7059	0.2593	0.7368

std. dev.	0.4556	0.4382	0.4403
weight sum	34	27	38
precision	1	1	1
Material_techo_otro			
mean	0.2941	0.7407	0.2632
std. dev.	0.4556	0.4382	0.4403
weight sum	34	27	38
precision	1	1	1
Material_piso_tierra			
mean	0.8235	0.7037	0.8947
std. dev.	0.3812	0.4566	0.3069
weight sum	34	27	38
precision	1	1	1
Material_piso_otro			
mean	0.1765	0.2963	0.1053
std. dev.	0.3812	0.4566	0.3069
weight sum	34	27	38
precision	1	1	1
Alumbrado_electricidad			
mean	0.9118	0.8148	0.4737
std. dev.	0.2836	0.3884	0.4993
weight sum	34	27	38
precision	1	1	1
Alumbrado_Vela			
mean	0.0882	0.0741	0.4737
std. dev.	0.2836	0.2619	0.4993
weight sum	34	27	38
precision	1	1	1
Alumbrado_Otro			
mean	0	0.1111	0.0526
std. dev.	0.1667	0.3143	0.2233
weight sum	34	27	38
precision	1	1	1
Agua_rio			
mean	0.9118	0.037	0.9211
std. dev.	0.2836	0.1889	0.2697
weight sum	34	27	38
precision	1	1	1
Agua_otro			
mean	0.0882	0.9259	0.0789
std. dev.	0.2836	0.2619	0.2697

weight sum	34	27	38
precision	1	1	1
SSHH_pozo_ciego			
mean	0.5294	0.5556	0.6053
std. dev.	0.4991	0.4969	0.4888
weight sum	34	27	38
precision	1	1	1
SSHH_otro			
mean	0.4118	0.3333	0.3158
std. dev.	0.4922	0.4714	0.4648
weight sum	34	27	38
precision	1	1	1
Combustible_cocina_leña			
mean	0.8824	0.4444	0.8947
std. dev.	0.3222	0.4969	0.3069
weight sum	34	27	38
precision	1	1	1
Combustible_cocina_otro			
mean	0.1471	0.4815	0.1316
std. dev.	0.3542	0.4997	0.338
weight sum	34	27	38
precision	1	1	1
Artefacto_Celular			
mean	0.6765	0.4815	0.2368
std. dev.	0.4678	0.4997	0.4251
weight sum	34	27	38
precision	1	1	1
Artefacto_No_Tiene			
mean	0.2941	0.2222	0.7105
std. dev.	0.4556	0.4157	0.4535
weight sum	34	27	38
precision	1	1	1
Artefacto_Otro			
mean	0.4412	1.2593	0.2105
std. dev.	0.8114	1.0747	0.6943
weight sum	34	27	38
precision	1	1	1
Nro_integrantes			
mean	3.9706	3.037	3.8421
std. dev.	1.7403	1.4268	1.77
weight sum	34	27	38

precision	1	1	1
Servicios			
mean	1.8529	2.2593	3.3684
std. dev.	1.3091	1.2647	1.223
weight sum	34	27	38
precision	1	1	1
Ingreso_familiar			
mean	1351.4329	3297.246	388.664
std. dev.	767.7091	1264.4039	467.7394
weight sum	34	27	38
precision	205.1282	205.1282	205.1282
ingreso_promedio_familiar			
mean	373.8337	1385.5683	110.3448
std. dev.	215.2702	911.9751	102.0596
weight sum	34	27	38
precision	65.5172	65.5172	65.5172
Cant_propiedades			
mean	0.7647	0	0.2895
std. dev.	0.4886	0.1667	0.4535
weight sum	34	27	38
precision	1	1	1
Cant_vehiculos			
mean	0.0294	0	0.0526
std. dev.	0.169	0.1667	0.2233
weight sum	34	27	38
precision	1	1	1
Lee_escribe			
mean	3.1176	2.3333	2.7632
std. dev.	1.5294	1.2172	1.5633
weight sum	34	27	38
precision	1	1	1
Discapacitados			
mean	0.1176	0.1852	0.0789
std. dev.	0.3222	0.3884	0.2697
weight sum	34	27	38
precision	1	1	1
No_Discapacitados			
mean	3.8529	2.8519	3.7632
std. dev.	1.7	1.6263	1.8699
weight sum	34	27	38
precision	1	1	1

Beneficiarios				
mean	1.7647	0.7778	1.7895	
std. dev.	1.646	1.0657	1.3408	
weight sum	34	27	38	
precision	1	1	1	
No_Beneficiarios				
mean	2.2059	2.2593	2.0526	
std. dev.	1.9217	1.4806	1.468	
weight sum	34	27	38	
precision	1	1	1	
Con_Seguro				
mean	3.6765	2.2222	3.6053	
std. dev.	1.8743	1.663	1.8714	
weight sum	34	27	38	
precision	1	1	1	
Sin_Seguro				
mean	0.2941	0.8148	0.2368	
std. dev.	0.5703	1.3344	0.7047	
weight sum	34	27	38	
precision	1	1	1	
Sector_agricola				
mean	1.3529	0.6667	1.0789	
std. dev.	1.1345	0.9027	0.9565	
weight sum	34	27	38	
precision	1	1	1	
Sector_Otros				
mean	2.6176	2.3704	2.7632	
std. dev.	1.3936	1.4944	1.7835	
weight sum	34	27	38	
precision	1	1	1	
Idioma_quechua				
mean	2.9244	0.6349	2.2256	
std. dev.	2.3056	0.9999	2.113	
weight sum	34	27	38	
precision	1.1429	1.1429	1.1429	
Idioma_castellano				
mean	1.1723	2.2857	1.7932	
std. dev.	1.6012	1.5779	2.3696	
weight sum	34	27	38	
precision	1.2857	1.2857	1.2857	


```

Idioma_Otro
  mean          0      0.2963      0
  std. dev.     0.1667    0.5316    0.1667
  weight sum    34      27      38
  precision     1       1       1

Time taken to build model: 0.11 seconds

=== Evaluation on training set ===

Time taken to test model on training data: 0.13 seconds

=== Summary ===

Correctly Classified Instances      86      86.8687 %
Incorrectly Classified Instances    13      13.1313 %
Kappa statistic                    0.801
Mean absolute error                 0.0837
Root mean squared error             0.2556
Relative absolute error             18.9981 %
Root relative squared error         54.4793 %
Total Number of Instances          99

=== Detailed Accuracy By Class ===

      TP Rate  FP Rate  Precision  Recall   F-Measure  MCC      ROC Area  PRC
Area  Class
      0,882    0,123    0,789    0,882    0,833     0,741    0,968    0,941
Pobre
      0,926    0,000    1,000    0,926    0,962     0,949    0,993    0,988
No pobre
      0,816    0,082    0,861    0,816    0,838     0,742    0,962    0,938
Pobre Extremo
Weighted Avg.  0,869    0,074    0,874    0,869    0,870     0,798    0,972    0,952

=== Confusion Matrix ===

  a  b  c  <-- classified as
30  0  4 |  a = Pobre
 1 25  1 |  b = No pobre
 7  0 31 |  c = Pobre Extremo

```

Fuente: Elaboración propia

De la tabla 23 se observa la distribución de los hogares con este algoritmo:

- 34 hogares clasificados como Pobres
- 27 hogares clasificados como No pobres

- 38 hogares clasificados como Pobres Extremo.

El modelo presenta una precisión promedio de 0.874 (84% de precisión).

4.3.4.4 EVALUACIÓN DEL MODELO

En esta etapa de la metodología, se evaluarán los resultados del modelo desde el punto de vista de los objetivos de la minería de datos.

Objetivo1: En el análisis descriptivo, se busca identificar los patrones de los distintos grupos de clasificación socioeconómica, para lo cual se identificarán las variables que tienen mayor impacto en el proceso de clasificación socioeconómica de los hogares.

Para poder realizar esta evaluación, utilizaremos la opción de evaluación de atributos de la herramienta Weka, la cual nos permite determinar las variables de mayor correlación e impacto para disminuir la entropía del modelo.

Utilizamos 3 métodos de evaluación de los atributos:

- **CorrelationAttributeEval**

Evalúa el valor de un atributo midiendo la correlación (Pearson) entre este y la clase.

Figura 62. Evaluación de Atributos CorrelationAttributeEval – Mayor relevancia

Fuente: Elaboración propia

Figura 63. Evaluación de Atributos CorrelationAttributeEval – Menor relevancia

Fuente: Elaboración propia

- **GainRatioAttributeEval**

Evalúa un atributo basado en la relación de ganancia.

Figura 64. Evaluación de Atributos GainRatioAttributeEval – Mayor relevancia

Fuente: Elaboración propia

Figura 65. Evaluación de Atributos GainRatioAttributeEval – Menor relevancia

Fuente: Elaboración propia

- **InfoGainAttributeEval**

Evalúa el valor de un atributo midiendo la ganancia de información con respecto a la clase.

Figura 66. Evaluación de Atributos InfoGainAttributeEval – Mayor relevancia

Fuente: Elaboración propia

Figura 67. Evaluación de Atributos InfoGainAttributeEval – Menor relevancia

Fuente: Elaboración propia

- Como se puede notar, las variables o atributos que contribuyen de mayor manera en el modelo de clasificación corresponden al ingreso familiar y a la forma de abastecimiento de agua en los hogares.

Realizando un análisis acerca de estos resultados, podemos indicar que la información obtenida es relevante ya que ambos factores son obviamente determinantes en el desarrollo para un hogar. Si el ingreso familiar varía, la posibilidad para adquirir mejores servicios se ve ampliamente afectados; dependerá también de la cantidad de miembros que tengan en la familia, afectando al ingreso promedio familiar. Por otro lado, si la familia cuenta con mejores accesos al agua, tendrá mayor posibilidad de desarrollo.

En tal sentido, con un mayor análisis se podrá recomendar las acciones a tomar que puedan impactar directamente a estas variables con la finalidad de modificar la clasificación que tenga un hogar y pasarlo de pobre extremo a pobre y de pobre a no pobre eventualmente.

- Por otro lado, se evalúan también las variables que no tienen mayor incidencia en el modelo o clasificación, encontrándose los accesos a los servicios higiénicos, el

acceso a los programas sociales (beneficiarios), el número de discapacitados dentro del hogar y el número de pisos de la vivienda.

Consideramos que estas variables deben de ser revisadas, otros indicadores mejores u otra forma de medir su impacto en el grupo socioeconómico podría mejorar el proceso de clasificación.

- Utilizando este análisis, se puede analizar los patrones de la información en los distintos cluster, a continuación, la evaluación de algunas de estas variables.

Cluster 0: No pobre

Cluster 1: Pobre Extremo

Cluster 2: Pobre

Ingreso_promedio_familiar

En la figura 68 podemos observar que los hogares clasificados como No pobres tienen un ingreso promedio familiar mayor que los otros dos clusters, el cual varía desde los 473.33 a 3,800 soles, mientras que el cluster de pobreza tiene un promedio que va desde los 300 a 400 soles y el cluster de extrema pobreza inicia en 50 soles.

Figura 68. Análisis Ingreso promedio familiar

Fuente: Elaboración propia

Agua_rio: Podemos observar que los hogares clasificados como No pobres, no utilizan como principal fuente de acceso al agua los ríos.

Figura 69. Análisis acceso al agua del rio

Fuente: Elaboración propia

Material_pared_adobe: Se puede observar en la figura 70 que la mayor concentración de hogares con material de adobe en la pared se encuentra en los clusters de pobre y pobre extremo.

Figura 70. Análisis Material de la pared tipo adobe

Fuente: Elaboración propia

Cantidad_propiedades: Según los resultados obtenidos, como se puede apreciar en la figura 71, los hogares que se encuentran clasificados como No pobres, tienen la menor concentración de cantidad de propiedades, lo cual indica un comportamiento no esperado, el cual deberá ser analizado.

Figura 71. Análisis cantidad de propiedades

Fuente: Elaboración propia

De esta manera se podrá continuar con el análisis detallado de cada una de las variables utilizadas para el modelo a fin de determinar su relevancia.

Objetivo2: En el análisis predictivo, se busca que, en base al modelo definido, se pueda determinar el clúster de clasificación socioeconómico al que pertenece un hogar. Es decir, se pueda predecir en base al comportamiento de una familia, cuál debería ser el grupo socioeconómico al que pertenece.

En base a la ejecución del modelo utilizando la data de prueba (189 hogares), se puede observar que es posible identificar la clasificación de estos.

Figura 72. Resultado Ejecución Data Prueba

Fuente: Elaboración propia

Los resultados del detalle de la clasificación se almacenan en un archivo ARTF, propio de la herramienta utilizada, como se muestra en la figura 73.

Figura 73. Resultados de Ejecución de Prueba en formato ARTF de Weka

Fuente: Elaboración propia

Dichos resultados pueden ser exportados a formato CSV para su posterior análisis, tal como se muestra en la figura 69.

AE	AF	AG	AH	AI	AJ	AK	AL	AM	AN	AP
No_Discapitados	Beneficiarios	No_Beneficiarios	Con_Seguro	Sin_Seguro	Sector_agricola	Sector_Otros	Idioma_quechua	Idioma_castellano	Idioma_Otro	Cluster
2	0	4	1	3	1	3	1	3	0	cluster0
5	2	3	5	0	1	4	5	0	0	cluster2
5	3	2	5	0	1	4	4	1	0	cluster2
5	3	2	5	0	1	4	5	0	0	cluster1
3	0	3	3	0	0	3	3	0	0	cluster2
2	0	2	2	0	1	1	0	2	0	cluster1
4	2	2	4	0	1	3	0	4	0	cluster1
3	0	3	3	0	0	3	1	2	0	cluster2
3	2	1	3	0	1	2	0	3	0	cluster1
3	1	2	3	0	1	2	0	3	0	cluster1
2	1	1	2	0	1	1	0	2	0	cluster1
3	1	2	3	0	1	2	0	3	0	cluster1
4	0	4	4	0	0	4	4	0	0	cluster2
6	2	4	6	0	1	5	0	6	0	cluster1
4	1	3	4	0	1	3	0	4	0	cluster1
3	0	3	3	0	1	2	0	3	0	cluster1
3	1	2	3	0	2	1	0	3	0	cluster1
1	0	1	1	0	1	0	0	1	0	cluster1

Figura 74. Extracto de resultados de Ejecución Data de Prueba en CSV

Fuente: Elaboración propia

*Se identificó el cluster o clasificación que le corresponde a cada uno de los 189 hogares procesados, según el modelo de entrenamiento ingresado previamente.

Objetivo3: Identificar a los hogares que presenten alguna anomalía o que no puedan ser fácilmente agrupados a fin de determinar si corresponde a fraude o error.

Utilizaremos una de las variables de mayor trascendencia para realizar este análisis.

Como se observa en la figura 75, encontramos en el cluster 2 (Pobre), un hogar con un promedio de ingreso familiar que se aleja del promedio grupal, esto nos advierte de algún tipo de inconsistencia.

Figura 75. Información anómala según ingreso promedio familiar

Fuente: Elaboración propia

Al ingresar al detalle de los datos de este hogar, podemos notar que, teniendo un ingreso familiar de 8,000 soles (promedio 2,000), este hogar no cuenta con los servicios básicos como el agua potable, servicios higiénicos ni combustible. Esta data no es coherente, lo cual indica que la información de este hogar deberá ser revisada.

Attribute	Value
Instance_number	79.0
Nro_pisos_vivienda	1.0
Propiedad_PropiaPagada	1.0
Propiedad_otros	0.0
Material_pared_adobe	0.0
Material_pared_otro	1.0
Material_techo_calamina	1.0
Material_techo_otro	0.0
Material_piso_tierra	0.0
Material_piso_otro	1.0
Alumbrado_electricidad	1.0
Alumbrado_Vela	0.0
Alumbrado_Otro	0.0
Agua_rio	1.0
Agua_otro	0.0
SSH pozo ciego	1.0
SSH_otro	0.0
Combustible_cocina_leña	1.0
Combustible_cocina_otro	0.0
Artefacto_Celular	1.0
Artefacto_No_Tiene	0.0
Artefacto_No_Tiene	0.0
Artefacto_Otro	0.0
Nro_integrantes	4.0
Servicios	1.0
Ingreso_familiar	8000.0
ingreso_promedio_familiar	2000.0
Cant_propiedades	0.0
Cant_vehiculos	0.0
Lee_escribe	3.0
Discapacitados	0.0
No_Discapacitados	4.0
Beneficiarios	2.0
No_Beneficiarios	2.0
Con_Seguro	4.0
Sin_Seguro	0.0
Sector_agricola	2.0
Sector_Otros	2.0
Idioma_quechua	1.0
Idioma_castellano	3.0
Idioma_Otro	0.0
Clasificación	No pobre
Cluster	cluster2

Figura 76. Posible filtración de información

Fuente: Elaboración propia

Conclusión: Del análisis realizado, podemos mencionar que los criterios de éxito han sido cumplidos, ya que ha sido posible determinar las variables de menor impacto para que puedan ser reevaluadas y el modelo utilizado ha tenido una precisión del 81% (19 % de instancias mal clasificadas)

Aumentar el análisis de acuerdo al plan de pruebas

- **Análisis de acuerdo con el plan de pruebas**

Para el algoritmo K-means se ha definido como medida de calidad el porcentaje de error de clasificación, cuyo valor no debe superar el 25%. En la figura 77 se puede observar que el valor del error de clasificación alcanzado por el modelo.

```

kMeans
=====
Incorrectly clustered instances:  19.0      19.1919 %

```

Figura 77. Resultado SSE

Fuente: Elaboración propia

El error de clasificación alcanzado por el modelo es de 19%, un valor menor a lo esperado (25%).

Para el algoritmo Naive Bayes se ha definido como medidas de calidad: precisión y exhaustividad. En la figura 78 se puede observar el valor de ambas medidas.

Class	Precision	Recall
Pobre	0,789	0,882
No pobre	1,000	0,926
Pobre Extremo	0,861	0,816

Figura 78. Resultados Precisión y Exhaustividad

Fuente: Elaboración propia

La precisión nos indica que el modelo acierta el 78.9% en predecir la clase “Pobre”, 100% en predecir la clase “No pobre” y 86.1% en predecir la clase “Pobre Extremo”.

La exhaustividad nos indica que el modelo identifica correctamente el 88.2% de hogares pobres, el 92.6% de hogares no pobres y 81.6% de hogares pobres extremos.

Conclusión: Del análisis realizado, podemos mencionar que los criterios de éxito han sido cumplidos, ya que ha sido posible determinar las variables de menor impacto para que puedan ser reevaluadas y el modelo utilizado ha tenido una precisión del 81% (19 % de instancias mal clasificadas).

4.3.4.5 EVALUACION

Esta etapa de la metodología tiene como finalidad analizar los resultados desde el punto de vista del negocio, determinando si los objetivos fueron alcanzados a fin de continuar con

las siguientes etapas del proyecto o caso contrario, determinar si es necesario una nueva evaluación del proceso.

4.3.4.5.1 EVALUACIÓN DE RESULTADOS

Desde el punto de vista del negocio, se plantearon 2 objetivos principales:

Objetivo 1: Dirigir y supervisar acciones para la determinación de la clasificación socioeconómica de los potenciales usuarios de los programas y subsidios del Estado que se ejecutan bajo criterios de focalización individual

Establecer si la propuesta permite cumplir con este objetivo puede ser un poco subjetivo, por lo que es necesario apoyarnos en los objetivos de la minería de datos, la cual nos indica que ha sido posible clasificar a los hogares con una fiabilidad del 81%. Cabe recalcar en este punto, que la información con la cual se han realizado las pruebas de los modelos no contiene toda la información con la que cuenta el MIDIS, por lo tanto, es posible que los resultados varíen. Sin embargo, lejos de ser una problemática para confirmar que este objetivo ha sido cumplido, abre la posibilidad de realizar acciones para mejorar el modelo enfocándose en el análisis de la información y no solo en la construcción de una herramienta.

Objetivo 2: Fortalecer la focalización y la gestión de usuarios contribuyendo al logro de objetivos de los programas sociales, con efectividad, articuladamente y con calidad.

De igual manera que el objetivo anterior, uno de los objetivos de la minería de datos nos ayuda a mostrar cómo se puede alcanzar este objetivo del negocio, ya que a partir de la identificación de los hogares que presentan anomalías en su información y que podrían estar presentando una clasificación incorrecta, será posible corregir o prevenir el uso de recursos del Estado en hogares a quienes no les corresponde, fortaleciendo a la focalización para una mejor efectividad.

4.3.4.5.2 REVISIÓN DEL PROCESO

Con la información trabajada hasta el momento en función a los datos recopilados de los formatos físicos, se ha podido construir un modelo que nos permite determinar la clasificación socioeconómica con una fiabilidad del 81%.

La mayor dificultad del proceso estuvo en la determinación inicial de la clasificación socioeconómica, debido a que los datos con los que se contaba no contenían la totalidad de

información que tiene el MIDIS, se espera que con una base de datos más completa, se puedan obtener resultados más detallados.

4.3.4.5.3 DETERMINACIÓN DE LOS PASOS O ACCIONES A SEGUIR

Analytics nos permite realizar la integración de diversas fuentes de información para el análisis y creación de modelos. Por lo tanto, aunque en la presente propuesta no se elaboren modelos más completos, se espera que, en una ejecución real, se pueda adicionar mucha más información que permita establecer patrones generales para los hogares.

4.4 EVALUACIÓN DE BENEFICIOS

4.4.1 EVALUACIÓN ECONÓMICA

4.4.1.1 AHORRO ESPERADO

En este punto se considera el presupuesto que el Estado destina para la reducción de la pobreza. Analizaremos el presupuesto general de los programas que utilizan la información del proceso de clasificación socioeconómico y el impacto esperado como resultado de la mejora del proceso en el MIDIS.

- **Costo de la Filtración Actual:** Considerando el porcentaje de filtración que tenemos como referencia, en la tabla 32 se muestran el costo asociado, el cual alcanza los S/ 478 millones de soles.

Tabla 32

Costo asociado a la Filtración Actual

Programa Social	Presupuesto	Beneficiarios	Filtraciones (como % de beneficiarios)	Filtraciones (absoluto)	Costo de filtraciones
SIS	647,420,454	9,503,409	39.70%	3,775,277	267,916,923
Juntos	852,844,592	518,339	17.80%	92,346	151,940,426
Pensión 65	264,357,731	133,099	22.10%	29,106	58,404,859
Total					478,262,208

Fuente: Elaboración propia, se tomó como fuente de información ENAHO 2012 – INEI

- **Supuesto:** Si consideramos una reducción en el porcentaje de filtración de al menos el 0.5%, obtendremos el resultado que se muestra en la tabla 33 donde se puede visualizar el costo esperado de filtración por un monto de S/469 millones de soles.

Tabla 33

Costo asociado a la Filtración Propuesta

Programa Social	Presupuesto	Beneficiarios	Filtraciones (como % de beneficiarios)	Filtraciones (absoluto)	Costo de filtraciones
SIS	647,420,454	9,503,409	39.200%	3,725,336	264,498,879
Juntos	852,844,592	518,339	17.30%	89,673	147,511,504
Pensión 65	264,357,731	133,099	21.60%	28,749	57,096,277
Total					469,106,660

Fuente: Elaboración propia, se tomó como fuente de información ENAHO 2012 – INEI

- **Ahorro Esperado:** Con este supuesto, tenemos un ahorro de S/ 9.1 millones anual.

4.4.1.2 PRESUPUESTO DEL PROYECTO

Como marco preliminar para el análisis del presupuesto, se considera que el costo del proyecto será asumido por el MIDIS, que, al ser una institución pública, se rige bajo normas y presupuestos asignados por parte del Estado.

En la tabla 35 se muestra el presupuesto utilizado para el análisis del proyecto planteado, considerando los siguientes factores:

- Se considera el costo total del equipo de trabajo, software y hardware necesario, a excepción de los servidores debido a que la institución ya cuenta con los mismos.
- Respecto a los gastos de Gestión Administrativa, dentro de la ley de presupuesto que se asigna al MIDIS cada año, los mismos son asignados a un centro de costo específico y no directamente a los proyectos en curso, motivo por el cual tampoco han sido considerados en el análisis. Sobre este punto, hacemos referencia a la información proporcionada por el Ministerio de Economía y Finanzas respecto a estas partidas presupuestales, como se muestra en la tabla 34.
- No existen costos asociados por local (compra ni alquiler) debido a que se considera el desarrollo del proyecto dentro de las instalaciones propias del MIDIS.

Tabla 34
Partidas presupuestales

Año de Ejecución: 2019								
Unidad Ejecutora 001-1424: MINISTERIO DE DESARROLLO E INCLUSION SOCIAL- SEDE CENTRAL- MIDIS								
Actividad/Acción de Inversión/Obra 5000003: GESTION ADMINISTRATIVA								
Meta 00005-1424: ACCIONES DE LOGISTICA								
Detalle Especifica	PIA	Presupuesto asignado	Certificación	Compromiso Anual	Ejecución			Avance %
					Atención de Compromiso Mensual	Devengado	Girado	
1: SERVICIO DE SUMINISTRO DE ENERGIA ELECTRICA	973,037	723,037	462,785	462,785	459,748	68,546	68,206	9.5
2: SERVICIO DE AGUA Y DESAGUE	94,767	94,767	48,780	48,780	47,819	6,840	6,631	7.2
Meta 00005-1424: ACCIONES DE TECNOLOGÍAS DE LA INFORMACIÓN								
Detalle Especifica	PIA	Presupuesto asignado	Certificación	Compromiso Anual	Ejecución			Avance %
					Atención de Compromiso Mensual	Devengado	Girado	
1: SERVICIO DE TELEFONIA MOVIL	268,073	191,339	152,514	88,458	88,458	75,027	75,027	39.2
2: SERVICIO DE TELEFONIA FIJA	46,367	449,301	123,100	123,100	123,100	54,376	54,376	12.1
3: SERVICIO DE INTERNET	250,386	310,386	286,993	286,993	286,993	47,378	47,378	15.3

Fuente: Consulta amigable del Ministerio de Economía y Finanzas – 2019

Tabla 35
Presupuesto de Proyecto

PRESUPUESTO DE LA SOLUCIÓN ANALYTICS				
Hardware	Medida	Cantidad	Precio Unitario	Total
Servidores de Aplicaciones	Unidad	0	S/. 0	S/. 0
Servidores de Base de Datos	Unidad	0	S/. 0	S/. 0
Servidor de Almacenamiento	Unidad	1	S/. 680,000	S/. 680,000
PCs	Unidad	9	S/. 9,100	S/. 81,900
Multifuncional	Unidad	1	S/. 3,200	S/. 3,200
Sub Total				765,100
Software	Cantidad	Meses	Precio Unitario	Total
Suscripción Minería de Datos	Unidad	12	S/. 3,000	S/. 36,000
Sub Total				36,000
Gastos administrativos - Personal	Cantidad	Meses	Precio Unitario	Total
Administrador de Base de Datos	1	12	S/. 12,000	S/. 144,000
Analistas Junior de Minería de Datos	1	12	S/. 7,000	S/. 84,000
Analistas Senior de Minería de Datos	1	12	S/. 10,000	S/. 120,000
Jefe de Proyecto	1	12	S/. 12,000	S/. 144,000
Analista de QA	2	12	S/. 8,000	S/. 192,000
Analista de Sistemas	2	12	S/. 8,000	S/. 192,000
Analista de Seguridad	1	12	S/. 9,000	S/. 108,000
Sub Total		9		984,000
Mobiliario	Unidad	Cantidad	Precio Unitario	Total
Escritorios	Unidad	9	S/. 780	S/. 7,020
Sillas	Unidad	9	S/. 520	S/. 4,680
Sub Total				S/. 11,700
Gastos administrativos - Otros	Unidad	Meses	Precio Unitario	Total
Local		12	S/. 0	S/. 0
Electricidad y agua		12	S/. 0	S/. 0
Telefonía e internet		12	S/. 0	S/. 0
Útiles de oficina		12	S/. 250	S/. 3,000
Sub Total				S/. 3,000
TOTAL PRESUPUESTO INICIAL DEL PROYECTO				S/. 1,796,800

Fuente: Elaboración Propia

4.4.1.3 FLUJO DE CAJA DEL PROYECTO

Para este análisis, se tomará en cuenta el ahorro esperado por concepto de filtración por la ejecución de la propuesta.

Se consideran los siguientes supuestos:

- Tiempo de duración del proyecto: 1 año (Consideramos ésta como la etapa inicial de la evaluación de la información, sin descartar posteriores actualizaciones o mejoras).
- Se realizará un análisis para 4 años.

- No existe deuda debido a que la institución recibe un presupuesto del Estado.
- Se utilizará como referencia el costo de una herramienta de minería de datos comercial. Esto puede variar ya que la propuesta no pretende indicar la herramienta a utilizar para la implementación de la solución.
- Tasa Social de Descuento: Se considera una TDS de 9%.
- En el Sector Público, el Instructivo Contable N° 2 “Criterio de valuación de los bienes de activo fijo, método y porcentaje de depreciación y amortización de los bienes de activo fijo e infraestructura pública”, aprobado mediante R.C. N° 067-97-EF/93.01 de fecha 31.12.97 y modificatorias, señala en el artículo 5.2 los porcentajes anuales de depreciación que son mostrados en la Tabla 36:

Tabla 36
Porcentaje Depreciación

BIENES	%
Edificios.	3
Infraestructura pública.	3
Maquinarias, Equipo y otras unidades para la producción.	10
Equipos de transporte.	25
Muebles y enseres.	10
Equipos de cómputo.	25

Fuente: Ministerio de Economía y Finanzas, 2019

ESTADO DE RESULTADOS

Para el análisis del Estado de Resultados consideramos los siguientes aspectos:

1. **Determinación de las ventas (Ahorro proyectado):** Al ser un proyecto orientado a disminuir el costo por filtración en el proceso de clasificación socioeconómico, este punto se analizará como el ahorro esperado anual con el supuesto de la disminución en el porcentaje de filtración de 0.5% que se analizó en el punto de ahorro esperado.
2. **Costo de Ventas (Costo software):** Consideramos el costo asociado al pago anual por el servicio de herramienta de minería de datos.
3. **Gastos administrativos:** Consideramos las remuneraciones del personal asignado al desarrollo del proyecto, el mantenimiento del mismo y los útiles de oficina.
4. **Depreciación y/o amortización:** Para obtener este valor consideramos el costo de inversión inicial del proyecto por la adquisición del mobiliario a utilizar por el equipo

de proyecto, el cual consta de escritorios, sillas, máquinas e impresora multifuncional, también se considera dentro de este rubro el costo por adquisición de equipo de almacenamiento. El detalle lo podemos observar en la tabla 37.

Tabla 37
Inversión Inicial del Proyecto

Inversión inicial	Valor Venta	IGV 18%	Inversiones
Mobiliario	9,915	1,785	11,700
PC's	69,407	12,493	81,900
Almacenamiento	576,271	103,729	680,000
Multifuncional	2,712	488	3,200
TOTAL	658,305	118,495	776,800

Fuente: Elaboración propia

Utilizaremos también la información de la tabla de porcentaje de depreciación establecida por el MEF para el cálculo de la depreciación anual como se puede observar en la tabla 38.

Tabla 38
Cálculo Depreciación

Inversión	Valor de Venta	% depreciación	Año 1	Año 2	Año 3	Año 4	Depreciación acumulada	Valor residual
Mobiliario	9,915	10%	991.53	991.53	991.53	991.53	3,966.10	5,949.15
PC's	69,407	25%	17,351.69	17,351.69	17,351.69	17,351.69	69,406.78	-
Almacenamiento	576,271	25%	144,067.80	144,067.80	144,067.80	144,067.80	576,271.19	-
Multifuncional	2,712	25%	677.97	677.97	677.97	677.97	2,711.86	-
Depreciación anual			163,088.98	163,088.98	163,088.98	163,088.98	652,355.93	5,949.15

Fuente: Elaboración propia

- 5. Gastos Financieros:** No se consideran gastos financieros debido a que el presupuesto para el proyecto se asigna según el presupuesto anual, no existiendo deudas a considerar.
- 6. Impuesto a las Ganancias:** Debido a que el proyecto corresponde a una inversión pública, no está gravado a impuestos. Adicionalmente, el proyecto no tiene una ganancia directa, sino que el análisis se representa como un beneficio por ahorro del presupuesto asignado para los programas sociales asociados al proceso de clasificación socioeconómico.

En la tabla 39 podemos observar el estado de resultados proyectados a 4 años, donde se observa que el ahorro neto anual se encuentra cercano a los S/ 8 Millones.

Tabla 39
Estado de Resultados

Nro	Estado de Resultados Proyectado	Año 0	Año 1	Año 2	Año 3	Año 4
1	Ahorro proyectado de presupuesto		-	9,155,547.8	9,155,547.8	9,155,547.8
2	Costo software		36,000.0	36,000.0	36,000.0	36,000.0
	Ahorro (Utilidad) Bruta		(36,000.0)	9,119,547.8	9,119,547.8	9,119,547.8
3	Gastos administrativos		987,000.0	987,000.0	987,000.0	987,000.0
4	Depreciación y/o Amortización		163,089.0	163,089.0	163,089.0	163,089.0
	Ahorro (Utilidad) antes de interés e impuestos (EBIT)		(1,186,089.0)	7,969,458.8	7,969,458.8	7,969,458.8
5	Gastos Financieros		-	-	-	-
	Ahorro (Utilidad) imponible		(1,186,089.0)	7,969,458.8	7,969,458.8	7,969,458.8
6	Impuesto a las ganancias (0%)		-	-	-	-
	Ahorro Neto (Utilidad neta)		(1,186,089.0)	7,969,458.8	7,969,458.8	7,969,458.8

Fuente: Elaboración propia

FLUJO DE CAJA PROYECTADO

Para elaborar el Flujo de Caja proyectado, utilizamos la información obtenida del estado de resultados.

1. **EBIT (Beneficio antes de intereses e impuestos):** Para el proyecto propuesto, este monto corresponde a la utilidad neta o ahorro neto, ya que como se observó en el punto anterior, no consideramos gastos financieros ni impuesto a las ganancias.
2. **NOPAT (Beneficio neto después de impuestos):** Este monto se mantendrá igual al EBIT debido a la ausencia de impuestos.
3. **Depreciación y amortización:** Consideramos los mismos montos utilizados en el estado de resultados.
4. **Cambio en capital de trabajo:** Para esta evaluación no se considera ninguna inversión adicional.
5. **Valor residual:** Tomaremos este valor del cálculo realizado en la tabla 33, donde se especifica el valor residual al año 4.
6. **Inversiones en Activos:** Para la evaluación de caja libre, consideramos como inversión en la etapa inicial del proyecto a los costos asociados a mobiliario y equipo del personal, estos costos se muestran en la tabla 37.
7. **Flujo de caja libre:** Como se puede observar en la tabla 40, al año 4 tenemos un flujo de caja de más de S/ 8 Millones.

Tabla 40
Flujo de Caja Libre

Nro	FLUJO DE CAJA LIBRE	Año 0	Año 1	Año 2	Año 3	Año 4
1	Ahorro (Beneficio) antes de interés e impuestos (EBIT)		(1,186,089.0)	7,969,458.8	7,969,458.8	7,969,458.8
	Impuestos a las ganancias (sin escudo tributario)		-	-	-	-
2	EBIT - impuestos = NOPAT		(1,186,089.0)	7,969,458.8	7,969,458.8	7,969,458.8
3	(+) Depreciación y amortización		163,089.0	163,089.0	163,089.0	163,089.0
4	(-) Cambio en capital de trabajo	-	-	-	-	-
5	(+) Valor residual					5,949.2
6	Inversiones en activos	(776,800.0)			-	
7	FLUJO DE CAJA LIBRE	(776,800.0)	(1,023,000.0)	8,132,547.8	8,132,547.8	8,138,496.9

Fuente: Elaboración propia

ANALISIS DE INDICADORES FINANCIEROS Y SOCIALES

Para analizar la viabilidad financiera del proyecto, analizaremos los indicadores financieros y sociales que se obtuvieron como resultado del análisis previo realizado.

Al ser un proyecto de inversión social, se considera una TDS (Tasa de descuento social) de 9.0%, y se obtienen los siguientes resultados:

- **VANS** (Valor actual neto social): S/ 17,175,006.09
- **TIRS** (Tasa interna de retorno social): 223.3%
- **Índice costo beneficio:** 23.11

Estos resultados son ampliamente favorables, lo cual confirma la viabilidad financiera del proyecto.

Tabla 41
Indicadores Financieros y sociales del Proyecto

Indicador	Valor
TDS	9.0%
VANS	17,175,006.09
Índice beneficio costo	23.11
TIRS	223.3%

Fuente: Elaboración propia

4.4.2 BENEFICIOS SOCIALES

El MIDIS es una institución del Estado, creada con la finalidad de mejorar la calidad de vida de la población. Sin embargo, dentro de su último análisis FODA, una de las debilidades identificadas es la de presentar limitaciones en el proceso de focalización, lo cual genera filtraciones y sub-coberturas en los programas sociales.

Como pudimos observar en la situación actual, el porcentaje de filtración alcanzado en el año 2012 es de 35%, representando una pérdida o mala asignación de los recursos del Estado por el valor de S/.478'262,208.

Como parte de la evaluación de un proyecto de inversión pública, tomaremos en cuenta la rentabilidad social, la cual involucra la comparación de los beneficios sociales y los costos sociales atribuibles a la propuesta.

Entre los beneficios sociales que podemos identificar para la presente propuesta se encuentran:

- Mayor participación de la población vulnerable a las distintas intervenciones públicas focalizadas al poder ser clasificadas correctamente. Esta participación le permitirá al poblador contar con servicios básicos que le ayuden a mejorar su calidad de vida. La ampliación de la cobertura de los programas y subsidios del Estado permite la atención de aquellas poblaciones que actualmente no están priorizadas.
- Disminución de costos asociados a la atención de salud, la población con menores posibilidades podrá tener una mejor alimentación lo que previene la generación de enfermedades que posteriormente requieran tratamiento, generando costos adicionales.
- Efectividad del proceso de clasificación socioeconómica al disminuir los porcentajes de filtración y subcobertura.
- Finalmente, se logra la reducción de la brecha de pobreza, al lograr que los más pobres puedan acceder a los servicios que proporcionan los programas y subsidios del Estado.

5 CONCLUSIONES

El Perú es un país con muchas necesidades sociales, y el Estado Peruano no puede cubrir todas las expectativas de la población. Es en este escenario que la correcta identificación de los hogares o familias que requieren de manera prioritaria la asistencia mediante los diferentes programas sociales con los que cuenta el gobierno peruano cobran mayor relevancia. Debido a que esta identificación se realiza a través del proceso de clasificación socioeconómica, los esfuerzos que se realicen para mejorar dicho proceso deben formar parte fundamental en los objetivos estratégicos del MIDIS.

No pretendemos indicar que el proceso del MIDIS es incorrecto, lo que se busca con la presente propuesta es brindarle a la institución, una alternativa que le permita enfocar sus esfuerzos en la identificación de información relevante para el proceso de clasificación socioeconómica. Tal como se pudo observar en las pruebas realizadas, las herramientas analíticas proporcionan un soporte simple para la revisión de cada variable dentro de los modelos utilizados, lo que permite a su vez reorientar dichas variables en caso se identifique que no contribuye en los resultados deseados.

Sin embargo, debe tenerse en cuenta que la herramienta no es la solución, lo que es realmente importante será el análisis que se desprenda de los resultados obtenidos en las diferentes ejecuciones. Podrán evaluarse diversos algoritmos hasta encontrar el que mejor se adecue a los objetivos de la institución.

El uso de una metodología de analytics como CRISP-DM permite orientar los esfuerzos de una manera más ordenada, teniendo en cuenta que las metodologías han sido probadas por diferentes instituciones y cuentan con respaldo de múltiples empresas.

El conocimiento de los patrones de un grupo de personas podrá también abrir la posibilidad de orientar esfuerzos para no solo realizar una correcta clasificación socioeconómica, sino poder identificar los factores que alteran esta condición y buscar alternativas que contribuyan a los cambios necesarios para la disminución de la pobreza en el Perú.

6 RECOMENDACIONES

- Si bien, a través de la selección de las variables y la información a la que se tuvo acceso para el presente estudio, se lograron alcanzar los objetivos de negocio y los objetivos de minería de datos planteados, dicha información fue limitada; por tanto, se recomienda que el MIDIS utilice toda la información a la cual tiene acceso, tanto interna como de otras instituciones, a fin de poder evaluar y determinar nuevos patrones en los datos.
- Es importante contar con personal idóneo para el análisis de la información. Se recomienda continuar con las pruebas de las diferentes técnicas de analytics para hallar el que genere los mejores resultados para la institución.
- Con fines de evaluación de la metodología, puede utilizar cualquier herramienta gratuita que le permita tener una rápida visión de los resultados que se pueden obtener.
- La necesidad de contar con una información confiable es la base para cualquier proyecto de análisis de datos, por lo cual se recomienda realizar trabajos para mejorar la información. Contar con otras fuentes de información será beneficioso para la determinación de patrones cada vez más confiables.
- No se debe considerar que el proyecto es de una sola ejecución y que no es necesario un monitoreo, por el contrario, el análisis constante de los modelos servirá para tener cada vez mejores resultados.

7 REFERENCIAS

Affeldt, F. S., & da Silva Júnior, S. D. (2013). INFORMATION ARCHITECTURE ANALYSIS USING BUSINESS INTELLIGENCE TOOLS BASED ON THE INFORMATION NEEDS OF EXECUTIVES. *Journal of Information Systems and Technology Management*, 10(2), 251–270.

Ariza-López, F. J., Rodríguez-Avi, J., & Alba-Fernández, V. (2018). CONTROL Estricto de Matrices de Confusión por medio de distribuciones multinomiales. *GeoFocus. Revista Internacional de Ciencia y Tecnología de la Información Geográfica*, (21), 215–226.

Brown, C., Ravallion, M., & Van de Walle, D. (2016). A poor means test? Econometric targeting in Africa. The World Bank.

Bustamante Martínez, A., Galvis Lista, E. A., & Gómez Flórez, L. C. (2013). Técnicas de modelado de procesos de ETL: una revisión de alternativas y su aplicación en un proyecto de desarrollo de una solución de BI. *Scientia et Technica*, 18(1).

Camargo-Vega, J. J., Camargo-Ortega, J. F., & Joyanes-Aguilar, L. (2015). Knowing the big data. *Facultad de Ingeniería*, 24(38), 63–77.

Castañeda, T., Lindert, K., de la Brière, B., Fernandez, L., Hubert, C., Larrañaga, O., ... Viquez, R. (2005). Designing and implementing household targeting systems: lessons from Latin America and the United States. *World Bank Social Protection Discussion Paper Series*, 526.

Chen, H., Chiang, R. H., & Storey, V. C. (2012). Business intelligence and analytics: from big data to big impact. *MIS quarterly*, 1165–1188.

Davis, E. P., & Sanchez-Martinez, M. (2015). Economic theories of poverty. Joseph Rowntree Foundation.

Gómez, A. A. R., & Bautista, D. W. R. (2010). Inteligencia de negocios: Estado del arte. *Scientia et Technica*, 1(44), 321–326.

Hernández-Leal, E. J., Duque-Méndez, N. D., & Moreno-Cadavid, J. (2017). Big Data: una exploración de investigaciones, tecnologías y casos de aplicación. *Tecno Lógicas*, 20(39), 17–24.

Hindle, G. A., & Vidgen, R. (2018). Developing a business analytics methodology: A case study in the foodbank sector. *European Journal of Operational Research*, 268(3), 836–851.

Lara, Y. T. A., Zepahua, B. A. O., Rodríguez-Mazahua, L., Alor-Hernández, G., & Contreras, H. M. (2016). Propuesta de arquitectura para un módulo de inteligencia de negocios basado en minería de datos. *Research in Computing Science*, 128, 47–56.

Laureano, R. M., da Silva Laureano, L. M., & Grencho, A. R. R. R. (2016). Framework to implement business analytics: Phases and critical success factors. En *Information Systems and Technologies (CISTI)*, 2016 11th Iberian Conference on (pp. 1–6). IEEE.

LaValle, S., Lesser, E., Shockley, R., Hopkins, M. S., & Kruschwitz, N. (2011). Big data, analytics and the path from insights to value. *MIT sloan management review*, 52(2), 21.

Lindert, K., Linder, A., Hobbs, J., & De la Brière, B. (2007). The nuts and bolts of Brazil's Bolsa Família Program: implementing conditional cash transfers in a decentralized context. *Social Protection Discussion Paper*.

Mahmoud, F. Z. M. (2017). The Application of Predictive Analytics: Benefits, Challenges and how it can be improved, 7(5), 18.

Martínez Álvarez, J. A. (2003). Economía de la pobreza.

Mohmmad, R., Butt, M. A., & Zaman, M. (2017). Tools for Predictive Analytics: An Overview, 6(7), 3.

Năstase, P., & Stoica, D. (2011). A new business dimension-business analytics. *Accounting & Management Information Systems/Contabilitate Si Informatica De Gestione*, 9(4).

Parra, O. J. S., Galeano, R. M., & Rodriguez, L. G. (2010). Metodología crisp para la implementación Data Warehouse. *Tecnura*, 14(26), 35–48.

Perez, P., & Rodriguez, A. (2015). «El ejercicio de medir la pobreza en el Perú» - Buscar con Google. Centro de Investigación de la Universidad del Pacífico. Recuperado de

<https://www.google.com/search?client=firefox-b-ab&q=%22El++ejercicio++de++medir++la++pobreza++en++el++Per%C3%BA%22>

Peyre Dutrey, A. (2007). Reporting Efficiency and Costs in Targeted Poverty Alleviation Programmes. Social Policy and Development Programme Paper, (35).

Rantzer, M. ans. (2016). Predictive analysis at Kronofogden: Classifying first-time debtors with an uplift model.

Redjeki, S., Guntara, M., & Anggoro, P. (2015). Naive Bayes Classifier Algorithm Approach for Mapping Poor Families Potential. Informatics Engineering Department, 4, 5.

Riquelme Santos, J. C., Ruiz, R., & Gilbert, K. (2006). Minería de datos: Conceptos y tendencias. Inteligencia artificial: Revista Iberoamericana de Inteligencia Artificial, 10(29), 11–18.

Rosemberg, J. C. (2017). Propuesta de aplicación de analítica de negocios para la optimización de recursos en la concesión coviperú sa, a partir de la predicción del comportamiento del flujo vehicular.

Sabates-Wheeler, R., Hurrell, A., & Devereux, S. (2015). Targeting social transfer programmes: Comparing design and implementation errors across alternative mechanisms. Journal of International Development, 27(8), 1521–1545.

Schniederjans, M. J., Schniederjans, D. G., & Starkey, C. M. (2014). Business analytics principles, concepts, and applications: what, why, and how. Pearson Education.

Shafique, U., & Qaiser, H. (2014). A comparative study of data mining process models (KDD, CRISP-DM and SEMMA). International Journal of Innovation and Scientific Research, 12(1), 217–222.

Sharma, A., Mansotra, V., & Shastri, S. (2015). An Exploratory Analysis of Public Healthcare Data: A Case Study of Jammu & Kashmir State. Asian Journal of Computer and Information Systems (ISSN: 2321–5658), 3(05).

Sharma, R., Mithas, S., & Kankanhalli, A. (2014). Transforming Decision-Making Processes: A Research Agenda for Understanding the Impact of Business Analytics on Organisations.

Tamayo, M., & Moreno, F. J. (2006). Análisis del modelo de almacenamiento MOLAP frente al modelo de almacenamiento ROLAP Ingeniería e Investigación, vol. 26, núm. 3, diciembre, 2006, pp. 135-142 Universidad Nacional de Colombia Bogotá, Colombia. Ingeniería e Investigación, 26(3), 135–142.

Vicente, I. C. (2008). Factores que afectan la implementación exitosa de un Data Warehouse. Revista vínculos, 5(1), 11–22.

World Bank Group. (s. f.). The State of Social Safety Nets 2018. International Bank for Reconstruction and Development, 43. <https://doi.org/10.1596/978-1-4648-1254-5>

8 ANEXOS

ANEXO 01:

Datos del informe técnico del INEI 2017

- **La pobreza y la edad:**

De acuerdo con lo indicado por el INEI (2017), la mayor cantidad de pobreza agrupa a la población infantil y adolescente. Se atribuye a este factor las situaciones de riesgo como desnutrición, abandono, poco acceso a servicios médicos, entre otros. Estas condiciones reducen las oportunidades de la población infantil y adolescente al ser difícilmente rebatibles.

Podemos observar en la siguiente figura 79, la incidencia de la pobreza por edades.

Figura 79. Pobreza por grupos de edades

Fuente: INEI – 2017

- **La pobreza y la lengua materna:**

El mayor porcentaje de la población del Perú tiene como lengua materna el castellano, y como se puede observar en la figura 80, la población cuya lengua materna ha sido una lengua nativa como el quechua, aymara u otra, tiene el mayor porcentaje de pobreza.

Figura 80. Pobreza según lengua materna

Fuente: INEI - 2017

Según estas cifras del INEI (2017), del porcentaje de la población que se encuentra en la clasificación de pobreza o pobreza extrema, el 33% de la misma tiene como lengua materna una lengua nativa

la pobreza afectó al 33% En el año 2017, la pobreza afectó al 33,0% (6,9% pobre extremo y 26,1% pobre no extremo) de las personas que mencionaron tener como lengua materna una lengua nativa, siendo casi el doble de la incidencia en la población con lengua materna castellano, 18,6% (2,9% pobre extremo y 15,7% pobre no extremo).

- **La pobreza y la Región Natural:**

En el año 2017, el 54,1% de los pobres se localiza en el área urbana y el 45,9% en el área rural. A nivel de región natural, el 46,2% están en la Sierra, el 36,9% en la Costa y el 16,9% en la Selva.

En cambio, el pobre extremo, mayoritariamente se encuentran en el área rural del país, llegando al 76,0%. Según regiones naturales, el 66,7% están en la Sierra, el 21,1% en la Selva y el 12,2% en la Costa.

Estos datos son mostrados en la tabla 42.

Tabla 42

Localización territorial de pobreza

PERU: LOCALIZACIÓN TERRITORIAL DE LA POBLACION POBRE, SEGÚN ÁREA DE RESIDENCIA, 2007 - 2017 (Distribución porcentual)											
Niveles de pobreza/ Área de residencia	2007	2008	2009	2010	2011	2012	2013	2014	2015	2016	2017
Pobre											
Área de residencia	100,0	100,0	100,0	100,0	100,0	100,0	100,0	100,0	100,0	100,0	100,0
Urbana	51,2	49,3	46,5	47,9	48,1	48,2	50,7	51,2	51,0	51,5	54,1
Rural	48,8	50,7	53,5	52,1	51,9	51,8	49,3	48,8	49,0	48,5	45,9
Región natural	100,0	100,0	100,0	100,0	100,0	100,0	100,0	100,0	100,0	100,0	100,0
Costa	37,0	36,3	33,3	34,7	34,5	34,7	35,9	34,5	34,8	34,2	36,9
Sierra	45,9	47,4	48,4	48,5	49,0	49,0	47,2	48,1	47,9	48,9	46,2
Selva	17,1	16,3	18,4	16,8	16,5	16,3	16,9	17,4	17,2	16,9	16,9
Pobre extremo											
Área de residencia	100,0	100,0	100,0	100,0	100,0	100,0	100,0	100,0	100,0	100,0	100,0
Urbana	18,5	18,1	15,8	17,9	16,6	17,3	16,7	17,7	19,1	19,2	24,0
Rural	81,5	81,9	84,2	82,1	83,4	82,7	83,3	82,3	80,9	80,8	76,0
Región natural	100,0	100,0	100,0	100,0	100,0	100,0	100,0	100,0	100,0	100,0	100,0
Costa	9,0	9,3	8,4	10,4	10,3	9,7	9,2	11,9	10,3	7,3	12,2
Sierra	74,1	72,0	69,9	68,3	71,3	72,5	71,9	69,7	69,0	70,6	66,7
Selva	16,9	18,7	21,7	21,3	18,4	17,7	18,9	18,4	20,6	22,2	21,1

Fuente: Instituto Nacional de Estadística e Informática - Encuesta Nacional de Hogares, 2007 - 2017.

Fuente: INEI - 2017

- Evolución de la pobreza según Región Natural**

Podemos observar en la figura 81 la variación de la pobreza según región entre los años 2016 y 2017.

Figura 81. Evolución de la pobreza 2016-2017 por Región Natural

Fuente: INEI - 2017

- **Pobreza y el número de miembros por familia:**

Como se observa en la figura 82, los hogares pobres son de mayor tamaño y tienen un mayor número de niñas y niños que el hogar promedio de los hogares no pobres.

Según los resultados observados, el hogar pobre tiene en promedio un miembro más que los hogares no pobres del total nacional. Por área de residencia, los hogares urbanos pobres tienen 1,4 miembros más que un hogar no pobre y en el área rural la diferencia es de 1,1 miembros.

Figura 82. Miembros de hogar en condición de pobreza, 2017

Fuente: INEI – 2017

- **Pobreza y material predominando en paredes exteriores:**

Como se observa en la figura 83, el material que predomina en las paredes exteriores de la vivienda de los pobres es el adobe /quincha (42,4%), mientras que, en las paredes de la vivienda de los no pobres, predomina el ladrillo o bloque de cemento (60,2%).

Con respecto al año 2016, en las paredes de la vivienda de los pobres se han reducido el uso de material adobe/quincha en 3,6 puntos porcentuales al pasar de 46,0% a 42,4%.

Figura 83. Material predominante en paredes de viviendas en condición de pobreza.

Fuente: INEI - 2017

- **Pobreza y Piso de la vivienda**

En el piso de la vivienda de los hogares pobres, predomina el piso de tierra con 58,8% seguido por el cemento con 32,0%, la madera (entablado) con el 6,8% de viviendas. En cambio, en el piso de la vivienda de los hogares no pobres en el 48,3% predomina el cemento, en el 24,9% el parquet o madera pulida, el 20,6% tiene piso de tierra, entre los principales.

Con respecto al 2016, se ha reducido el piso de tierra en la vivienda de los pobres en 3,8 puntos y el uso de piso de madera en 1,2 puntos porcentuales.

Figura 84. Material predominante en piso de viviendas en condición de pobreza, 2017

Fuente: INEI - 2017

- **Pobreza y Formas de abastecimiento de agua para consumo humano**

En el año 2017, el 69,9% de los hogares pobres se abastecieron de agua para consumo humano proveniente de red pública dentro de la vivienda. En el caso de los hogares no pobres fue el 83,6%, siendo mayor en 13,7 puntos porcentuales. El 11,7% de los hogares pobres se abastece de agua de una fuente proveniente del río, acequia o manantial; en los hogares no pobres el 3,2%.

Figura 85. Abastecimiento de agua en viviendas en condición de pobreza, 2017

Fuente: INEI – 2017

- **Pobreza y Área de residencia (rural/urbana)**

Como se observa en la figura 86, por área de residencia, la pobreza afectó más a los residentes del área rural, que incidió en el 44,4% de su población, siendo tres veces más que en el área urbana (15,1%).

Al comparar con el año 2016, la pobreza aumentó en 1,2 puntos porcentuales en el área

urbana, en tanto, el área rural lo hizo en 0,6 punto porcentual.”

Figura 86. Evolución de la pobreza monetaria, según área de residencia 2007-2017

Fuente: INEI - 2017

Con respecto a la pobreza extrema, se observa en la figura 87 “Según área de residencia, la pobreza extrema afectó al 12,8% de la población del área rural, mientras que en el área urbana al 1,2%. Comparada con la registrada en el 2016, la pobreza extrema disminuyó en el área rural en 0,4 punto porcentual mientras que en el área urbana se incrementó en 0,3 punto porcentual.”

Figura 87. Evolución de la pobreza extrema, según área de residencia 2007-2017

Fuente: INEI - 2017

• Pobreza y Analfabetismo

En el año de análisis, el 12,9% de la población pobre de 15 y más años de edad no sabía leer ni escribir, es decir, era analfabeto. Este fenómeno afectó más a los pobres extremos ya que el 19,4% eran iletrados. Entre la población no pobre, la tasa de analfabetismo se ubicó en 4,3%.

A nivel de área de residencia, la tasa de analfabetismo de la población pobre del área urbana se ubicó en 8,5% y en el área rural en 18,4%. Entre la población no pobre la incidencia del analfabetismo en el área urbana fue de 2,8% y en el área rural de 12,5%.

Tabla 43

Evolución del analfabetismo según condición de pobreza, 2007-2017

PERÚ: EVOLUCIÓN DE LA TASA DE ANALFABETISMO, SEGÚN CONDICIÓN DE POBREZA Y ÁREA DE RESIDENCIA, 2007-2017 (Porcentaje)													
Área de residencia/ Condición de pobreza	2007	2008	2009	2010	2011	2012	2013	2014	2015	2016	2017	Diferencia (en puntos porcentuales)	
												2017-2016	2017-2007
Total													
Pobre	15,8	16,4	16,3	16,3	15,8	14,0	14,0	14,0	13,5	13,1	12,9	-0,2	-2,9
Pobre extremo	26,1	26,7	25,7	27,3	24,4	21,9	22,3	23,1	20,1	20,6	19,4	-1,2	-6,7
Pobre no extremo	12,6	12,7	13,0	13,1	13,6	11,9	12,1	12,1	12,1	11,6	11,7	0,1	-0,9
No pobre	4,1	4,3	4,1	4,3	4,4	4,0	4,2	4,5	4,3	4,4	4,3	-0,1	0,2
Urbana													
Pobre	9,1	9,3	9,6	9,2	8,9	7,4	8,1	8,8	8,6	8,1	8,5	0,4	-0,6
Pobre extremo	18,4	15,8	19,6	17,9	14,6	12,1	16,2	17,4	12,9	12,6	12,3	-0,3	-6,1
Pobre no extremo	8,3	8,7	8,7	8,4	8,5	7,0	7,5	8,2	8,3	7,7	8,2	0,5	-0,1
No pobre	3,0	3,1	2,9	3,0	3,1	2,6	2,7	2,9	2,8	3,1	2,8	-0,3	-0,2
Rural													
Pobre	23,1	23,6	22,4	23,3	22,7	20,8	20,5	19,8	18,8	18,8	18,4	-0,4	-4,7
Pobre extremo	27,7	29,0	26,8	29,4	26,4	24,2	23,6	24,3	21,8	22,6	21,6	-1,0	-6,1
Pobre no extremo	19,9	19,5	19,3	19,8	20,8	19,0	19,0	17,8	17,6	17,3	17,3	0,0	-2,6
No pobre	12,0	12,1	11,8	12,0	11,8	11,5	12,1	12,7	12,1	11,8	12,5	0,7	0,5

Fuente: Instituto Nacional de Estadística e Informática - Encuesta Nacional de Hogares, 2007-2017.

Fuente: INEI - 2017