



**Modelo de Gestión de información basado en Inteligencia de
Negocios para plantear recomendaciones que mejoren la frecuencia de recepción de las
transferencias monetarias del programa Mtukula Pakhomo en Malawi**

Suárez Peñafiel, Mónica Cristina

Vicerrectorado de Investigación, Innovación y Transferencia de Tecnología

Centro de Posgrados

Maestría en Gestión de Sistemas de Información e Inteligencia de Negocios

Trabajo de titulación, previo a la obtención del título de Maestría en Gestión de Sistemas de
Información e Inteligencia de Negocios

Ing. Vinuesa Trujillo, Jaime Augusto, MBA.

29 de noviembre de 2022



Tesis_Suarez Peñafiel Monica Cristina_desde resumen202...

Scanned on: 17:28 December 15, 2022 UTC



Overall Similarity Score



Results Found



Total Words in Text

Identical Words	389
Words with Minor Changes	149
Paraphrased Words	560
Omitted Words	641



Website | Education | Businesses

Firma:

.....

Ing. Vinueza Trujillo, Jaime Augusto, MBA.

Director



ESPE

UNIVERSIDAD DE LAS FUERZAS ARMADAS
INNOVACIÓN PARA LA EXCELENCIA

Vicerrectorado de Investigación, Innovación y Transferencia de Tecnología

Centro de Posgrados

Certificación

Certifico que el trabajo de titulación “**Modelo de Gestión de información basado en Inteligencia de Negocios para plantear recomendaciones que mejoren la frecuencia de recepción de las transferencias monetarias del programa Mtukula Pakhomo en Malawi**”, fue realizado por la Señorita **Suárez Peñafiel, Mónica Cristina**, el mismo que cumple con los requisitos legales, teóricos, científicos, técnicos y metodológicos establecidos por la Universidad de las Fuerzas Armadas ESPE, además fue revisado y analizado en su totalidad por la herramienta de prevención y/o verificación de similitud de contenidos; razón por la cual me permito acreditar y autorizar para que se lo sustente públicamente.

Sangolquí, 29 de noviembre de 2022

Firma:

.....
Ing. Vinuesa Trujillo, Jaime Augusto, MBA.

Director

C.C.: 1716028509



ESPE
UNIVERSIDAD DE LAS FUERZAS ARMADAS
INNOVACIÓN PARA LA EXCELENCIA

Vicerrectorado de Investigación, Innovación y Transferencia de Tecnología

Centro de Posgrados

Responsabilidad de Autoría

Yo, **Suárez Peñafiel, Mónica Cristina**, con cédula de identidad No. 1718008103, declaro que el contenido, ideas y criterios del trabajo de titulación: “**Modelo de Gestión de información basado en Inteligencia de Negocios para plantear recomendaciones que mejoren la frecuencia de recepción de las transferencias monetarias del programa Mtukula Pakhomo en Malawi**” es de mi autoría y responsabilidad, cumpliendo con los requisitos legales, teóricos, científicos, técnicos y metodológicos establecidos por la Universidad de Fuerzas Armadas ESPE, respetando los derechos intelectuales de terceros y referenciando las citas bibliográficas.

Sangolquí, 29 de noviembre de 2022

Firma (s)

Una firma manuscrita en tinta azul que parece leer 'Mónica Suárez Peñafiel'. Debajo de la firma hay una línea de puntos.

Suárez Peñafiel, Mónica Cristina

C.C.: 1718008103



ESPE
UNIVERSIDAD DE LAS FUERZAS ARMADAS
INNOVACIÓN PARA LA EXCELENCIA

Vicerrectorado de Investigación, Innovación y Transferencia de Tecnología

Centro de Posgrados

Autorización de Publicación

Yo, **Suárez Peñafiel, Mónica Cristina**, autorizo a la Universidad de las Fuerzas Armadas ESPE publicar el trabajo de titulación: **“Modelo de Gestión de información basado en Inteligencia de Negocios para plantear recomendaciones que mejoren la frecuencia de recepción de las transferencias monetarias del programa Mtukula Pakhomo en Malawi”** en el Repositorio Institucional, cuyo contenido, ideas y criterios son de mi responsabilidad.

Sangolquí, 29 de noviembre de 2022

Firma (s)

Suárez Peñafiel, Mónica Cristina

C.C.: 1718008103

Dedicatoria

A mis hijos Alaia Kristen y Erick David, mis mellizos, y a mi esposo Fredy David porque son el motor que me impulsa e inspira a ser mejor persona y profesional día tras día.

A mis padres Mónica Elizabeth y Carlos Alberto por haberme dado la vida y un hogar en el que prevalece el amor, el respeto y la honestidad. Por su apoyo constante en todas las etapas transcurridas de mi vida, por sus consejos, enseñanzas, regaños y ejemplos dignos de superación y entrega.

A mi hermano Carlos David quien me ha demostrado que con esfuerzo y dedicación se puede alcanzar todo lo que se proponga.

Agradecimiento

A Dios por guiar mi camino, y darme la perseverancia para lograr cada uno de mis objetivos y metas planteadas. A mis hijos y esposo por su amor y sobre todo por la comprensión y paciencia durante el tiempo que emplee para el desarrollo del presente trabajo de titulación. A mi mamá, papá y hermano por su amor, paciencia y por la confianza depositada en mí, porque sé que siempre van a brindarme su apoyo incondicional.

A la Universidad de las Fuerzas Armadas ESPE por darme la oportunidad de continuar preparándome académicamente y lograr así una meta más en mi vida. Al cuerpo docente de la Unidad de Postgrados – Maestría en Gestión de Sistemas de Información e Inteligencia de Negocios por los conocimientos compartidos. A mi director y a mi oponente de tesis por ser unos excelentes profesionales, por el apoyo, la dedicación y sobre todo por los conocimientos impartidos para el desarrollo del presente trabajo de titulación.

A todo el personal de la compañía Ayala Consulting Cia. Ltda., Especialmente a su presidente y gerente general, por brindarme el acceso necesario a la información requerida para la ejecución de esta tesis y por el apoyo y comprensión durante la culminación de mi posgrado.

Y finalmente a todas las personas que de alguna u otro manera estuvieron presentes brindándome su apoyo y confiando en mí.

Índice de Contenido

Dedicatoria.....	6
Agradecimiento.....	7
Resumen	14
Abstract.....	15
Capítulo I: Análisis Del Estado Actual	16
Antecedentes	16
Planteamiento del problema.....	18
Objetivos de la Investigación.....	22
<i>Objetivos Específicos</i>	23
Justificación y Alcance del proyecto	23
<i>Preguntas de Investigación</i>	25
Hipótesis	28
Capítulo II: Marco Referencial.....	29
Categorización de variables	29
<i>Variable Dependiente</i>	29
<i>Variable Independiente</i>	29
Marco Teórico	30
<i>Fundamentación de la Variable Independiente</i>	30
<i>Fundamentación de la Variable Dependiente</i>	35
Estado del Arte.....	38
<i>Criterios de Inclusión</i>	38
<i>Criterios de Exclusión</i>	39
<i>Grupo de Control</i>	39
<i>Palabras clave y coincidencias</i>	41

<i>Cadena de búsqueda y resultados obtenidos</i>	42
<i>Selección de Estudios</i>	43
<i>Resultados de la Revisión</i>	44
Capítulo III: Marco Metodológico.....	50
Introducción	50
Metodología de Investigación.....	50
<i>Fases de la Metodología</i>	53
Capítulo IV: Construcción de la Solución	56
Transaccional Data Base (Base de datos transaccional)	57
Proceso de Extracción Transformación y Carga de datos (ETL) y Construcción de la Máster Data Table (MDT)	59
Proceso de descubrimiento de datos (Data Discovery)	60
<i>Member collector age</i>	61
<i>Member collector gender</i>	62
<i>Member collector relationship</i>	63
<i>Member collector marital status</i>	64
<i>Member collector disability</i>	65
<i>Member collector chronic illness</i>	66
<i>Member collector attended education level</i>	67
<i>Transfer agency type</i>	69
<i>Member who collects the payments</i>	70
<i>Was transfer collected</i>	71
<i>To be delivered</i>	72
Proceso de limpieza depuración y asignación de valores numéricos	73
Proceso de agrupamiento de datos (Data clustering)	77
<i>Variables aceptadas por Rapid Minner para iniciar la clasificación</i>	77

<i>Método del codo para determinar el número de clusters idóneo</i>	78
<i>Clústeres para familias que cobran el beneficio</i>	79
<i>Clústeres para familias que no cobran el beneficio</i>	81
Capítulo V: Análisis y Presentación de Resultados	83
Selección de Clusters más representativos para los grupos (cobra o no cobra el beneficio)	83
<i>No cobra el beneficio</i>	83
<i>Cobra el beneficio</i>	84
Comparación de características no discriminatorias	85
Perfilamiento Cluster NO cobra el beneficio	86
Perfilamiento Cluster SI cobra el beneficio	87
Presentación de resultados a los responsables de la implementación del programa Mtukula Pakhomo	88
<i>Análisis concluyente de los resultados obtenidos:</i>	91
<i>Recomendación que permitirá mejorar la frecuencia de recepción por parte de los beneficiarios que actualmente lo cobran de manera irregular o no lo cobran.</i>	92
Capítulo VI: Conclusiones y Recomendaciones	95
Conclusiones	95
Recomendaciones	97
Bibliografía	99

Índice de Tablas

Tabla 1 <i>Estudios del grupo de control y términos clave</i>	40
Tabla 2 <i>Palabras clave - Frecuencia</i>	41
Tabla 3 <i>Construcción de la Cadena de búsqueda</i>	42
Tabla 4 <i>Selección de estudios</i>	43
Tabla 5 <i>Selección de Estudios Primarios</i>	44
Tabla 6 <i>Estudios Primarios</i>	45
Tabla 7 <i>Variables demográficas representativas</i>	61
Tabla 8 <i>Asignación de valores a variables categóricas</i>	74
Tabla 9 <i>Concentración de beneficiarios por clusters (no cobra el beneficio)</i>	84
Tabla 10 <i>Concentración de beneficiarios por clusters</i>	85
Tabla 11 <i>Variables no discriminantes</i>	86
Tabla 12 <i>Resumen de Variables utilizadas durante el proceso ejecutado</i>	89

Índice de Figuras

Figura 1 <i>Árbol de problemas</i>	20
Figura 2 <i>Categoría de variables</i>	30
Figura 3 <i>Cuadrante de Gartner para Plataformas de BI y Analíticas</i>	33
Figura 4 <i>Metodología de la Investigación</i>	52
Figura 5 <i>Esquema de la solución propuesta</i>	57
Figura 6 <i>Esquema transaccional de base de datos</i>	58
Figura 7 <i>Creación MDT</i>	60
Figura 8 <i>Member collector age</i>	62
Figura 9 <i>Pareto Member collector gender</i>	63
Figura 10 <i>Pareto Member collector relationship</i>	64
Figura 11 <i>Pareto Member collector marital status</i>	65
Figura 12 <i>Pareto Member collector disability</i>	66
Figura 13 <i>Pareto Member collector chronic illness</i>	67
Figura 14 <i>Pareto Member collector attended education level</i>	68
Figura 15 <i>Pareto District</i>	69
Figura 16 <i>Pareto Transfer agency type</i>	70
Figura 17 <i>Pareto Member who collects the payments</i>	71
Figura 18 <i>Pareto Was transfer collected</i>	72
Figura 19 <i>Pareto To be delivered</i>	73
Figura 20 <i>Estandarización y calidad de datos</i>	74
Figura 21 <i>Variables descartadas para la clasificación</i>	78
Figura 22 <i>Diagrama método del codo</i>	79
Figura 23 <i>Clusters para familias que cobran el beneficio</i>	80
Figura 24 <i>Centroides para clusters de familias que cobran el beneficio</i>	80

Figura 25 <i>Clusters para familias que no cobran el beneficio.....</i>	81
Figura 26 <i>Centroides para clusters de familias que no cobran el beneficio</i>	82
Figura 27 <i>Perfil de los beneficiarios que no cobran</i>	90
Figura 28 <i>Perfil de los beneficiarios que si cobran</i>	90

Resumen

Los proyectos de beneficio social, en este caso las transferencias de dinero en efectivo, están direccionadas a las personas más pobres o aquellas que se encuentran por debajo del umbral de pobreza definido, con el objetivo de cubrir necesidades básicas como la alimentación, la salud y la educación. Por este motivo es fundamental que los beneficiarios de este tipo de programas reciban el beneficio regularmente según lo estipulan las normas y reglas de cada proyecto (cada mes, cada dos meses, cada trimestre, etc). El programa Mtukula Pakhomo es un programa social de transferencias monetarias direccionado al 10% de los hogares ultra pobres de Malawi (África). Se ha analizado la información del proceso de pagos desde el año 2016 y se ha determinado que alrededor de un 3 al 5% de los beneficiarios están cobrando el beneficio de manera irregular; para que estos beneficiarios no afecten su bienestar o vuelvan a sus condiciones de vida previas, con el presente trabajo de titulación se plantea implementar un modelo de gestión de información utilizando técnicas y soluciones de inteligencia de negocios para emitir recomendaciones que mejoren la frecuencia de recepción del beneficio por parte de este tipo de beneficiarios. Se aplicó una metodología de investigación propia con el fin de abarcar cada uno de los objetivos y preguntas de investigación propuestas; esta metodología comprende las fases: i) definición de la situación actual, ii) estudio de viabilidad, iii) construcción de la solución y iv) validación. Con la implementación de las fases de la metodología mencionada se ha dado respuesta a las preguntas de investigación planteadas, y se han obtenido los patrones y/o perfiles que rigen el comportamiento de los beneficiarios que cobran de manera irregular el beneficio otorgado por el gobierno de Malawi (África) y otros donantes, de este modo se han planteado también las recomendaciones que permitirán mejorar la frecuencia de recepción del beneficio monetario por parte de este tipo de beneficiarios.

Palabras clave: inteligencia de negocios, patrones de comportamiento, agrupamiento de datos, minería de datos, protección social.

Abstract

Social protection projects, in this case cash transfers, are aimed at the poorest people or those who are below the defined poverty line, with the objective of covering basic needs such as food, health, and education. For this reason, it is essential that the beneficiaries of this type of program regularly receive the benefit as stipulated in the norms and rules of each project (every month, every two months, every quarter, etc.). The Mtukula Pakhomo program is a social cash transfer program that targets 10% of the ultra-poor households in Malawi (Africa). The program's payment process information since 2016 has been analyzed and it has been determined that about 3 to 5% of the beneficiaries are collecting the benefit irregularly. It is worrisome as these beneficiaries' welfare may be affected and may return to their previous living conditions. It is proposed to implement an information management model using business intelligence techniques and solutions to issue recommendations to increase the frequency of the benefit being collected by this type of beneficiaries. The author carried out a detailed research methodology in order to cover each of the proposed research objectives and questions; this methodology consists of the following phases: i) definition of the current situation, ii) feasibility study, iii) construction of the solution and iv) validation. After implementing the phases of the aforementioned methodology, the posed research questions have been answered, and the patterns and/or profiles that govern the behavior of beneficiaries who irregularly receive the benefit granted by the government of Malawi (Africa) and other donors have been obtained. Recommendations have also been made that would improve the frequency of receipt of the cash benefit for this type of beneficiaries.

Keywords: business intelligence, behavior patterns, data clustering, data mining, social protection.

Capítulo I: Análisis Del Estado Actual

Antecedentes

El trabajo de los organismos internacionales de ayuda social como el Banco Mundial, Unicef, Banco Interamericano de Desarrollo (BID), entre otros; está direccionado a brindar apoyo a leyes y medidas que beneficien a las personas (mujeres y hombres) pobres, vulnerables y relegados para que tengan igualdad de condiciones y oportunidades, así como también para que contribuyan al progreso económico y social de su nación y se puedan beneficiar de dicho progreso. (Kwesell, 2014)

El proceso de crecimiento y desarrollo económico impide invariablemente a un cierto sector de la población aprovechar al máximo los beneficios de este. Los sectores marginados incluyen a ancianos, enfermos crónicos, huérfanos y otros niños vulnerables, personas con discapacidades y familias indigentes que se vuelven vulnerables a los instrumentos de gestión de riesgos. En este sentido, el Gobierno de Malawi (África), con el apoyo de socios para el desarrollo, ha implementado el Programa Social de Transferencias en Efectivo “Mtukula Pakhomo”. El Programa se dirige al 10% de hogares ultra pobres de Malawi (África), con el objetivo de reducir la pobreza, el hambre y la inanición entre los hogares extremadamente pobres y con restricciones laborales. Además, el Programa tiene como objetivo aumentar la inscripción, retención y finalización de la escuela y mejorar el estado de salud y nutrición de los niños que viven en los hogares seleccionados. (Shawa, 2016)

Un estudio que analizó los efectos de los programas sociales de transferencia monetaria en África subsahariana, tomando como referencia 6 países, entre ellos Malawi

(África) con el SCTP “Mtukula Pakhomo” por sus siglas en inglés Social Cash Transfer Program, indica que las transferencias monetarias permitieron que los beneficiarios mejoren sus medios de subsistencia; ayudaron a satisfacer las necesidades inmediatas de los beneficiarios más pobres, lo cual les devolvió la esperanza y la confianza en el futuro y les dio un descanso de sus preocupaciones y angustias; permitieron a los beneficiarios poner fin o reducir su dependencia a estrategias de adaptación negativas como la mendicidad, el trabajo sexual, la venta de bienes en tiempos de dificultad o la disminución de la frecuencia de las comidas, así como el trabajo rural ocasional; también fomentaron la matriculación y asistencia escolar. El aumento de la asistencia escolar se tradujo en una disminución del trabajo infantil. (Pozarny & Davis, 2015)

Dicho esto, los retrasos en los cobros del beneficio por parte de los beneficiarios o los retrasos en los pagos por parte de los programas han llevado a algunos beneficiarios a reanudar su conducta anterior reduciendo el efecto positivo y poniendo en riesgo los logros alcanzados. Por otro lado se puede decir que los efectos positivos serían mayores si los pagos y cobros fueran regulares. (Pozarny & Davis, 2015).

Mtukula Pakhomo comenzó en el año 2006 como un plan piloto en un solo distrito, debido a su éxito se ha expandido a 9 distritos entre el año 2006 y 2012; a 9 distritos adicionales en 2014; y actualmente el programa se está implementando a nivel nacional, es decir en 28 distritos con una totalidad de 279.671 beneficiarios a marzo del 2019. Los beneficiarios del proyecto reciben una transferencia monetaria cada 2 meses, sin embargo, se ha comprobado a través de la revisión de reportes e indicadores del procesos de pagos, que el Porcentaje de transacciones no pagadas por ciclo cerrado de pago desde el año 2016 hasta marzo del 2019, oscila entre un 3% y 5%; esto quiere decir que aproximadamente un 5% de los

beneficiarios por ciclo de pago no recibe su beneficio y lo acumula para el siguiente periodo, en el que puede decidir cobrar o acumularlo una vez más.

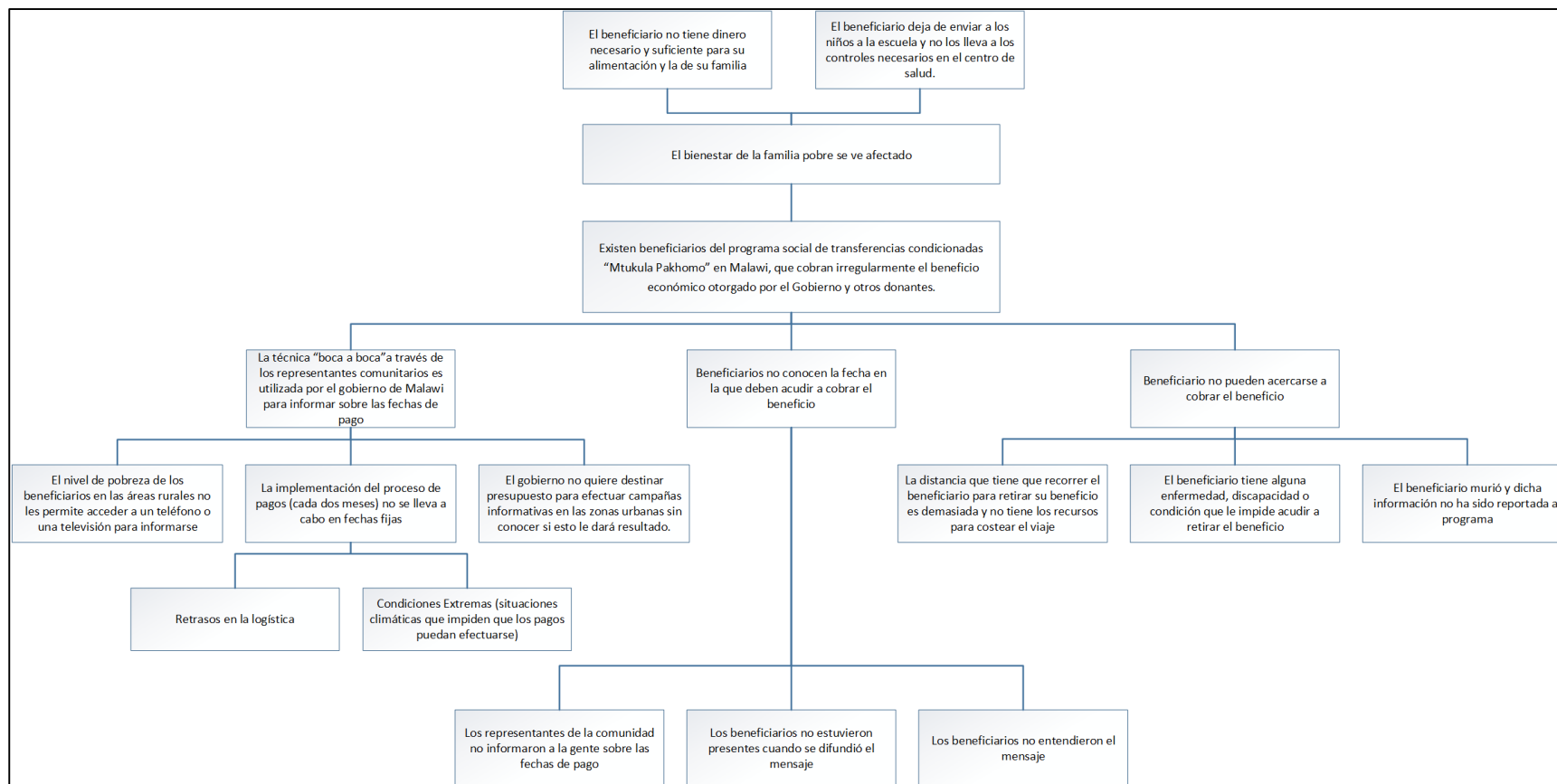
Planteamiento del problema

El problema que está experimentando el programa “Mtukula Pakhomo” es que existe un porcentaje de beneficiarios que cobran irregularmente el beneficio económico otorgado por el Gobierno y otros donantes (Banco Mundial, el Banco de Desarrollo Alemán, la Unión Europea, la Ayuda Irlandesa, Unicef).

A continuación, se presenta el árbol de problemas (Figura 1), el mismo que detalla las causas que ocasionan la problemática del programa, así como los efectos que este acarrea. Con el objetivo de identificar las causas y efectos del problema evidenciado, se efectuó una entrevista al director general del programa en la consultora encargada del diseño e implementación del mismo, debido a que su experiencia y su conocimiento sobre el tema, le permite exponer claramente la información del proceso de pago del beneficio a los beneficiarios del programa.

Figura 1

Árbol de problemas



Nota. Esta figura detalla el problema identificado, así como también las causas y efectos del mismo.

Que los beneficiarios de un programa social de transferencias monetarias no cobren el beneficio de manera regular en los periodos correspondientes, ocasiona un efecto directo en su bienestar y el de sus hogares, ya que en dicho periodo los hogares no contarán con el dinero necesario y suficiente para su alimentación y para solventar las actividades escolares y de salud de los niños existentes en el hogar.

El cobro irregular del beneficio por parte de un grupo de beneficiarios puede estar ocasionado por las siguientes causas: i) por el lado del programa, la falta de comunicación o una comunicación inadecuada para dar a conocer las fechas de pago y ii) por el lado de los beneficiarios, el desconocimiento de las fechas de pago y/o el impedimento para acercarse a cobrar el beneficio.

Falta de comunicación o comunicación inadecuada para dar a conocer las fechas de pago: La técnica que el gobierno de Malawi utiliza para dar a conocer las fechas de pago a los beneficiarios es la de “boca a boca” a través de los representantes comunitarios, tal vez esta técnica no es la mejor porque puede ocasionar que el mensaje no llegue a todos los beneficiarios, que el mensaje se pierda o no sea entendido; pero es aplicada debido a que el nivel de pobreza que existe en las áreas rurales de Malawi no les permite a los beneficiarios acceder a tecnología como celulares o televisiones para informarse sobre el programa, y en el caso de las áreas urbanas el gobierno no quiere destinar presupuesto para campañas informativas sin saber si esto le dará resultado o no. Adicionalmente, los pagos del beneficio no se lleva a cabo en fechas fijas ya sea por retrasos en la logística o por condiciones extremas; es por esto que las autoridades del programa han optado por esta técnica para informar sobre las fechas de pago del beneficio.

Desconocimiento de las fechas de pago: Los motivos por los que los beneficiarios no conocen la fecha en la que deben acudir a cobrar el beneficio pueden deberse a que i) los representantes de su comunidad no les informaron, ii) al momento que los representantes comunitarios difundieron el mensaje los beneficiarios no estuvieron presentes y/o iii) los beneficiarios no entendieron el mensaje entregado por parte de los representantes comunitarios.

Impedimento para acercarse a cobrar el beneficio: los impedimentos que pueden evidenciarse son i) distancias largas y falta de recursos económicos para trasladarse a cobrar su beneficio y ii) condiciones físicas que le impiden movilizarse como discapacidades, enfermedades crónicas, embarazo, etc. En el caso de que el beneficiario no acude a cobrar consecutivamente por varios periodos puede deberse a que ha fallecido ya la información no ha sido reportada al programa.

Objetivos de la Investigación

Esta sección presenta el objetivo general y los objetivos específicos que rigen el presente trabajo de investigación.

Objetivo General

Implementar un modelo de gestión de información utilizando técnicas y soluciones de inteligencia de negocios para plantear recomendaciones que mejoren la frecuencia de recepción del beneficio de los beneficiarios del SCTP “Mtukula Pakhomo” que actualmente lo cobran de manera irregular.

Objetivos Específicos

OE1: Identificar las causas y efectos que surgen de la problemática del cobro irregular del beneficio otorgado por el programa Mtukula Pakhomo mediante técnicas de recolección o adquisición de información.

OE2: Identificar herramientas tecnológicas orientadas a la selección de patrones de comportamiento que rigen un segmento de datos mediante una revisión preliminar de literatura (PRL), para determinar la propuesta que mejor se ajuste a las necesidades del proyecto.

OE3: Recolectar, organizar, limpiar, adecuar y homologar los datos disponibles mediante un proceso de extracción, transformación y carga (ETL) para que sean explotados a través de la solución BI seleccionada.

OE4: Analizar los datos a través de la solución BI para identificar los patrones y/o perfiles que rigen el comportamiento de los beneficiarios que cobran de manera irregular el beneficio y emitir recomendaciones operativas que permitan el cobro regular del mismo.

OE5: Validar los resultados y las recomendaciones emitidas del análisis de los datos a través de una entrevista al director del proyecto y stakeholders involucrados.

Justificación y Alcance del proyecto

El programa social de transferencias monetarias “Mtukula Pakhomo” fue implementado para reducir en el corto plazo la pobreza, el hambre de los hogares más pobres de Malawi (África) y también para promover la educación y el cuidado de la salud de los niños; por el

efecto positivo que ha tenido, actualmente el programa está siendo implementado a nivel nacional; por lo dicho, si los hogares beneficiarios no reciben su beneficio de manera regular volverán a su condición previa poniendo en peligro su bienestar y opacando el objetivo primordial del programa. Con el modelo de gestión de información basado en inteligencia de negocios propuesto en el presente estudio, se busca identificar el perfil o los patrones de comportamiento de los beneficiarios que no cobran de manera regular el beneficio otorgado por el programa para plantear recomendaciones que permitan mejorar la frecuencia de recepción de las transferencias monetarias en este tipo de beneficiarios, la aplicación de estas recomendaciones dependerá directamente del Gobierno de Malawi (África) y los donantes. Con una revisión preliminar de literatura se identificarán las mejores propuestas tecnológicas que permitan detectar los patrones de comportamiento que rigen a los segmentos de datos.

La información de los beneficiarios, misma que se almacena en la base de datos del sistema SCTP “Mtukula Pakhomo” será analizada y depurada, únicamente se tomarán los datos relevantes para el caso de estudio y se excluirán los valores atípicos que pueden afectar la identificación de los patrones de comportamiento; una vez que se haya seleccionado la información esta deberá pasar por un proceso de ETL respectivo. La información será analizada a través de las soluciones de BI identificadas con el propósito de clasificar, segmentar e identificar el perfil de los beneficiarios que no reciben regularmente el beneficio otorgado y de esta manera proponer recomendaciones operativas que mejoren la frecuencia de la recepción de los beneficios en estos hogares. Ej. Si el patrón o el perfil de estos beneficiarios indica que las personas que no cobran regularmente el beneficio son ancianos que viven solos o personas que tienen alguna discapacidad, se establecerá recomendaciones como el pago directamente en el hogar del beneficiario, si se determina que la gente que no cobra corresponde a una misma ubicación geográfica, recomendación estará enfocada a la técnica utilizada para difundir las fechas de pago.

Preguntas de Investigación

Para alcanzar el objetivo del proyecto, es necesario que se respondan las siguientes preguntas de investigación para cada objetivo específico:

OE1: Identificar las causas y efectos que surgen de la problemática del cobro irregular del beneficio otorgado por el programa Mtukula Pakhomo mediante técnicas de recolección o adquisición de información.

- **OE1-RQ1.1:** ¿Qué técnica de recolección o adquisición de información es la adecuada para identificar las causas y efectos de la problemática que presenta el programa “Mtukula Pakhomo”?
- **OE1-RQ1.2:** ¿Qué causas y efectos se pueden detectar de la problemática actual del programa “Mtukula Pakhomo” en Malawi (África) con respecto al cobro irregular del beneficio otorgado por parte de los beneficiarios?

OE2: Identificar herramientas tecnológicas orientadas a la selección de patrones de comportamiento que rigen un segmento de datos mediante una revisión preliminar de literatura (PRL), para determinar la propuesta que mejor se ajuste a las necesidades del proyecto.

- **OE2 –RQ2.1:** ¿Qué artículos hacen referencia al análisis y gestión adecuada de información en temas de identificación de perfiles o patrones de comportamiento que rigen a un segmento de datos?

- **OE2 – RQ2.2:** ¿Cuáles son las herramientas tecnológicas que permiten identificar perfiles o patrones de comportamiento que rigen a un segmento de datos y que se acoplan mejor a las necesidades del SCTP Mtukula Pakhomo?

OE3: Recolectar, organizar, limpiar, adecuar y homologar los datos disponibles mediante un proceso de extracción, transformación y carga (ETL) para que sean explotados a través de la solución BI seleccionada.

- **OE3 – RQ3.1:** ¿Cuáles son las fuentes y los tipos de datos con los que cuenta el SCTP Mtukula Pakhomo?
- **OE3 – RQ3.2:** ¿Qué tipo de datos se necesita para llevar a cabo la identificación de perfiles o patrones de comportamiento de los beneficiarios que no cobran regularmente el beneficio otorgado por el gobierno en el SCTP Mtukula Pakhomo?
- **OE3 – RQ3.3:** ¿Existen valores atípicos que puedan afectar la identificación de perfiles o patrones de comportamiento de este tipo de beneficiarios?
- **OE3 – RQ3.4:** ¿Cuál es la mejor herramienta para realizar un ETL de la información recopilada?

OE4: Analizar los datos a través de la solución BI para identificar los patrones y/o perfiles que rigen el comportamiento de los beneficiarios que cobran de manera irregular el beneficio y emitir recomendaciones operativas que permitan el cobro regular del mismo.

- **OE4 – RQ4.1:** ¿Cuáles son los perfiles o patrones de comportamiento obtenidos de los beneficiarios que no cobran regularmente el beneficio otorgado por el gobierno en el SCTP Mtukula Pakhomo?
- **OE4 – RQ4.2:** ¿Cuáles son las recomendaciones obtenidas del estudio, que deberían aplicarse al programa, para que los beneficiarios que cobran irregularmente el beneficio lo reciban de forma regular?

OE5: Validar los resultados y las recomendaciones emitidas del análisis de los datos a través de una entrevista al director del proyecto y stakeholders involucrados.

- **OE5 - RQ5.1:** ¿Las recomendaciones resultantes de la implementación del modelo de gestión de información basado en BI, mejorarán la frecuencia de recepción del beneficio otorgado por el programa en los beneficiarios que actualmente lo cobran de manera irregular?
- **OE5 – RQ5.2:** ¿Es factible la aplicación de las recomendaciones en el programa Mtukula Pakhomo?

Nota: Debido al marco legal gubernamental, el presente estudio no cubrirá la aplicación de las recomendaciones en campo para verificar si efectivamente la frecuencia de recepción del beneficio mejora, ya que esto depende de las decisiones político-administrativas del gobierno de Malawi (África), mismas que están fuera del alcance del estudio.

Hipótesis

La implementación de un modelo de gestión de información utilizando técnicas y soluciones de BI para el programa Mtukula Pakhomo permitirá establecer recomendaciones para mejorar la frecuencia de recepción del beneficio por parte de los beneficiarios que actualmente lo cobran de manera irregular.

Capítulo II: Marco Referencial

Categorización de variables

Esta sección se presenta la variable dependiente e independiente identificadas durante el análisis de la hipótesis de la presente investigación.

Variable Dependiente

Existen beneficiarios que no cobran regularmente la transferencia monetaria no condicionada (beneficio) otorgada por el gobierno de Malawi (África) y otros donantes.

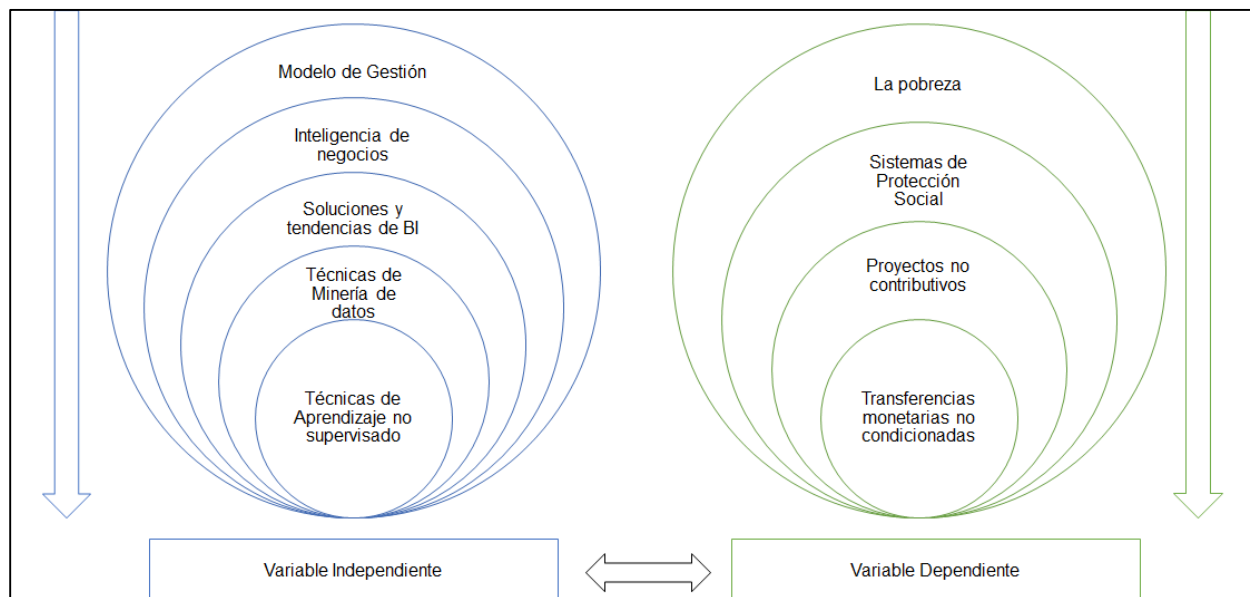
Variable Independiente

Implementar un modelo de gestión de información utilizando soluciones de BI que permita establecer recomendaciones.

Partiendo de la variable independiente y la variable dependiente identificadas en el problema de investigación, se propone la siguiente jerarquía de estudio (ver figura 2):

Figura 2

Categoría de variables



Nota. Esta figura muestra la categorización de la variable independiente y de la variable dependiente.

Marco Teórico

Fundamentación de la Variable Independiente

Modelo de Gestión. Un Modelo de Gestión es definido por sus dos palabras, modelo que se considera como un marco teórico que integra diferentes actividades y gestión que se refiere al sistema utilizado para organizar una actividad. Una vez dicho lo anterior, Navarro conceptualiza al Modelo de Gestión como una representación teórica mediante la cual se lleva a término un proceso (Navarro, 2017).

Pérez por su parte define al modelo de gestión como un marco de referencia para la administración de una entidad, e indica que estos modelos pueden ser aplicados en negocios privados y en la administración pública; el modelo de gestión de las empresas privadas se basa en la obtención de ganancias económicas mientras que el público se basa en el bienestar de la población (Pérez, 2008).

Inteligencia de Negocios. Según Peña, el fin o propósito de la Inteligencia de Negocios es mejorar el desempeño de las instituciones, maximizando sus ventajas competitivas en su área de inmersión a través de la toma de decisiones eficientes y eficaces (Peña, 2006). El uso de métodos, procesos y herramientas tecnológicas permiten lograr lo anteriormente dicho, ya que con ellas se puede condensar, limpiar y transformar la data objeto de análisis para finalmente obtener resultados que aporten conocimiento (Parr, 2000), los resultados pueden ser datos estructurados mismos que identifican y detallan las características de un área de interés (Stackowiak, Rayman, & Greenwald, 2007), y generan conocimiento sobre las oportunidades y amenazas de la institución, esto con el fin de que sean aprovechadas y abordadas respectivamente (Ballard, y otros, 2006).

En toda organización existen tres tipos de necesidades de información: Operacional, Táctica y Estratégica (Colmenares, 2011)

- A nivel operativo, la aplicación de la inteligencia de negocios, permite que los trabajadores recibirán información oportuna, exacta y adecuada. La información operativa por lo general se presenta en un único formato a través de herramientas de reporte u hojas de Excel que se actualizan constantemente.

- A nivel táctico, es decir a nivel de analistas de datos y gerencias medias, la Inteligencia de Negocios tiene el fin de que estos trabajadores tengan acceso directo (sin la involucración de terceras personas) a la información a través de herramientas analíticas y de consulta.
- A nivel estratégico (dirección de la empresa), la Inteligencia de Negocios ayuda a los directores en el análisis y monitoreo de las tendencias, patrones, metas y objetivos estratégicos de la institución.

Soluciones y Tendencias de BI. La Inteligencia de Negocios actualmente es el as bajo la manga de las organizaciones con empoderamiento y éxito, pues estas herramientas tecnológicas les permiten obtener ventajas para obtener los resultados deseados. Para las pequeñas o medianas empresas (PYME) con un presupuesto ajustado, el software de inteligencia empresarial de código abierto y gratuito es su manera de obtener los beneficios de los datos y análisis de forma gratuita. A continuación, se mencionan soluciones de BI de código abierto o gratuita según (Hope, 2018).

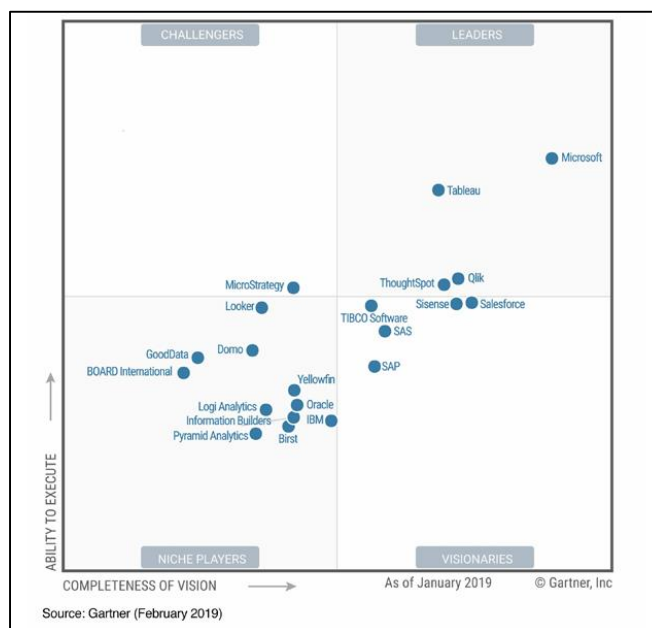
- Birt
- Jaspersoft Community
- Knime
- Metabase
- Pentaho Business Analytics
- Report Server Community Edition
- Knowage
- Dataiku DSS Free Edition

- Microsoft Power BI Desktop
- Qlik Sense Cloud Basic
- Qlik View Personal Edition
- RapidMiner
- Style Scope AE
- Tableau Public
- Zoho Reports

Muchas de las soluciones mencionadas se encuentran contempladas en el cuadrante de Gartner, como se observa en la figura 3.

Figura 3

Cuadrante de Gartner para Plataformas de BI y Analíticas



Nota. La figura muestra las compañías que forman parte del cuadrante mágico de Gartner para plataformas de BI y Analíticas. Fuente: <https://www.qlik.com/es-es/gartner-magic-quadrant-business-intelligence>.

La compañía IDC en su reporte: Worldwide Business Analytics Software 2012–2016 Forecast and 2011 Vendor Shares, manifiesta un crecimiento anual del 9,8% de la Inteligencia de Negocios, que en su mayoría va estar marcado por las nuevas tendencias que están entrando con fuerza en este ámbito (Ej. Real Time BI, Big Data, Mobile BI e integración Social Media).

Técnicas de Minería de Datos. La identificación de información procesable dentro de grandes o extensos conjuntos de datos se define como minería de datos (Microsoft, 2019). La minería de datos, a través de un análisis matemático identifica patrones y tendencias resaltantes en los datos. Este tipo de patrones y tendencias no pueden ser detectados mediante una exploración o un análisis tradicional de datos debido a la complejidad o a la gran cantidad de datos existentes. Los modelos de minería de datos pueden aplicarse en los siguientes escenarios:

- **Previsión:** Ej. Cuando se quiere calcular el nivel de ventas que tendrá la empresa, Para predecir el número de cargas reportadas al servidores o a su vez el tiempo de inactividad del mismo.
- **Riesgo y probabilidad:** Ej. Cuando se requiere hacer una elección de los mejores clientes que tiene la compañía para el envío de correos directos promocionales. Cuando se desea determinar el punto de equilibrio para escenarios de riesgo, y las probabilidades de otros resultados.
- **Recomendaciones:** Ej. Al momento de identificar productos que pueden ser vendidos en conjunto y establecer las recomendaciones necesarias.
- **Buscar secuencias:** Ej. Para identificar los productos que los clientes han enviado al carrito de compras y así predecir los futuros posibles eventos.

- **Agrupación:** Ej. Clusterizar los clientes, eventos, de elementos relacionados, analizar y predecir semejanzas e intereses.

Se han identificado técnicas de minería de datos supervisadas y no supervisadas.

Técnicas de aprendizaje no supervisado. Según (Recuero, 2017) las técnicas de minería de datos con aprendizaje no supervisado deben aplicarse cuando no existen disponibles datos etiquetados para el entrenamiento; es decir únicamente se conocen los datos de entrada. Tiene un carácter exploratorio y suele usarse en problemas de clustering, agrupamientos de co-ocurrencia y perfilamiento de datos.

Los tipos de algoritmo más habituales en aprendizaje no supervisado son i) Algoritmos de clustering, ii) Análisis de componentes principales, iii) Descomposición en valores singulares (singular value decomposition) y iv) Análisis de componentes independientes (Independent Component Analysis).

Fundamentación de la Variable Dependiente

La Pobreza. Según la **Organización de las Naciones Unidas (ONU)**, la pobreza no solo puede definirse tomando en cuenta los ingresos y recursos para garantizar unos medios de vida sostenibles, la pobreza va más allá de eso; se debe incluir un análisis del hambre, la malnutrición, el acceso limitado a la educación y a otros servicios básicos, la discriminación, la exclusión social y la falta de participación en la toma de decisiones.

Los avances que se han logrado en cuanto a la reducción de la pobreza han sido notables en las últimas décadas. El mundo cumplió la meta del primer objetivo de desarrollo del

milenio, el mismo que buscaba para el año 2015 reducir al 50% la tasa de pobreza registrada en 1990, esto se logró en el año 2010, cinco años antes de lo previsto. Pero pese a estos avances, la cantidad de personas que viven en pobreza extrema en el mundo sigue siendo inaceptablemente alta. En la actualidad, comparando los valores con el año 1990, alrededor de 1100 millones de personas menos viven en la pobreza. Aunque las tasas de pobreza se han reducido en todas las regiones, los avances no han sido equitativos, más del cincuenta por ciento de la población muy pobre es de África al sur del Sahara. De hecho, la cantidad de pobres de esa región aumentó en 9 millones, por lo que en 2015 había 413 millones de personas que vivían con menos de USD 1,90 al día, cifra superior a la de todas las demás regiones en conjunto. Si esta tendencia se mantiene, para 2030 prácticamente 9 de 10 personas extremadamente pobres vivirán en África al sur del Sahara (Howton, 2018).

Sistemas de Protección Social. El **Banco Mundial** manifiesta que los sistemas de protección social ayudan a las personas y las familias en condiciones de pobreza y vulnerabilidad a enfrentarse a las crisis, encontrar un empleo digno, mejorar la productividad, destinar recursos en la salud y educación de sus hijos y brindar protección a la población de la tercera edad. Para lograr los objetivos de poner fin a la pobreza e impulsar la prosperidad compartida, el Banco Mundial considera un aspecto importante el apoyo que da para el acceso a la protección social universal.

La protección social universal incluye el apoyo social a través de transferencias de dinero en efectivo a las personas más necesitadas (niños; embarazadas, discapacitados, personas víctimas de accidentes laborales o desempleo, y ancianos). La asistencia se otorga mediante mecanismos de seguro social que están financiados con impuestos, programas de obras públicas y otros sistemas que garantizan la seguridad del ingreso básico. Los países al contar con sistemas de protección social debidamente diseñados e implementados, pueden

fortalecer su capital humano y mejorar así la productividad; disminuir las desigualdades existentes; mejorar la resiliencia, y poner fin al ciclo de la pobreza intergeneracional (Banco Mundial, 2018)

Proyectos no contributivos. Los regímenes no contributivos enfocados a los pobres son regímenes sujetos a comprobación de recursos, suelen servirse de mecanismos de focalización que garantizan que los programas cubran solo a las personas cuyos medios de subsistencia (por lo general sus activos e ingresos) se encuentran por debajo de un umbral determinado. Pueden variar en cuanto a diseño y características, y esta diversidad puede manifestarse a través de los métodos de focalización utilizados, las condiciones complementarias necesarias para que los beneficiarios accedan a las prestaciones y a otras intervenciones además de la transferencia de ingresos. Los regímenes sujetos a la comprobación de recursos incluyen: i) Apoyo a los ingresos (transferencias no condicionadas), ii) Transferencias Monetarias Condicionadas y iii) Garantía del empleo (OIT, s.f.)

Transferencias monetarias no condicionadas. El dinero en efectivo es más apetecible a ojos de los beneficiarios que las transferencias en especie por algunas razones. (Tabor, Peppiatt, & Holzmann, 2002). En primer lugar, las transferencias en efectivo son más eficientes para lograr el impacto deseado. El apoyo en especie como alimentos o insumos agrícolas, tienen asociados altos costos de transacción y puede en cierta manera perjudicar a los mercados locales en cuanto a los productos que forman parte de la ayuda. Por ejemplo, los productores agrícolas pueden verse afectados de forma negativa por la distribución gratuita de alimentos en su región. En segundo lugar, se ha comprobado que las transferencias en efectivo han mejorado el bienestar de los hogares considerablemente, ya que las familias saben cómo dar un mejor uso al dinero. Sin embargo algunos programas restringen el gasto del beneficio otorgado a ciertos rubros, particularmente a los alimentos (FAO, s.f.)

Las transferencias monetarias pueden ser condicionadas o no condicionadas, las primeras son otorgadas con la condición de que los beneficiarios cumplan con ciertos requisitos de conducta. Esto puede consistir en asegurarse de que sus hijos asistan regularmente a la escuela o en utilizar los servicios preventivos de alimentación básica y de salud (OIT, s.f.); mientras que las segundas son proporcionadas para aliviar la pobreza en el corto plazo pues son proporcionadas a personas en situación de pobreza y pobreza extrema

Estado del Arte

Se realizó una revisión de literatura preliminar (PRL), para lo cual se utilizó como fuente principal de búsqueda de información, el repositorio académico IEEEExplore. Se han determinado los siguientes criterios de inclusión y exclusión:

Criterios de Inclusión

Se han identificado los siguientes criterios de inclusión:

- Artículo relacionado al uso de herramientas tecnológicas para la identificación de patrones de comportamiento que rigen a un segmento de datos.
- El artículo se refiere al análisis predictivo aplicado a variables categóricas.
- El artículo se refiere a la identificación de patrones de comportamiento utilizando datos categóricos.
- Se tomaron en cuenta únicamente aquellos artículos que han sido publicados desde el año 2011 en adelante.

Criterios de Exclusión

Se han identificado los siguientes criterios de exclusión:

- Artículos que no hagan referencia a la identificación de patrones de comportamiento sobre datos categóricos.
- Artículos que hagan referencia a Machine Learning
- Artículos que hagan referencia al internet de las cosas
- Se excluyeron artículos que han sido publicados previo al año 2011

Grupo de Control

Posterior a la identificación de criterios de inclusión y exclusión, se procedió a buscar estudios relacionados al uso de herramientas tecnológicas para la identificación de patrones de comportamiento o a fines, con el propósito de detectar y seleccionar palabras clave (ver tabla 1).

Tabla 1*Estudios del grupo de control y términos clave*

No.	Documento	Palabras Clave
GC1	Minería de Datos Aplicada a la Detección de Patrones Delictivos en Argentina (Valenga, et al.)	Patterns, association, clustering, classification, behavior patterns
GC2	Técnicas de minería de datos para la detección y prevención del lavado de activos y la financiación del terrorismo (LA/FT) (Unidad de Información y Análisis Financiero, 2014)	Behavior, prediction, detect, patterns, grouping, methods, knowledge
GC3	Identificación de técnicas de minería de datos para apoyar la toma de decisiones en la solución de problemas empresariales (Gutiérrez & Molina, 2016)	Grouping, classification, behavior, knowledge
GC4	Técnicas de minería de datos para identificar patrones de colaboración de los estudiantes que hacen uso del EVA de la UTPL (Pulla, 2011)	Techniques, Identification, behavior patterns, grouping

Nota. Esta tabla muestra el título de los cuatro estudios seleccionados para el grupo de control y las palabras clave identificadas en cada uno.

Palabras clave y coincidencias

Posterior a la identificación de palabras claves, se verificó la frecuencia con la que se repiten y su cobertura¹, con la finalidad de que la cadena de búsqueda sea más afinada (ver tabla 2).

Tabla 2

Palabras clave - Frecuencia

No.	Palabras clave	GC1	GC2	GC3	GC4	Total	Cobertura
1	Patterns	12	12	5	27	56	100%
2	Association	5	5	2	6	18	100%
3	Clustering	8	2	0	61	71	75%
4	Classification	15	7	6	12	40	100%
5	Behavior patterns	2	1	0	6	9	75%
6	Behavior	4	6	4	42	56	100%
7	Prediction	0	2	1	0	3	50%
8	Detect	4	2	0	1	7	75%
9	Methods	1	8	0	19	28	75%
10	Knowledge	14	26	12	39	91	100%
11	Grouping	2	0	2	20	24	75%
12	Techniques	8	28	37	36	109	100%
13	Identification	5	6	12	14	37	100%

Nota. Esta tabla muestra el número de repeticiones de las palabras clave en cada uno de los estudios pertenecientes al grupo de control.

¹ Porcentaje de artículos del grupo de control donde aparece la palabra clave

Cadena de búsqueda y resultados obtenidos

Las palabras clave identificadas se utilizaron para la conformación de las cadenas de búsqueda y así determinar la cadena de búsqueda ideal para el presente estudio (ver tabla 3). Para efectuar esta actividad se utilizó el repositorio académico IEEExplore.

Tabla 3

Construcción de la Cadena de búsqueda

No.	Cadena de búsqueda	No. estudios	Estudios del Grupo de Control
CB1	("Patterns" or "behavior patterns") AND ("Clustering" or "classification" or "association")	34	GC1, GC2, GC3, GC4
CB2	"behavior patterns" and ("prediction" or "detect" or "identification")	17	GC1, GC2, GC3, GC4
CB3	("behavior patterns") AND ("Clustering" or "classification" or "grouping") AND "techniques"	10	GC1, GC2, GC3, GC4
CB4	"techniques" and ("clustering" or "grouping") and "behavior patterns"	38	GC1, GC2, GC3, GC4

Nota. Esta tabla muestra las cadenas de búsqueda que fueron probadas y la seleccionada como más idónea.

La cadena de búsqueda (CB4) es la adecuada para determinar los estudios primarios. En la siguiente sección se lleva a cabo el proceso de selección correspondiente.

Selección de Estudios

La cadena de búsqueda CG4 tuvo un resultado de 38 estudios candidatos, de los cuáles 32 fueron descartados en una primera revisión debido a que eran estudios previos al año 2011 y otros no hacían referencia al uso de herramientas tecnológicas para la identificación de patrones de comportamiento que rigen a un segmento de datos. Seis (6) fueron los estudios seleccionados como se puede evidenciar en la tabla 4.

Tabla 4

Selección de estudios

Estudios	Cantidad	Porcentaje
Estudios Seleccionados	6	16%
Estudios Descartados	32	84%
Total	38	100%

Nota. Esta tabla muestra el número de estudios seleccionados y descartados durante la primera revisión.

Una vez revisados los 6 estudios seleccionados anteriormente, se determinó que 3 son estudios primarios debido a que se relaciona con el manejo de herramientas informáticas en la identificación de patrones a través de técnicas de agrupación y 3 fueron descartados por no enfocarse directamente en la identificación de patrones o perfilamiento de datos. (ver tabla 5).

Tabla 5*Selección de Estudios Primarios*

Estudios	Cantidad	Porcentaje
Estudios Primarios	3	50%
Estudios Descartados	3	50%
Estudios Sin analizar	0	0%
Total	6	100%

Nota. Esta tabla muestra los estudios primarios y los estudios descartados durante la segunda revisión.

Resultados de la Revisión

Esta sección detalla los resultados de la revisión de literatura efectuada, dando respuesta a las preguntas de investigación planteadas para el objetivo 2 del presente trabajo de titulación

Objetivo específico 2: Identificar herramientas tecnológicas orientadas a la selección de patrones de comportamiento que rigen un segmento de datos mediante una revisión preliminar de literatura (PRL), para determinar la propuesta que mejor se ajuste a las necesidades del proyecto.

- **OE2 –RQ2.1: ¿Qué artículos hacen referencia al análisis y gestión adecuada de información en temas de identificación de perfiles o patrones de comportamiento que rigen a un segmento de datos?**

La tabla 6 muestra los 3 estudios primarios seleccionados que hacen referencia a la identificación de perfiles o patrones de comportamiento que rigen a un segmento de datos utilizando herramientas tecnológicas para Inteligencia de Negocios. Los artículos fueron publicados entre los años 2014 y 2018 y se encuentran en idioma inglés.

Tabla 6

Estudios Primarios

No.	Título	Autor	Año publicación
1	Analysis of Malware Behaviour: Using Data Mining Clustering Techniques to Support Forensics Investigation	Edem Inang ; Chafika Benzaïd ; Ameer Al-Nemrat ; Paul Watters	2014
2	A two-stage pattern recognition method for electric customer classification in smart grid	Bo Peng ; Can Wan ; Shufeng Dong ; Jin Lin ; Yonghua Song ; Yi Zhang ; Jun Xiong	2016
3	Formalistic Modelling Based on Pattern Recognition Applied to the Knowledge and Human Talent Sector in Ecuador	Guayasmín, Adrián; Fuertes, Walter; Campaña, Mauricio; Toulkeridis, Theofilos	2018

Nota. Esta tabla contiene información de los estudios primarios seleccionados.

En el siguiente apartado se detalla un resumen de los resultados obtenidos del proceso de revisión preliminar de literatura efectuado a los artículos detallados en la Tabla 6.

- **OE2 – RQ2.2: ¿Cuáles son las herramientas tecnológicas que permiten identificar perfiles o patrones de comportamiento que rigen a un segmento de datos y que se acoplan mejor a las necesidades del SCTP Mtukula Pakhomo?**

Antes de proceder con la respuesta a la pregunta de investigación previamente citada, es importante mencionar que los estudios primarios analizados no están directamente relacionados con el tema de la presente tesis, sin embargo, por la no existencia de estudios relacionados se han considerado los que se enfocan en técnicas de agrupamiento de datos para la identificación de patrones o perfiles de comportamiento, como soporte a los problemas que han presentado.

Un resumen de cada uno de los estudios primarios analizados se presenta a continuación, mencionando como resultado las propuestas tecnológicas utilizadas en el presente trabajo de titulación.

Los autores Edem, Benzaïd, Al-Nemrat, y Watters en su artículo **Analysis of Malware Behaviour: Using Data Mining Clustering Techniques to Support Forensics Investigation** (Edem, Benzaïd, Al-Nemrat, & Watters, 2014) proponen un modelo de marco de investigación de malware automatizado, un enfoque basado en componentes que está diseñado para respaldar la investigación al integrar tanto el análisis de malware como las técnicas de agrupamiento de minería de datos, sin embargo, el análisis dinámico de las enormes muestras de malware solo producirá un gran volumen de datos de informes, lo que podría abrumar al investigador. Por lo tanto, la necesidad de adoptar técnica de agrupamiento de los datos como un proceso no supervisado ya que se quiere encontrar grupos homogéneos de puntos en un conjunto de datos que se puede considerar que comparten características similares. El agrupamiento implementado fue particional utilizando el algoritmo k-medias a través de la

herramienta rapid miner, misma que también fue utilizada para la transformación de datos. En general, el algoritmo de K-medias fue la mejor opción para este estudio, se agruparon 11.591 características en 8 grupos en 2 minutos y 32 segundos. Lo que supone una ganancia de tiempo considerable durante la investigación de malware.

Los autores Peng, y otros en su artículo **A two-stage pattern recognition method for electric customer classification in smart grid** (Peng, y otros, 20016) manifiestan que identificar los patrones de consumo de los clientes eléctricos y agruparlos en clases según sus características de carga puede ser muy significativo para el suministro de energía y la gestión del lado de la demanda en redes inteligentes. Es por esto que proponen un algoritmo de agrupación en dos etapas que combina métodos de aprendizaje supervisado y no supervisado. En primer lugar, los resultados de la agrupación se obtienen según el método de aprendizaje no supervisado donde se ejecuta la comparación de algoritmos K-medias, Fuzzy c-Means y Ward, resultando el mejor el algoritmo k-medias por ser la rodilla de todos los resultados de la agrupación. En segundo lugar, los clientes se reclasifican utilizando algoritmos de aprendizaje supervisado como Classification and Regression Tree (CART), Random Forest, Support Vector Machine (SVM) y K Nearest Neighbors, se elige SVM como el algoritmo de reclasificación de la segunda etapa debido a su desempeño confiable en la clasificación. En general, el método propuesto puede mejorar la agrupación de curvas de carga de dimensiones completas en la similitud de forma de carga. La carga se puede reclasificar de forma más razonable que los métodos clásicos.

Los autores Guayasmín, Fuertes, Campaña, & Toulkeridis en su artículo **Formalistic Modelling Based on Pattern Recognition Applied to the Knowledge and Human Talent Sector in Ecuador** (Guayasmín, Fuertes, Campaña, & Toulkeridis, 2018) manifiestan su propósito en analizar el conjunto de datos sobre la educación recibida por los graduados de

secundaria en Ecuador, para reconocer patrones de comportamiento y la relación correspondiente entre su empleo y las actividades económicas asociadas. Para cumplir con su objetivo han realizado varios algoritmos de clasificación, así como técnicas de asociación, utilizando el proceso de minería de datos CRISP-DM donde mencionan herramientas como R-Studio, WEKA, Rapid Miner, sin embargo, para su estudio utilizaron la herramienta R-studio debido a que cumple con todos los requerimientos para la minería de datos del presente estudio. Para la agrupación de datos, utilizaron el algoritmo K-means y su calibración con el método experimental del codo. Aplicaron la plataforma de integración de datos Pentaho y PostgreSQL para temas de ETL y la implementación misma del estudio. Los resultados demuestran la generación dinámica de la composición social, con algunos problemas detectados, como el desempleo, el subempleo, la ocupación informal e incluso un excedente de profesionales en ciertas áreas, dando una conclusión de la relación con su actividad económica. La información obtenida puede permitir el desarrollo de políticas públicas que podrían mejorar la gestión del Conocimiento, así como la matriz productiva del país.

Como se puede ver en los párrafos anteriores, los estudios primarios hacen referencia a investigaciones que buscan o proponen el agrupamiento de los datos para conocer patrones de comportamiento en diferentes áreas, los autores han implementado técnicas de inteligencia de negocios como la minería de datos con métodos de aprendizaje no supervisado, algoritmos de agrupación K-medias y para la calibración del número de cluster idóneo el método experimental del codo, herramientas de minería de datos como Rapid Miner, R studio, Weka y para ETL herramientas como Rapid Miner y Pentaho Data Integration con PostgreSQL. Estas herramientas, respectivamente, aportaron positivamente cada una de estas investigaciones permitiéndoles llegar así a los resultados deseados, brindándoles también ahorros de tiempo en la investigación.

Los artículos previamente analizados son la base para la determinación de que herramientas y métodos se van a aplicar en el presente trabajo de titulación. Se considera que, las herramientas y métodos que se acoplan de mejor manera a las necesidades del SCTP Mtukula Pakhomo son: i) Pentaho Data Integration para efectuar el ETL ya que es una herramienta de libre acceso, amigable para el usuario y permite trabajar con grandes cantidades de datos para transformarlos según las necesidades del programa, adicionalmente porque brinda confiabilidad en los datos que en el presente caso serán el input para la determinación de los patrones o perfiles de comportamiento de los beneficiarios que no cobran regularmente el beneficio del programa, ii) RapidMiner para el análisis y el agrupamiento de datos utilizando el algoritmo K-means (método de aprendizaje no supervisado, particional) ya que lo que se busca en la presente tesis es determinar patrones o perfiles de los beneficiarios que no cobran regularmente el beneficio, es decir agrupar los datos existentes para así obtener los resultados; y en base al análisis de los estudios primarios este algoritmo fue el que dio los mejores resultados. RapidMiner tiene una licencia gratuita por un tiempo determinado y tiene una interfaz que facilita su uso. Adicionalmente para determinar el número de cluster idóneo a utilizar se empleará el método del codo ya que permite identificar fácilmente el punto de la curva (codo) en el que se obtendrán los mejores resultados.

Capítulo III: Marco Metodológico

Introducción

Basado en la hipótesis del presente trabajo de titulación, donde se recalca la implementación de un modelo de gestión de información utilizando técnicas y soluciones de BI, se ha seleccionado un enfoque cualitativo² para regir la investigación. Esta investigación es considerada como una investigación exploratoria y de análisis, que ayudará a identificar el perfil o los patrones que rigen el comportamiento de los beneficiarios que no cobran de manera regular el beneficio otorgado por el programa Mtukula Pakhomo para de esta manera establecer recomendaciones que permitan mejorar la frecuencia de recepción del beneficio en este tipo de beneficiarios.

El enfoque cualitativo me permitirá identificar además nuevos fenómenos que pueden surgir en el desarrollo de la investigación, así como también me proporcionará una comprensión profunda del objeto de estudio.

Metodología de Investigación

La presente investigación estará orientada por una metodología propia. La metodología propuesta está compuesta por 4 fases:

1. Identificación de la situación actual

² La investigación cualitativa genera principalmente palabras, es principalmente una investigación exploratoria. Se utiliza sobre todo para obtener una comprensión de las opiniones y motivaciones. (<https://www.questionpro.com/blog/es/cualitativa-cuantitativa-o-ambas-haz-investigacion-de-mercados/>)

2. Estudio de viabilidad de la solución
3. Construcción de la solución
4. Validación de la solución

Esta metodología permitirá establecer un orden en el desarrollo de la investigación, asegurando que cada uno de los objetivos y preguntas de investigación planteadas sean alcanzados (ver figura 4).

Figura 4

Metodología de la Investigación

FASES	<div style="display: flex; justify-content: space-between;"> <div style="width: 22%;">Definición de la Situación Actual</div> <div style="width: 22%;">Estudio de viabilidad de la solución</div> <div style="width: 22%;">Construcción de la solución</div> <div style="width: 22%;">Validación de la solución</div> </div>			
MÉTODOS	<div style="display: flex; justify-content: space-between;"> <div style="width: 22%;">Entrevista al Director del proyecto Análisis de la información de pagos Planteamiento de Solución</div> <div style="width: 22%;">Revisión Preliminar de Literatura</div> <div style="width: 22%;">ETL Implementación de la solución de BI/minería de datos (Selección de algoritmo) Identificación del perfil/patrones de los beneficiarios</div> <div style="width: 22%;">Obtención de resultados (recomendaciones del estudio) Entrevista al Director del proyecto</div> </div>			
OBJETIVOS CUBIERTOS	<div style="display: flex; justify-content: space-between;"> <div style="width: 22%;"> <p>OE1: Identificar las causas y efectos que surgen de la problemática del cobro irregular del beneficio otorgado por el programa Mtukula Pakhomo mediante técnicas de recolección o adquisición de información</p> </div> <div style="width: 22%;"> <p>OE2: Identificar herramientas tecnológicas orientadas a la selección de patrones de comportamiento que rigen un segmento de datos mediante una revisión preliminar de literatura (PRL), para determinar la propuesta que mejor se ajuste a las necesidades del proyecto.</p> </div> <div style="width: 22%;"> <p>OE3: Recolectar, organizar, limpiar, adecuar y homologar los datos disponibles mediante un proceso de extracción, transformación y carga (ETL) para que sean explotados a través de la solución BI seleccionada</p> <p>OE4: Analizar los datos a través de la solución BI para identificar los patrones y/o perfiles que rigen el comportamiento de los beneficiarios que cobran de manera irregular el beneficio y emitir recomendaciones operativas que permitan el cobro regular del mismo.</p> </div> <div style="width: 22%;"> <p>OE4: Analizar los datos a través de la solución BI para identificar los patrones y/o perfiles que rigen el comportamiento de los beneficiarios que cobran de manera irregular el beneficio y emitir recomendaciones operativas que permitan el cobro regular del mismo..</p> <p>OE5: Validar los resultados y las recomendaciones emitidas del análisis de los datos a través de una entrevista efectuada al director del proyecto y stakeholders involucrados.</p> </div> </div>			
PREGUNTAS DE INVESTIGACIÓN CUBIERTAS	<div style="display: flex; justify-content: space-between;"> <div style="width: 22%;"> <p>OE1-RQ1.1 ¿Qué técnica de recolección o adquisición de información es la adecuada para identificar las causas y efectos de la problemática que presenta el programa “Mtukula Pakhomo”?</p> <p>OE1-RQ1.2 ¿Qué causas y efectos se pueden detectar de la problemática actual del programa “Mtukula Pakhomo” en Malawi con respecto al cobro irregular del beneficio otorgado por parte de los beneficiarios?</p> </div> <div style="width: 22%;"> <p>OE2 – RQ2.1: ¿Qué artículos hacen referencia al análisis y gestión adecuada de información en temas de identificación de perfiles o patrones de comportamiento que rigen a un segmento de datos?</p> <p>OE2 – RQ2.2: ¿Cuáles son las herramientas tecnológicas que permiten identificar perfiles o patrones de comportamiento que rigen a un segmento de datos y que se acoplan mejor a las necesidades del SCTP Mtukula Pakhomo?</p> </div> <div style="width: 22%;"> <p>OE3 – RQ3.1: ¿Cuáles son las fuentes y los tipos de datos con los que cuenta el SCTP Mtukula Pakhomo?</p> <p>OE3 – RQ3.2: ¿Qué tipo de datos se necesita para llevar a cabo la identificación de perfiles o patrones de comportamiento de los beneficiarios que no cobran regularmente el beneficio otorgado por el gobierno en el SCTP Mtukula Pakhomo?</p> <p>OE3 – RQ3.3: ¿Existen valores atípicos que puedan afectar la identificación de perfiles o patrones de este tipo de beneficiarios?</p> <p>OE3 – RQ3.4 ¿Cuál es el mejor método y herramienta para realizar un ETL de la información recopilada?</p> <p>OE4 – RQ4.1: ¿Cuáles son los perfiles o patrones obtenidos de los beneficiarios que no cobran regularmente el beneficio otorgado por el gobierno en el SCTP Mtukula Pakhomo?</p> </div> <div style="width: 22%;"> <p>OE4 – RQ4.2: ¿Cuáles son las recomendaciones obtenidas del estudio, que deberían aplicarse al programa, para que los beneficiarios que cobran irregularmente el beneficio lo reciban de forma regular?</p> <p>OE5 - RQ5.1: ¿Las recomendaciones resultantes de la implementación del modelo de gestión de información basado en BI, mejorarán la frecuencia de recepción del beneficio otorgado por el programa en los beneficiarios que actualmente lo cobran de manera irregular?</p> <p>OE5 – RQ5.2: ¿Es factible la aplicación de las recomendaciones en el programa Mtukula Pakhomo?</p> </div> </div>			

Nota. Esta figura muestra la metodología utilizada en la presente investigación.

Fases de la Metodología

Esta sección describe cada una de las fases identificadas para la metodología propuesta.

Identificación de la Situación Actual. A través del análisis de la información de pagos almacenada en el sistema de manejo de información se evidencia el problema del SCTP Mtukula Pakhomo con respecto al cobro irregular del beneficio por un grupo de beneficiarios. Entrevistas que contemplan preguntas abiertas para la obtención de información, efectuadas al director del proyecto de la consultora contratada para la elaboración del diseño e implementación del programa, nos ayudan a determinar las causas y efectos de la problemática existente. Este tipo de información nos llevará a plantear una solución clara al proyecto, con lo cual se identifican los objetivos generales, específicos e hipótesis del presente trabajo de titulación.

En esta fase metodológica se aborda el primer objetivo específico de la investigación y por ende se da respuesta a sus preguntas de investigación.

Estudio de Viabilidad de la Solución. Durante esta fase se efectúa una Revisión Preliminar de Literatura (RPL) con el objetivo de identificar propuestas tecnológicas que permitan identificar los perfiles o patrones que rigen el comportamiento de los beneficiarios que cobran de manera irregular el beneficio otorgado por el programa Mtukula Pakhomo.

El segundo objetivo específico de la investigación es cubierto en esta etapa, dando respuesta a sus preguntas de investigación. Los artículos científicos confiables, resultado del estado del arte efectuado permiten dar respuesta a las preguntas de investigación planteadas.

Construcción de la Solución. En esta fase se lleva a cabo la recopilación de los datos que el programa tiene y almacena de sus beneficiarios, se da un tratamiento de limpieza, adecuación y homologación de los mismos; se identifica los tipos de datos requeridos y la relevancia de los mismos para el caso de estudio, así como también los valores atípicos que pueden afectar la solución. Para esto se solicita al director del proyecto, los permisos de acceso a la fuente de datos.

Adicionalmente se define la solución BI y el algoritmo a ser implementado, con el propósito de llevar a cabo el análisis de los datos y la obtención de los perfiles o patrones que rigen el comportamiento de los beneficiarios que cobran de manera irregular el beneficio del programa.

Esta fase es ejecuta sobre la base de la metodología CRISP-DM, ya que es flexible y puede ser personalizada según el requerimiento del presente trabajo de titulación. Esta metodología consta de procesos que abarcan una visión de lo macro a lo detallado (Chapman, y otros, 1999). La metodología CRISP-DM a aplicarse en la construcción de la presente solución está compuesta de 5 fases:

- Comprensión del negocio
- Comprensión de los datos
- Preparación de los datos
- Modelado
- Resultados

En esta etapa se cubre el tercer objetivo específico dando respuesta a todas sus preguntas de investigación y el cuarto objetivo específico dando respuesta a una de sus preguntas de investigación.

Validación de la Solución. En esta etapa se analizan los resultados obtenidos, es decir los perfiles o los patrones que rigen el comportamiento de los beneficiarios que cobran el beneficio de manera irregular; en base a estos resultados se plantearán las recomendaciones que deberían ser aplicadas en el programa con el objetivo de mejorar la frecuencia en la recepción de las transferencias monetarias de este tipo de beneficiarios.

Adicionalmente se procederá a presentar los resultados y las recomendaciones del estudio al director del proyecto, para que sea el quien los valide y determine si es factible la aplicación de la solución en el programa Mtukula Pakhomo. También se capacitará a los interesados en el uso de la solución para así determinar el nivel de satisfacción que tienen con respecto a su usabilidad y a su aporte en el desempeño del negocio.

En esta etapa se cubre el cuarto objetivo específico dando respuesta a una de sus preguntas de investigación y el quinto objetivo específico de igual manera dando respuesta a sus preguntas de investigación.

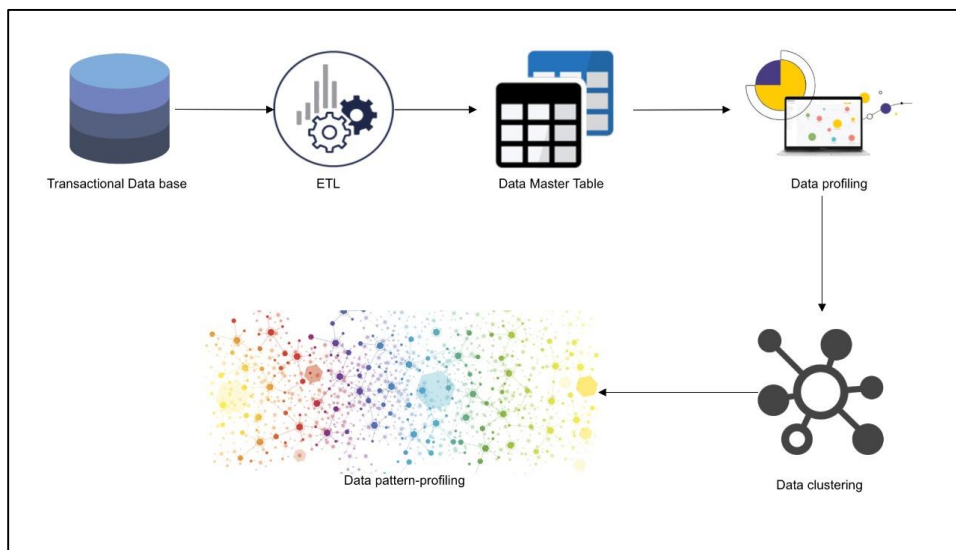
Capítulo IV: Construcción de la Solución

En la *Figura 5* se presenta el esquema de la solución propuesta, se inicia con la revisión del esquema de base de datos transaccional (**Transactional Data base**) facilitado por la empresa **Ayala Consulting Cia. Ltda**, a continuación, esta data estructurada es transformada a un modelo unificado por **una MDT (master data table)** a través de un proceso **ETL** que permitió agrupar todos los datos en una sola tabla de información estructurada, sin relaciones ni claves; adicionalmente se tipificó cada una de las variables categóricas a un esquema numérico detallado posteriormente en la sección 4.4 - tabla 8: Asignación de valores a variables categóricas.

Estos datos se someten a un proceso de **data profiling** donde se obtiene un vistazo general de alto nivel de la MDT tanto para variables categóricas como numéricas, el siguiente paso es conocer el perfil o los perfiles de aquellos beneficiarios que cobran y que no cobran el beneficio otorgado por el programa Mtukula Pakhomo, para lo cual se utilizó un proceso de **data clustering** no supervisado con el algoritmo K-means para determinar clusters o grupos con características comunes y los pesos por grupo para cada variable que intervino en el estudio; finalmente en el proceso de **data pattern-profiling** se utilizó un nuevo proceso de perfilamiento de los datos para conocer las características comunes de cada agrupación y así determinar los perfiles requeridos, tanto para el beneficiario que cobra como para el que no cobra.

Figura 5

Esquema de la solución propuesta



Nota: Esta figura representa el diseño de la solución propuesta.

Transaccional Data Base (Base de datos transaccional)

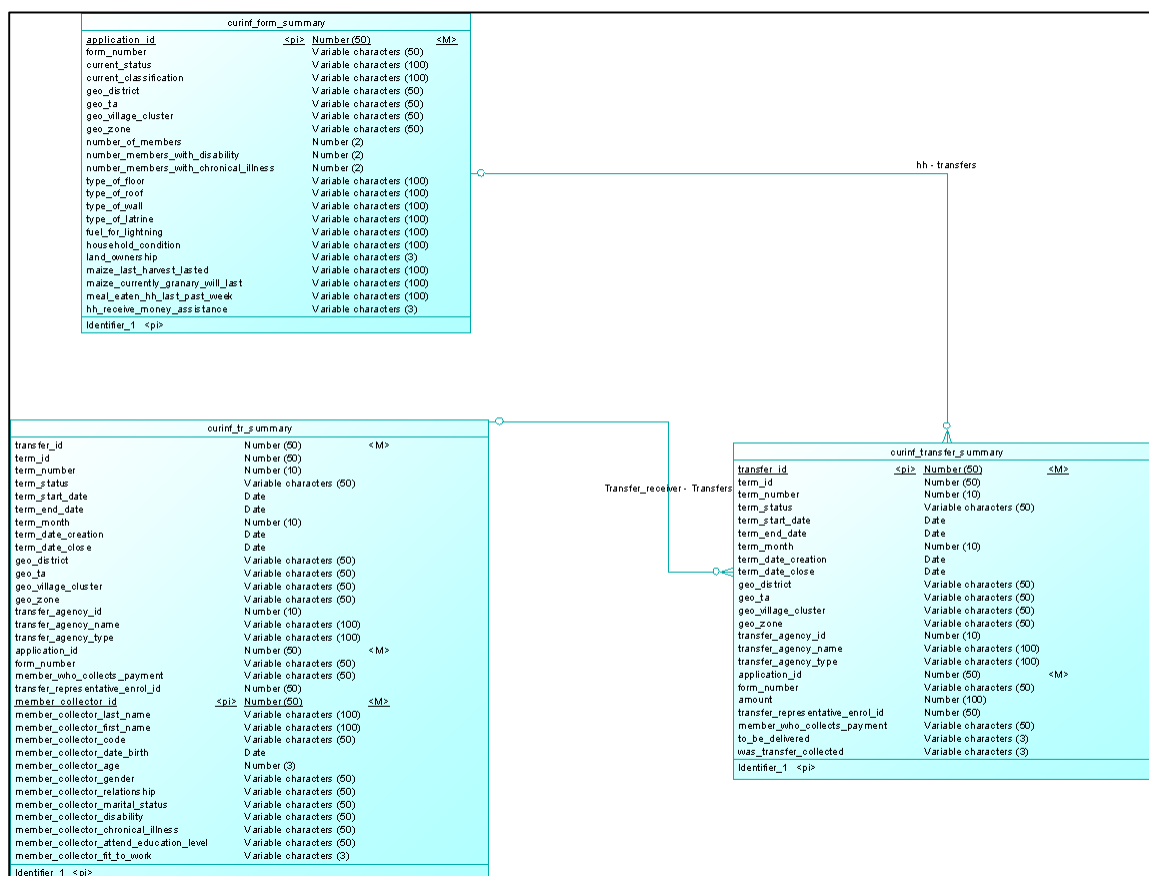
La empresa Ayala Consulting Cia. Ltda, entregó un respaldo de las tablas transaccionales del programa Mutkula Pakhomo donde se almacena actualmente la información de los beneficiarios, conjuntamente con un histórico de datos desde el año 2013 hasta el 2019 con un aproximado de 1M de registros.

En la *Figura 6. Esquema transaccional de base de datos*, se presenta el modelo entidad relación de estas tablas donde encontramos: **i) currinfo_form_summary** que contiene la información recolectada durante el proceso de enrolment de los hogares beneficiarios del programa; el proceso de enrolment se lleva acabo al inicio de la implementación del programa; **ii) curinf_TR_summary** que muestra información de la persona que recolectó las transferencias del hogares beneficiarios; y **iii) curinf_transfers_summary** esta tabla refleja

la información de las transferencias efectuadas a los hogares beneficiarios, el diccionario de datos se puede revisar en el Anexo 1.

Figura 6

Esquema transaccional de base de datos



Nota: Esta figura muestra el esquema transaccional de las tres tablas utilizadas de la base de datos.

Proceso de Extracción Transformación y Carga de datos (ETL) y Construcción de la Máster Data Table (MDT)

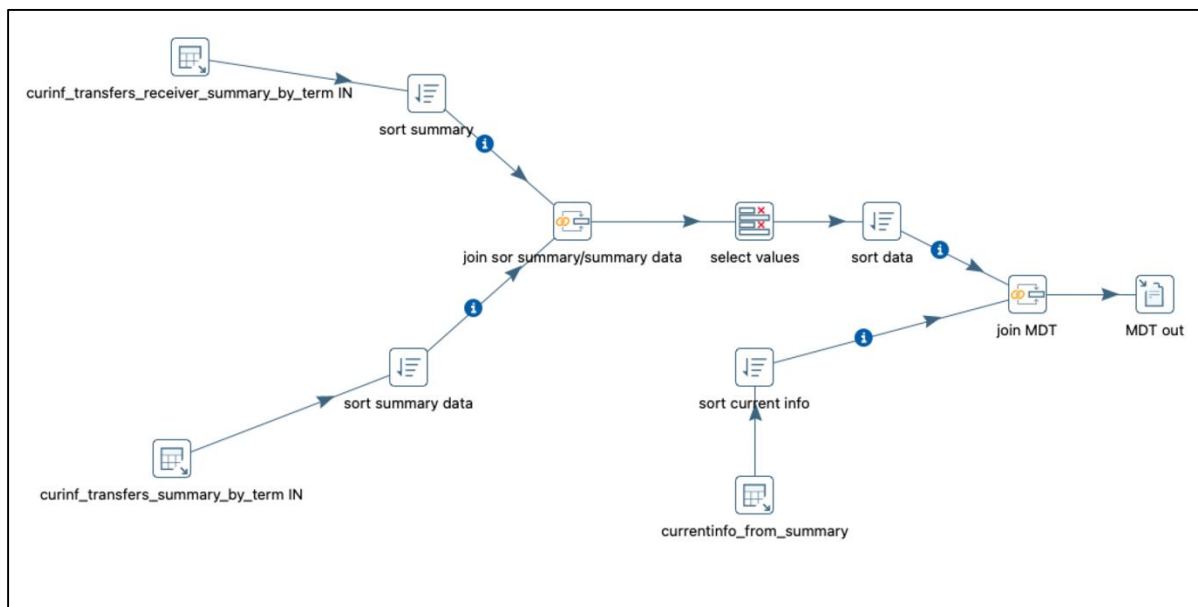
Se seleccionó el proceso ETL (extracción, transformación y carga), debido a su rápida curva de aprendizaje, su fácil comprensión y explicación y debido a que el proceso ELT (extracción, carga y transformación) que pudo haberse empleado también, requiere de un manejo un poco más detallado y avanzando del lenguaje SQL dificultando de cierta manera este trabajo.

Se utilizó la herramienta Pentaho Data Integration con su componente Spoon debido a que es open source y cuenta con una interfaz amigable y de fácil entendimiento para el usuario; es un gran componente ETL mismo que permite crear conexiones a los datos, efectuar transformaciones sencillas y complejas, insertar fórmulas sin necesidad de recurrir a la programación con código. (itop Academy, 2019)

Con este componente se creó una MDT cruzando los datos de las tres tablas del modelo transaccional detallado en la sección anterior, en la Figura 7. Creación MDT, se presenta el esquema del ETL realizado donde iniciamos el proceso extrayendo la información de dos tablas principales `curinf_TR_summary` y `curinf_transfers_summary`, a continuación, se efectuó el cruce por las claves primarias; finalmente, se enlazó estos datos con la tabla `currinfo_form_summary`. Al finalizar el procesamiento desembocó en 1M registros procesables.

Figura 7

Creación MDT



Nota: Esta figura muestra el proceso que se siguió en la herramienta Spoon para la creación de la Master Data Table.

Proceso de descubrimiento de datos (Data Discovery)

Una vez construida la MDT con las variables demográficas que permitirán el perfilamiento, se estudió individualmente cada una de ellas utilizando la herramienta Rapid Minner y su complemento “Turbo prep”; debido a que otorga una licencia gratuita por treinta días y porque permite efectuar la preparación y análisis de datos con un entorno gráfico que facilita la comprensión de los resultados obtenidos instantáneamente (Rapidminer, 2021). En la tabla 7 se presenta el listado de las doce variables que fueron determinadas conjuntamente con los expertos de negocio como las más representativas.

Tabla 7*Variables demográficas representativas*

No.	Nombre de la variable
1	Member_collector_age
2	Member_collector_gender
3	Member_collector_relationship
4	Member_collector_marital_status
5	Member_collector_disability
6	Member_collector_chronical_illness
7	Member_collector_attend_education
8	Geo_district
9	Transfer_agency_type
10	Member_who_collects_payment
11	Was_transfer_collected
12	To_be_delivered

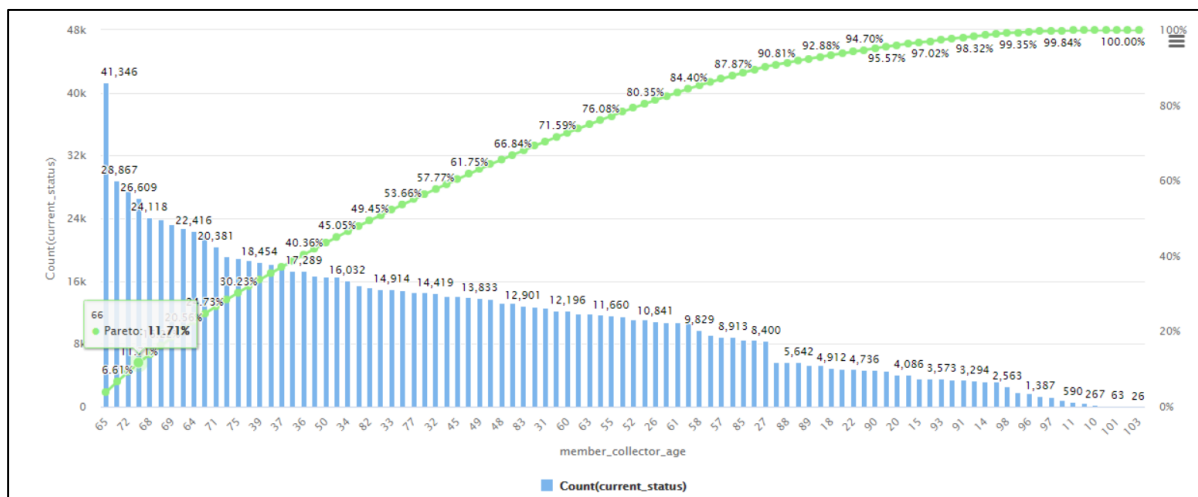
Nota. Esta tabla muestra las variables demográficas más representativas para ser utilizadas en el presente trabajo de investigación.

A continuación, se detalla el estado actual de cada una de estas variables.

Member collector age

Este campo representa la edad de los miembros que cobra el beneficio, por regla de negocio se conoce que las edades pueden oscilar entre los 11 y 90 años.

Figura 8

Member collector age

Nota: Esta figura muestra el comportamiento actual de la variable member collector age.

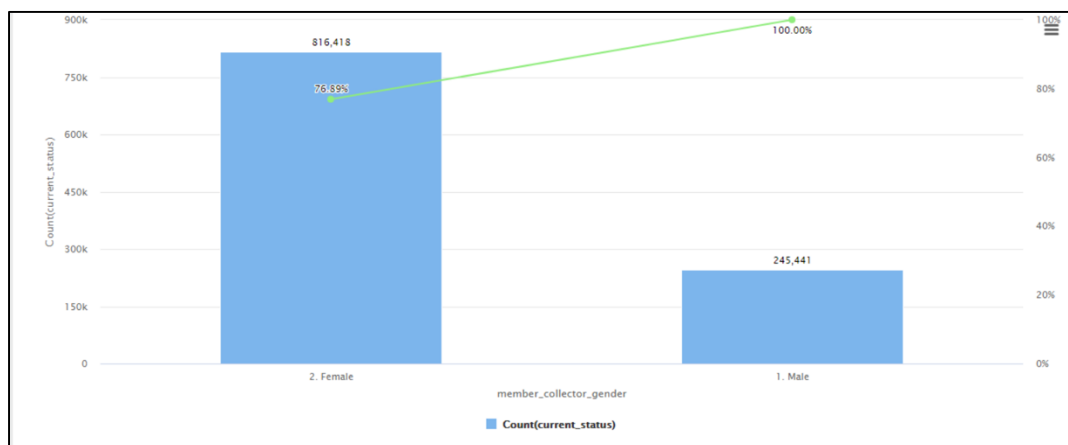
Según la Figura 8, el 52.91% de los registros de transferencia están asociados a receptores cuyas edades están entre 19 y 64 años, el 44.86% de registros están asociados a receptores mayores a 64 años y un 2.23% de registros están asociados a receptores menores de 19 años.

Member collector gender

Esta variable hace referencia al género de las personas que cobran el beneficio, para este caso particular esta variable únicamente cuenta con dos valores, hombre y mujer.

Figura 9

Pareto Member collector gender

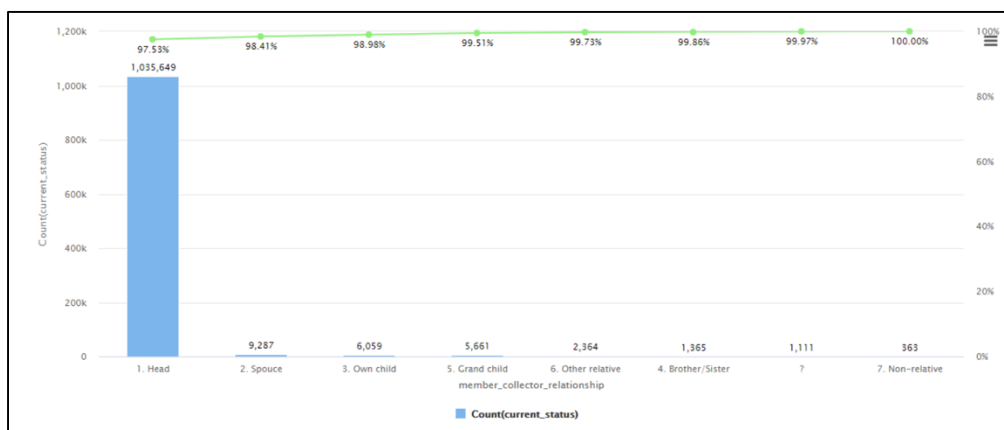


Nota: Esta figura muestra el comportamiento actual de la variable member collector gender.

Actualmente según la Figura 9, del total de registros de transferencia, un 77% está asociado a receptores mujeres, mientras que el 23% restante está asociado a receptores hombres.

Member collector relationship

Esta variable representa el parentesco del receptor de la transferencia con el jefe del hogar, los valores que esta puede tomar son head (cabeza de hogar), spouce (conyuge), own child (hijo propio), gran child (nieto), other relative (otro pariente), brother / sister (hermano/hermana), non-relative (sin parentezco).

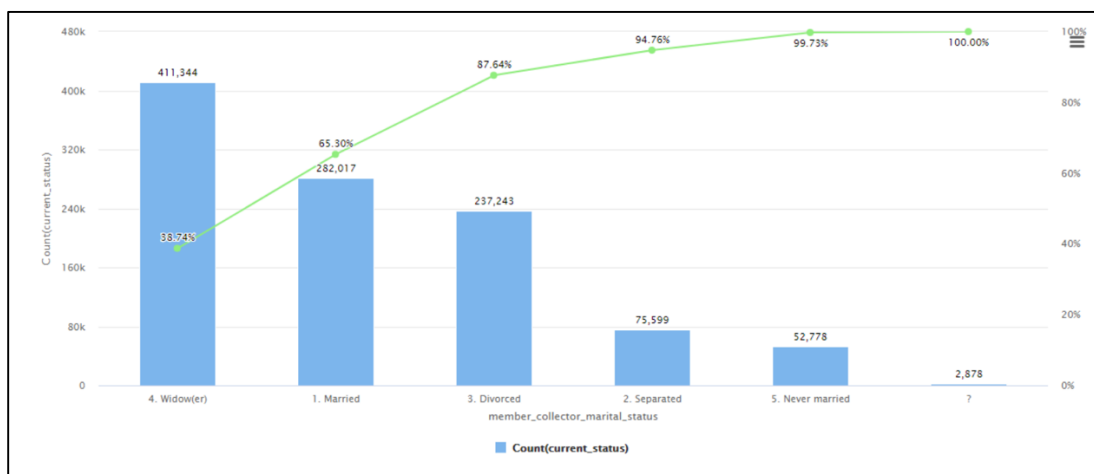
Figura 10***Pareto Member collector relationship***

Nota: Esta figura muestra el comportamiento actual de la variable member collector relationship.

Según la Figura 10, el 97% de los registros de transferencia están asociados a un receptor que es el jefe de hogar, el 2.33% está asociado a un receptor q forma parte del hogar y un 0.03% está asociado a receptores que no forman parte del hogar. Finalmente, el 0.10% de los registros no cuenta con este tipo de información

Member collector marital status

Este campo representa el estado civil del miembro del hogar que cobra el beneficio, actualmente esta variable puede tomar cualquiera de estos valores, widow(er) (viudo/da), married (casado), divorced (divorciado), separated (separado) o never married(soltero).

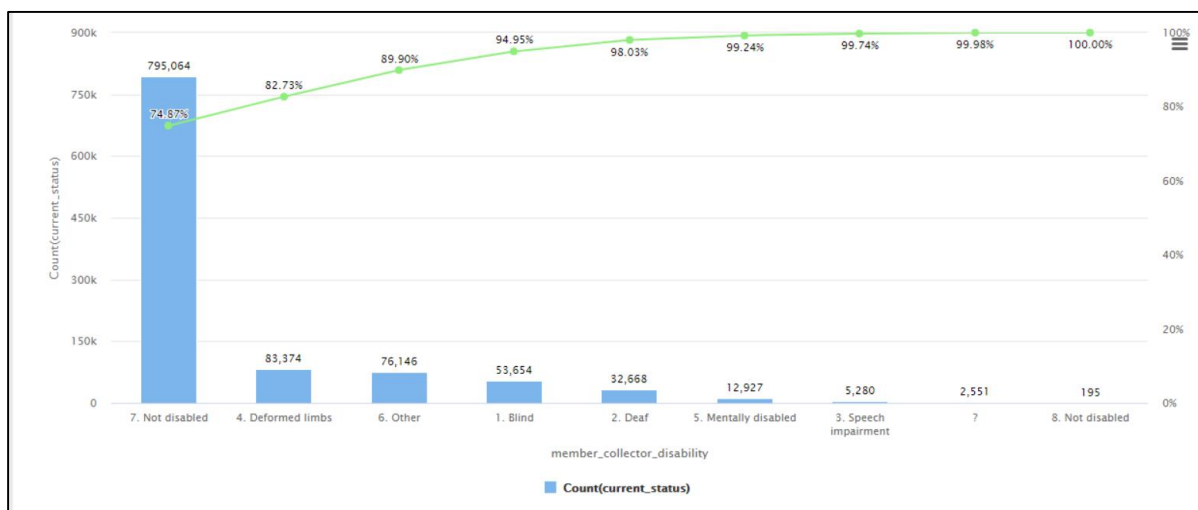
Figura 11***Pareto Member collector marital status***

Nota: Esta figura muestra el comportamiento actual de la variable member collector marital status.

Según la Figura 11, del total de registros de transferencia, el 39% corresponde o está asociado a receptores divorciados; el 27% está asociado a receptores casados. Un 0.27% no cuenta con esta información.

Member collector disability

Esta variable representa el tipo de discapacidad del receptor de la transferencia.

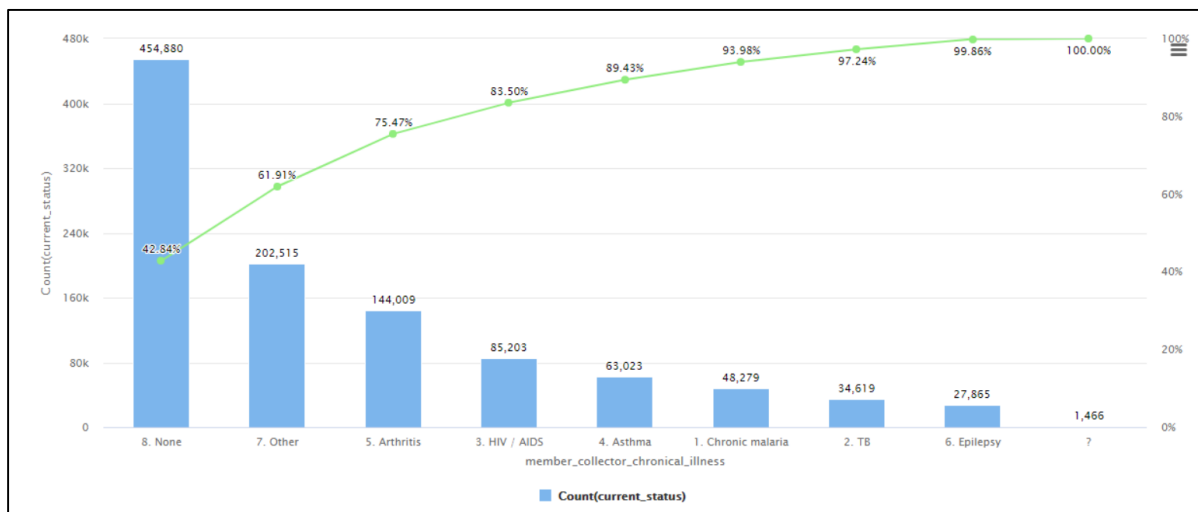
Figura 12***Pareto Member collector disability***

Nota: Esta figura muestra el comportamiento actual de la variable member collector disability.

En la Figura 12, el 75% de los registros de las transferencias están asociados a receptores que no tienen ningún tipo de discapacidad. El 24.87, está asociado a receptores que presentan una discapacidad, y el 0.23% no cuenta con esta información.

Member collector chronic illness

Esta variable representa el tipo de enfermedad crónica del receptor de la transferencia.

Figura 13***Pareto Member collector chronic illness***

Nota: Esta figura muestra el comportamiento actual de la variable member collector chronic illness.

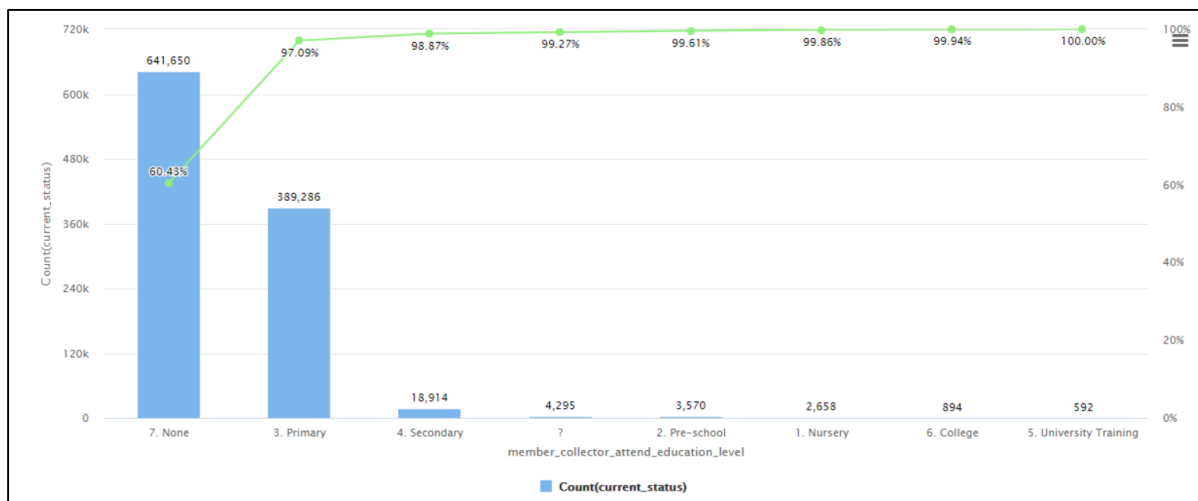
En la Figura 13, se observa que el 42.84% de los registros de las transferencias están asociados a receptores que no tienen ningún tipo de enfermedad crónica. El 57.02%, está asociado a receptores que presentan una enfermedad crónica y el 0.14% no cuenta con esta información.

Member collector attended education level

Esta variable indica el nivel educativo del receptor de la transferencia.

Figura 14

Pareto Member collector attended education level

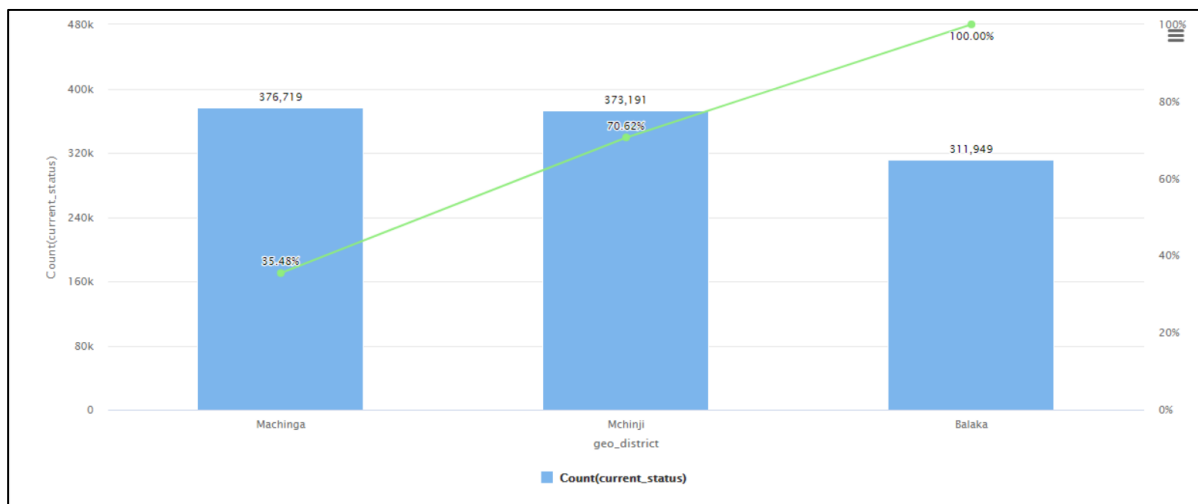


Nota: Esta figura muestra el comportamiento actual de la variable member collector attended education level.

Actualmente según la Figura 14, podemos observar que, del total de registros de transferencia, el 60.43% está asociado a receptores que no asistieron a la escuela y apenas el 0.40% de los registros no cuenta con información.

District

Esta variable corresponde al distrito donde vive el hogar.

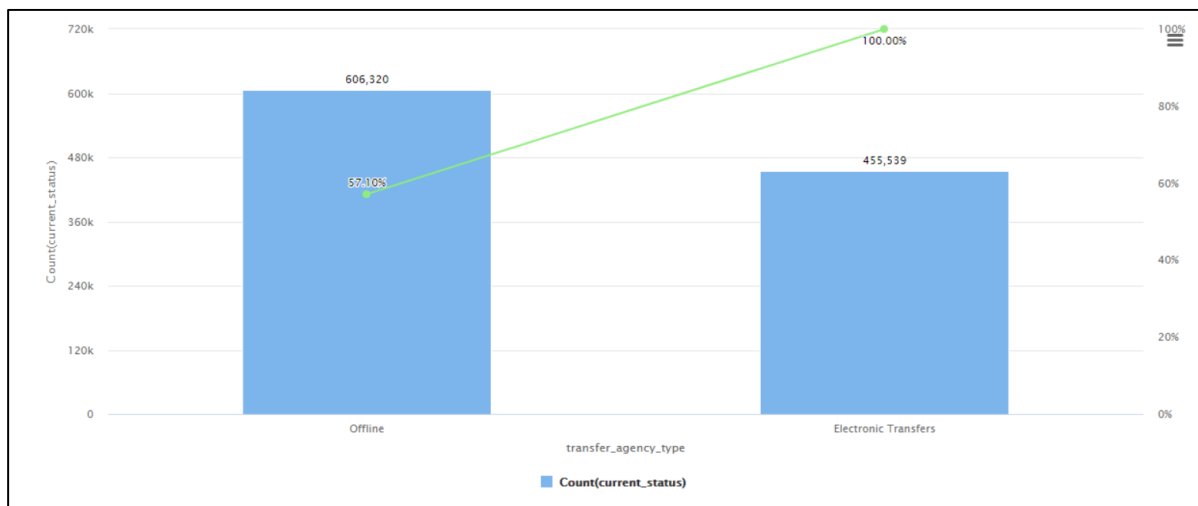
Figura 15*Pareto District*

Nota: Esta figura muestra el comportamiento actual de la variable district.

En la Figura 15 podemos observar que el 36% de los registros de transferencia corresponden a hogares del distrito de Machinga, mientras que el 35% corresponden a hogares del distrito de Mchinji y finalmente, el 29% corresponden a hogares del distrito de Balaka.

Transfer agency type

Esta variable representa al tipo de agencia donde se realiza el pago.

Figura 16*Pareto Transfer agency type*

Nota: Esta figura muestra el comportamiento actual de la variable transfer agency type.

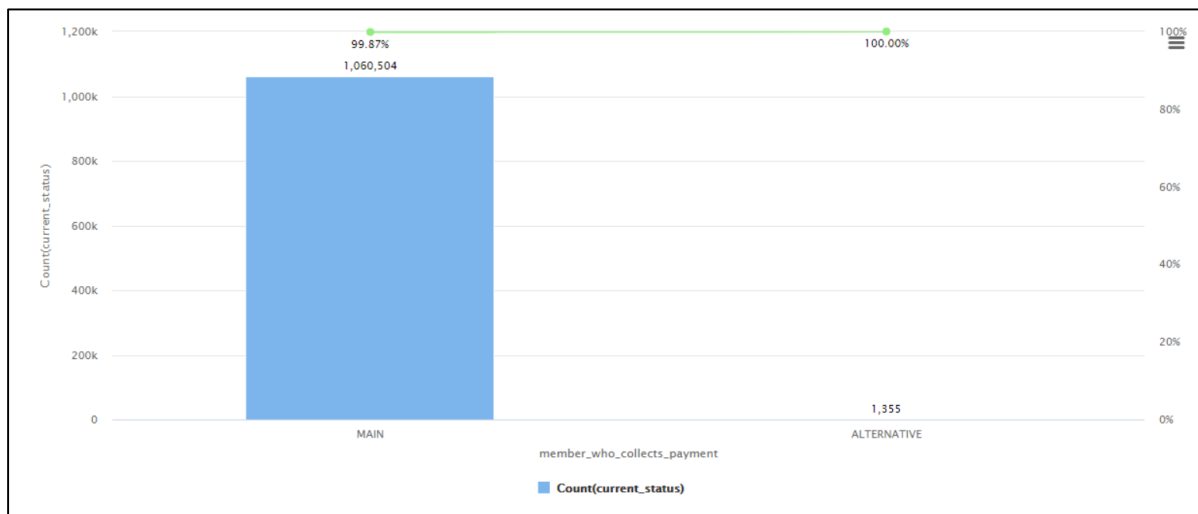
Según la Figura 16, del total de registros de transferencia, el 57% son transferencia offline es decir se entrega el dinero en efectivo, mientras que el 43% corresponden a transferencias electrónicas.

Member who collects the payments

Esta variable indica si el miembro del hogar que cobra el beneficio es el principal o el alterno.

Figura 17

Pareto Member who collects the payments



Nota: Esta figura muestra el comportamiento actual de la variable member who collects the payments.

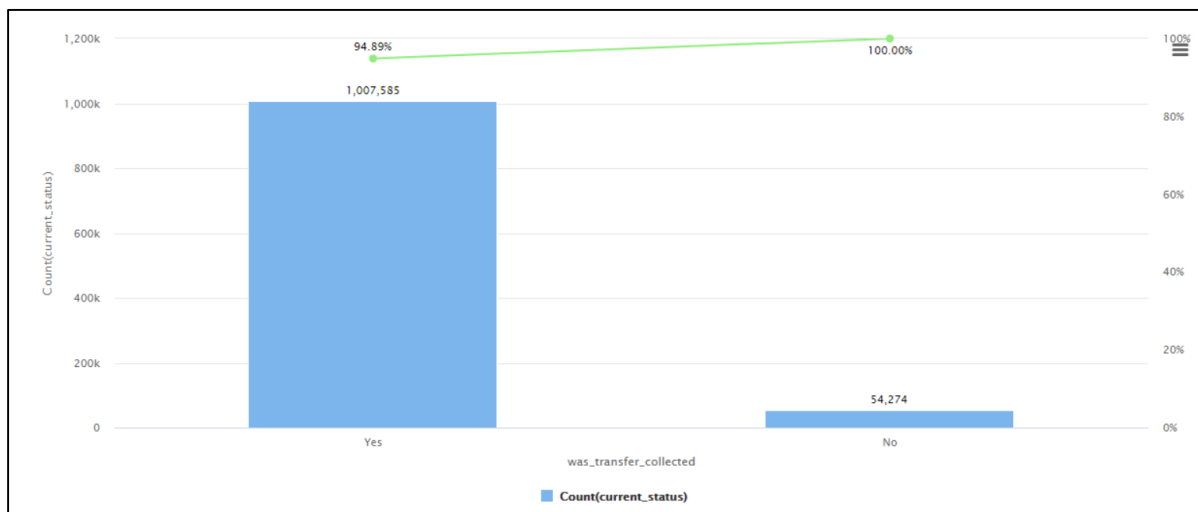
En la Figura 17 podemos observar que el 99.87% de los registros de transferencia están asociados a receptores principales, mientras que un 0.13% de los registros están asociados a receptores alternativos.

Was transfer collected

Esta variable indica si el beneficio fue cobrado o no por el receptor de la transferencia.

Figura 18

Pareto Was transfer collected



Nota: Esta figura muestra el comportamiento actual de la variable was transfer collected.

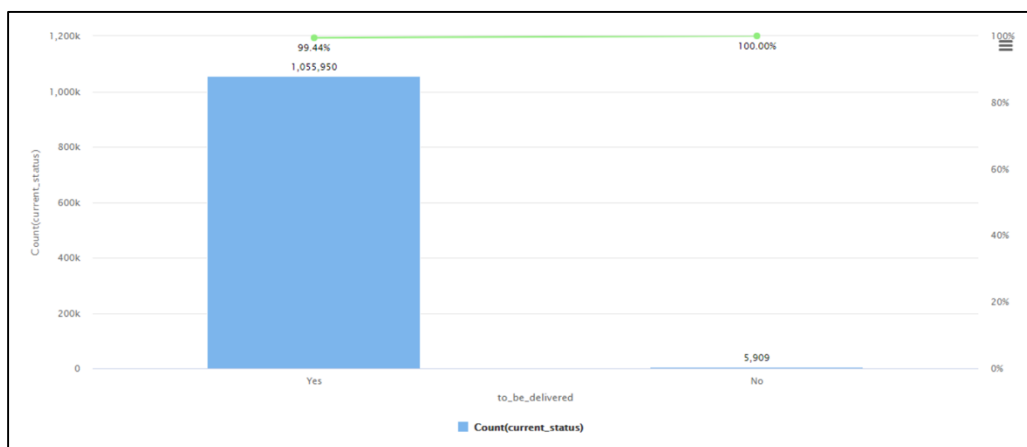
En la Figura 18 podemos observar que el 95% de las transferencias si han sido colectadas, mientras que un 5% no fue colectado.

To be delivered

Esta variable indica si la transferencia de un período específico debía o no ser entregada al hogar.

Figura 19

Pareto To be delivered



Nota: Esta figura muestra el comportamiento actual de la variable to be delivered.

El 99% de las transferencias si debían ser entregadas en el tiempo correspondiente (ver Figura 19), y únicamente 1% no debía ser transferido por algún motivo en específico, como por ejemplo la suspensión del beneficiario para recibir la transferencia.

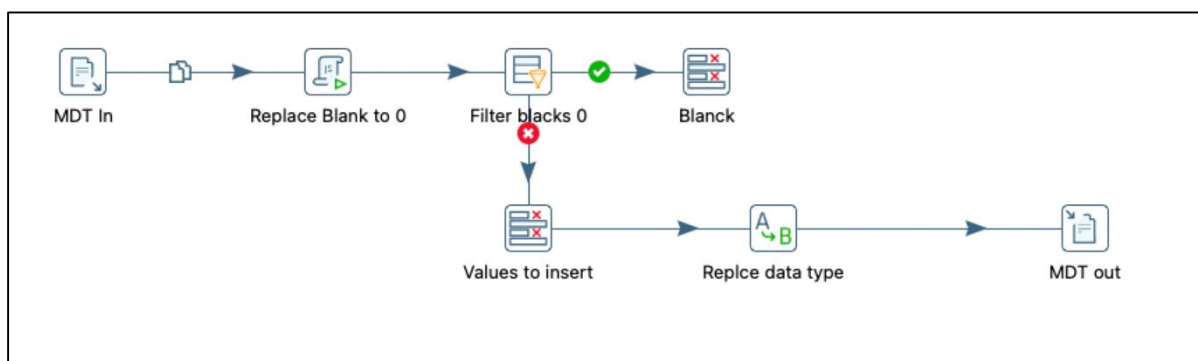
Proceso de limpieza depuración y asignación de valores numéricos

Dado que las tablas presentan únicamente datos demográficos, se procedió a transformar los valores de estos, asignándoles pesos a cada una de las variables. En la Figura 20. Estandarización y calidad de datos, se presenta el proceso realizado en la herramienta Pentaho con la finalidad de cambiar estos valores por un equivalente numérico; iniciando el proceso filtrando todos los valores nulos ya que podrían causar un sesgo innecesario en la clasificación, Finalmente, se reemplazaron los datos categóricos por numéricos obteniendo un total de 553,548 registros. En la Tabla 8. Asignación de valores a variables categóricas, se

encuentran las variables y los pesos asignados con los que se reemplazaron los valores mencionados anteriormente.

Figura 20

Estandarización y calidad de datos



Nota: Esta figura muestra el proceso de estandarización y calidad de datos efectuado a través de la herramienta Spoon.

Tabla 8

Asignación de valores a variables categóricas

Variable Categorica	Valor Asignado	Observación
Distrito Balaka	1 = Distrito más lejano y de difícil acceso	
Distrito Machinga	2 = Distrito más lejano.	District – distrito o localización actual /3
Distrito Mchinji	3 = Distritito que se encuentra en las mejores condiciones para los beneficiarios.	Se clasifica desde el peor al mejor

Nursery	1	Member collector attend education level school /6 Se califica desde el nivel más bajo de instrucción hasta el más alto. El "None" representa ninguna instrucción educativa por tanto se categoriza como 0.
Pre-school	2	
Primary	3	
Secondary	4	
University Training	5	
College	6	
None	0	
Chronic malaria	1	Member collector chronic illness Se califica como 1 o 0 al no poder determinar su gravedad exacta El "None" indica que los beneficiarios no presentan ningún tipo de chronic illness por tanto se categoriza como 0
TB	2	
HIV / AIDS	3	
Asthma	4	
Arthritis	5	
Epilepsy	6	
Other	7	
None	0	
Blind	1	Member collector disability Se califica como 1 o 0 al no poder determinar su gravedad exacta. El "Not disabled" indica que los beneficiarios no presentan ningún tipo de discapacidad por tanto se categoriza como 0
Deaf	2	
Speech impairment	3	
Deformed limbs	4	
Mentally disabled	5	
Other	6	
Not disabled	0	
		Member collector gender
Male	0	Se categoriza como mujer 1 y 0 a los hombres
Female	1	

Married	3 = Familia completa	
Widow(er)	2 = Viudos que perdieron parte de su núcleo familiar	
Never married	1= Si nunca ha formado una familia	Member collector marital status/ 0-3 Se lo califica en función del núcleo familiar.
Separated	0 = Aquellos que no presentan estabilidad familiar	
Divorced		
Head	3 = primer grado de consanguinidad con el jefe del hogar	
Spouce		
Own child		
Brother/Sister	2 = segundo grado de consanguinidad con el jefe del hogar	Member collector relationship/ 0-3
Grand child		
Other relative	1 = tercer u otro grado de consanguinidad con el jefe del hogar	Se lo califica en función del grado de consanguinidad.
Non-relative	0 = miembros que no tienen relación de consanguinidad con el jefe del hogar	
Main	2	Member who collects the payments
Alternative	1	Se califica como 2 si el beneficio lo cobra el titular

		To be delivered
Yes	1	
No	0	Se califica como 1 si debía cobrar el beneficio en un tiempo determinado
Offline	1	Transfer agency type / medio de cobro
Electronic Transfers	2	Se penaliza offline ya que consume más recursos por cobro físico.
Yes	1	Was transfer collected / cobro o no el beneficio
No	0	Si cobró se asigna un valor de 1

Nota. Esta tabla muestra los valores numéricos asignados a las variables categóricas.

Proceso de agrupamiento de datos (Data clustering)

Para el proceso de data clustering se dividió el conjunto de datos en dos: **i)** aquellos que cobran (90% de los valores del conjunto de datos) y **ii)** aquellos que no cobran el beneficio (10% de los registros totales); esto permitirá identificar los perfiles relacionados a los dos subconjuntos de datos, de esta manera en un futuro se podrá en función de los pesos que deriven de esta clasificación, determinar un modelo que permita predecir si un beneficiario cobra o no el beneficio, para este proceso se utilizó una vez más la herramienta Rapid Minner conjuntamente con el algoritmo K-means.

Variables aceptadas por Rapid Minner para iniciar la clasificación

En ambos casos tanto para el que cobra como para el que no cobra, la herramienta nos sugirió que descartemos las variables member_collector_relationship, member_who_collects_payment, was_transfer_collect y to_be_delivered (ver Figura 21),

debido a su alta estabilidad, en otras palabras, estas variables tienen una tendencia que supera el 90% de tomar un solo valor, por lo que no es representativo para realizar una clasificación.

La variable `was_transfer_collect` se descarta ya que previamente dividimos el conjunto de datos en 2.

Figura 21

Variables descartadas para la clasificación

Selected	Status ↑	Quality	Name	Correlation	ID-ness	Stability	Missing	Text-ness
<input checked="" type="checkbox"/>	●		<code>member_collector_relationship</code> ?		0.00%	97.66%	0.00%	0.00%
<input type="checkbox"/>	●		<code>member_who_collects_payme...</code> ?		0.00%	99.97%	0.00%	0.00%
<input type="checkbox"/>	●		<code>was_transfer_collected</code> ?		0.00%	100.00%	0.00%	0.00%
<input type="checkbox"/>	●		<code>to_be_delivered</code> ?		0.00%	100.00%	0.00%	0.00%

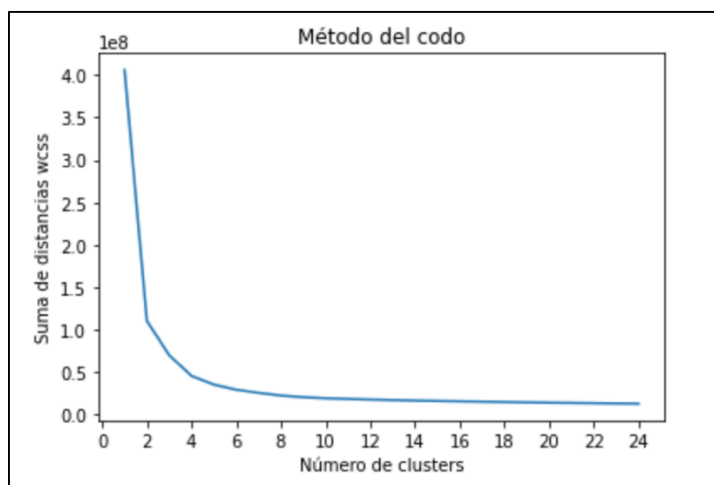
Nota: Esta figura muestra las variables sugeridas para ser descartadas como resultado del proceso ejecutado en la herramienta Rapid Minner.

Método del codo para determinar el número de clusters idóneo

Utilizando la librería Sklearn de Python, se determinó para ambos conjuntos de datos cual es el número correcto de agrupaciones que se deberían realizar para la clusterización, se probaron desde 1 a 25 clasificaciones, dando como resultado que cuatro es un número razonable para clasificar estos datos (ver Figura 22. Diagrama método del codo).

Figura 22

Diagrama método del codo



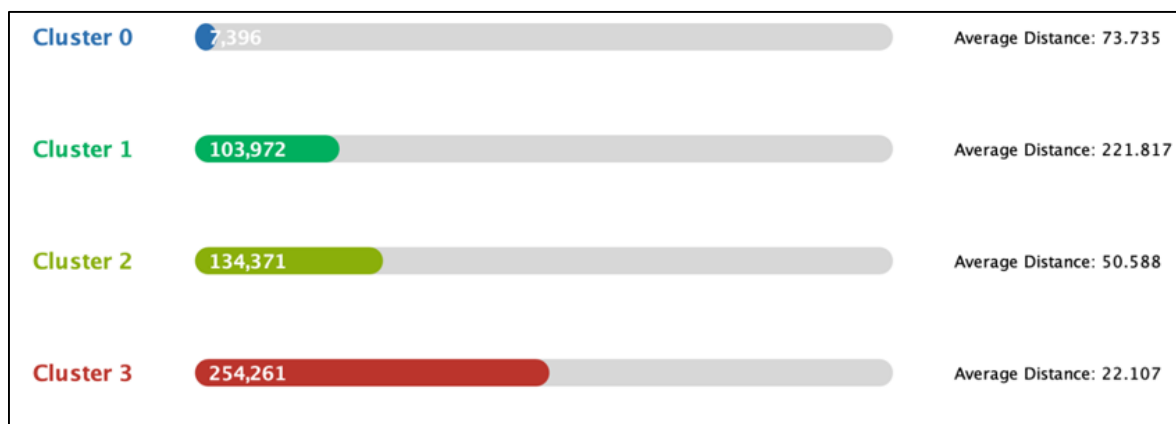
Nota: Esta figura muestra el número idóneo de clústeres en los que se deberán clasificar los datos, utilizando el método del codo.

Clústeres para familias que cobran el beneficio

Utilizando el algoritmo kmeans con cuatro clústeres, se obtuvieron las clasificaciones detalladas en la Figura 23. Clústeres para familias que cobran el beneficio, donde encontramos que la concentración de registros se distribuye de mayor a menor iniciando con el clúster 3 hasta el clúster 0.

Figura 23

Clústeres para familias que cobran el beneficio



Nota: Esta figura muestra los resultados de la agrupación de datos de las familias que si cobran el beneficio como resultado del proceso ejecutado en la herramienta Rapid Minner.

En la Figura 24, se presenta la tabla de los centroides para cada variable según los clústeres especificados anteriormente, para este algoritmo; posteriormente con la data separada por clústeres se determinará el perfilamiento de los datos utilizando un proceso de data profiling.

Figura 24

Centroides para clústeres de familias que cobran el beneficio

Cluster	ge... ↑	member_collector_age	member_collector_att...	member_collector_chroni...	member_collector_disability	member_collector_gender	member_collec...	transfer...
Cluster 0	1.534	28.937	5.963	7.107	6.555	1.641	4.355	1.582
Cluster 2	2.002	49.470	3.017	6.102	6.391	1.985	3.094	1.503
Cluster 3	2.059	59.439	6.998	6.275	6.183	1.976	3.253	1.376
Cluster 1	2.141	62.687	4.789	6.061	5.934	1.000	1.602	1.455

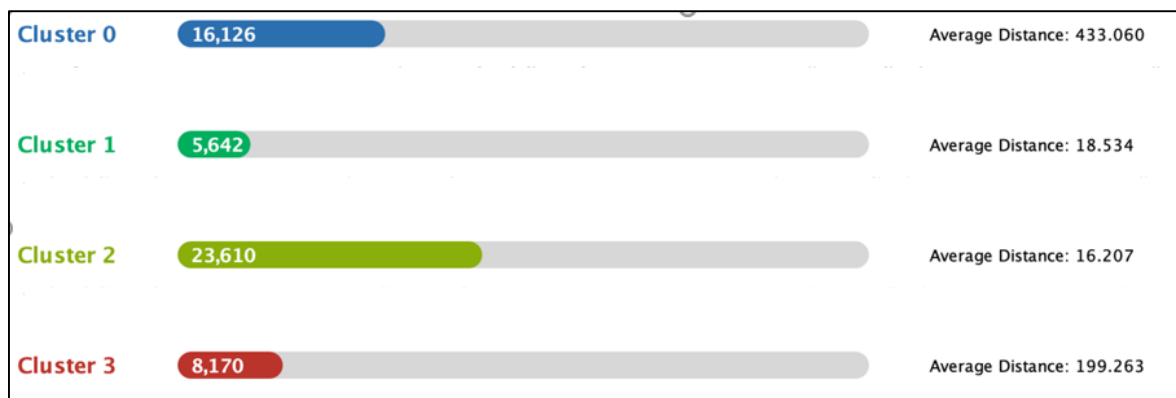
Nota: Esta figura muestra los centroides de los clústeres de las familias que si cobran el beneficio como resultado del proceso ejecutado en la herramienta Rapid Minner.

Clústeres para familias que no cobran el beneficio

Se repitió el ejercicio para la segunda parte de los datos, donde se encontró que a diferencia del caso anterior la mayor concentración de datos está en el clúster 2, seguido del clúster 0, el clúster 3 y finalmente el clúster 1 (ver Figura 25. Clústeres para familias que no cobran el beneficio).

Figura 25

Clusters para familias que no cobran el beneficio



Nota: Esta figura muestra los resultados de la agrupación de datos de las familias que no cobran el beneficio como resultado del proceso ejecutado en la herramienta Rapid Miner.

En la Figura 26, se presenta la tabla de centroides de los clusters anteriormente enunciados, posteriormente con la data clasificada se procederá a analizar los resultados.

Figura 26

Centroides para clústeres de familias que no cobran el beneficio

Cluster	geo_distr...	member_collector_age	member_collector_atten...	member_collector_chron...	member_collector_disabili...	member_collector_gender	member_collector_marital_stat...	transfer_ag...
Cluster 0	1.002	54.940	5.107	6.064	6.133	1.757	2.913	2
Cluster 1	1.517	56.991	5.235	6.062	6.066	1.719	2.944	1.966
Cluster 2	2.828	58.660	5.589	6.249	6.209	1.971	3.335	1.001
Cluster 3	2.872	63.921	4.834	5.986	5.917	1.001	1.732	1.001

Nota: Esta figura muestra los centroides de los clústeres de las familias que no cobran el beneficio como resultado del proceso ejecutado en la herramienta Rapid Minner.

Capítulo V: Análisis y Presentación de Resultados

En el capítulo anterior se realizó un análisis de variables, se efectuó una primera discriminación de importancia y finalmente se llevó a cabo un ejercicio de clustering utilizando el algoritmo K-means donde se estableció para cada caso (cliente cobra o cliente que no cobra el beneficio) 4 grupos. A continuación, se presentará el análisis detallado del perfilamiento más acertado para los beneficiarios del programa Mtukula Pakhomo.

Selección de Clusters más representativos para los grupos (cobra o no cobra el beneficio)

No cobra el beneficio

Dentro de los 4 grupos obtenidos por el algoritmo k-means se obtuvo que la concentración más alta se formaba en el cluster #2 seguido del #0. En la Tabla 9. Concentración de beneficiarios por clusters (no cobra el beneficio), se detalla la concentración porcentual y su representación. Dado que la concentración más grande se encuentra en los clusters #0 y # 2 con un 30% y un 44.1% respectivamente, se tomará a estos dos para determinar cuál es la definición del beneficiario que posiblemente no cobre el beneficio. Por otro lado, el resto de los resultados se descartan por su baja representatividad inferior al 20%.

Tabla 9

Concentración de beneficiarios por clúster (no cobra el beneficio)

Cluster	Total Beneficiarios	Representatividad
0	16,126	30%
1	5,642	10%
2	23,610	44.1%
3	8,170	15.9%

Nota. Esta tabla muestra la concentración de beneficiarios en los clústeres identificados para los beneficiarios que no cobran el beneficio.

Cobra el beneficio

De los 4 grupos obtenidos anteriormente se determina que la concentración más alta se formaba en el cluster #3, seguido del #2 y así sucesivamente, en la Tabla 10. Concentración de beneficiarios por clusters (si cobra el beneficio), se detalla la concentración porcentual y su representación. Donde encontramos que más de la mitad de los beneficiarios se concentran en el cluster #3 por lo que se decide focalizar el análisis de las características de este grupo para el perfilamiento; consecuentemente, los clusters #2 y #1 representan en conjunto la segunda mayor concentración, pero dado que individualmente representan menos del 30% de la representatividad total, se decide excluirlos del siguiente análisis.

Tabla 10*Concentración de beneficiarios por clusters*

Cluster	Total Beneficiarios	Representatividad
0	7,396	1.4%
1	103,972	20.8%
2	134,371	26.9%
3	254,261	50.9%

Nota. Esta tabla muestra la concentración de beneficiarios en los clústeres identificados para los beneficiarios que si cobran el beneficio.

Comparación de características no discriminatorias

Luego de analizar los 3 clusters para el beneficiario que cobra y que no cobra se tiene que variables como edad, genero, discapacidad y nivel de instrucción académica no representan diferencia predominante en ninguno de los clusters, en la Tabla 11. Variables no discriminantes, se presenta este ejercicio; donde, para la variable edad, en los 3 clusters se evidencia que el 75% de las personas van de 8 a 70 años, en la variable género predomina las mujeres, en la variable discapacidad no hay diferencia ya que en todos los casos predominan aquellas personas que no son discapacitadas y en la variable nivel de instrucción académica en todos los clusters predominan beneficiarios que no han recibido ninguna instrucción escolar, por lo que estas variables no se consideran en el perfilamiento.

Tabla 11*Variables no discriminantes*

Cluster	Q3-si	Q0-no	Q2-no
Edad	8-70 años	9-70 años	10-65 años
Genero	75%-mujeres	72%-mujeres	80%-mujeres
collector_disability	Not disabled – 75%	Not disabled – 72%	Not disabled – 77%
member_collector_attendance	None-52%	None-65%	None-60%

Nota. Esta tabla muestra las variables que han sido consideradas como no discriminantes y el resultado obtenido para cada una en cada uno de los clústeres.

Perfilamiento Cluster NO cobra el beneficio

Para los cluster Q0 y Q2 existen datos muy similares en la mayoría de las variables como se puede apreciar en los centroides de la Figura 26. Por lo tanto, las variables que determinan el perfil de una persona que no cobra el beneficio y su porcentaje de representatividad sobre el total de las observaciones en el modelo son las siguientes:

- member_collector_marital_status: Widower 58%, otras opciones 42%
- member_collector_chronical_illnesses: other (otra fuera de la lista) 63%, otras opciones 37%
- geo_district: Balaka 99%, Machinga y Mchinji 1%
- transfer_agency_type: electrónica 99%, manual 1%

Entonces, se puede afirmar que las personas que viven en el Distrito Balaka, que son viudas, que tienen una enfermedad crónica no especificada y que tienen asignado un método electrónico de pago, no cobran de manera regular el beneficio otorgado por el programa.

Adicionalmente, si tomamos también como referencia las variables no discriminantes más representativas del modelo “género” y “nivel de instrucción académica”, podemos decir que las personas que no cobran de manera regular el beneficio otorgado por el programa son mujeres viudas del distrito Balaka, que no tienen un nivel de instrucción académica, que presentan una enfermedad crónica no especificada y que tienen asignado un método electrónico de pago.

Perfilamiento Cluster SI cobra el beneficio

Como se puede ver en los centroides de la Figura 24, la mayor concentración de los datos (beneficiarios) se encuentran en casi todas las variables del cluster Q3. El análisis realizado nos indica que las variables que determinan el perfil de una persona que si cobra el beneficio y su porcentaje de representatividad sobre el total de observaciones en el modelo son las siguientes:

- member_collector_marital_status: Married 57%, Separated 35%. Otras opciones 8%
- member_collector_chronical_illnesses: None 98%, otras opciones 2%
- geo_district: Machinga 42%, Mchinji 47%, Balaka 11%
- transfer_agency_type: manual 99%, electronica 1%

Entonces, se puede afirmar que las personas que viven en el Distrito Machinga o Mchinji, que son married o separated, que no tienen una enfermedad crónica y que tienen asignado un método manual de pago, si cobran de manera regular el beneficio otorgado por el programa.

Presentación de resultados a los responsables de la implementación del programa

Mtukula Pakhomo

Los resultados fueron expuestos mediante correo electrónico y llamada telefónica con el director general del programa y los stakeholders involucrados. Se explicó a través de una tabla resumen, las variables que fueron analizadas durante el proceso, las seleccionadas inicialmente, aquellas que fueron descartadas por sugerencia de la herramienta Rapid Minner previo a la definición de clusters, las que fueron descartadas por no ser discriminantes entre clusters y finalmente aquellas que determinan el perfil de los beneficiarios que si cobran y que no cobran el beneficio (ver tabla 12).

Tabla 12

Resumen de Variables utilizadas durante el proceso ejecutado

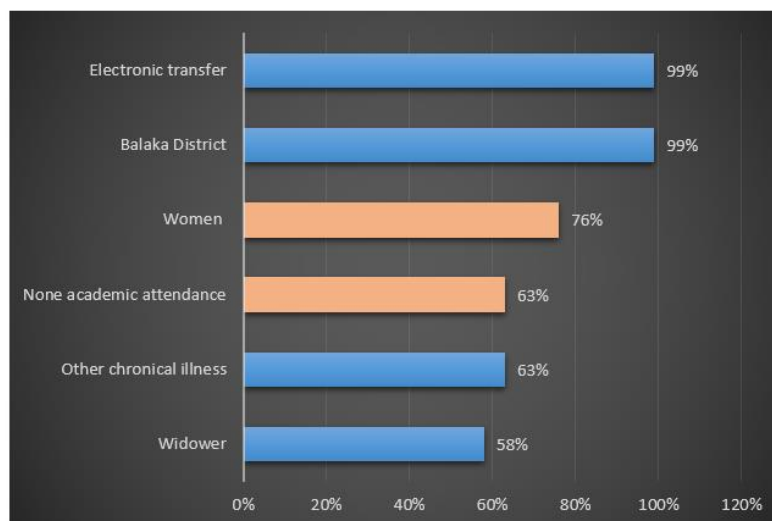
Variables (Seleccionadas para el análisis)	Descartadas (sugerencia de Rapid Minner)	Descartadas (no discriminantes)	Utilizadas (determinan el perfil)
Member_collector_age		X	
Member_collector_gender		X	
Member_collector_relationship	X		
Member_collector_marital_status			X
Member_collector_disability		X	
Member_collector_chronical_illness			X
Member_collector_attend_education		X	
Geo_district			X
Transfer_agency_type			X
Member_who_collects_payment	X		
Was_transfer_collected	X		
To_be_delivered	X		

Nota. Esta tabla muestra un resumen de las variables seleccionadas para el análisis, variables descartadas y variables utilizadas.

También se presentó las gráficas en la que se muestra el perfil de los beneficiarios que no cobran regularmente el beneficio económico otorgado por el programa Mtukula Pakhomo (ver figura 27) y el perfil de los beneficiarios que si cobran el beneficio (ver figura 28); en la primera grafica se muestra en color naranja las variables no discriminantes más relevantes debido a que aportan valor al presente análisis.

Figura 27

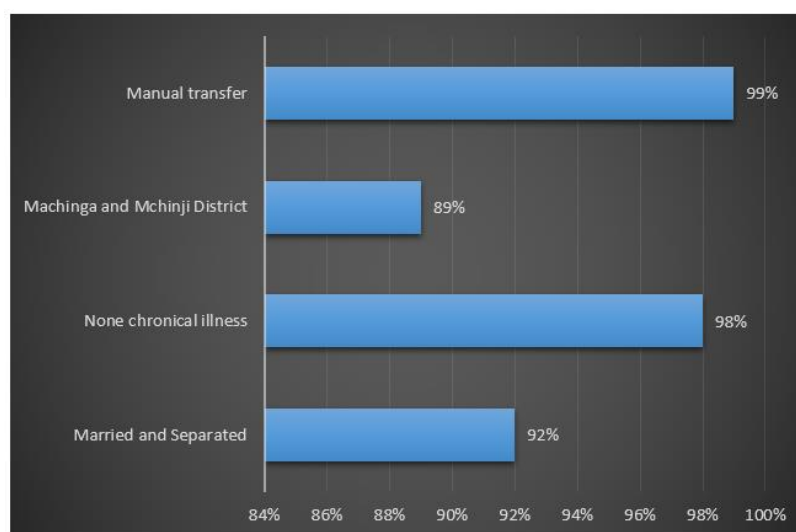
Perfil de los beneficiarios que no cobran



Nota: Esta figura muestra el perfil o características de los beneficiarios que no cobran el beneficio.

Figura 28

Perfil de los beneficiarios que si cobran



Nota: Esta figura muestra el perfil o características de los beneficiarios que si cobran el beneficio.

Análisis concluyente de los resultados obtenidos:

Operativamente, no es factible un pago directo en efectivo (manual) para aquellos beneficiarios que son vulnerables, es decir que tienen enfermedades y/o que son adultos mayores, porque se les dificulta acudir al lugar céntrico donde se efectúan los pagos; por este motivo las autoridades de los programas sociales han buscado soluciones de pago que permita que todos los beneficiarios reciban el beneficio económico. Sin embargo, no siempre la implementación de tecnología (pagos electrónicos) es la mejor solución.

Los pagos electrónicos en el Programa Mtukula Pakhomo son una solución implementada a través del uso de celulares con SIM card, el beneficiario recibe el dinero en el celular y debe realizar pagos en tiendas para comprar comida principalmente. Los beneficiarios que no cobran el beneficio regularmente no saben leer ni escribir y si a eso se le suma las enfermedades crónicas no especificadas que poseen, y su edad avanzada, se tiene un cuadro de beneficiarios con capacidad limitada para entender el sistema de pago. Adicionalmente si se toma en cuenta el estado civil de este tipo de beneficiarios “viudos” se puede decir que son personas que están solas y no cuentan con el apoyo de un familiar que les ayude a utilizar la tecnología para hacer uso del dinero que el programa les otorga.

Este tipo de beneficiarios necesita controlar su dinero, verlo físicamente. La no familiarización con la tecnología los lleva a pagar una comisión para que una tercera persona les ayude a realizar las transacciones, exponiéndose a ser víctimas de estafa.

Las implementaciones de este tipo de sistemas de pago resuelven problemas logísticos de la entrega del beneficio en efectivo a nivel de gobiernos es decir de las autoridades del

programa, pero por otro lado afecta a este tipo de beneficiarios que por sus condiciones no usan la tecnología.

Se requiere un análisis a fondo para determinar todas las causas por la que los beneficiarios no cobran, siempre se podría pensar que hace falta una verdadera campaña de comunicación efectiva que llegue a todos los lugares incluso a los más remotos, pero según el perfil obtenido se debería prestar atención también a la accesibilidad del Distrito Balaka; ¿es un lugar muy remoto?; ¿llegan los medios de comunicación?, ¿tienen acceso a internet?

Se concluye que los sistemas de pagos electrónicos no son viables para los beneficiarios que son adultos mayores, que tienen alguna enfermedad crónica, que no tienen un nivel de instrucción académica o que viven en lugares que no cuentan con buena conexión a internet.

Recomendación que permitirá mejorar la frecuencia de recepción por parte de los beneficiarios que actualmente lo cobran de manera irregular o no lo cobran.

La recomendación general va a nivel de las autoridades gubernamentales, estas deberían plantear ajustes en la política con el fin de implementar sistemas de pago que faciliten a los beneficiarios el uso del beneficio económico otorgado por el programa, principalmente para este tipo de beneficiarios que no cobran y/o usan regularmente su dinero.

La entrega de dinero en efectivo sería la mejor opción, porque como se puede evidenciar en el análisis del perfil de los beneficiarios que, si cobran el beneficio, el 99% de ellos tienen asignado el mecanismo manual de pago. Sin embargo, el costo de logística

incrementaría radicalmente para el programa, ya que gente del programa debería acudir al hogar para entregar el beneficio económico en manos de los beneficiarios.

Por otro lado, si el programa busca la innovación y la inmersión de la tecnología durante su implementación, debería pensar en un proceso progresivo flexible para pasar de un sistema de pagos manual a un sistema de pagos electrónico; siempre dejando la opción al beneficiario de mantenerse en pago manual si al final de haber sido capacitado y/o entrenado decide quedarse en pago manual porque no entiende la tecnología. Esto implicaría la inversión de recursos en temas de capacitación y entrenamiento.

En los dos escenarios existe un incremento de costos. Operativamente hablando, la aplicación de estas recomendaciones dependerá de los gobiernos y los organismos multilaterales que financian este tipo de programas sociales. El costo beneficio que representen deberá ser analizado tomando en cuenta la situación de vida de los beneficiarios.

El director general del programa, basado en su experiencia en implementación de proyectos sociales, siempre ha manifestado la teoría de que los métodos de pago electrónicos no funcionan para todos los beneficiarios de un programa; la aplicación de esta solución y los resultados obtenidos confirman dicha aseveración.

Con respecto a las recomendaciones emitidas, indicó que la aplicación de estas, si mejorarán la frecuencia de recepción del beneficio y se estima que del 5% de beneficiarios que no cobran regularmente el beneficio, el 4% lo hará si el pago es en efectivo. Siempre la entrega del dinero en efectivo directamente en las manos del beneficiario y en su hogar es la mejor opción para los beneficiarios con el perfil obtenido, ya que tienen el control del dinero y pueden utilizarlo de acuerdo a sus necesidades. el 1% restante es considerado como un margen de

eventualidades adversas que pueden impedir al beneficiario no cobrar regularmente el beneficio, como por ejemplo una calamidad domestica no prevista.

Capítulo VI: Conclusiones y Recomendaciones

Este capítulo detalla las conclusiones y recomendaciones que son el resultado del desarrollo de la presente tesis.

Conclusiones

Al efectuar el análisis de la situación actual del programa Mtukula Pakhomo, a través de la revisión de la información y la entrevista al director del proyecto en la consultora contratada, se evidencia el cobro irregular del beneficio por un grupo de beneficiarios, lo que refleja una implementación del proceso de pagos ineficiente; se identificaron como causas del problema la falta de comunicación o comunicación inadecuada por parte del programa para dar a conocer las fechas de pago, desconocimiento de los beneficiarios acerca de las fechas de pago y el impedimento que tienen los beneficiarios para acercarse a cobrar el beneficio, lo que desemboca en un efecto inmediato en el bienestar de las familias pobres.

En la revisión de la literatura pre-liminar se encontraron trabajos referentes al uso de herramientas tecnológicas y técnicas de agrupamiento en temas de identificación de perfiles o patrones de comportamiento, como son las técnicas de aprendizaje no supervisado con la aplicación del algoritmo K-means y las herramientas de inteligencia de negocios y Data Mining, Pentaho Data Integration y Rapidminer respectivamente, mismas que sirvieron para la construcción de la solución propuesta en la presente investigación por analizar grandes cantidades de datos, transformarlos y utilizarlos de manera eficiente. Estas herramientas fueron seleccionadas porque son las que dieron mejores resultados en los estudios analizados.

La información de los beneficiarios es almacenada en una base de datos que tiene información de cada uno de los procesos que conforman el ciclo del proyecto Identificación, Registro, Condicionalidades, Pagos, Manejo de Casos, Indicadores. Para cumplir el propósito del presente trabajo de titulación se ha trabajado con información del proceso de Registro ya que es información actualizada con respecto a la información del proceso de identificación, e información del proceso de pagos misma que indica el cobro o no cobro del beneficio. Para efectuar el ETL de la información mencionada se utilizó la herramienta Pentaho Data Integration con su componente Spoon.

Para identificar los perfiles de los beneficiarios que no cobran y que si cobran regularmente el beneficio del programa se ha utilizado la metodología CRISP-DM conjuntamente con la herramienta Rapidminer, lo que ha dado como resultado nuevos conocimientos útiles que permitirán una mejor toma de decisiones durante la implementación del proceso de pagos de proyectos de beneficio social en Malawi. El diagrama de la solución propuesta está conformado por 6 componentes, mismos que están contenidos en las fases de la metodología aplicada: las fuentes de información, el proceso de extracción, transformación y carga, la creación de una Master Data Table, el data profiling, el data clustering y el data pattern-profiling.

El Data profiling realizado permitió identificar el comportamiento actual de las variables obtenidas y priorizarlas; de esta manera fue más sencillo discriminar aquellas variables que no aportan en el modelamiento por su nivel de granularidad (ej. geolocation a nivel de village, zone) o por algún sesgo encontrado durante el análisis (registros sin información para las variables representativas fueron eliminados).

Los resultados obtenidos durante el desarrollo de la solución son factibles ya que brindan conformidad a los interesados, puesto que permiten evaluar la situación y tomar decisiones que mejoren la recepción del beneficio, principalmente en aquellos con el perfil que no cobran regularmente el mismo.

Si el programa Mtukula Pakhomo implementa únicamente el método de pago manual para la entrega del beneficio económico directamente en el hogar del beneficiario que no lo cobra regularmente, se estima que el porcentaje de beneficiarios que no cobran de manera regular disminuiría del 5% al 1%. Se considera un 1% de beneficiarios que no cobren el beneficio regularmente porque pueden existir eventos inesperados que ocasionen el no cobro del beneficio.

Recomendaciones

Se recomienda la implementación de esta solución, la cual permitirá a los responsables del programa, analizar la información a nivel nacional de manera ágil, para identificar el perfil de los beneficiarios que no cobran de manera regular el beneficio económico y tomar decisiones que ayuden a mejorar la eficiencia del proceso de pagos. Esta solución podría aplicarse a futuros programas

Es importante realizar algunas mejoras al proceso de recolección de la información y la clasificación de variables como por ejemplo en el caso de la variable `member_collector_chronical_illnesses` (enfermedades crónicas), ya que, en la mayoría de los casos, existen muchas personas que no tienen especificada el tipo de enfermedad crónica que afecta su diario vivir.

Se sugiere que para efectuar la clusterización de los datos, las variables categóricas sean transformadas a numéricas (se cambia el valor categórico por un numérico, ejemplo: Mujer = 1), ya que las herramientas de modelado y clasificación trabajan mejor con variables numéricas. Esto es un proceso común que se efectúa en la analítica avanzada de datos.

Para futuras aplicaciones de la solución, se sugiere incluir un nivel más bajo de ubicación geográfica en los datos a analizar, el resultado será más amplio y el número de variables no discriminantes podría ser menor.

Al conocer el perfil de los beneficiarios que no cobran regularmente el beneficio, se recomienda que la entrega del mismo, para este tipo de personas, sea a través del método de pago manual directamente en el hogar del beneficiario, de esta manera el objetivo de la implementación del proyecto se cumpliría ya que los beneficiarios no retomarían conductas negativas anteriores como mendicidad, venta de bienes, trabajo infantil entre otros, y se mantendría así la estabilidad y maximización de los logros alcanzados por el impacto del proyecto en los beneficiarios.

Se sugiere la aplicación de las recomendaciones propuestas en base a los resultados del modelo planteado, ya que de esta manera el cobro del beneficio por parte de los beneficiarios que no lo cobran regularmente mejoraría en un 99%, (es decir, del 5% que no cobrara el beneficio, el 4% lo hará). El 1% se maneja como margen de error debido a que pueden existir causas de fuerza mayor que impidan que el beneficiario no reciba el beneficio de forma regular.

Bibliografía

- Ballard, Abdel-Hamid, A., Frankus, R., Hasegawa, F., Larrechart, J., Leo, P., & Ramos, J. (2006, Agosto). *IBM*. Retrieved from Improving Business Performance Insight with Business Intelligence and Business Process Management:
<http://www.redbooks.ibm.com/redbooks/pdfs/sg247210.pdf>
- Banco Mundial. (2018, Abril 05). *Banco Mundial*. Retrieved from Entendiendo la Pobreza, Protección Social: <https://www.bancomundial.org/es/topic/socialprotection/overview>
- Chapman, P., Clinton, J., Keber, R., Khabaza, T., Reinartz, T., Shearer, C., & Wirth, R. (1999). *CRISPDM*. Retrieved from CRISP-DM 1.0 Step by Step BI guide:
http://www.crispdm_org/CRISPWP-0800.pdf.
- Colmenares, R. (2011, Enero 6). Retrieved from Sistema de Información:
<http://ronaldcolmenares.blogspot.com/2011/01/inteligencia-de-negocios.html>
- Edem, I., Benzaïd, C., Al-Nemrat, A., & Watters, P. (2014). Analysis of Malware Behaviour: Using Data Mining Clustering Techniques to Support Forensics Investigation. *2014 Fifth Cybercrime and Trustworthy Computing Conference* (p. 10). Auckland, New Zealand: IEEE.
- FAO. (n.d.). *Organización de las Naciones Unidas para la Alimentación y la Agricultura*. Retrieved from Instrumentos políticos innovadores y evaluación en el desarrollo agrícola y rural en América Latina y el Caribe - Benjamín Davis:
<http://www.fao.org/3/y4940s/y4940s08.htm#TopOfPage>
- Guayasmín, A., Fuertes, W., Campaña, M., & Toulkeridis, T. (2018). Formalistic Modelling Based on Pattern Recognition Applied to the Knowledge and Human Talent Sector in Ecuador. *2018 ICAI Workshops (ICAIW)* (p. 6). Bogota, Colombia: IEEE.

- Gutiérrez, J., & Molina, B. (2016). Identificación de técnicas de minería de datos para apoyar la toma de decisiones en la solución de problemas empresariales. 33-52.
- Hope, G. (2018, Julio 12). *Capterra*. Retrieved from The Top 15 Free and Open Source Business Intelligence Software: <https://blog.capterra.com/top-8-free-and-open-source-business-intelligence-software/>
- Howton, E. (2018, Septiembre 24). *Banco Mundial*. Retrieved from Entendiendo la pobreza: <https://www.bancomundial.org/es/topic/poverty/overview>
- itop Academy. (2019, junio 26). Retrieved from <https://www.itop.academy/blog/item/spoon-componente-pentaho-data-integration-kettle.html#:~:text=Una%20de%20estas%20herramientas%20muy,mandos%2C%20anal%C3%ADticas%20y%20mucho%20m%C3%A1s.>
- Kwesell, A. (2014, Abril 09). Retrieved from Banco Mundial: <http://www.bancomundial.org/es/results/2013/04/14/social-development-results-profile>
- Microsoft. (2019, 01 08). *Microsoft*. Retrieved from Documentación SQL, Conceptos de minería de datos: <https://docs.microsoft.com/es-es/sql/analysis-services/data-mining/data-mining-concepts?view=sql-server-2017>
- Navarro, J. (2017, Febrero). *Definición de Modelo de Gestión*. Retrieved from Definición ABC: <https://www.definicionabc.com/economia/modelo-gestion.php>
- OIT. (n.d.). *Organización Internacional del Trabajo OIT*. Retrieved from Protección Social, Introducción a los instrumentos de políticas: transferencias sociales: <https://www.social-protection.org/gimi/ShowTheme.action?id=11>
- Parr, O. (2000). Retrieved from Data Mining Cookbook Modeling Data for Marketing, Risk, and Customer Relationship Management: <https://books.google.com.co/books?id=L3w0loZrcU0C&printsec=frontcover&dq=Data+Mining+Cookbook#v=one%20page&q=&f=false>

- Peña, A. (2006). *Inteligencia de Negocios: Una Propuesta para su Desarrollo en las organizaciones*. México: Instituto Politécnico Nacional.
- Peng, B., Wan, C., Dong, S., Lin, J., Song, Y., Zhang, Y., & Xiong, J. (2016). A two-stage pattern recognition method for electric customer classification in smart grid. *2016 IEEE International Conference on Smart Grid Communications (SmartGridComm)*. Sydney, NSW, Australia: IEEE.
- Pérez, J. (2008). *Definición de Modelo de Gestión*. Retrieved from Definición.DE: <https://definicion.de/modelo-de-gestion/>
- Pozarny, P., & Davis, B. (2015, Mayo). Efectos de los programas de transferencias monetarias en las dinámicas comunitarias en África subsahariana. *10 years International Policy Centre for Inclusive Growth*.
- Pulla, C. (2011). *Técnicas de minería de datos para identificar patrones de colaboración de los estudiantes que hacen uso del EVA de la UTPL*. Loja - Ecuador: Universidad Técnica Particular de Loja.
- Rapidminer. (2021). *rapidminer documentation*. Retrieved from <https://docs.rapidminer.com/latest/studio/guided/turbo-prep/>
- Recuero, P. (2017, Noviembre 16). *LUCA AI Powered Decisions*. Retrieved from Los 2 tipos de aprendizaje en Machine Learning: supervisado y no supervisado: <https://data-speaks.luca-d3.com/2017/11/que-algoritmo-elegir-en-ml-aprendizaje.html>
- Shawa, M. (2016). *Operation Manual "Mtukula Pakomo Social Cash Transfer Programme"*. Malawi.
- Stackowiak, R., Rayman, J., & Greenwald, R. (2007). Retrieved from Oracle Data Warehousing and Business Intelligence Solutions: https://books.google.com.co/books?id=Gxy6_drRWRgC&dq=%22Oracle+Data+Warehousing+and+Business+Intelligence+Solutions%22&printsec=frontcover&source=bn&hl=e

s&ei=W0uJSqmGsqItgewwtjnDA&sa=X&oi=book_result&ct=result&resnum=4#v=onepage&q=&f=false

Tabor, Peppiatt, M., & Holzmann. (2002). *Analiza la experiencia de transferencias en efectivo en situaciones de emergencia y de no emergencia.*

Unidad de Información y Análisis Financiero. (2014). *Universidad del Rosario*. Retrieved from Técnicas de minería de datos: https://www.urosario.edu.co/observatorio-de-lavado-de-activos/Archivos_Lavados/Tecnicas-de-mineria-de-datos-para-la-prevencion-de.pdf

Valenga, F., Fernández, E., Merlino, H., Rodríguez, D., Procopio, C., Britos, P., & García-Martínez, R. (n.d.). Minería de Datos Aplicada a la Detección de Patrones Delictivos. *Semanticscholar*, 9.