ESPE

UNIVERSIDAD DE LAS FUERZAS ARMADAS

INNOVACIÓN PARA LA EXCELENCIA

VICERRECTORADO DE INVESTIGACIÓN, INNOVACIÓN Y TRANSFERENCIA DE TECNOLÓGÍA

CENTRO DE POSGRADOS

MAESTRÍA EN GESTIÓN DE SISTEMAS DE INFORMACIÓN E INTELIGENCIA DE NEGOCIOS

TRABAJO DE TITULACIÓN PREVIO A LA OBTENCIÓN DEL TÍTULO DE MAGÍSTER EN: GESTIÓN DE SISTEMAS DE INFORMACIÓN E INTELIGENCIA DE NEGOCIOS

MODELO DE DATOS PARA DETERMINAR LOS PATRONES DE COMPORTAMIENTO DE LOS BENEFICIARIOS DEL BONO DE DESARROLLO HUMANO (BDH) Y DE PENSIONES.

AUTOR: ING. SOSA ZÚÑIGA, DAVI D ISMAEL

DIRECTORA: ING.PARRAGA VILLAMAR VIVIANA CRISTINA, MSC

SANGOLQUÍ

2019

VICERRECTORADO DE INVESTIGACIÓN, INNOVACIÓN Y TRANSFERENCIA DE TECNOLOGÍA

CENTRO DE POSGRADOS

CERTIFICACIÓN

Certifico que el trabajo de titulación, "***MODELO DE DATOS PARA DETERMINAR LOS PATRONES DE COMPORTAMIENTO DE LOS BENEFICIARIOS DEL BONO DE DESARROLLO HUMANO (BDH) Y DE PENSIONES***" fue realizado por el señor ***Sosa Zúñiga, David Ismael*** el mismo que ha sido revisado en su totalidad, analizado por la herramienta de verificación de similitud de contenido; por lo tanto cumple con los requisitos teóricos, científicos, técnicos, metodológicos y legales establecidos por la Universidad de Fuerzas Armadas ESPE, razón por la cual me permito acreditar y autorizar para que lo sustente públicamente.

Sangolquí, 20 Mayo 2019

Firma:

ING. VIVIANA CRISTINA PARRAGA VILLAMAR, MSC

C.C:1721903407

UNIVERSIDAD DE LAS FUERZAS ARMADAS
INNOVACIÓN PARA LA EXCELENCIA

VICERRECTORADO DE INVESTIGACIÓN, INNOVACIÓN Y TRANSFERENCIA DE TECNOLOGÍA
CENTRO DE POSGRADOS

AUTORÍA DE RESPONSABILIDAD

Yo, ***Sosa Zúñiga, David Ismael***, con cédula de ciudadanía n° 1721734281, declaro que el contenido, ideas y criterios del trabajo de titulación: ***MODELO DE DATOS PARA DETERMINAR LOS PATRONES DE COMPORTAMIENTO DE LOS BENEFICIARIOS DEL BONO DE DESARROLLO HUMANO (BDH) Y DE PENSIONES*** es de mi autoría y responsabilidad, cumpliendo con los requisitos teóricos, científicos, técnicos, metodológicos y legales establecidos por la Universidad de Fuerzas Armadas ESPE, respetando los derechos intelectuales de terceros y referenciando las citas bibliográficas. Consecuentemente el contenido de la investigación mencionada es veraz.

Sangolquí, 20 Mayo 2019

Firma:

ING. DAVID ISMAEL SOSA ZÚÑIGA
C.C:1721734281

VICERRECTORADO DE INVESTIGACIÓN, INNOVACIÓN Y TRANSFERENCIA DE TECNOLOGÍA
CENTRO DE POSGRADOS

AUTORIZACIÓN

Yo, ***Sosa Zúñiga, David Ismael*** *autorizo a la Universidad de las Fuerzas Armadas ESPE publicar el trabajo de titulación:* ***MODELO DE DATOS PARA DETERMINAR LOS PATRONES DE COMPORTAMIENTO DE LOS BENEFICIARIOS DEL BONO DE DESARROLLO HUMANO (BDH) Y DE PENSIONES*** *en el Repositorio Institucional, cuyo contenido, ideas y criterios son de mi responsabilidad.*

Sangolquí, 20 Mayo 2019

Firma:

ING. DAVID ISMAEL SOSA ZÚÑIGA
C.C:1721734281

DEDICATORIA

A mi seres queridos como mi padre Miguel Sosa, mi madres Mónica Zuñiga, Mama Eulalia, Sra. Bertita mis 3 hermanos (Francia, Mónica, Miguelito) ,mi esposa Angélica Molina, familiares y amigos por siempre apoyarme.

Gracias por ser el pilar fundamental de mi vida y ayudarme alcanzar mis metas planteadas.

David S.

AGRADECIMIENTO

Primeramente a Dios por darme la vida y las fuerzas necesarias para seguir adelante día a día

A todos mis familiares por ser parte fundamental de mi vida.

A mi esposa Angie que me apoyaba moralmente para no desmayar en el camino.

A mi Tutora, profesores y amigos que siempre me dieron un aliento de esperanza cuando las cosas se ponían complicadas.

.

David S

ÍNDICE DE CONTENIDOS

ÍNDICE DE TABLAS

ÍNDICE DE FIGURAS

RESUMEN

Una Institución Pública a través de 7 Sistemas Auxiliares de pago (Financoop, Banco Desarrollo de los Pueblos S.A., Coonecta, Exsersa, Representaciones Ordoñez y Negrete, Pacifico y Banred) realiza la entrega del Bono de Desarrollo Humano (BDH) y de Pensiones a personas de extrema pobreza y vulnerabilidad. La información de tipos de bonos, denuncias, ubicación, puntos de pago, beneficiarios y pagos se encuentra almacenada en una base de datos; sin embargo, estos datos no habían sido utilizados por las autoridades para la toma de decisiones, detectándose vulnerabilidades en el sistema de cobros. La presente investigación utilizó la metodología de investigación científica orientada al diseño y el modelo de minería de datos basándose en la metodología CRISP-DM. En la fase de rigor se realizó un proceso ETL para crear una bodega de datos, los cuales fueron sometidos al modelo Naive Bayes y Decission Tree que identificó los patrones de comportamiento de los beneficiarios. Se identificó en los últimos 3 años, denuncias sobre cobros indebidos y desequilibrio de los puntos de pago a nivel nacional, además se detectó que la Pensión Adulto Mayor y Mis Mejores Años, correspondiente a adultos mayores y el BDH tienen tendencia a cobros indebidos,por lo que es necesario establecer campañas informativas. Así también, se determinó las zonas geográficas que presenta mayor número de beneficiarios y se identificó que la cantidad de puntos de pago no satisface la demanda de usuarios.

PALABRAS CLAVES:

- **MODELO DE MINERÍA DE DATOS**
- **METODOLOGÍA DE INVESTIGACIÓN CIENTÍFICA ORIENTADA AL DISEÑO**
- **PATRONES DE COMPORTAMIENTO**

ABSTRACT

A public institution through 7 auxiliary payment systems (Financoop, Banco de Desarrollo de los Pueblos SA, Coonecta, Exsersa, Representaciones Ordoñez and Negrete, Pacifico and Banred) makes the delivery of the Human Development Bonus (BDH) and pensions to people of extreme poverty and vulnerability.Information on bonus types, complaints, location, payment points, beneficiaries and payments is stored in a database; however, these data have not been used by the decision-making authorities, detecting vulnerabilities in the collection system.This research used the design-oriented scientific research methodology and Data Mining model based on the CRISP-DM methodology. In the rigor phase, an ETL process was carried out in order to create a data warehouse, which were submitted to the model Naive Bayes and Decission Tree, which identified the behavior patterns of the beneficiaries.During the past 3 years, allegations about non due charges and disorganization of the points of payment were identified all over the country.The research also found that the Elderly pension and My best years pension and BDH have a tendency to collect inappropriately, so it is necessary to implement informative campaings in order to educate people.On the other hand,the geographical areas with the largest number of beneficiaries do not count with enough payment pints wich causes more delays and lack of information.

KEYWORDS:

- **DATA MINING MODEL**
- **DESIGN-ORIENTED SCIENTIFIC RESEARCH METHODOLOGY**
- **BEHAVIOR PATTERNS**

CAPITULO I

INTRODUCCION

1.1. Antecedentes

Con la reforma del sector social en el 2003, el programa de Bono Solidario se convirtió en el Programa del Bono de Desarrollo Humano (BDH), siendo una transferencia monetaria condicionada, induciendo la responsabilidad y participación activa de los padres/madres en el cuidado de la salud y educación de sus hijos.

A partir de marzo de 2013, la Institución Pública es el autorizador de pagos de las transferencias monetarias relacionadas al BDH y de Pensiones.

La Institución Pública tiene a cargo la administración de la Plataforma Transaccional del Pago de las Transferencias Monetarias, la cual brinda el servicio del Pago del Bono de Desarrollo Humano (BDH) y de Pensiones con los siguientes bonos:

- Bono Desarrollo Humano
- Pensión Adulto Mayor
- Bono de Emergencia
- Persona con Discapacidad
- Menor Con Discapacidad
- Mis Mejores Años
- Bono Variable

La Plataforma transaccional administrada por la Institución Pública trabaja con 7 Sistemas Auxiliares de Pago para el pago de bonos y pensiones, los cuales son:

- Financoop
- Banco Desarrollo de los Pueblos S.A.
- Coonecta
- Exsersa
- Representaciones Ordoñez y Negrete
- Pacifico
- Banred

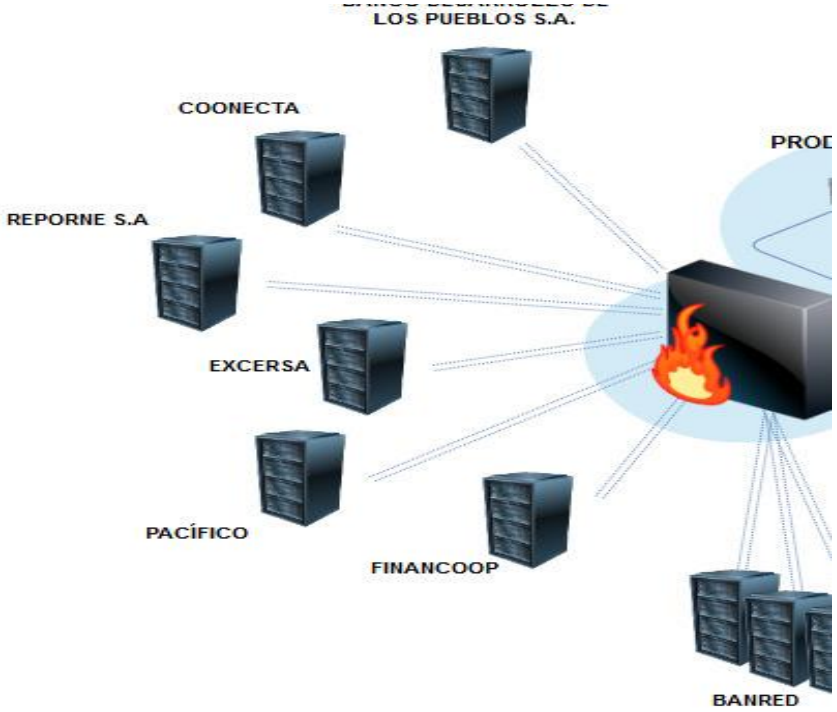


Figura 1 Enlaces con Auxiliares de Pago

A través de sus autoridades, la Institución Pública fomenta la inclusión económica y social con énfasis en los grupos de extrema pobreza y vulnerabilidad para que puedan desarrollarse y que se dé más igualdad en el ámbito social y económico.

Por este motivo las autoridades de la Institución Pública quieren obtener conocimiento de los grandes volúmenes de datos que se maneja en la Plataforma transaccional para poder tomar las mejores decisiones.

1.2. Justificación e Importancia

La Institución Pública actualmente se encuentra enfocado en mejorar y generar cambios en el Ecuador, por este motivo ejecutan estrategias, proyectos, servicios entre otros, para enfocarse en las en las personas que se encuentran en situación de extrema pobreza y vulnerabilidad durante su ciclo de vida a nivel nacional.

Por cuanto este proyecto desarrolló un modelo de minería de datos aplicado a los beneficiarios del BDH y de Pensiones para cumplir con la misión de la Institución Pública y brindar un servicio de calidad.

El presente proyecto entregó a las autoridades de la Institución Pública un panorama real del comportamiento de los beneficiarios del BDH y de Pensiones. Para cumplir con este objetivo se analizó, depuro, reformateo, entre otros, grandes volúmenes de datos, almacenados en la BDD de la Institución Pública para crear un modelo de minería de datos que determine los patrones de comportamiento de los beneficiarios del BDH y de Pensiones, con esto se podrán tomar acciones pertinentes para reducir las vulnerabilidades detectadas.

La presente investigación realizó la recolección y análisis de los datos del BDH y de Pensiones y determinó:

- Tipos de bonos y de pensiones que son más propensos a ser vulnerados por cobros indebidos
- Desequilibrio del número de puntos de pago

Con estos resultados obtenidos se evaluó el nivel de confianza del modelo de minería de datos. Además, con todo este conocimiento se puede establecer estrategias, planes, programas, proyectos y brindar servicios de calidad y con calidez a los ecuatorianos.

1.3. Planteamiento del problema

A junio de 2018 la pobreza extrema en el Ecuador alcanza un 9 % de acuerdo al Instituto Nacional de Estadística y Censo (INEC). El gobierno del Ecuador a través de una Institución Pública ha buscado reducir este porcentaje mediante la entrega del Bono de Desarrollo Humano a las personas de escasos recursos económicos para alcanzar una igualdad social en el Ecuador.

La Institución Pública recibe denuncias sobre cobros indebidos de los diferentes tipos bonos y pensiones que entrega el Estado a personas en situación de pobreza y extrema pobreza. Todo esto debido a que los usuarios desconocen de su beneficio, por cuanto son más propensos a ser vulnerados por cobros indebidos.

Además de acuerdo a la normativa vigentes mensualmente se determina que beneficiarios están habilitados o deshabilitados para recibir el Bono Desarrollo Humano, por cuanto es difícil determinar qué zonas del Ecuador requieren mayor o menor puntos de pago en base al número de beneficiarios que lo requieran. Lo que en algunos casos presenta un desequilibrio de puntos de pagos para la entrega del BDH y de Pensiones.

Existen datos de beneficiarios (sexo, status, subsidio), cadenas, por el tipo de bono (subsidio), país, región, Provincia, Ciudad, Parroquia, periodos (fecha), transacción, denuncias receptadas entre otros, todos estos datos podrían ser utilizados para crear un modelo de datos para reducir estas vulnerabilidades.

Es por esto que se quiere saber cuáles son los comportamientos más usuales de los beneficiarios en los últimos 3 años como: denuncias sobre cobros indebidos de los diferentes tipos bonos y pensiones que entrega el Estado y en qué zonas del país se encuentran los usuarios que realizan transacciones recurrentes para así clasificarlos. Con esto se pretende detectar las zonas donde se podrían establecer campañas informativas de quienes son beneficiarios. Además, identificar cuáles son los lugares con mayor tendencia de beneficiarios para determinar el número de puntos de pago requeridos.

Con todo este conocimiento las autoridades de la Institución Pública tendrán un panorama real de cómo reducir estas vulnerabilidades y tomar acciones pertinentes.

1.4. Objetivo general

Construir un modelo de datos mediante el análisis de información de los beneficiarios del Bono de Desarrollo Humano y de Pensiones, para identificar sus patrones de comportamiento y reducir vulnerabilidades detectadas.

1.5. Objetivos específicos

OE1: Realizar un análisis de la literatura mediante una revisión inicial, para determinar los algoritmos de minería de datos y herramientas, más adecuadas para crear el modelo de clasificación del BDH y de Pensiones.

OE2: Recolectar y depurar los datos existentes en la base de datos de la Institución Pública para estructurar una nueva BDD, con la cual se procederá a formatear los datos para garantizar el correcto funcionamiento del modelo de minería de datos a ser implementado.

OE3: Crear y configurar un modelo de minería de datos que determine los patrones de comportamiento de los beneficiarios del BDH y de Pensiones, para clasificarlos de acuerdo a las vulnerabilidades detectadas siguiendo los lineamientos de la metodología.

OE4: Evaluar el modelo analítico a través del uso de técnicas de validación implementadas en minería de datos, para determinar el nivel de confianza del modelo.

1.6. Formulación del problema

El proyecto de investigación busca responder las preguntas para cada objetivo específico e identificar los patrones de comportamiento de los beneficiarios del BDH y de Pensiones.

Las preguntas a resolver mediante este modelo son las siguientes:

OE1 – RQ1.1 ¿Cuáles son los estudios más actuales sobre algoritmos de minería de datos y herramientas para crear un modelo de clasificación del BDH y de Pensiones?

OE1 – RQ1.2: ¿Cuáles son las técnicas más factibles para crear un modelo de clasificación de acuerdo a las necesidades institucionales?

OE2 – RQ2.1: ¿Cuáles son las fuentes de datos con las que trabaja la institución para detectar vulnerabilidades de los diferentes bonos y de pensiones?

OE2 – RQ2.2: ¿Que herramienta es la más factible para depurar y migrar los datos desde un motor de BDD?

OE3 – RQ3.1: ¿Cuál es el modelo de minería de datos que determina los patrones de comportamiento de los beneficiarios del Bono de Desarrollo Humano y de Pensiones para reducir vulnerabilidades detectadas?

OE3 – RQ3.2: ¿Es factible acoplar los datos existentes para que se ajuste al modelo de minería de datos seleccionado?

OE4 – RQ4.1: ¿Qué criterios se debe considerar para validar el modelo y determinar un nivel de confianza en los resultados obtenidos?

OE4 – RQ4.2: ¿Qué margen de error se debe manejar al evaluar un modelo?

CAPÍTULO II

FUNDAMENTACIÓN TEÓRICA

2.1. Marco teórico

El marco teórico busca demostrar en forma ordenada la hipótesis, para esto motivo se realizó un análisis de la teoría utilizando las variables del problema planteado, con esto se demuestra jerárquicamente cada categoría hasta llegar a explicar las variables dependientes e independientes, que se muestra a continuación:

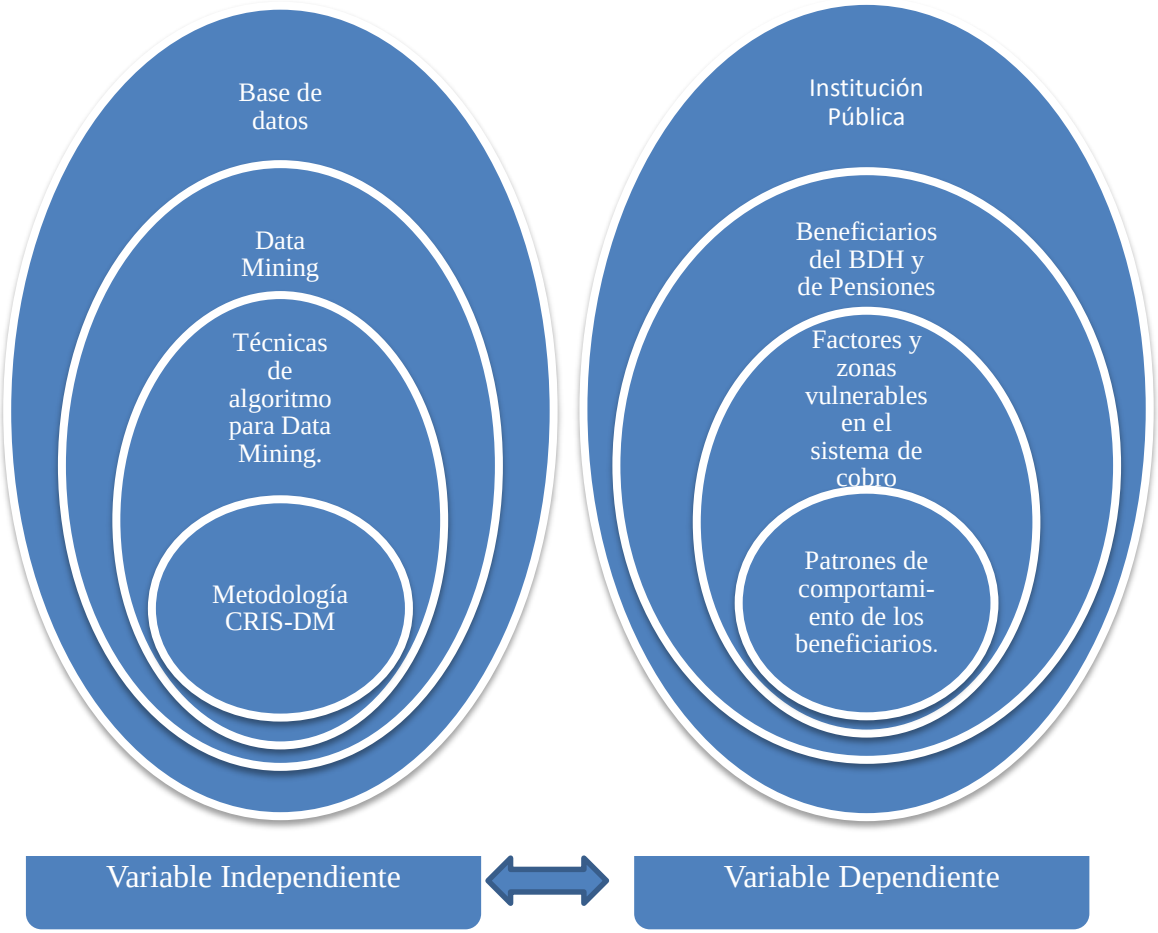


Figura 2 Categorización de variables

2.1.1. Fundamentación de las variables Independientes.

Base de Datos: Una Base de Datos es un conjunto de datos no estructurado, los cuales se encuentran ordenados e implementados en máquinas, los usuarios pueden acceder a ellos en tiempo real para poder obtener sus propios análisis.

Las bases de datos constituyen una parte fundamental de los sistemas de información en las que están integradas. El estado actual de la tecnología de bases de datos en el mundo, es el resultado de la evolución que a lo largo de décadas ha tenido lugar en el procesamiento de los datos y en la gestión de información. Esta tecnología se ha ido desarrollando a lo largo del tiempo desde los métodos más primitivos de los años cincuenta, hasta los potentes sistemas de hoy en día, estipulada por un lado por la demanda y las necesidades de las gestiones de la información y restringida por las limitaciones de la tecnología del momento (PANDORAFMS, 2018 ARTICA ST).

Minería de datos: La minería de datos (DM) es una analogía en grandes volúmenes de datos, que se encuentran almacenados en una BDD. Las empresas buscan tener la ventaja competitiva frente a otras, por este motivo indagan en sus datos para poder obtener comportamientos, patrones, tendencias que le permitan ser más competitivos.

La minería de datos descubre relaciones, tendencias, desviaciones, comportamientos atípicos, patrones y trayectorias ocultas, con el propósito de soportar los procesos de toma de decisiones con mayor conocimiento. La Minería de Datos se puede ubicar en el nivel más alto de la evolución de los procesos tecnológicos de análisis de datos (Martínez, 2013).

Técnicas de algoritmo para Data Mining: Según el problema específico que se necesite resolver se tendrá que escoger la técnica de minería de datos más adecuada, entre estas tenemos:

Redes neuronales artificiales. -Radica en el aprendizaje secuencial, el hecho de utilizar transformaciones de las variables originales para la predicción y la no linealidad del modelo.

Árboles de decisión. -Permiten obtener de forma visual las reglas de decisión bajo las cuales operan los consumidores, a partir de datos históricos almacenados.

Agrupamiento (Clustering). -Agrupa un conjunto de observaciones en un número dado de clusters o grupos, está basado en la idea de similitud de los grupos.

Algoritmo Jerárquico. -Se debe calcular la distancia entre los pares de objetos o clúster, se busca los dos clúster más cercanos éstos se juntan y constituyen uno solo, se repite los pasos hasta que no quedan pares de comparación.

Regla de Inducción. -Consiste en derivar un conjunto de reglas para clasificar casos, generan un conjunto de reglas independientes que permiten contrastar árboles de decisión y patrones a partir de los datos de entrada.

La información de entrada será un conjunto de casos en que se ha asociado una clasificación a un conjunto de variables o atributos (Evaluando Software, 2016).

Agrupamiento (K-Means). - Esta técnica de minería de datos no supervisada divide un conjunto de datos en subgrupos llamados clases que a su vez se diferencian por su máxima distancia de separación entre ellas. Se considera una clase cuando todos sus elementos tienen la mínima separación de distancia posible entre ellos (Erendira Rendon & Abundez Barrera, 2016).

Reglas de asociación. - Según Haydeé Gommez y María de los Angeles Cerón resaltan que: "El método de asociación detecta eventos que ocurren de manera simultánea" (Gommez Díaz & Cerón Reyes, 2010). El objetivo de esta técnica es encontrar patrones de asaciones que se relacionen de alguna forma o que den pistas de nuevas relaciones (Morales & Gonzáles, 2012).

Regresión lineal. - Es una técnica predictiva netamente estadística utilizada en la minería de datos que estudia el cambio de la variable independiente en relación de la variable dependiente (Moral Peláez, 2012).

Metodología CRIS-DM.

CRISP-DM, que son las siglas de Cross-Industry Standard Process for Data Mining, es un método probado para orientar sus trabajos de minería de datos. Como metodología, incluye descripciones de las fases normales de un proyecto, las tareas necesarias en cada fase y una explicación de las relaciones entre las tareas. ν Como modelo de proceso, CRISP-DM ofrece un resumen del ciclo vital de minería de datos (BM Corporation, 2012.).

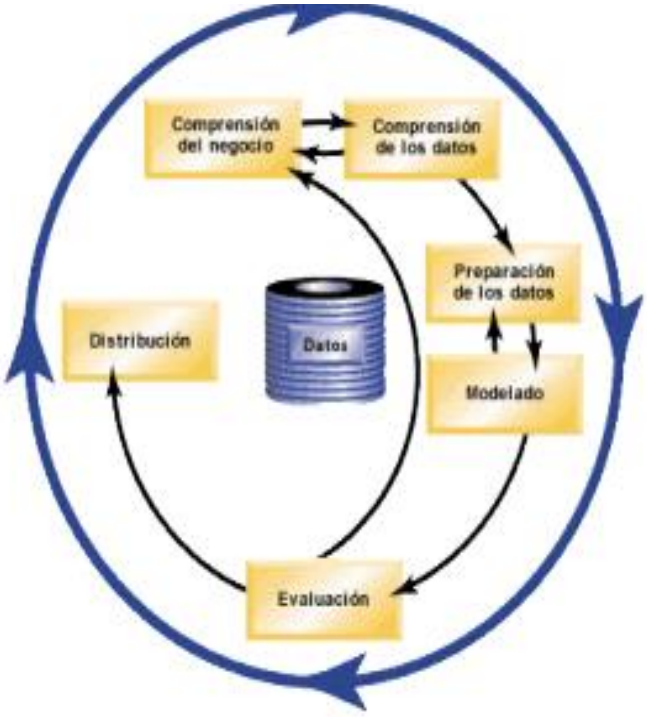


Figura 3. Metodología CRIS-DM
Fuente: (BM Corporation, 2012.)

1. Fase de comprensión del negocio o problema

Marcelo Barrios resalta que: en esta fase se determinan los objetivos y requerimientos del proyecto desde una perspectiva del negocio, definiendo el problema de minería y el plan de trabajo (Barrios, 2010). Esta fase es de extrema importancia debido a que el explorador debe mentalizar y comprender el negocio para poder definir junto con la empresa los objetivos de la minería. Como se mencionó con anterioridad, los objetivos deben ser medibles y alcanzables a corto plazo porque eso busca una empresa, resolver sus problemas en el menor tiempo posible.

2. Comprensión de los datos

En esta fase se debe recolectar los datos que sean de utilidad para la investigación, en otras palabras, irían en relación a los objetivos definidos, luego se procede a describirlos para tener una idea clara de los datos para pasar posteriormente a explorarlos utilizando herramientas o técnicas estadistas para ver su distribución, comportamiento entre otros aspectos (Folgueiras Bertomeu, 2010).

3. Preparación de los datos

Del conjunto de datos proporcionados por el departamento de tecnología, se debe seleccionar que parte de ellos se utilizaran para el análisis y luego estos pasaron por el proceso ETL, por ejemplo si se tiene en un campo fecha valores como “00-00-00” estos deberán ser limpiado o ignorados.

El siguiente paso a seguir es armar la estructura donde se almacenará el resultado y posteriormente unificarlos. Algunas veces se tiene a colocar los datos en Tablas des normalizadas para evitar tener tantas Tablas (Rigeiro, 2012).

4. Modelado.

En esta etapa tendremos claro cómo funciona el negocio, cual es la problemática, los objetivos que tendremos que alcanzar, con toda esta información no tendremos problema en modelar el modelo multidimensional, la cual comprende 6 dimensiones (dimensión Bono, dimensión Ubicación, dimensión Tiempo, dimensión Punto Pago, dimensión Denuncias, dimensión Beneficiario y una tabla de hecho (Transacción), se comparó la mejor técnica de modelado, para finalmente evaluarlo.

5. Evaluación

En esta fase vamos a comparar el resultado de los datos del modelo que previamente se ha escogido con los objetivos que plantearon del negocio, se evaluó los resultados utilizando una matriz de confusión, mediante el operador de validación cruzada que es propio de la herramienta Rapidminer, con este proceso se pide identificar si los resultados son aplicables o no.

6. Implantación

Con el nuevo conocimiento que se tiene se tendrá que realizar implementaciones o modificaciones en el negocio, con el objetivo de incrementar conocimiento.

En general, la fase de distribución de CRISP-DM incluye dos tipos de actividades:

- Planificación y control de la distribución de los resultados.
- Finalización de tareas de presentación como la producción de un informe final y la revisión de un proyecto.

Dependiendo de las necesidades de su organización, es posible que necesite completar una o varias fases (BM Corporation, 2012.).

2.1.2. Fundamentación de la variable dependiente.

Institución Pública.

La Institución Pública tiene a cargo la inclusión económica y social, con énfasis en los grupos de atención prioritaria y la población que se encuentra en extrema pobreza y vulnerabilidad a nivel nacional.

La Institución Publica dentro de sus procesos tiene a cargo la administración de la Plataforma Transaccional del Pago de las Transferencias Monetarias, la cual brinda el servicio del Bono de Desarrollo Humano y de Pensiones en todo el Ecuador a través de un Switch transaccional, el cual se enlaza a los sistemas auxiliares de pago.

Actualmente, la plataforma transaccional de la Institución Pública trabaja con 7 sistemas auxiliares de pago: Financoop, Banco Desarrollo de los Pueblos S.A., Red Transaccional Cooperativa S.A, Exsersa, Representaciones Ordoñez y Negrete, Pacifico, Banred, los mismos que abarcan a todas las entidades financieras que realizan en el pago de las transferencias monetarias, fomentando de esta manera el incremento de puntos de pago y el desarrollo de la economía popular y solidaria.

A partir de marzo de 2013 hasta la presente fecha, la Institución Pública es el concentrador de comunicaciones y autorizador de pagos de las transferencias monetarias relacionadas al BDH y de Pensiones, con esto se busca promover actividades productivas y poder disminuir la dependencia de transferencias monetarias que se otorgan.

Beneficiarios del Bono Desarrollo Humano.

Es un subsidio de 50 dólares que reciben las familias más pobres del país, corresponde a una transferencia no contributiva. Los beneficiarios a este subsidio son:

- Representante de las familias que viven en condiciones de extrema pobreza y vulnerabilidad. De preferencia, se lo suele dar a las mujer jefa de hogar, pero también puede ser al cónyuge o a la persona que tenga como responsabilidad de decisiones de compra
- Adultos mayores
- Personas con discapacidad

La condición de pobreza se lo mide según el índice de clasificación socioeconómica del Registro Social y varía dependiendo del beneficiario (familias, adultos mayores, personas con discapacidad).

Factores de los beneficiarios del BDH y de Pensiones.

La Institución Pública ha detectado algunos factores en los beneficiarios del BDH y de Pensiones como son: falta de información si es beneficiario del BDH en algunos lugares del país, revisión de puntos pago, siendo estos los más principales.

Es por esto que se quiere saber cuáles son los comportamientos más usuales de los beneficiarios y así clasificarlos, con esto se determinara el porcentaje de ocurrencia y cómo afecta.

Mediante un análisis se obtuvo datos de denuncias y transaccional, con los cuales se trabajó en el proyecto:

- Denuncias
- Beneficiarios(sexo, status, subsidio)
- Cadenas
- Subsidio
- Evento

- Lugar
- Tiempo
- Transacción

Que permitirán determinar los patrones de comportamiento de los beneficiarios del BDH y de Pensiones.

Patrones de comportamiento.

Cuando un usuario tiene un mismo patrón de conducta, comportamiento para realizar una acción en un determinado tiempo o periódicamente.

Cuenta con un catálogo de patrones de comportamiento bien identificados a través de momentos o puntos del tiempo (los puntos pueden ser días, semanas, meses, años u otra unidad de tiempo) (Arenas & Luna).

2.2. Antecedentes del estado del arte

Se definió y creó un Mapeo Sistemático de Literatura (SMS)[1] para el estado del arte, con el objetivo de obtener los criterios de inclusión, exclusión y la estrategia de búsqueda. Para esta investigación se usaron los repositorios académicos Scopus, Springer, IEEE Xplore.

Definición el objetivo: El objetivo del estudio del estado del arte fue relacionar las preguntas de investigación planteadas para que cumplan con los objetivos específicos.

Definición de criterios de inclusión y exclusión: En esta fase el investigador realizó la búsqueda en las bases digitales de acuerdo al tema planteado, con lo cual retornó varios artículos

[1] Systematic Mapping Study: Es el estudio de artículos, publicaciones entre otros, que se encuentran almacenadas en diferentes repositorios y se busca revisar los temas relacionados al trabajo que se propone.

relacionados. Es de suma importancia definir las características apropiadas de inclusión y exclusión referente al tema propuesto.

Criterios de inclusión:

- Que el artículo contenga información con algoritmos de minería de datos orientados al manejo de bonos
- Que el artículo sea sobre minería de datos aplicada al manejo de información de bonos
- Artículo sobre clasificación de patrones de comportamiento basado en indicadores de cobro de bonos y de Pensiones
- El artículo debe contener técnicas de validación de minería de datos

Criterios de exclusión:

- Artículos sobre minería de datos que no estén aplicados al manejo de información de bonos
- Artículos sobre el aprendizaje de máquina
- Artículos sobre modelos de predicción
- Artículos que hablen sobre big data

Definición de la estrategia de búsqueda.

Revisión inicial: Se realizó una indagación previa con palabras claves, resumen, contenidos, entre otros para poder constatar estudios relacionados al tema planteado.

Validación cruzada de estudios: Con la validación cruzada se pudo avalar que se cumplan los criterios de inclusión y exclusión para obtener como resultado información veraz y oportuna en un grupo de control.

Integración de grupo de control: El grupo de control cumple con los criterios de incluso y exclusión, con lo cual se realizó una análisis inicial del título, contenido, palabra clave, como resultado se obtiene:

Tabla 1

Grupo de Control

Grupo de control	Título	Palabra clave
EC1	*Research of Pension Fund Market Risk Model Based on Data Mining*	*Pensions, Data mining, Reactive power, Forward contracts, Educational institutions, Economic forecasting, Frequency, Data security, Random variables, Risk analysis.*
EC2	*The decision support system of auditing social insurance funds on-line based on data warehouse*	*Insurance, Data warehouses, Pensions, Decision support systems, Databases, Security, Data mining.*
EC3	*Perceptions of postretirement benefit obligations by bond rating analysts.*	*Postretirement, benefit obligations, bond ratings, investor, default risk.*
Ec4	*Training neural networks for deriving bond rating formulas.*	*Neural networks, Bonding, Feedforward systems, Space technology, Management training,*

Continúa

		Computer science, *Technology management,* *Data mining,* *Backpropagation,* *Spatial databases.*
Ec5	*A Recommendation of Pension Service Based on Trusted Network*	*Pensions,* *Indexes,* *Reliability,* *Data mining,* *Medical services,* *Correlation,* *Cleaning.*

Construcción de la cadena de búsqueda: Para poder crear la cadena de búsqueda se tomó en cuenta el grupo de control, las palabras claves de la investigación propuesta, con lo cual se clasificó en 3 grupos: Minería de datos, bonos y pensiones, Indicadores.

Tabla 2

Construcción de la cadena de búsqueda.

		EC1	EC2	EC3	EC4	EC5	Número de repeticiones
Data Mining.	*Data mining*	*X*	*X*		*X*	*X*	*4*
	Data security	*X*	*X*				*2*
	Data warehouses		*X*			*X*	*2*
	Decision support systems		*X*				*1*
	Databases		*X*		*X*		*2*
Bonos y Pensiones.	*Pensions*	*X*	*X*			*X*	*3*
	Risk analysis	*X*		*X*			*2*
	Insurance		*X*				*1*
	benefit obligations			*X*			*1*
	bond ratings			*X*	*X*		*2*

Continúa

Indicadores.	*Reactive power*	*X*		*X*	*2*
	Forward contracts	*X*			*1*
	Economic forecasting	*X*			*1*
	Frequency	*X*			*1*
	Postretirement		*X*		*1*

La cadena de búsqueda se clasifica con las palabras claves que más se repitan en cada contexto, con AND para concatenar y con OR para unir la cadena, con lo que se establece la siguiente cadena:

(Data Mining OR Data Security OR Data Warehouse OR Databases) AND (Pensions OR Risk Analysis OR Bond Rating) AND (Reactive power).

La cadena de búsqueda fue aplicada en el repositorio IEEE, con los criterios de inclusión y exclusión, temas relacionados, área de interés, tendencias existentes para que el investigador tome ese conocimiento y pueda guiarse a una solución efectiva.

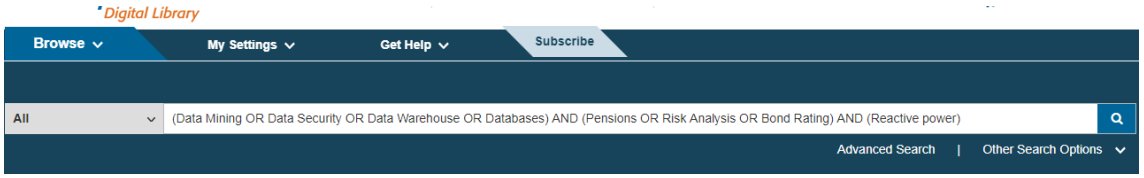


Figura 4. Cadena IEEE Xplore

Research of Pension Fund Market Risk Model Based on Data Mining.

Brought forward a novel algorithm to measure VAR based on Data Mining, and considered of decay and magnify attribute of financial time series to optimize risk market model. First, VAR estimation model was established with the thought of quantile plot, and different time segment's VAR was calculated under given confidence level. Secondly, VAR's Failure Frequency was got statistically according to portfolio's real profit or loss value, which is used to construct the discriminant of the best decay and magnify factor. Finally VAR was gained.

This novel algorithm was adopted by Chinese Social Security Fund invest management and control system. The experiment results show that the VAR's Failure Frequency is between 2.65%-5.56% under given confidence level 95%, is close to 5% and the algorithm is accurate and reliable (Xianlin Zhuo, 2007).

Study and applications of Data Mining to the structure risk analysis of customs declaration cargo

In order to solve the confliction between rapid increasing of cargo quantities and the customs limited inspection force, a study and application of Data Mining is used in the structure risk analysis of customs declaration cargo. The cluster method of Data Mining is used in this paper to divide the cargo into seven types, thus customs can put the mainly inspection force to the high risk level cargo. The results show that this kind of method can be used to reform the operation mode of customs inspection (Yan-hai & lin-yan, 2005).

Conclusión.

Se realizó la revisión de literatura con la cadena de buscada seleccionada y como resultado se tuvo artículos relacionados a la minería de datos en fondos de pensiones (Research of Pension Fund Market Risk Model Based on Data Mining), análisis de riesgos (Study and applications of Data Mining to the structure risk analysis of customs declaration) pero no enfocados a determinar los patrones de comportamiento de los beneficiarios del Bono de Desarrollo Humano (BDH) y de Pensiones. Por este motivo es un tema innovador y pionero ya que se entregó información real y confiable a las autoridades de la Institución Pública, para que tomen las mejores decisiones en el Ecuador.

2.3. Marco conceptual

Este proyecto se desarrolló tomando en cuenta las herramientas más adecuadas con el fin de cumplir con los objetivos planteados, por lo cual se tuvo:

PowerDesigner

Es una herramienta que permite analizar metadatos para poder desarrollar técnicas de modelación de las cuales tenemos: modelo conceptual, lógico y físico y obtener una arquitectura de información óptima.

Ofrece una mejora tecnológica continua en el negocio con el objetivo de disminuir la falta de comunicación, errores de duplicación de procesos, falta de datos, entre otros.

PowerDesigner tiene la facilidad de trabajar con diferentes motores de base de datos y se acopla a las funcionalidades de cualquier empresa.

Con la data extraída se pudo analizar y crear en PowerDesigner el modelo entidad-relación y modelo multidimensional (PowerDesigner, 2015).

SQL Server

Gestor de Base de Datos que fue desarrollado por la empresa Microsoft con el objetivo de que los usuarios puedan manipular, crear bases de datos, tablas, vistas, relaciones, procedimientos almacenados entre otros.

SQL Server es un sistema de gestión de base de datos relacional que incluye un entorno grafico de administración por comandos Lenguaje definición de datos (DDL) y Lenguaje manipulación de datos (DML).

En SQL Server se creó la BDD, tablas y se almaceno todos los datos para que posteriormente fueran procedentes para el proceso de extracción transformación y carga (ETL) (Microsoft , 2018)**.**

Power BI

Power BI es una solución de análisis empresarial que permite visualizar sus datos y compartir información en su organización o incorporarlos en su aplicación o sitio web. Con Power BI sus datos cobran vida con paneles e informes en vivo (Microsoft, 2019).

Se puede tomar decisiones al instante. Conectar, modelar y luego explorar datos con informes visuales que puede colaborar, publicar y compartir. Power BI se integra con otras herramientas, incluido Microsoft Excel, para obtener conocimiento rápidamente y trabajar sin problemas con soluciones existentes (Microsoft, 2019).

RAPIDMINER

Rapidminer está desarrollado en java con entorno gráfica y su distribución es bajo la licencia pública general de Affero **(**AGPL).

Es una herramienta intuitiva que permite realizar ETL, minería de datos, cluster, clasificación por análisis exploratorio, operadores de entrada y salida, multiplataforma, se puede integrar con diferentes módulos, entre otras características.

RapidMiner fue nombrado Líder en el Cuadrante Mágico 2019 de Gartner para las ciencias de la información y las plataformas de aprendizaje automático por sexto año consecutivo (RapidMiner, 2019).

Sencillez sofisticada: características como Auto Model, capacidades analíticas aumentadas como Turbo Prep y una UI superior a la media hacen de RapidMiner Studio uno de los favoritos de los científicos de datos ciudadanos (2018 Gartner, Inc. y / o sus Afiliados, 2018).

Características avanzadas: La facilidad de uso no excluye la presencia de energía. Más allá del aprendizaje profundo y la compatibilidad con GPU, la plataforma de RapidMiner ahora incluye la

funcionalidad de aumento de datos y características mejoradas de series de tiempo. (2018 Gartner, Inc. y / o sus Afiliados, 2018).

Plataforma coherente de extremo a extremo: los clientes de referencia hicieron muchos comentarios complementarios sobre la coherencia de la experiencia del usuario de RapidMiner, desde su gestión de repositorio escalable hasta su puntuación en tiempo real. Los elementos que contribuyen al continuo incluyen RapidMiner Studio (para el desarrollo de modelos); RapidMiner Server (para compartir, colaborar, implementar y mantener modelos); RapidMiner Cloud (incluidos los servicios de repositorio y ejecución destinados a alojar capacidades de automodeling); y RapidMiner Real-Time Scoring (introducido en 2018 para proporcionar un motor de ejecución de modelos de baja latencia) (2018 Gartner, Inc. y / o sus Afiliados, 2018).

CAPÍTULO III

MEMORIA TÉCNICA METODOLÓGICA

3.1. Metodología de Investigación

La presente investigación utilizó la metodología de investigación científica orientada al diseño mencionada en el análisis de minería de datos de los autores (Parraga & Zaldumbide, 2018).

Dicha metodología comprende tres ciclos: relevancia, diseño y rigor como se demuestra a continuación:

Figura 5. Metodología de investigación científica orientada al diseño
Fuente: (Gonzalez & Pomares Quimbaya, 2012)

Los tres ciclos están orientados a desarrollar la investigación que se plantea:

Entorno: Se realizó un análisis de la BDD de los beneficiarios del Bono de Desarrollo Humano para depurar los datos existentes e identificar los algoritmos de minería de datos que serán utilizados en el modelo a diseñar

Diseño: Con CRIS-DM se comprendió los datos del negocio, para luego ser preparados a través de una extracción, transformación y carga (ETL), luego de este proceso se escogió el modelo más adecuado y se determinó los patrones de comportamiento de los beneficiarios del BDH para poder reducir vulnerabilidades en el sistema de cobro y se obtuvo el nivel de confianza del resultado.

Base de Conocimiento: Se pudo obtener resultados reales de como un modelo de datos determina los patrones de comportamiento de los beneficiarios del Bono de Desarrollo Humano (BDH) y de Pensiones para poder mitigar las dos vulnerabilidad en el sistema de cobros.

Las investigaciones que se realizó en este proyecto son:

- Experimental: permite relacionar el modelo de datos que se ajuste a las necesidades planteadas y revisión de las variables para determinar las herramientas necesarias
- Descriptiva: con la cual se pudo revisar y guardar datos importantes sobre el tema planteado

CRIS-DM, la cual ayudó a explorar todos los datos que se requiere de la institución, esta metodología comprende 6 etapas:

1. **Comprensión del negocio.**

En esta etapa se vio cuáles son los objetivos de la institución, valorización de la situación actual y los objetivos de la minería de datos que se alcanzó.

2. **Comprensión de datos.**

Se empezó con la recolectando datos, análisis de estos para resolver las preguntas plantadas anteriormente.

3. **Preparación de datos.**

En esta etapa se pudo obtener la lista minable de los datos se obtuvo de la institución.

4. **Modelado.**

Se cubrió todo el proceso del modelado desde la selección del modelado hasta la evaluación de modelo.

5. **Evaluación.**

Se evaluaron los resultados obtenidos de los beneficiarios del BDH y de Pensiones para determinar si son útiles a las necesidades de la institución.

6. **Despliegue o distribución.**

Se tuvo una idea general de los resultados obtenidos para poder integrarlos en la toma decisiones de las autoridades.

3.2. Ejecución del proceso de investigación

El proceso se desarrolló con la metodología de investigación científica orientada al diseño, la cual comprende 3 fases que se ven a continuación:

ENTORNO: Se pudo revisar los procesos de denuncias y transacción, en base a esto se analizó la data y se logró diseñar con PowerDesigner el modelo entidad-relación y multidimensional con el propósito de almacenar estos datos de una manera óptima. Estos datos fueron tratados con la herramienta RapidMiner para que cumplan con los estándares que se maneja al momento que se manipula grandes volúmenes de datos, como resultado se obtuvo una nueva bodega de datos eficaz y de acuerdo a las necesidades que se plateo anteriormente en esta investigación.

DISEÑO: Una vez revisado los requerimientos de entorno, se procedió a crear el modelo de minería de datos y determinar los patrones de comportamiento de los beneficiarios del Bono Desarrollo Humano y de Pensiones, de este proceso se obtuvo un resultado, el cual fue evaluado

por la matriz de confusión con el fin de tener un óptimo nivel de confianza. El proyecto fue desarrollado usando la metodología CRIS-DM, la cual permitió revisar y comprender al negocio en sus 6 fases.

Fase I. Comprensión del negocio.

La Institución Pública administra y gestiona la Plataforma Transaccional de Pago del Bono de Desarrollo Humano (BDH) y de Pensiones. Para realizar el pago de las transferencias monetarias se tienen 7 Sistemas Auxiliares como son: Financoop, Banco Desarrollo de los Pueblos S.A., Coonecta, Exsersa, Representaciones Ordoñez y Negrete, Pacifico, y Banred.

Todas las transacciones y denuncias que realizan los beneficiarios del Bono Desarrollo Humano en diferentes partes del país son almacenadas en una Base de Datos, la cual no había sido utilizada para generar conocimiento.

La Institución Pública recibe denuncias sobre cobros indebidos del Bono Desarrollo Humano que entrega el estado a personas en situación de pobreza y extrema pobreza. Todo esto debido a que los usuarios desconocen de su beneficio en algunos lugares del país o qué tipo de bono y de pensión es el más vulnerable a cobros indebidos.

Es por esto que se determinó los comportamientos más usuales de los beneficiarios en los últimos 3 años como: denuncias sobre cobros indebidos de los diferentes tipos bonos y pensiones que entrega el estado y en qué zonas del país se encuentran los usuarios que realizan consultas, pagos y reversos recurrentes para así clasificarlos. Con esto se detectó las zonas donde se podrían establecer campañas informativas de quienes son beneficiarios. Además, se identificó cuáles son los lugares con mayor tendencia de beneficiarios habilitados para determinar el número de puntos de pago requeridos, con esto se determinó el porcentaje de ocurrencia y cómo afecta al estado.

Con todo este conocimiento las autoridades de la Institución Pública tendrán un panorama real de cómo reducir estas vulnerabilidades y tomar acciones pertinentes.

Fase II. Comprensión de los datos.

Se realizó la recopilación y verificación de calidad de datos relevantes de acuerdo a los objetivos planteados en el proyecto, con lo cual se obtuvo lo siguiente de los últimos 3 años:

- Beneficiario: Subsidio, Sexo, status
- Beneficiario Emergente: Sexo, Subsidio
- Cadena: Descripción, Estado
- Comercio: Identidad, Descripción, Tipo Comercio
- Comisionista: Descripción comisionista
- Denuncias: Tramite Fecha Creación Denuncia, Estado Denuncia, Comisionista, Institución financiera, Agencia, Provincia, Ciudad, Parroquia, Fecha de Transacción de Pago, Hora de Transacción Pago, Fecha Creación Pago, Hora Creación Pago, Fecha Envió Concentrador, Estado Denuncia Concentrador, Estado Denuncia Concentrador, Observación Denuncia, Tipo Subsidio.
- Evento: Descripción
- Provincia: Región, Provincia, descripción
- Ciudad: Ciudad, descripción
- Parroquia: Parroquia, descripción, estado
- Transacción: Tipo Transacción, Estado, Estado Cierre, Fecha Transacción, Hora Transacción, Provincia, Ciudad, Parroquia, Comisionista, Estado cadena, Institución financiera, Estado comercio, Agencia, Estado Agencia

Se escogió los siguientes datos: beneficiarios, cadenas financieras y sucursales, tipos de bono, evento, estado de cobro, Provincia, Ciudad, Parroquia, periodos de pago, fecha de transacción, tipo de transacción, denuncias receptadas, los cuales ayudaron a comprender el estado de almacenamiento de la data y estos fueron escogidos para solventar las vulnerabilidades detectadas en la fase de comprensión del negocio.

Fase III. Análisis y selección de datos

Se analizó los datos extraídos y se puedo clasificar en un modelo entidad-relación y multidimensional. Los datos se encontraban en diferentes formatos, contenían campos innecesarios, espacios nulos, con caracteres especiales, entre otros, con lo cual se procedió a realizar un proceso de extracción, transformación y carga (ETL) para poder reformatear los datos y poderlos integrar en una nueva base de datos con el objetivo de tener un formato estándar que permitió la manipulación más rápida de estos.

Fase IV. Modelado

Se seleccionó, construyó y evaluó el modelo de minería de datos que determine los patrones de comportamiento de los beneficiarios del BDH y de Pensiones, para clasificarlos de acuerdo a las vulnerabilidades detectadas y se obtuvo información confiable.

Fase V. Evaluación de resultados obtenidos.

Se procedió a evaluar el modelo analítico a través del uso de técnicas de validación implementadas en minería de datos, para determinar el nivel de confianza del modelo y si los resultados obtenidos son aplicables o no.

Fase VI. Despliegue.

Con los resultados obtenidos se incrementa el conocimiento de las vulnerabilidades detectadas y se genera un resumen comparativo para la toma de decisiones que se pueden realizar.

Se recomienda planificar la monitorización y mantenimiento si es el caso. Además, se contribuye con nuevo conocimiento para que se despliegue las acciones pertinentes del caso.

BASES DE CONOCIMIENTO: Con los resultados del modelo de datos obtenido se identificó los patrones de comportamiento de los beneficiarios de Bono Desarrollo Humano y de Pensiones en los últimos 3 años, se pudo observar las vulnerabilidades detectadas y entregar conocimiento para toma de decisiones.

CAPÍTULO IV

RESULTADOS

4.1 Informe de Resultados

4.1.1. Análisis y selección de datos

Se analizó los datos históricos con los que cuenta la Institución y se recolectó datos que aportaban a la investigación como son: beneficiarios que revisen el BDH y Pensión, los Sistemas Auxiliares de pago (Financoop, Banco Desarrollo de los Pueblos S.A., Coonecta, Exsersa, Representaciones Ordoñez y Negrete, Pacifico, y Banred los cuales interactúan con el beneficiario para entregar un aporte económico, ubicación es donde se encuentra a nivel nacional, las denuncias son registras por los balcones de servicios, los tipos de bonos y pensiones son los que otorga el estado a personas de extrema pobreza y vulnerabilidad.

Se procedió a seleccionar las variables beneficiario, denuncias, bonos, ubicación, transacción, puntos de pago, los cuales determinaron los patrones de comportamiento de los beneficiarios, teniendo en cuenta las vulnerabilidades detectadas. Se procedió a crear una bodega de datos, descritos en la tabla 3:

Tabla 3
Datos

DATOS	ATRIBUTO
BENEFICIARIO	Sexo
	Status
	Subsidios
DENUNCIAS	Tramite
	Fecha de creación de la denuncia
	Estado que se encuentra la denuncia

Continúa

	Fecha de Transacción de pago
	Hora de Transacción de pago
	Estado de la denuncia del concentrador
	Respuesta del estado de la denuncia del concentrador
	Observación que se tiene
	Subsidio de la denuncia
BONO	Subsidios
	Detalle del Bono
UBICACION	Provincia
	Ciudad
	Parroquia
PUNTO PAGO	Descripción del comisionista
	Estado del Comisionista
	Institución financiera
	Estado Institución Financiera
	Estado de la agencia
TRANSACCION	Tipo de transacción
	Fecha de creación
	Hora de creación
	Fecha de transacción
	Hora de transacción

4.1.2. Preparación de los datos

Se analizó los datos que contenían las tablas y las relaciones que tienen entre ellas, con esto se realizó un diagrama entidad-relación como se muestra en la figura 6 que permitió realizar una bodega de datos de acuerdo a las vulnerabilidades detectadas.

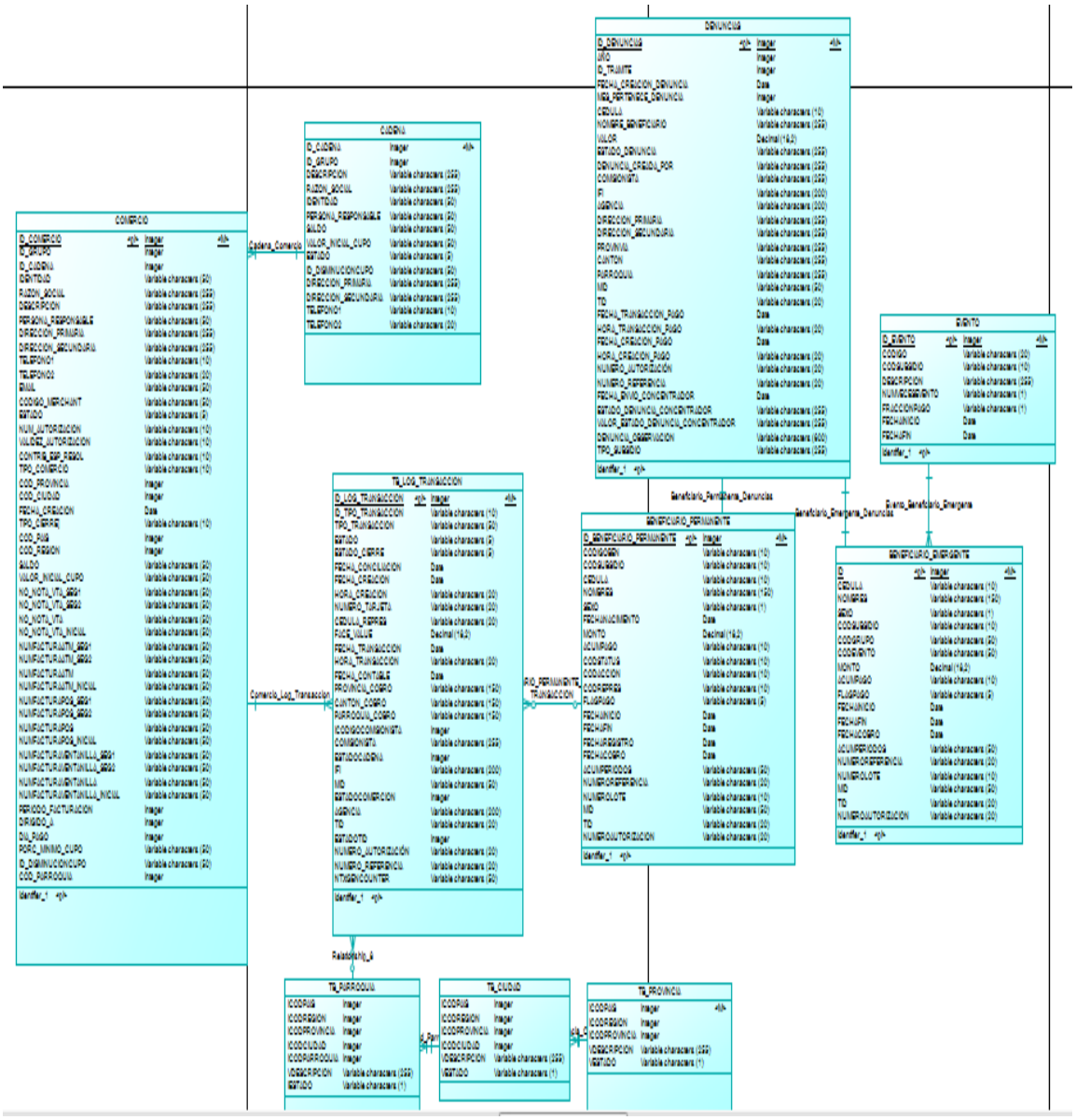


Figura 6. Diagrama entidad relación

Una vez que se obtuvo el diagrama entidad relación se pudo construir el modelo multidimensional, el cual está enfocado a resolver los objetivos que se planteó en este proyecto como se observa en la figura 7:

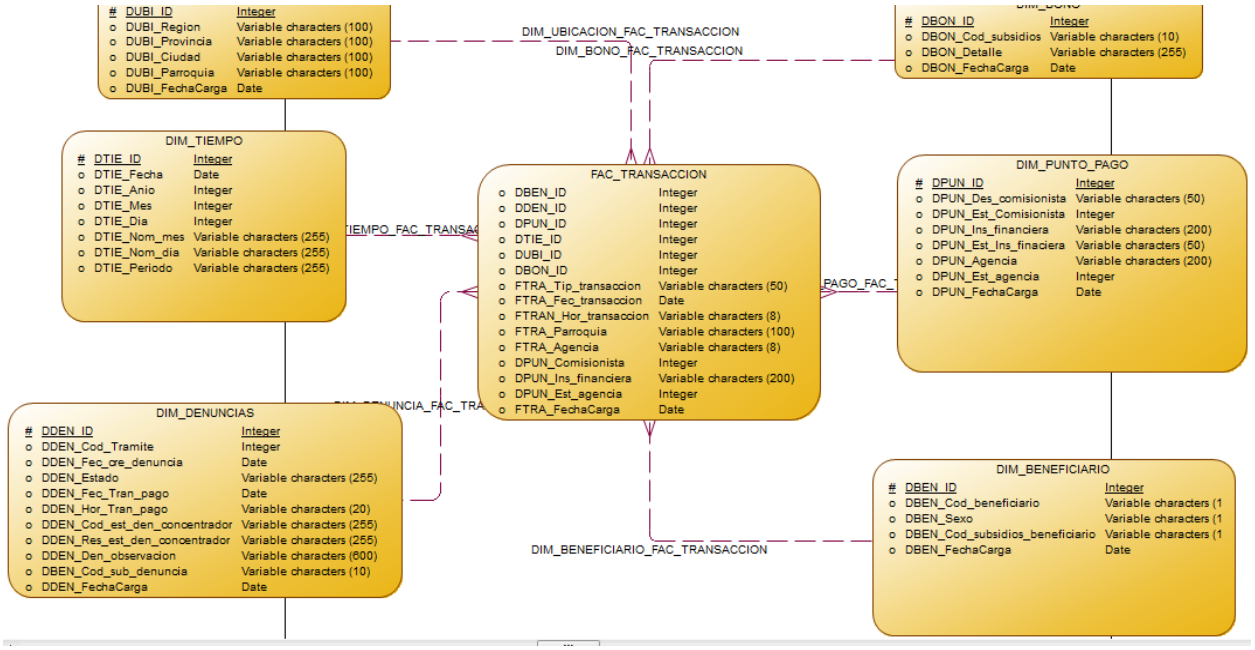


Figura 7. Modelo Multidimensional

De acuerdo al proyecto se realizó 6 dimensiones con una tabla de hechos como se describe a continuación:

Dimensión Bono (DIM_BONO). – En esta tabla se encuentra los diferentes tipos de bonos y pensiones que el estado otorga a las personas de escasos recursos económicos.

Dimensión Ubicación (DIM_UBICACION). – De acuerdo a los datos registrados en esta dimensión se tiene la ubicación de los puntos de pago a nivel nacional.

Dimensión Ubicación (DIM_TIEMPO). – Se encuentra registrado las líneas de tiempo de acuerdo a los periodos que se plantea.

Dimensión Punto Pago (DIM_PUNTO_PAGO). – Esta tabla registra los puntos de los Sistemas Auxiliares de Pago que encuentra a nivel nacional.

Dimensión Denuncias (DIM_DENUNCIAS). – Se encuentra el listado de las denuncias registradas a nivel nacional a través de los balcones de servicio.

Dimensión Beneficiario (DIM_BENEFICIARIO). – En esta dimensión se encuentra la información del sexo de los beneficiarios de acuerdo al bono o pensión que recibe.

Tabla de Hechos (FAC_TRANSACCION). – En esta tabla central se encuentran los indicadores y los datos cualitativos de la institución y esta es cargada con las dimensiones antes mencionadas.

4.1.3. Proceso ETL

Se realizó ETLs con la herramienta Rapidminer ya que es compatible con los procesos realizados, se procedió a extraer los datos de denuncias, beneficiarios, subsidios, puntos pago, transacciones, ubicación, los cuales se encontraban almacenados en una Base de Datos de Microsoft SQL Server, posterior a esto se carga los datos a un repositorio local de Rapidminer, como se demuestra en la figura 9:

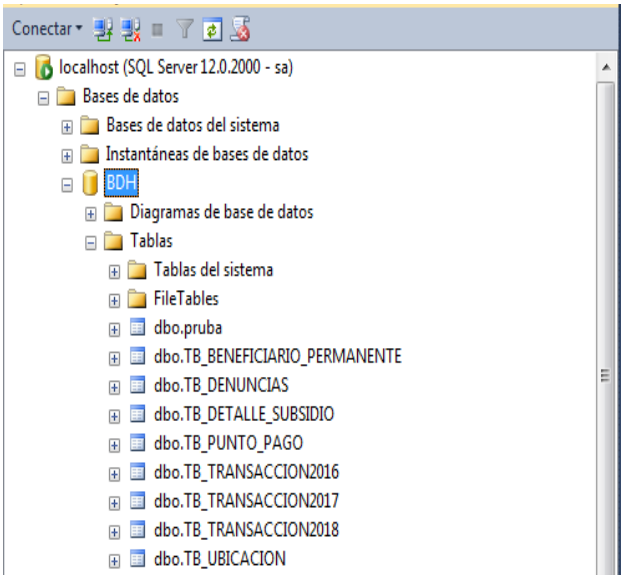


Figura 8. Origen de datos

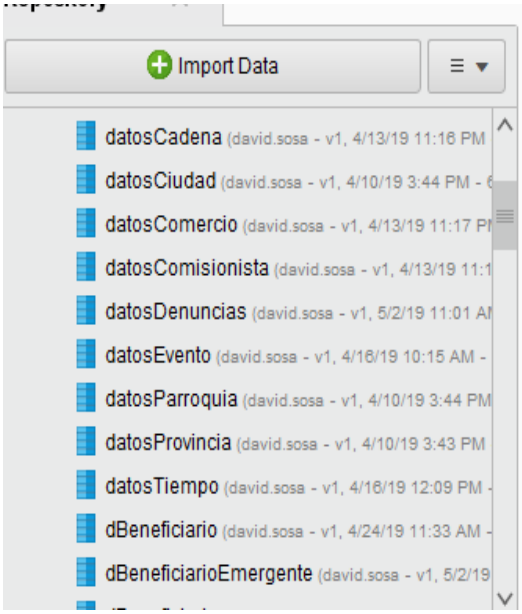


Figura 9. Repositorio local Rapidminer

Los datos almacenados en Rapidminer son depurados, limpiados y trasformados de acuerdo al modelo multidimensional. Se creó seis dimensiones y una tabla de hechos a través de procesos ETLs, estos datos son almacenados en una base de datos DWH_BDH.

Se creó la dimensión Bono (DIM_BONO) como los siguientes procesos:

- Se carga los datos del repositorio local
- Selección de atributos
- Se generó un ID
- Se renombro los atributos de acuerdo a la dimensión
- Se crea el atributo fecha carga
- Se ordena los atributos
- Se escribe en la base de datos SQL Server

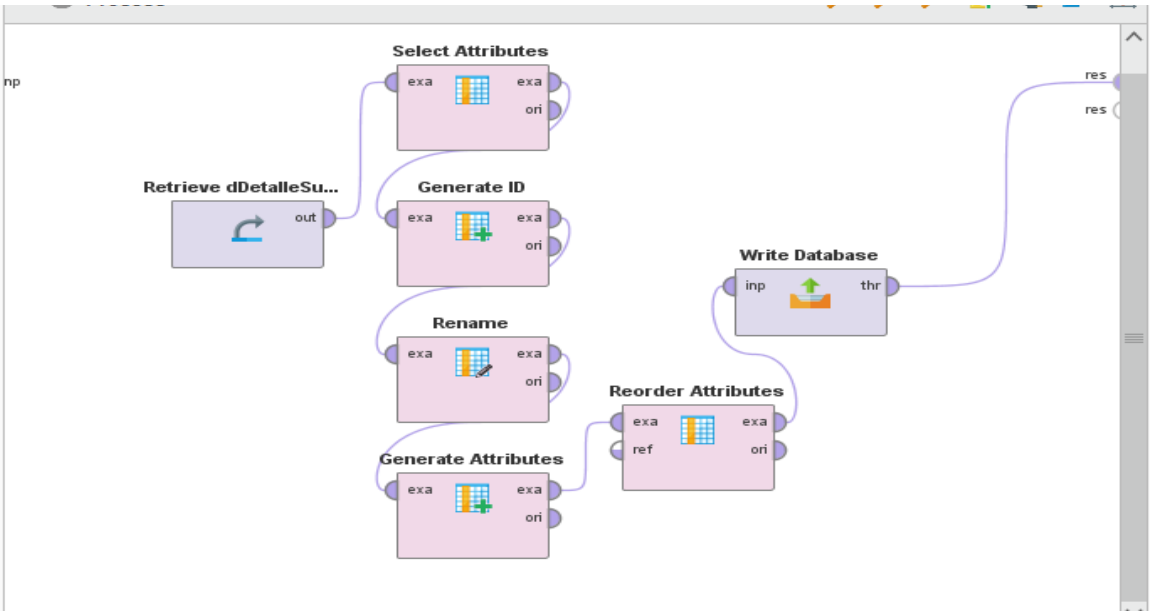


Figura 10. **Proceso ETL para la dimensión Bono.**

Se creó la dimensión Punto pago (DIM_PUNTO_PAGO) como los siguientes procesos:

- Se carga los datos del repositorio local
- Selección de atributos
- Unió de datos
- Se removió los duplicados
- Se generó un ID
- Se renombro los atributos de acuerdo a la dimensión
- Se crea el atributo fecha carga
- Se escribe en la base de datos SQL Server

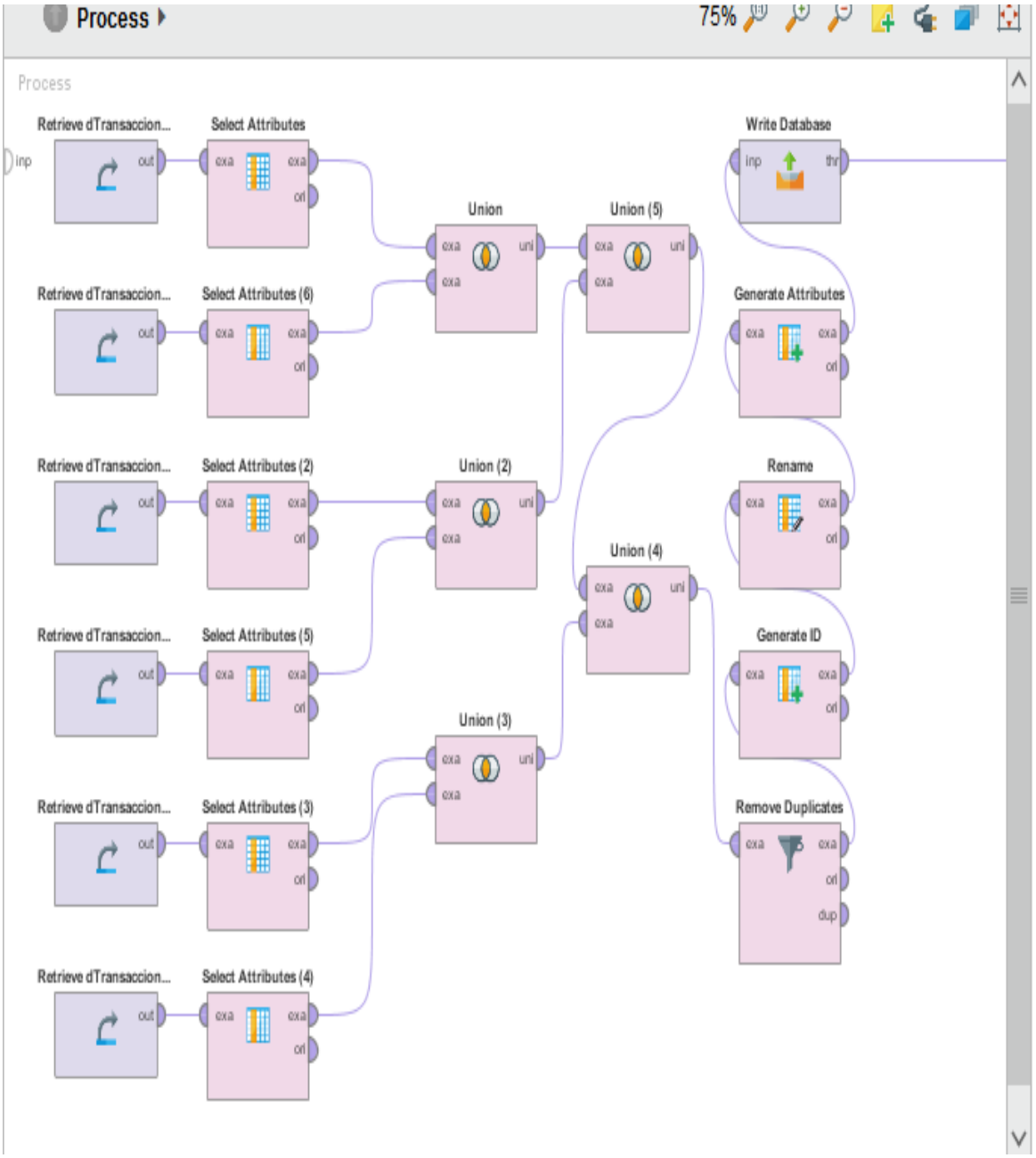


Figura 11. Proceso ETL para la dimensión Punto Pago

Se creó la dimensión Tiempo (DIM_TIEMPO) como los siguientes procesos:

- Se carga los datos del repositorio local
- Selección de atributos
- Se removió los duplicados
- Se generó un ID
- Se renombro los atributos de acuerdo a la dimensión
- Se escribe en la base de datos SQL Server

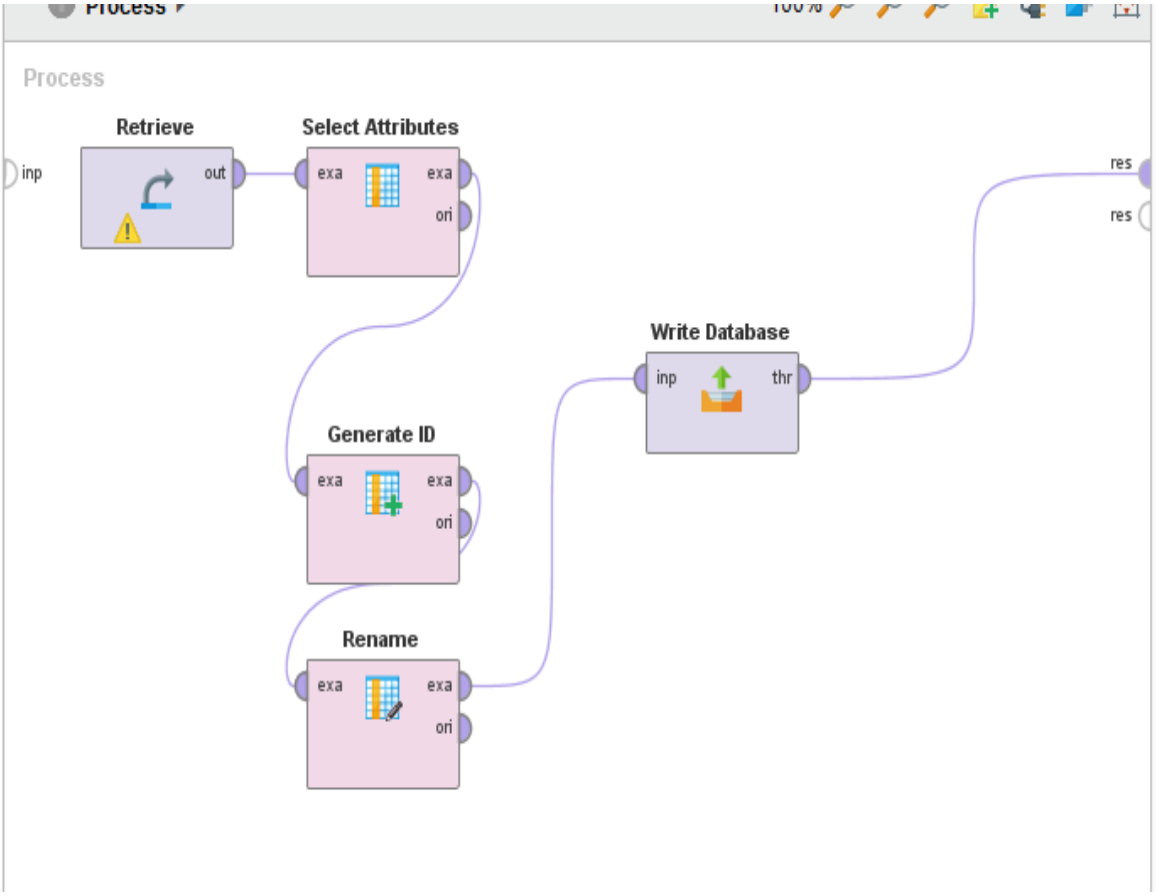


Figura 12. Proceso ETL para la dimensión Tiempo

Se creó la dimensión Ubicación (DIM_UBICACION) como los siguientes procesos:

- Se carga los datos del repositorio local
- Selección de atributos
- Se generó un ID
- Genero el atributo fecha carga
- Se renombro los atributos de acuerdo a la dimensión
- Se escribe en la base de datos SQL Server

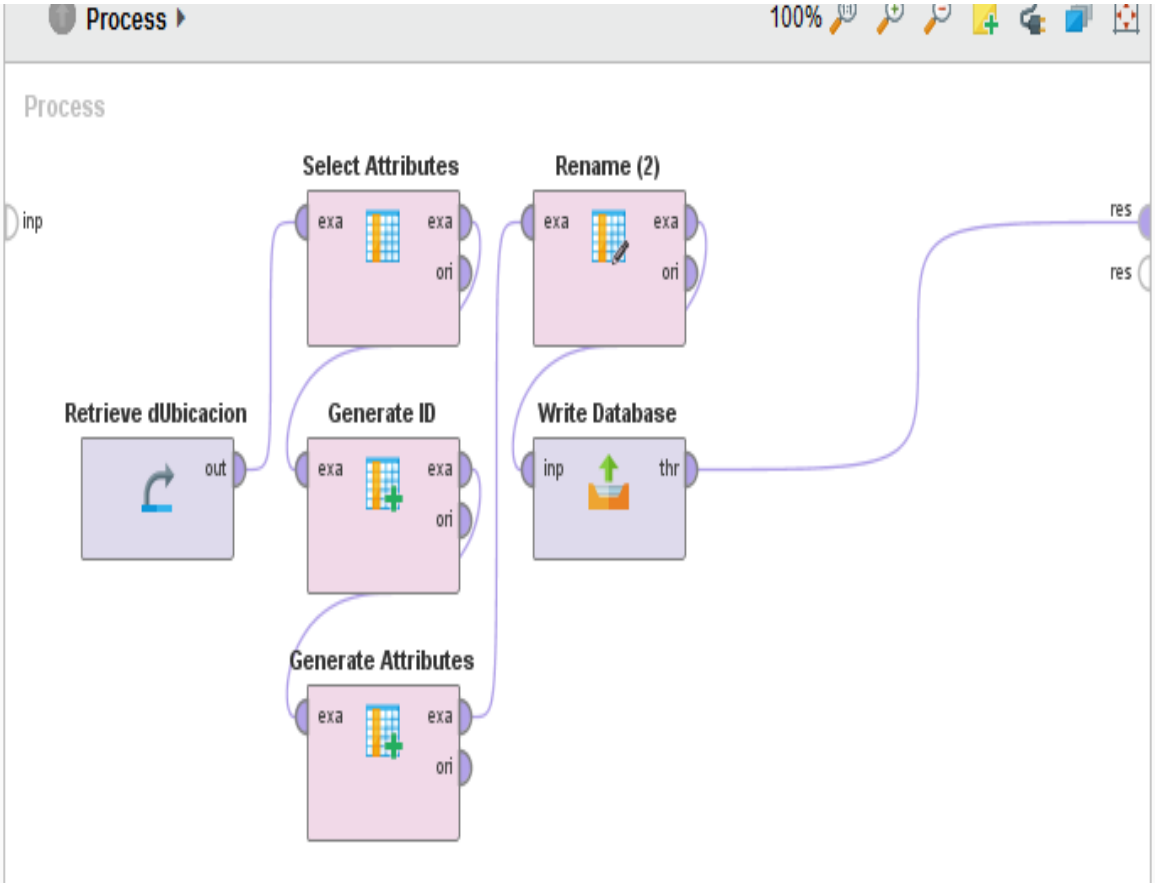


Figura 13. Proceso ETL para la dimensión Ubicación

Se creó la dimensión Beneficiario (DIM_BENEFICIARIO) como los siguientes procesos:

- Se carga los datos del repositorio local
- Selección de atributos
- Se generó un ID
- Se renombro los atributos según la dimensión
- Se removió los duplicados
- Reemplazo de valores faltantes
- Se ordenó los atributos
- Se escribe en la base de datos SQL Server

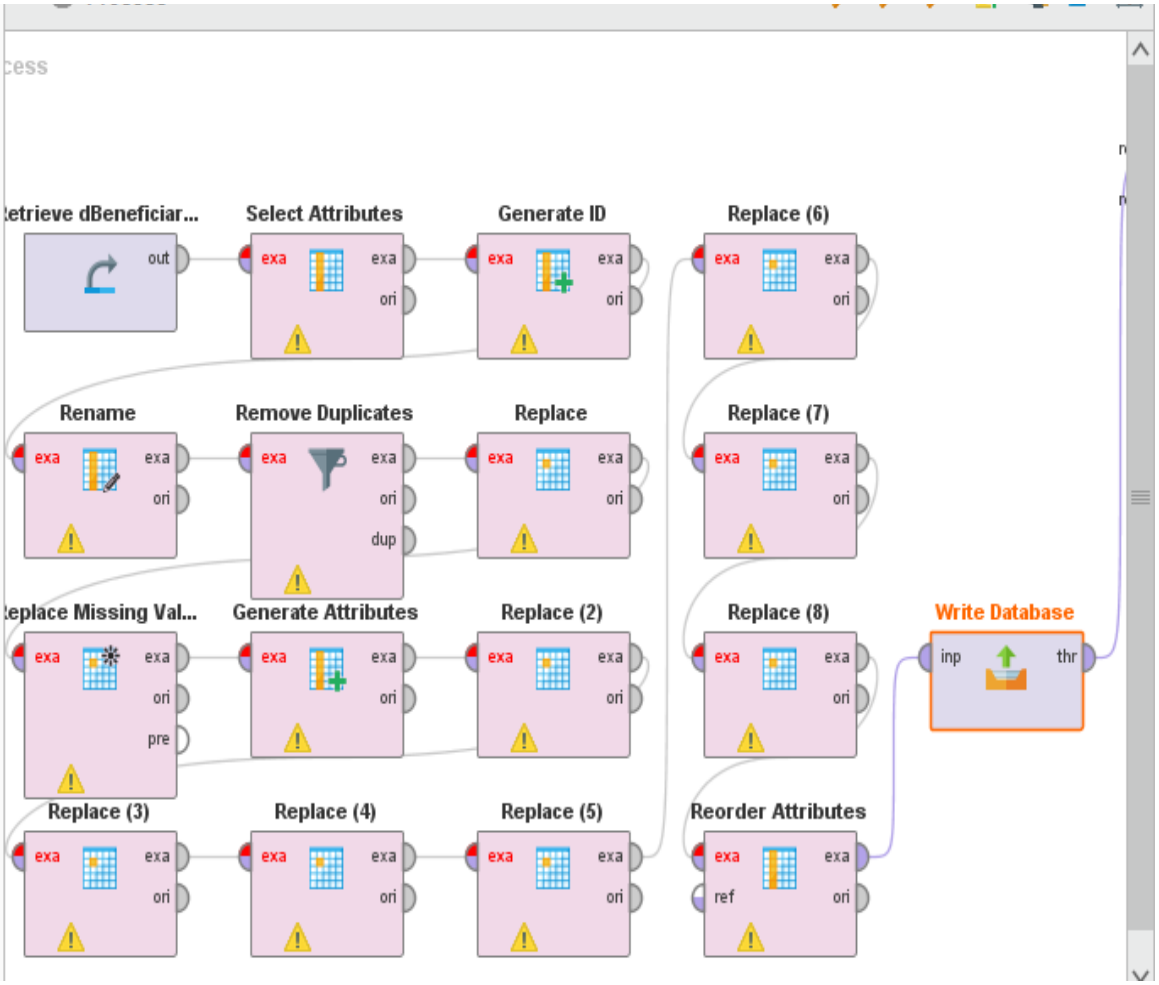


Figura 14. Proceso para ETL para la dimensión Beneficiario

Se creó la dimensión Denuncias (DIM_DENUNCIAS) como los siguientes procesos:

- Se carga los datos del repositorio local
- Selección de atributos
- Se generó un ID
- Se renombro los atributos según la dimensión
- Reemplazo de valores faltantes
- Reemplazo de valores
- Se ordenó los atributos
- Se escribe en la base de datos SQL Server

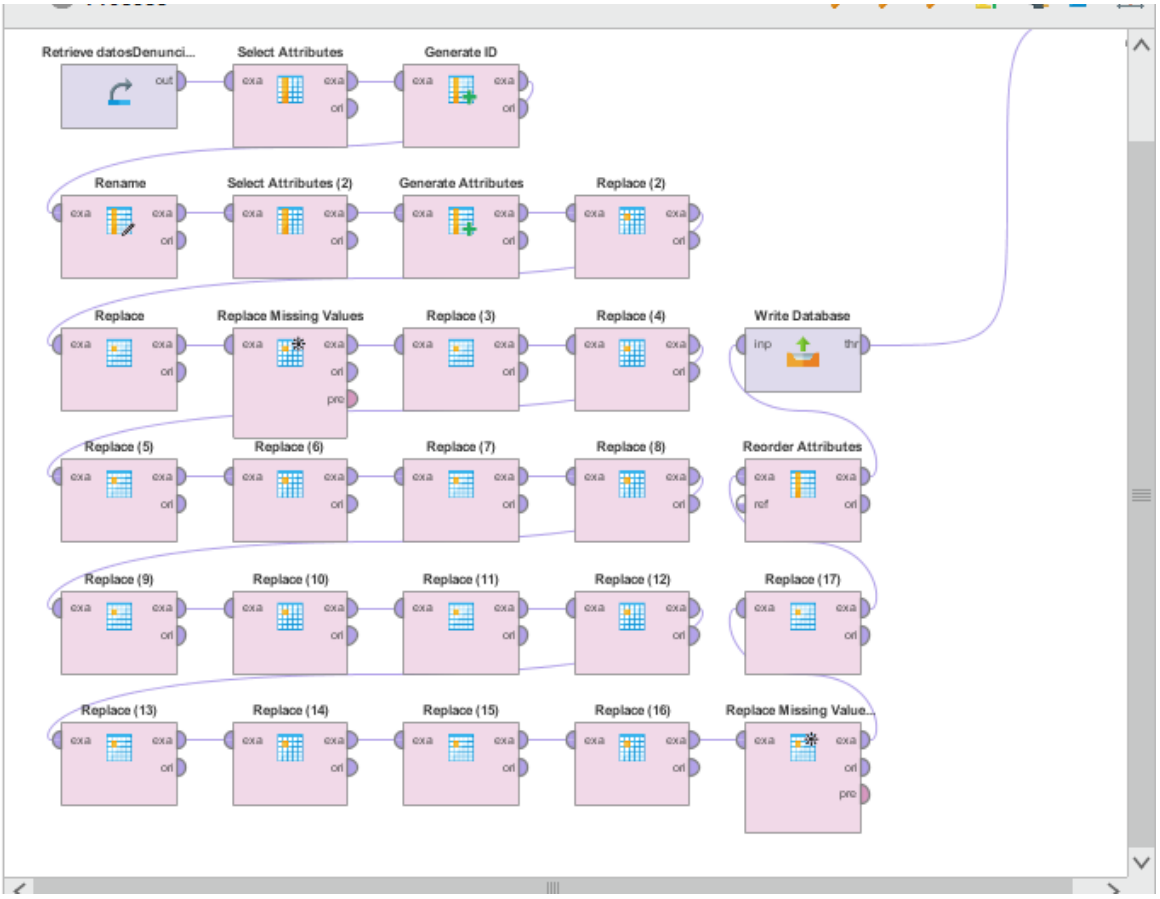


Figura 15. Proceso ETL para la dimensión Denuncias

Para la tabla de hechos se carga en tres fases Fac_Transaccion2016, Fac_Transaccion2017, Fac_Transaccion2018, debido a la cantidad de registros que se encuentran almacenados se realiza una muestra estratificada.

Se creó la tabla de hechos (FAC_TRANSACCION2016) como los siguientes procesos:

- Se carga los datos de las dimensiones
- Selección de atributos
- Filtro para la dimensión denuncias
- Se realiza uniones por atributos de las dimensiones
- Se escribe en la base de datos SQL Server

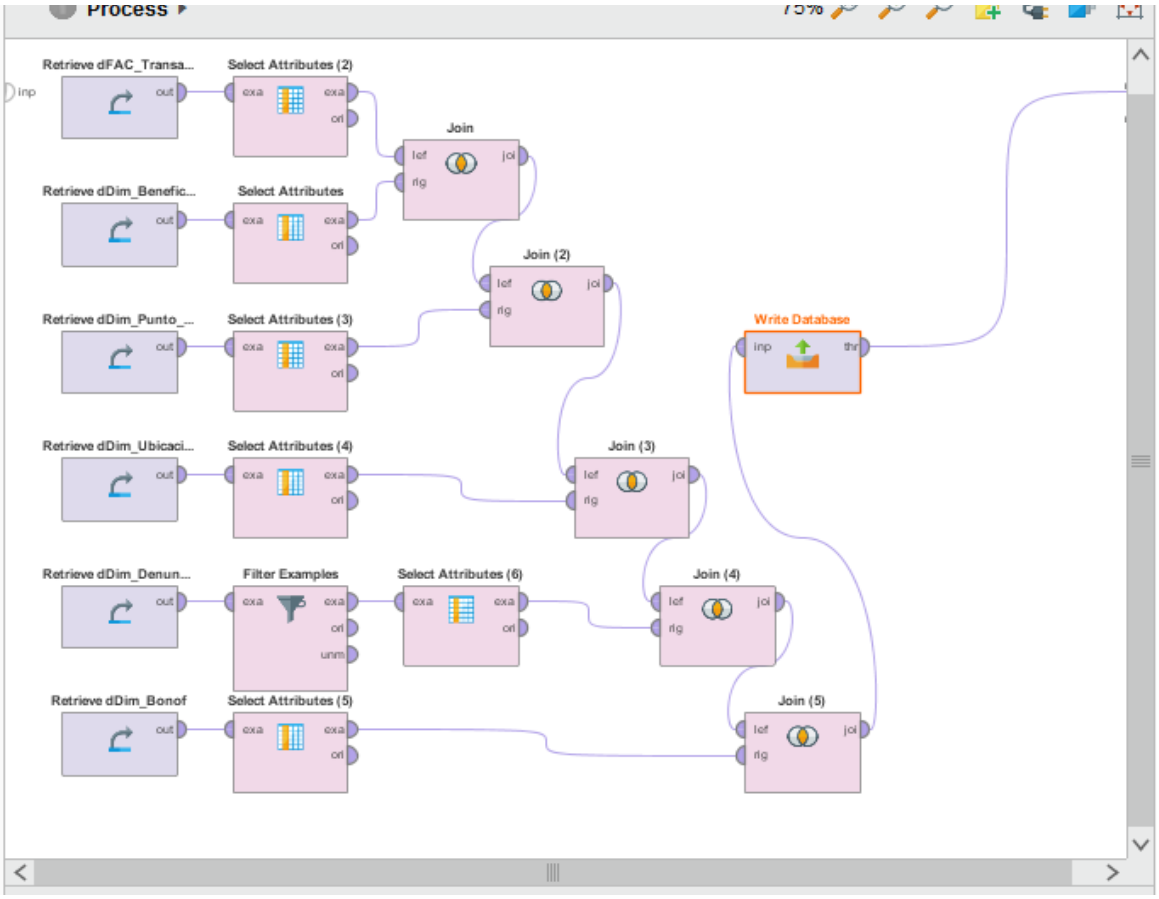


Figura 16. Proceso ETL para la carga de la Tabla de hechos 2016

Se creó la tabla de hechos (FAC_TRANSACCION2017) como los siguientes procesos:

- Se carga los datos de las dimensiones
- Selección de atributos
- Filtro para la dimensión denuncias
- Se realiza uniones por atributos de las dimensiones
- Se escribe en la base de datos SQL Server

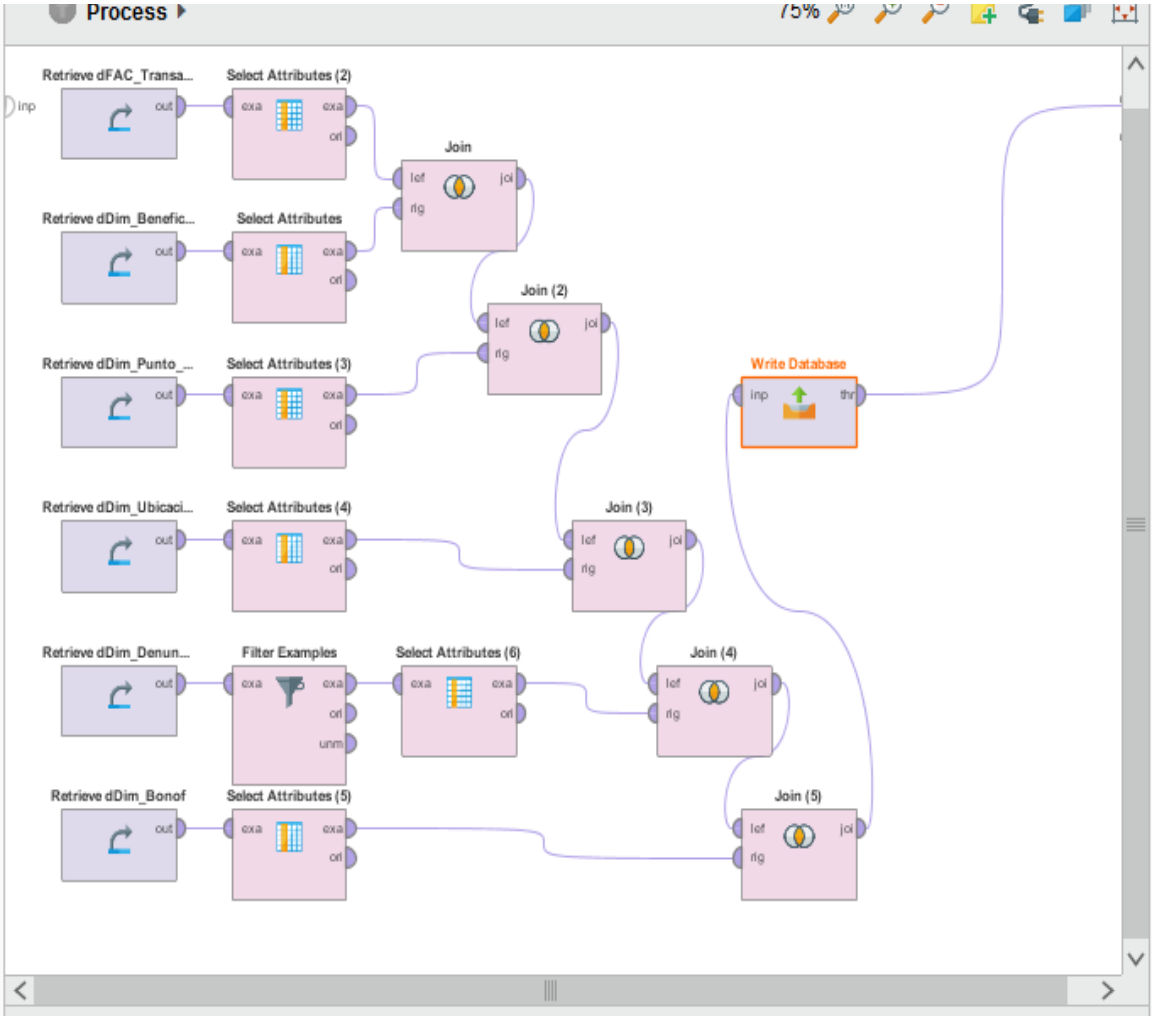


Figura 17. Proceso ETL para la carga de la Tabla de hechos 2017

Se creó la tabla de hechos (FAC_TRANSACCION_2018) como los siguientes procesos:

- Se carga los datos de las dimensiones
- Selección de atributos
- Filtro para la dimensión denuncias
- Se realiza uniones por atributos de las dimensiones
- Se escribe en la base de datos SQL Server

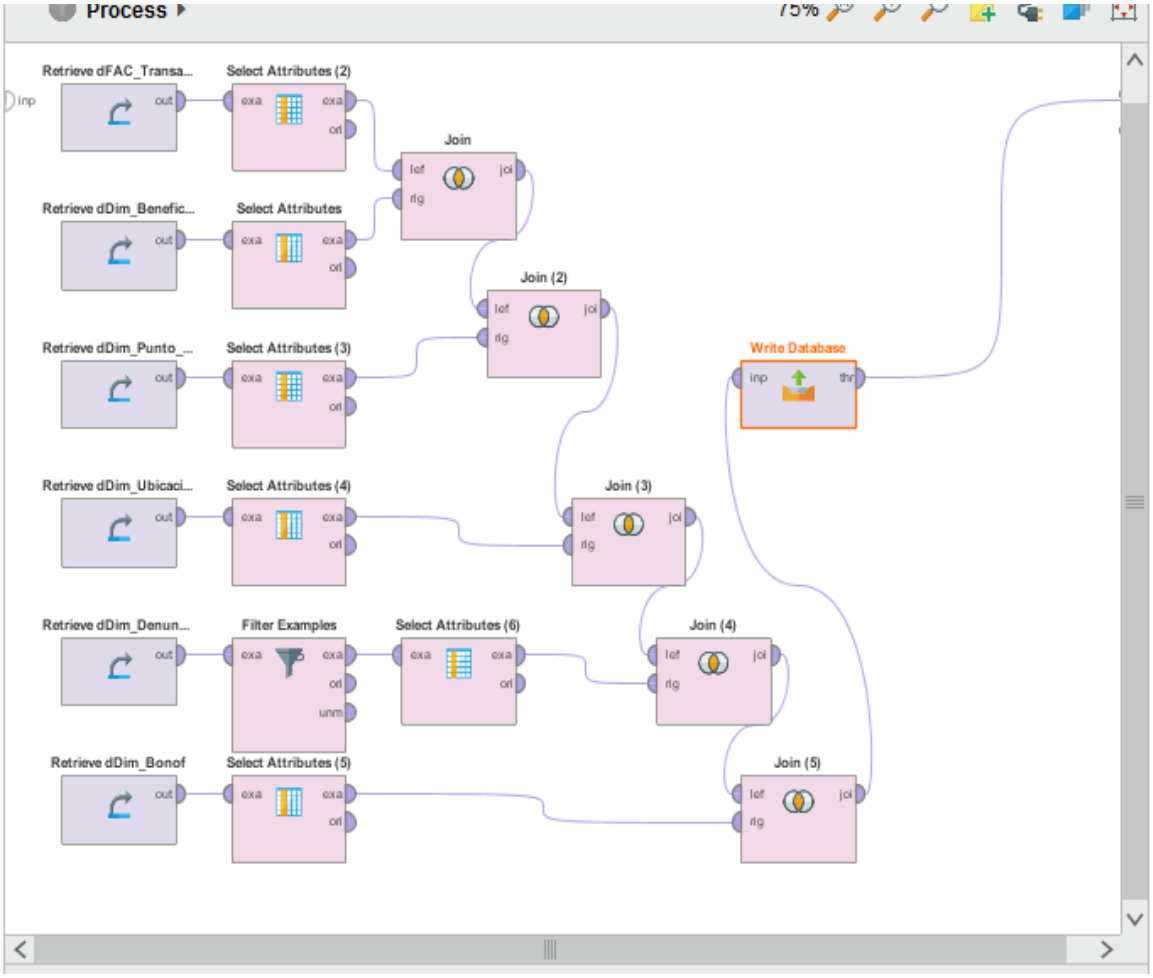


Figura 18.Proceso ETL para la carga de la Tabla de hechos 2018

4.1.4. Base de datos

Se utilizó el motor de base de datos SQL Server para poder crear la base de datos BDH la cual fue utilizada para cargar los datos de las tablas relacionales de acuerdo al objetivo planteado, como se muestra en la figura 19:

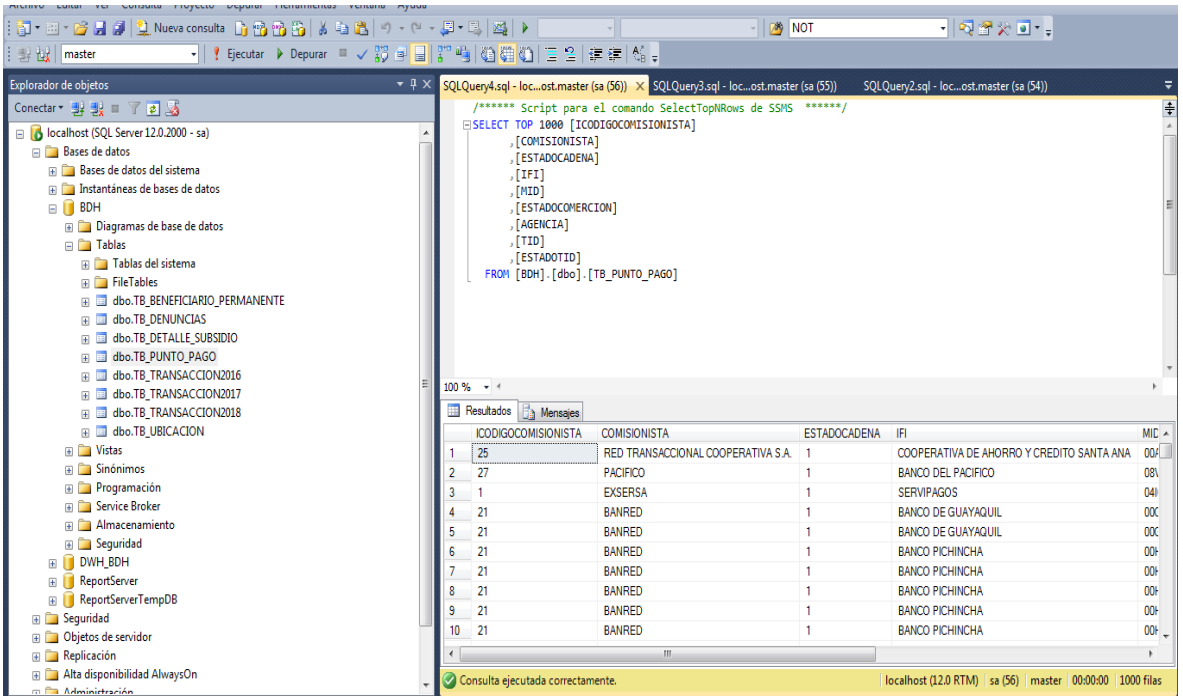


Figura 19. Base de Datos BDH

Se crea la conexión entre la herramienta Rapidminer y SQL Server con el nombre SQL_CONEXION_BD para poder visualizar y cargar los datos que se tiene en la base de datos BDH, como se muestra en la figura:

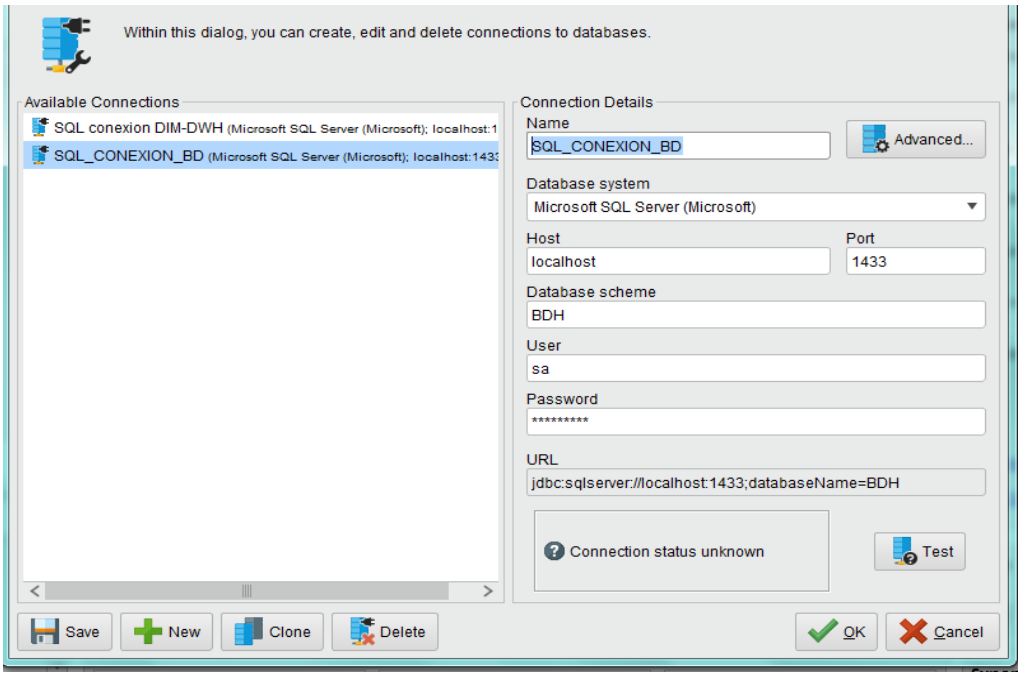


Figura 20. Conexión con la base de datos

Para poder almacenar los datos de las dimensiones y la tabla de hechos se crea la base de datos DWH_BDH y su respectiva conexión como se muestra en la figura 21:

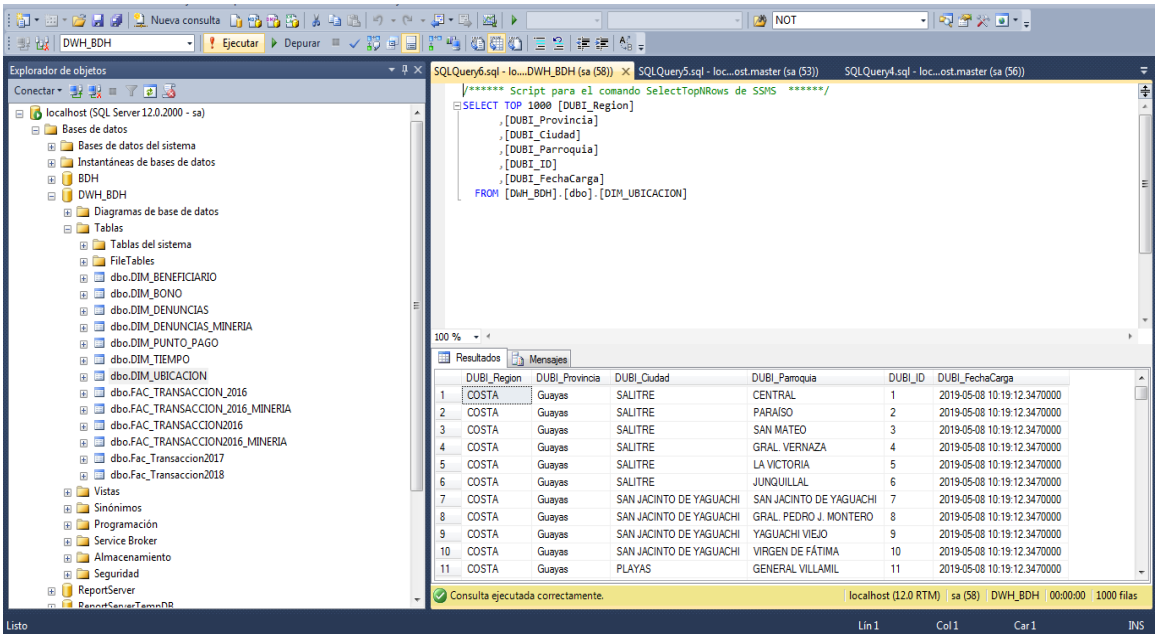


Figura 21. Data warehouse BDH

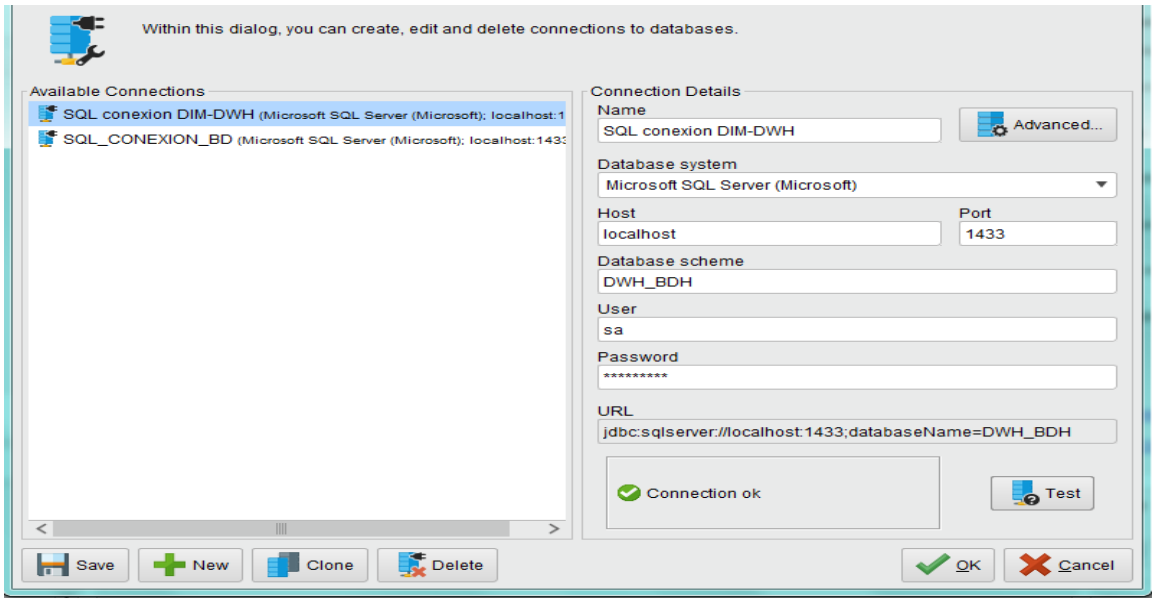


Figura 22. Conexión desde Rapidminer a SQL Server

La base de datos DWH_BDH fue cargada con todas las dimensiones y tabla de hechos, de igual manera se cargó el repositorio local con sus respetivos procesos ETL para poder crear esta bodega de datos y posteriormente fue utilizada para minería de datos, como se muestra en las figuras 23 y 24:

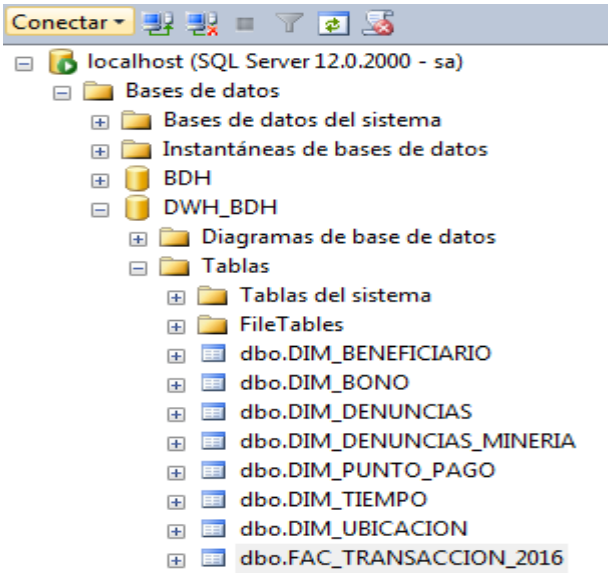


Figura 23. Dimensiones y tabla de hechos

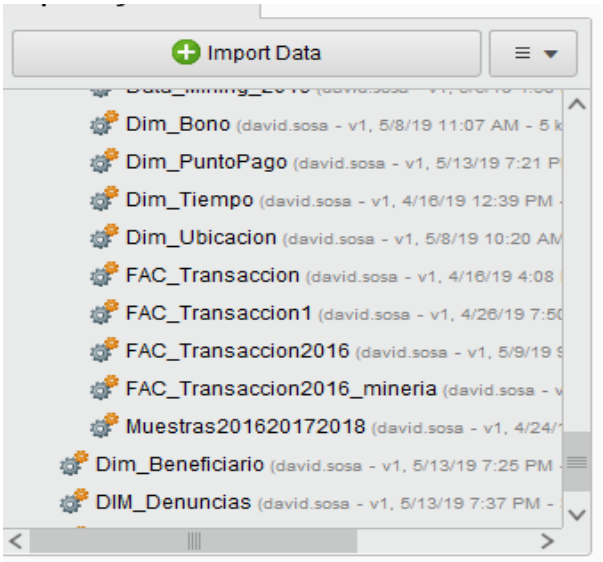


Figura 24. Repositorio local de Rapidminer

4.1.5. Auto Model RapidMiner

Rapidminer cuenta con la extensión Auto Model para poder crear y validar modelo.Con la bodega de datos creada se podrá:

- Cargar los datos de la bodega creada
- Seleccionarlos
- Prepararlos
- Seleccionar los factores de entrada
- Seleccionar el modelo que más se acople al proyecto.
- Finalmente se observara los resultados obtenidos y se podrá realizar simulaciones

Se procedió a cargar la bodega de datos con la cual se encuentra cargada con las dimensiones y tabla de hechos, en este paso se analizó las denuncia registradas en el año 2016, 2017,2018.

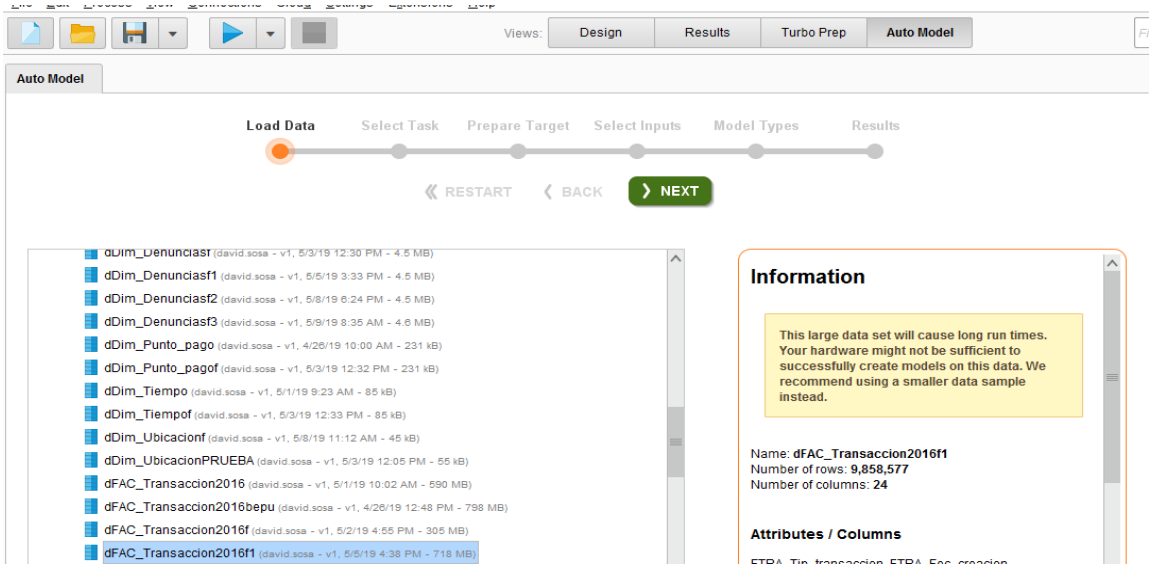
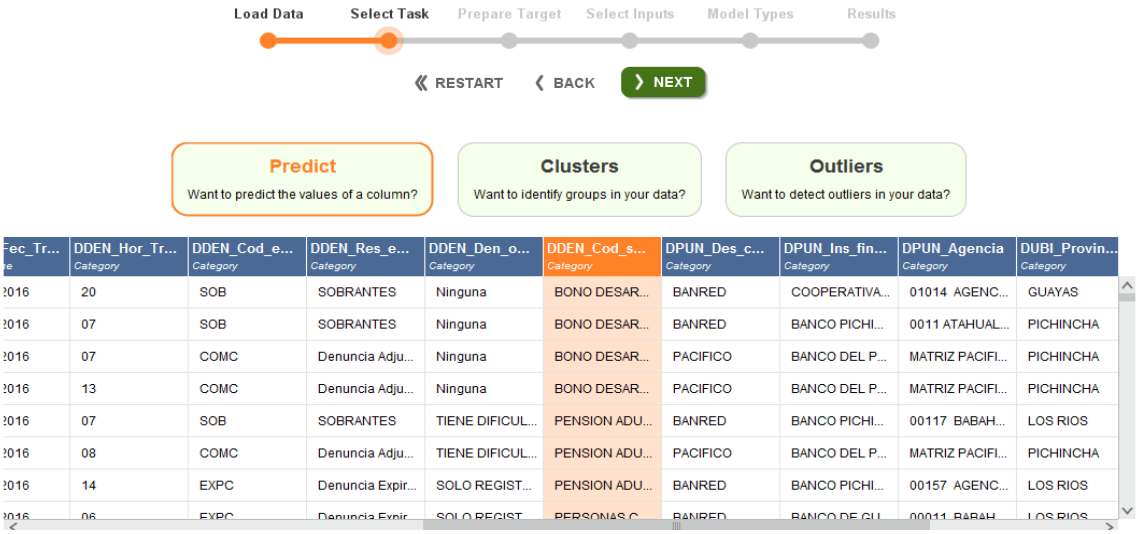


Figura 25. Auto Model para la cargar datos

Se escoge Predecir (Predict) los valores de la comuna DDEN_Cod_subsidio para poder desarrollar un modelo de aprendizaje automático y poder resolver el problema de las denuncias sobre cobros indebidos de los diferentes tipos bonos y pensiones que entrega el Estado a personas en situación de pobreza y extrema pobreza.



Fec_Tr...	DDEN_Hor_Tr...	DDEN_Cod_e...	DDEN_Res_e...	DDEN_Den_o...	DDEN_Cod_s...	DPUN_Des_c...	DPUN_Ins_fin...	DPUN_Agencia	DUBI_Provin...
	Category	Category	Category	Category	Category	Category	Category	Category	Category
2016	20	SOB	SOBRANTES	Ninguna	BONO DESAR...	BANRED	COOPERATIVA...	01014 AGENC...	GUAYAS
2016	07	SOB	SOBRANTES	Ninguna	BONO DESAR...	BANRED	BANCO PICHI...	0011 ATAHUAL...	PICHINCHA
2016	07	COMC	Denuncia Adju...	Ninguna	BONO DESAR...	PACIFICO	BANCO DEL P...	MATRIZ PACIFI...	PICHINCHA
2016	13	COMC	Denuncia Adju...	Ninguna	BONO DESAR...	PACIFICO	BANCO DEL P...	MATRIZ PACIFI...	PICHINCHA
2016	07	SOB	SOBRANTES	TIENE DIFICUL...	PENSION ADU...	BANRED	BANCO PICHI...	00117 BABAH...	LOS RIOS
2016	08	COMC	Denuncia Adju...	TIENE DIFICUL...	PENSION ADU...	PACIFICO	BANCO DEL P...	MATRIZ PACIFI...	PICHINCHA
2016	14	EXPC	Denuncia Expir...	SOLO REGIST...	PENSION ADU...	BANRED	BANCO PICHI...	00157 AGENC...	LOS RIOS
2016	06	EXPC	Denuncia Expir...	SOLO REGIST...	PERSONAS C...	BANRED	BANCO DE GU...	00011 BABAH...	LOS RIOS

Figura 26. Proceso de predicción con Subsidio

En este paso se trasforma el atributo a predecir en uno que categoriza los diferentes tipos de bono y pensiones. Mediante el diagrama de barras se puede visualizar que Pensión Adulto Mayor tiene mayor número de denuncias (6,751) y Bono Variable tiene un menor número de denuncias (1.290) a nivel nacional desde el año 2016 hasta el 2018.

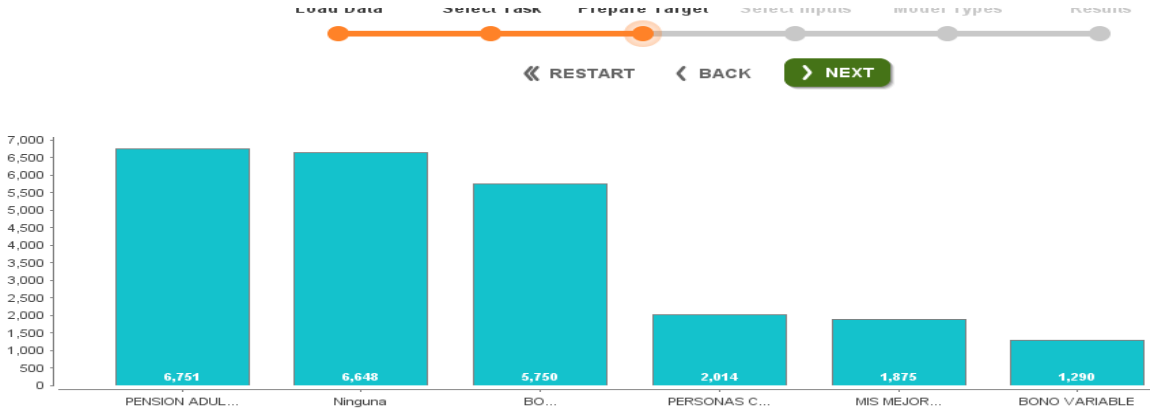


Figura 27. Margen de denuncias por tipos de bono y pensión

En este paso se seleccionó los factores de entrada y Rapidminer mostró un porcentaje de correlación que existe en las columnas, con lo cual se escoge los atributos con mayor porcentaje.

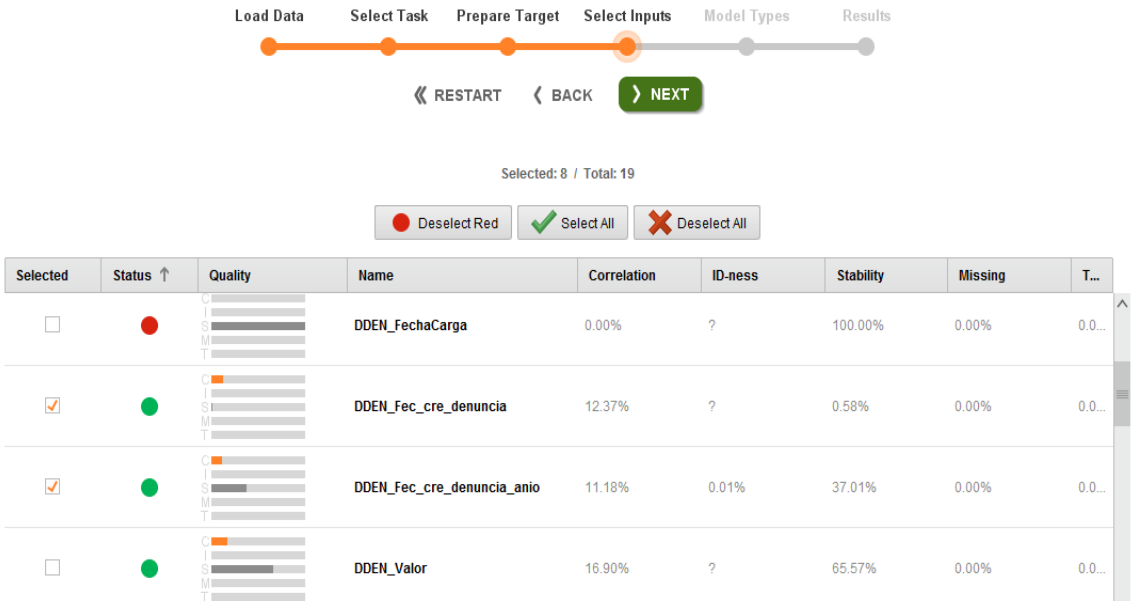


Figura 28. Selección de factores de entrada

Rapidminer presenta diferentes modelos para poderlos aplicar y determinar el porcentaje de precisión o error.

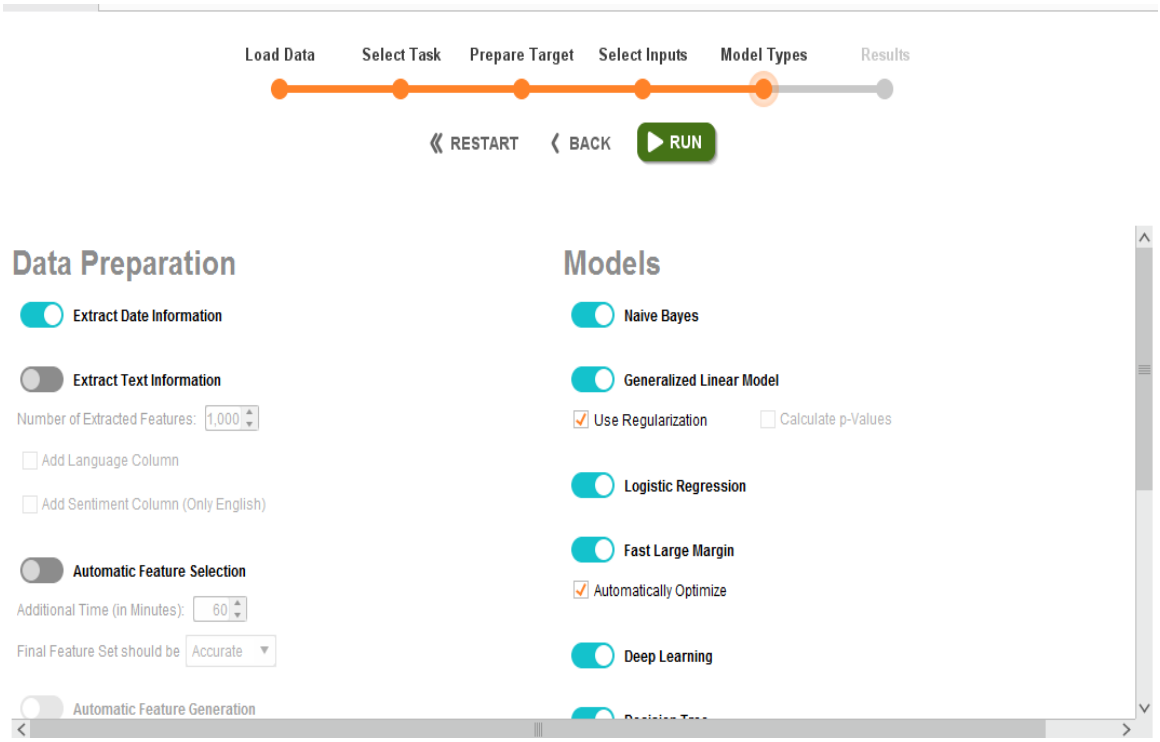


Figura 29. Selección de modelos

En los resultados de los modelos escogidos anteriormente se puede observar que el modelo Naive Bayes y Decision Tree tiene mayor exactitud frente a los otros modelos.

- Naive Bayes es una técnica de clasificación y predicción, ayuda a la categorización de datos.
- Decision Tree toma la forma de un árbol conformado por nodos los cuales se puede visualizar las diferentes decisiones de acuerdo a los atributos que se maneja

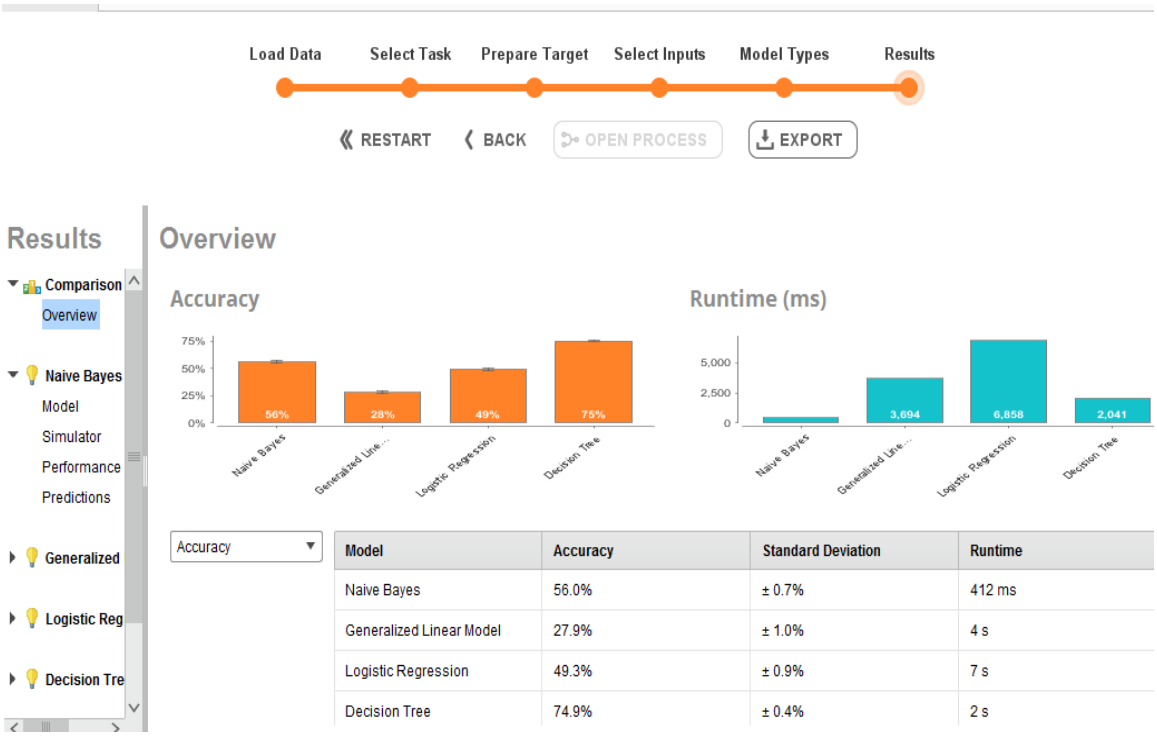


Figura 30. Comparación de los diferentes modelos

En el modelo de Naive Bayes se puede visualizar que en las Provincias de Pichincha, Guayas, Manabí, Los Ríos, Esmeraldas tiene mayor índice de denuncias por la entrega de bonos y pensiones a nivel nacional.

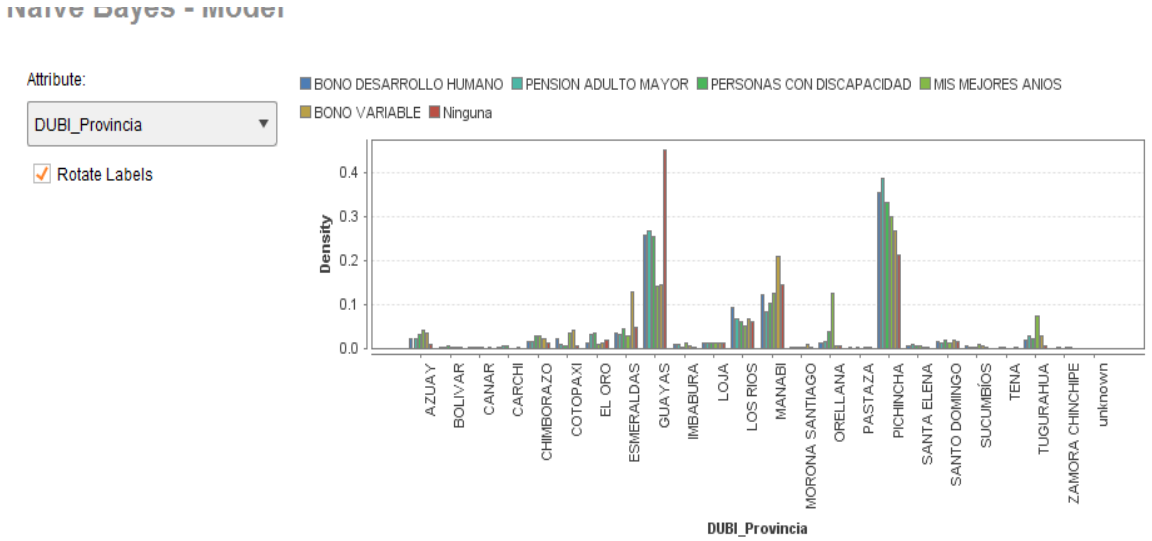


Figura 31. Provincia con mayor porcentaje de denuncias

Adicional, los Sistemas Auxiliares de Pago (concentradores) Banred, Pacifico y Exsersa tiene el índice más alto de denuncias por bonos y pensiones entregados y se confirma que Pensión Adulto Mayor y Bono Desarrollo Humano tienen una tendencia de seguir incrementando debido a la vulnerabilidad y desinformación de los usuarios que reciben este beneficio.

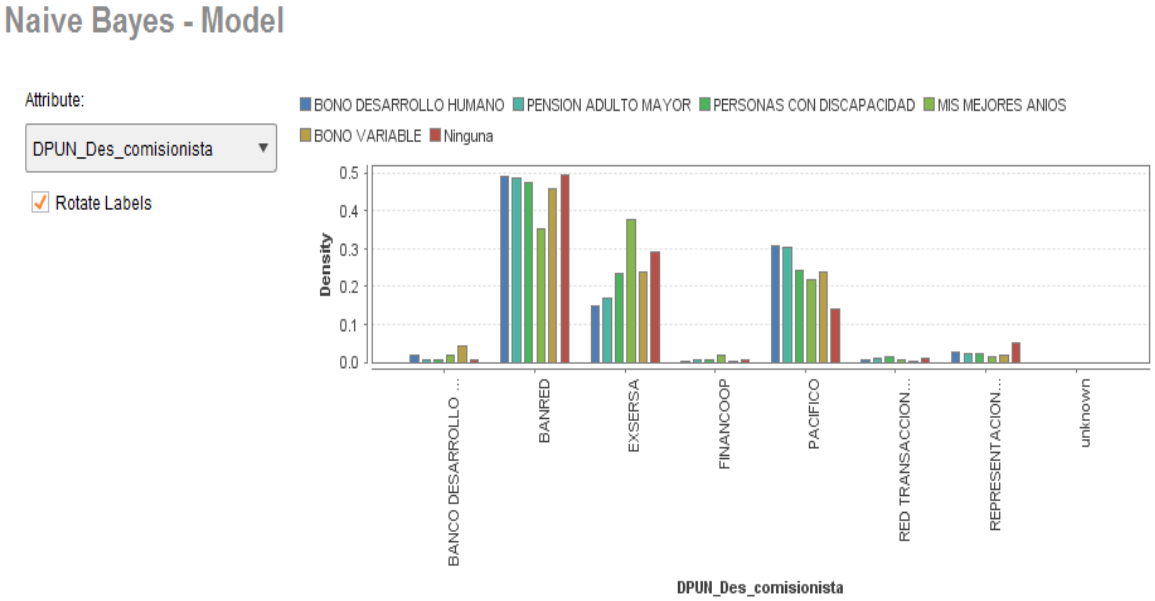


Figura 32. Número de denuncias por tipo de bono y pensión

Se analizó la hora de transacción de la entrega de los diferentes tipos de bonos y pensiones y se visualiza que a las 8:00 hasta las 10:00 am hay mayor transaccionabilidad, pero como dato adicional desde las 21:00 hasta las 06:00 se registran transacciones.

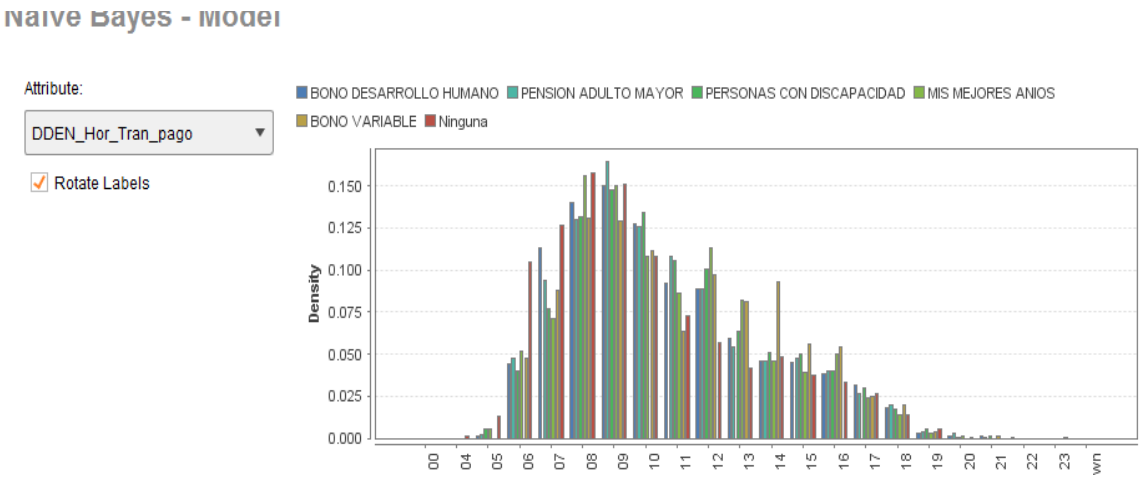


Figura 33. Hora en las cuales se transaccional

En el año 2016,2017 y 2018 se recepto a través de la mesa de servicio de la Institución Pública 24.328 denuncias a nivel nacional , en el año 2016 se registró 8.676 denuncias, en el año 2017 se registró 6.648 denuncias, en el año 2018 se registró 9.005 denuncias.

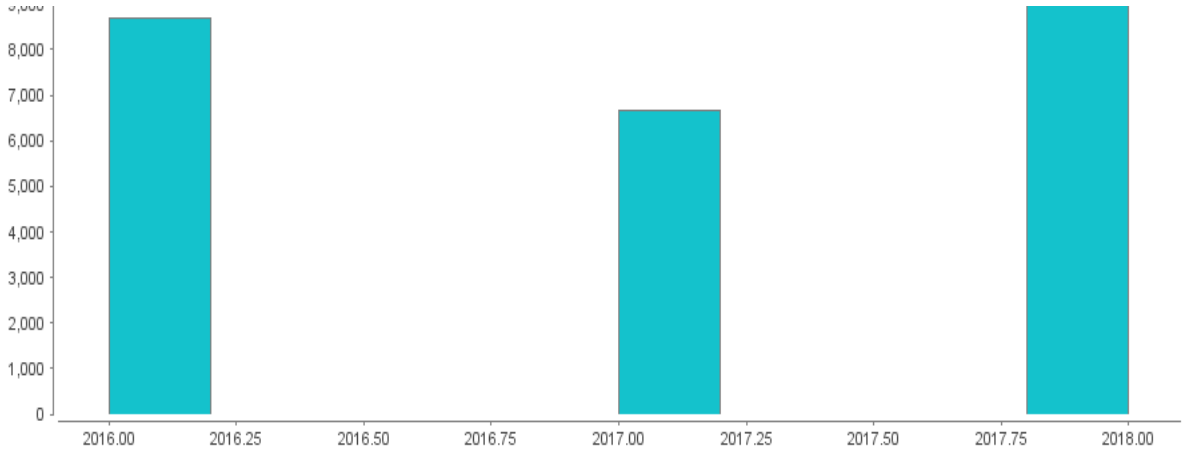


Figura 34. Número de denuncias por año

Con los datos obtenidos anteriormente se creó una simulación para el año 2016 y 2017 dando como resultado que el 51% son denuncias por Pensión Adulto Mayor y 40 % son denuncias del Bono Desarrollo Humano en Provincia de Pichincha con el concentrador Banred.

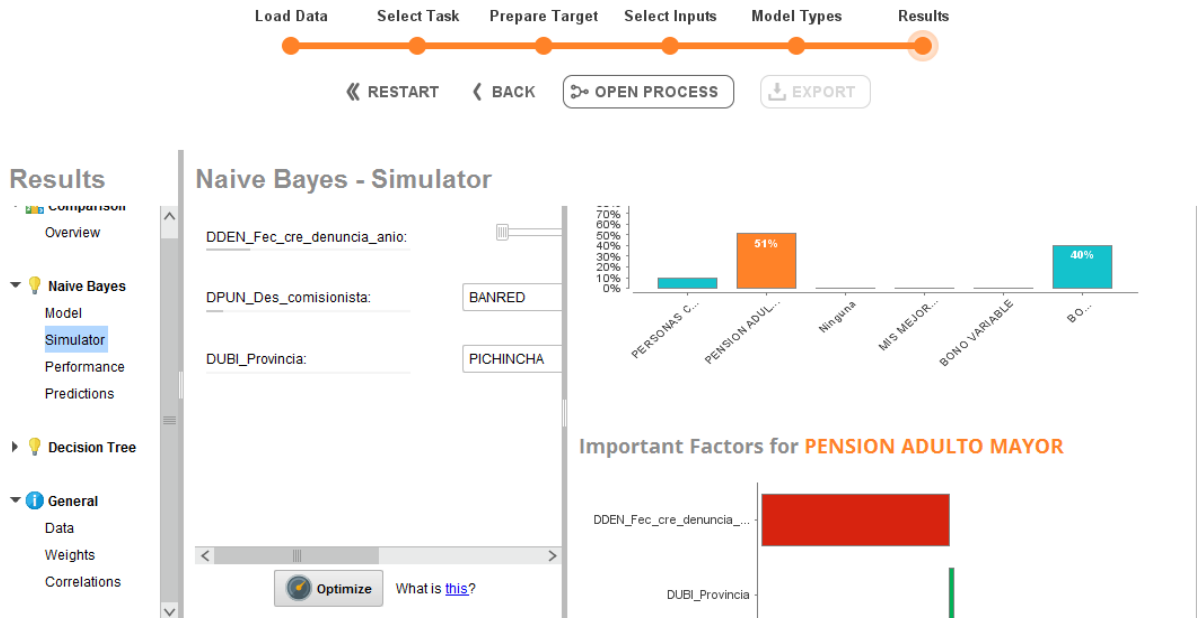


Figura 35. Tendencia por el tipo de bono y pensión

Mientras que en el año 2018 surge una variación con un incremento de denuncias del 56% para Mis Mejores Años y 44% para el Bono Variable. Dando nuevamente una tendencia del vulnerabilidad a los adultos mayores frente a los dos diferentes bonos que reciben (Pensión Adulto Mayor y Mis Mejores años).

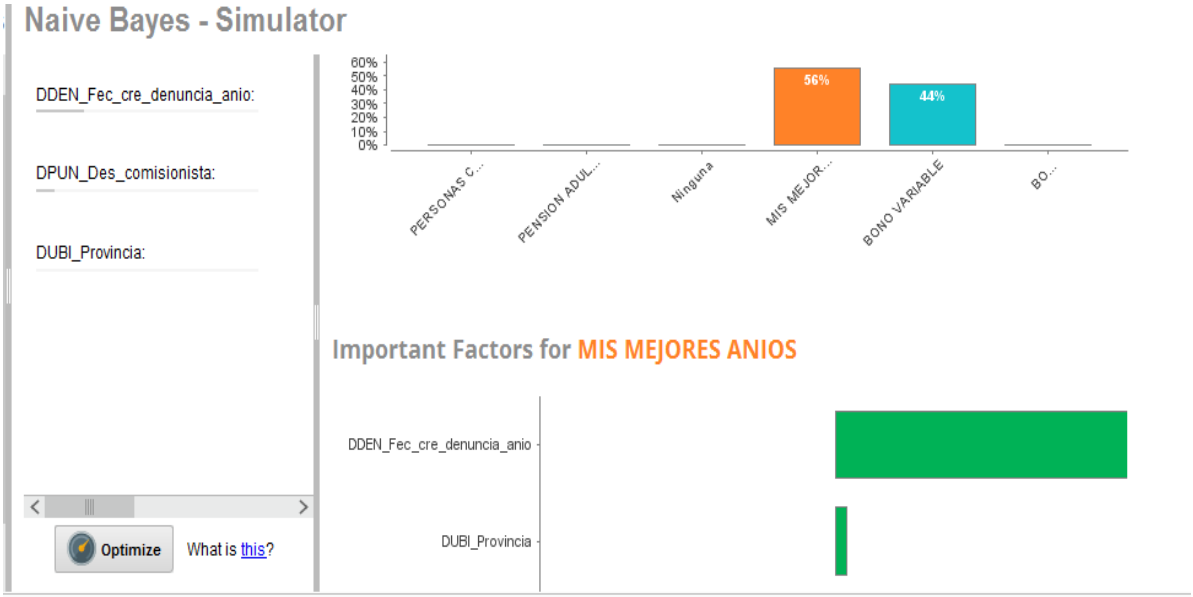
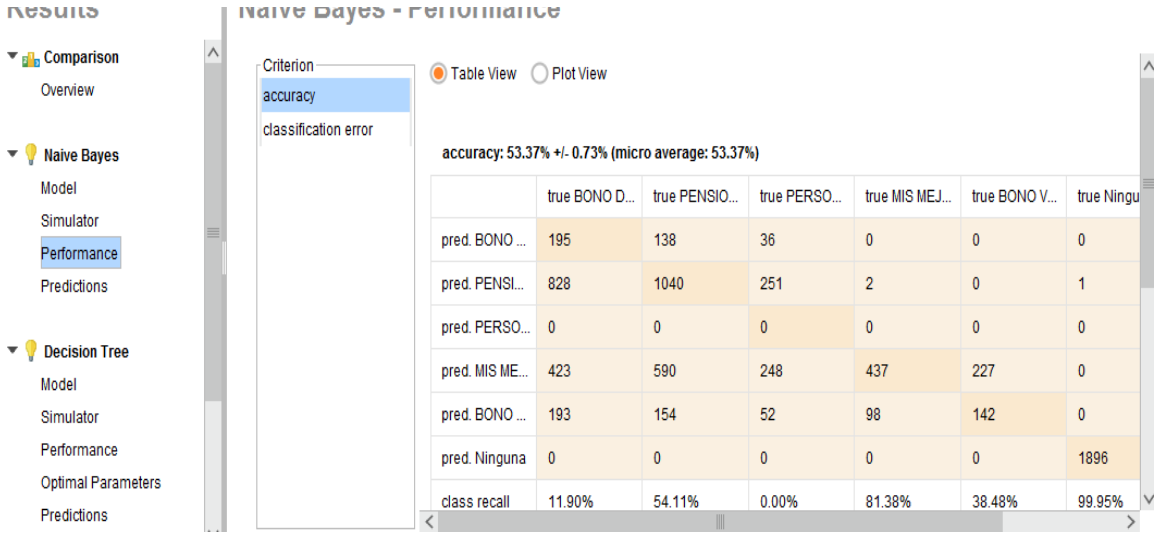


Figura 36. Tendencia por el tipo de bono y periodo

Se puede observar los resultados de acuerdo a la tabla de precisión.



	true BONO D...	true PENSIO...	true PERSO...	true MIS MEJ...	true BONO V...	true Ningu
pred. BONO ...	195	138	36	0	0	0
pred. PENSI...	828	1040	251	2	0	1
pred. PERSO...	0	0	0	0	0	0
pred. MIS ME...	423	590	248	437	227	0
pred. BONO ...	193	154	52	98	142	0
pred. Ninguna	0	0	0	0	0	1896
class recall	11.90%	54.11%	0.00%	81.38%	38.48%	99.95%

Figura 37. Resultados de la tabla de precisión

ANALISIS DE RESULTADOS DE MATRIZ

En el modelo de Decisión Tree se pudo observar que los nodos van por el año, fecha de transacción y los tipos de bonos y pensiones que se entrega.

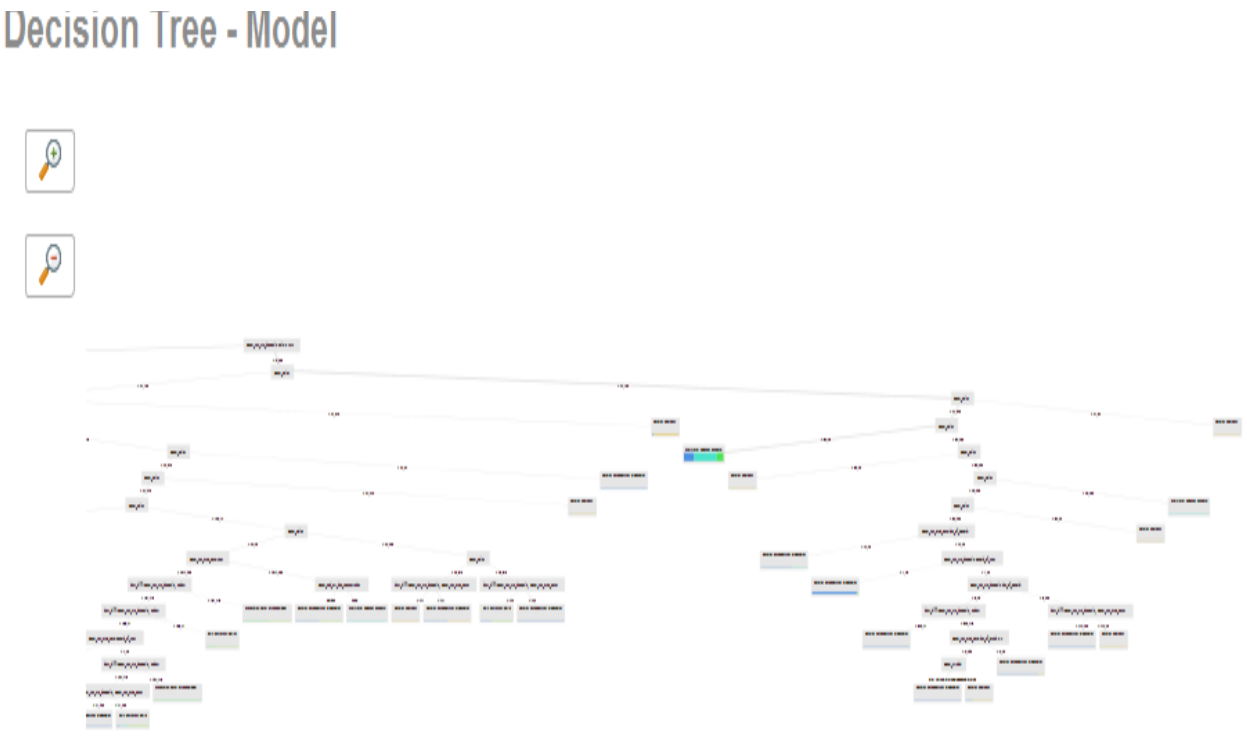


Figura 38. Resultados del modelo Decision Tree

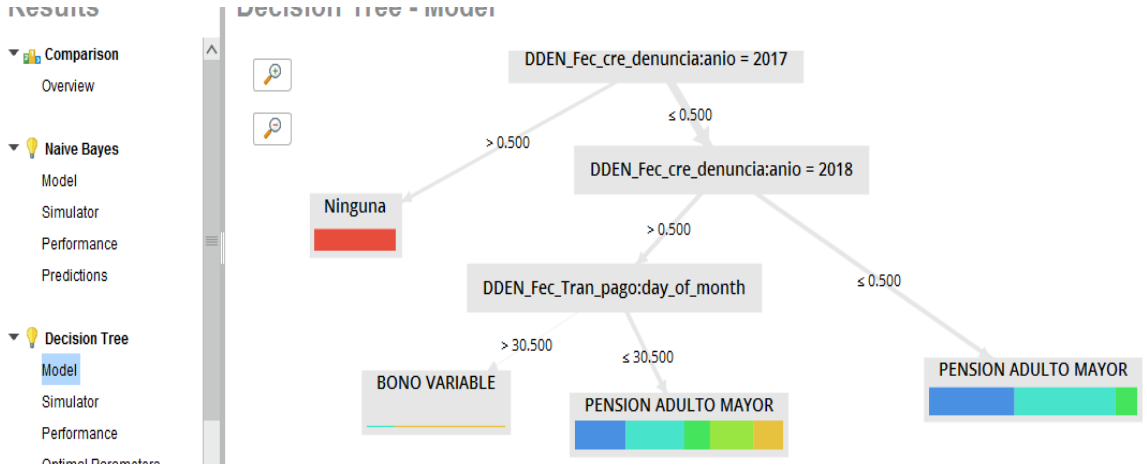


Figura 39. Resultados de Modelo Decision Tree

Además, con el modelo Desisicion Tree se pudo revisar el rendimiento, la clase de precisión de acuerdo al tipo de bono y de pensión.

	true BONO D...	true PENSIO...	true PERSO...	true MIS MEJ...	true BONO V...	true Ninguna	class pre
pred. BONO ...	747	70	18	9	10	0	87.47%
pred. PENSI...	785	1816	546	54	6	0	56.63%
pred. PERSO...	1	0	0	3	0	0	0.00%
pred. MIS ME...	46	73	22	478	30	0	73.65%
pred. BONO ...	19	0	0	1	319	0	94.10%
pred. Ninguna	0	0	0	0	0	1898	100.00%
class recall	46.75%	92.70%	0.00%	87.71%	87.40%	100.00%	

Figura 40. Resultados de precisión

Complementando en la predicción se puede visualizar que Pensión Adulto Mayor tiene una mayor tendencia a ir incrementando y detectando las vulnerabilidades que tiene frente a los otros bonos.

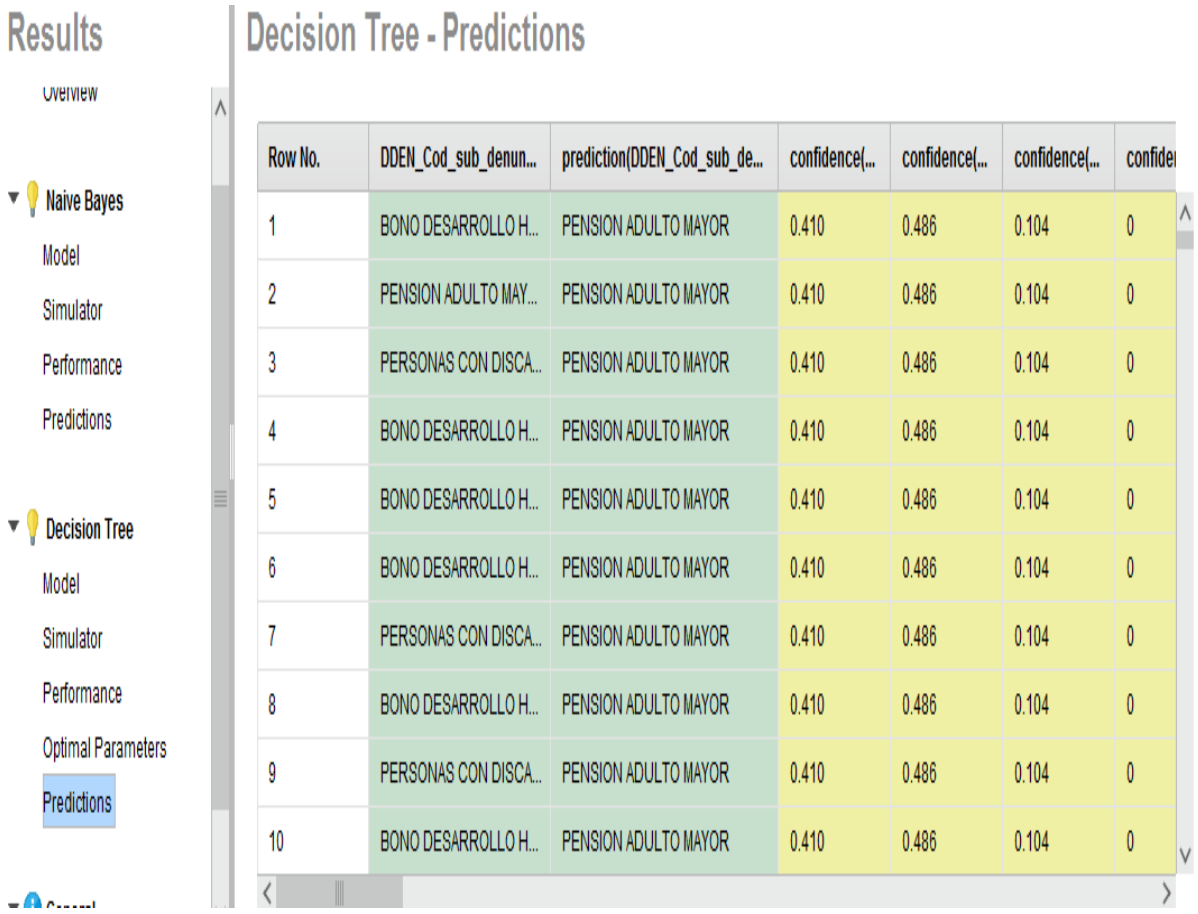


Figura 41. Predicción de tipo de bono o pensión

Se abre el proceso para visualizar el flujo que se creó y se verifica que los datos sean confiables para poder trabajar con este modelo. En la primera parte se carga los datos, se realiza la depuración, se selecciona la columna a predecir tipos de bonos y pensiones, entrada de factores ubicación, puntos de pago, hora, año con el modelo Naive Bayes, pasando por el modelo de simulación y posterior de predicción, como se demuestra a continuación:

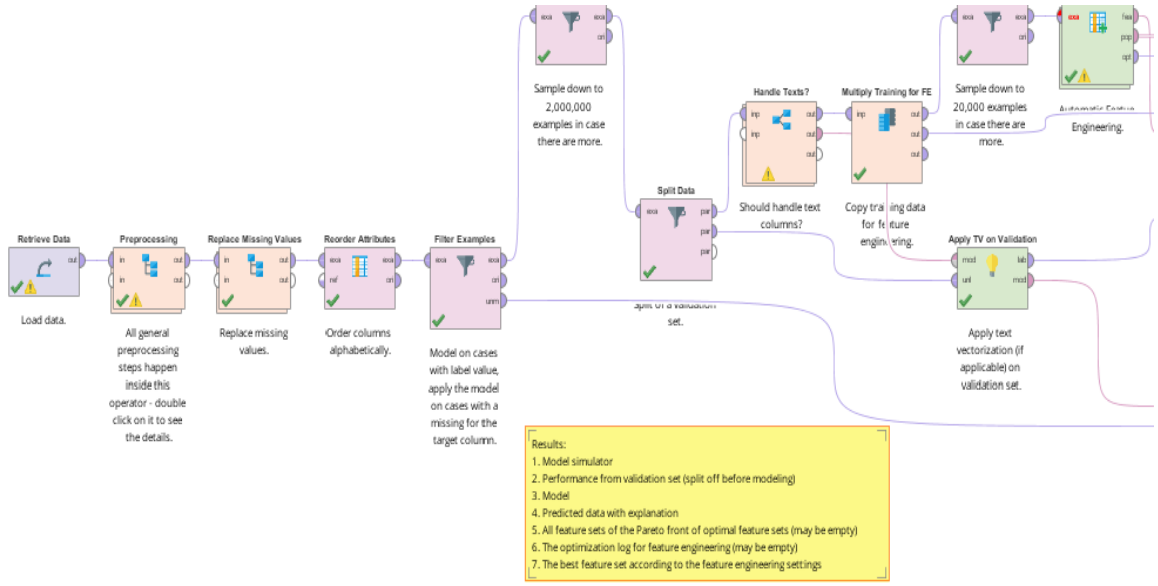


Figura 42. Proceso del modelo Naive Bayes en RapidMiner parte 1

Adicional, se escoge los dos tipos de modelos Naive Bayes y Decision Tree como los cuales se realizan las simulaciones, rendimiento y predicción.

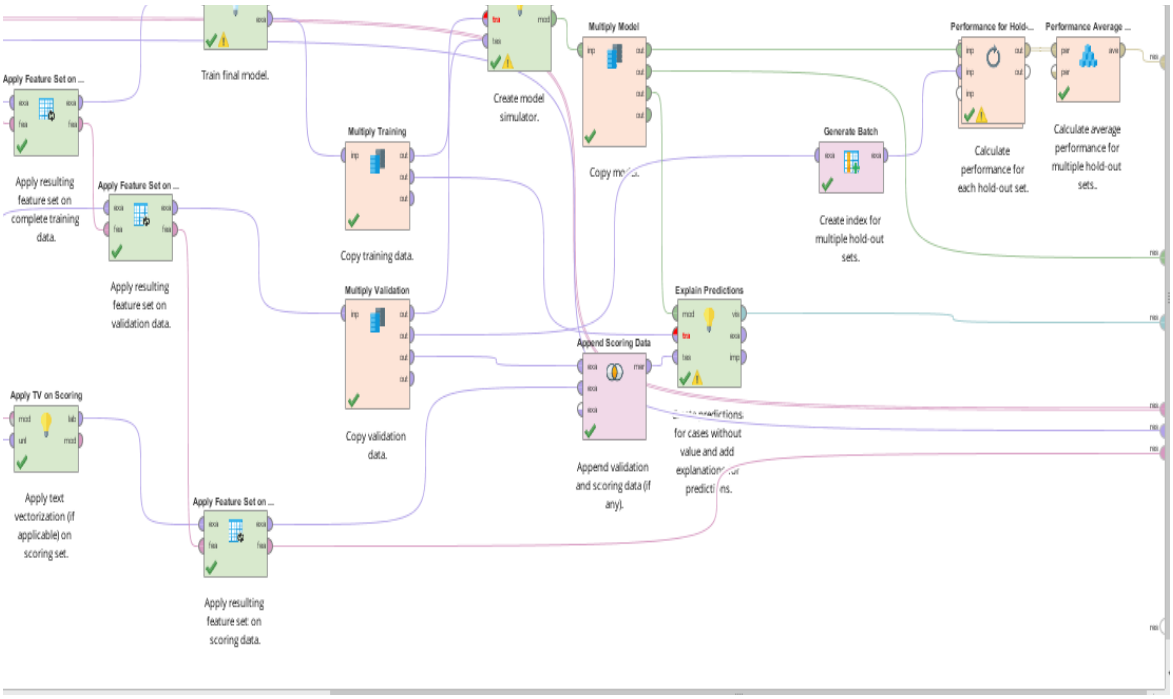


Figura 43. Proceso del modelo Naive Bayes en RapidMiner parte 2

Además, se determinó las zonas geográficas que presenta mayor número de beneficiarios y se identificó que la cantidad de puntos de pago no satisface la demanda de usuarios. Se cargó los datos a la extensión auto model con los puntos de pago, ubicaciones, beneficiarios, tipos de transacciones, como se muestra en la figura 42:

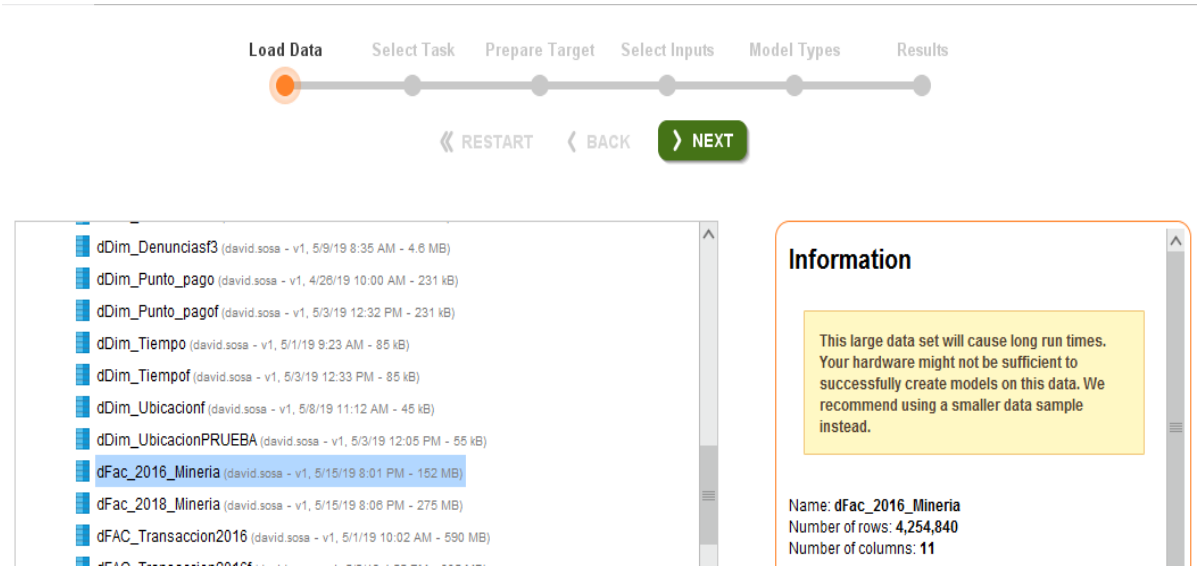
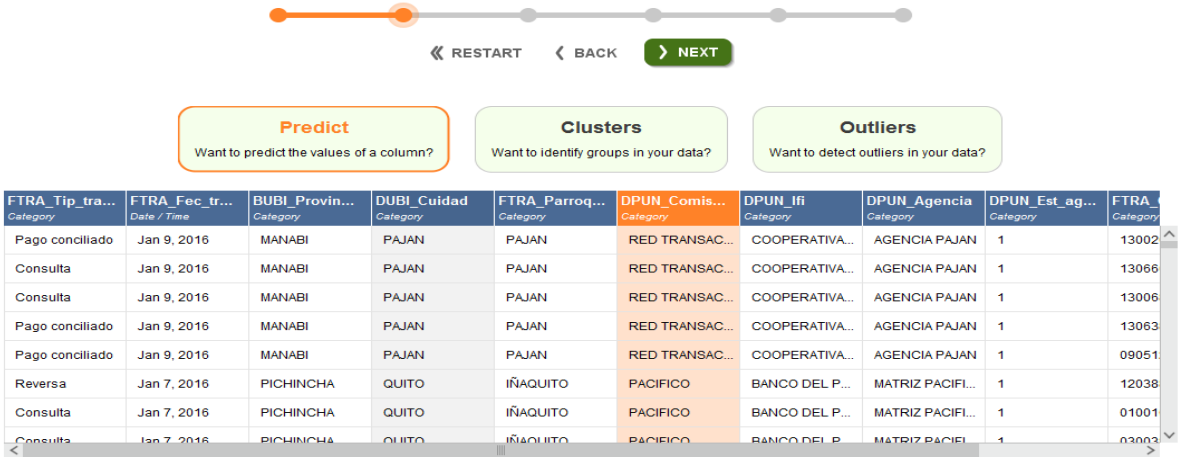


Figura 44. Carga de datos en la extensión Auto Model

Se escoge Predecir (Predict) los valores de la comuna DPUN_Comisionista para poder desarrollar un modelo de aprendizaje y de esta maneja visualizar el desequilibrio de puntos de pago.



FTRA_Tip_tra... Category	FTRA_Fec_tr... Date / Time	BUBI_Provin... Category	DUBI_Cuidad Category	FTRA_Parroq... Category	DPUN_Comis... Category	DPUN_Ifi Category	DPUN_Agencia Category	DPUN_Est_ag... Category	FTRA_ Category
Pago conciliado	Jan 9, 2016	MANABI	PAJAN	PAJAN	RED TRANSAC...	COOPERATIVA...	AGENCIA PAJAN	1	13002
Consulta	Jan 9, 2016	MANABI	PAJAN	PAJAN	RED TRANSAC...	COOPERATIVA...	AGENCIA PAJAN	1	13066
Consulta	Jan 9, 2016	MANABI	PAJAN	PAJAN	RED TRANSAC...	COOPERATIVA...	AGENCIA PAJAN	1	13006
Pago conciliado	Jan 9, 2016	MANABI	PAJAN	PAJAN	RED TRANSAC...	COOPERATIVA...	AGENCIA PAJAN	1	13063
Pago conciliado	Jan 9, 2016	MANABI	PAJAN	PAJAN	RED TRANSAC...	COOPERATIVA...	AGENCIA PAJAN	1	09051
Reversa	Jan 7, 2016	PICHINCHA	QUITO	IÑAQUITO	PACIFICO	BANCO DEL P...	MATRIZ PACIFI...	1	12038
Consulta	Jan 7, 2016	PICHINCHA	QUITO	IÑAQUITO	PACIFICO	BANCO DEL P...	MATRIZ PACIFI...	1	01001
Consulta	Jan 7, 2016	PICHINCHA	QUITO	IÑAQUITO	PACIFICO	BANCO DEL P...	MATRIZ PACIFI...	1	03003

Figura 45. Predicción de la columna Comisionista

Se procedió a seleccionar los factores de entrada como la ubicación, beneficiarios, tipos de transacción.

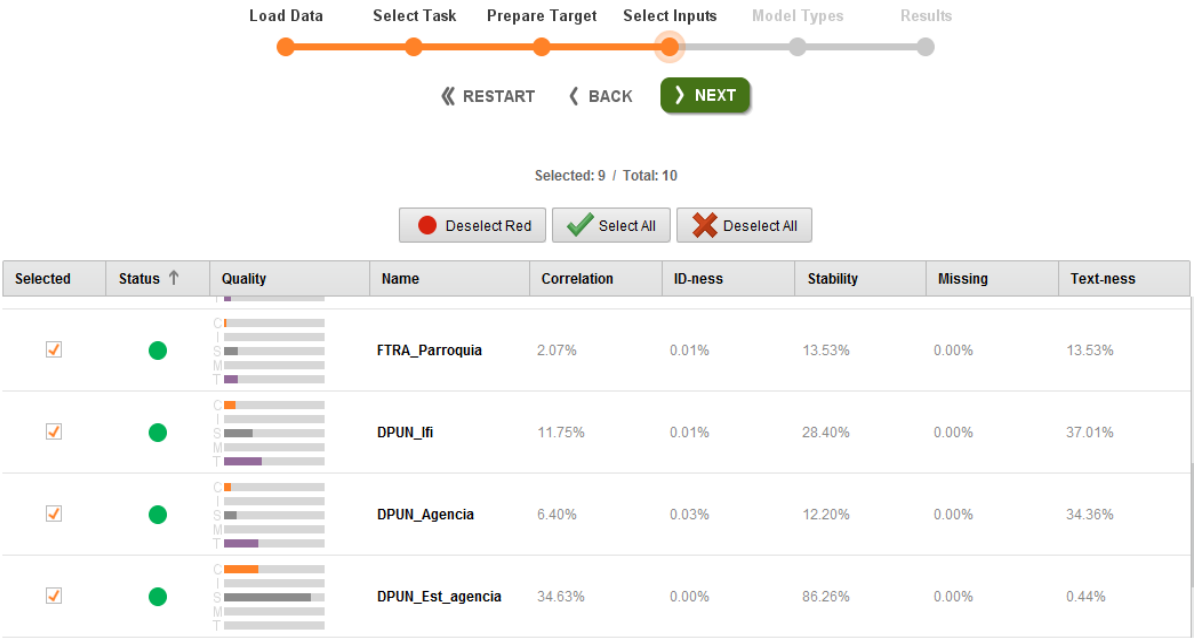


Figura 46. Factores de entrada para construcción del modelo

RapidMiner solo recomienda el modelo Naive Bayes debido al gran volumen de datos que se está manejando.

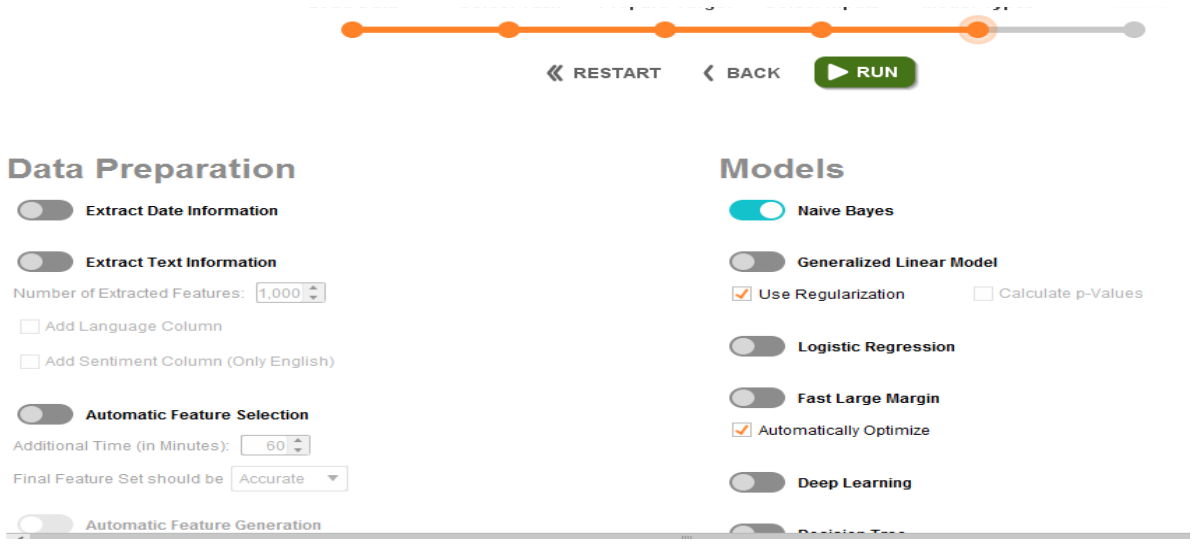


Figura 47. Modelo Naive Bayas recomendado por Rapidminer

En el año 2016 se verifica que existen mayor número de consulta a través del concentrador Pacifico, esto se debe a que en la fecha indicada existía una baja o nula transaccionalidad mediante los Corresponsales No Bancarios (CNB).Mientras que en referencia al concentrador que registra mayor número de pagos exitosos es Banco Desarrollo de los Pueblos que paga el subsidio económico a través de Cooperativas de Ahorro y Crédito.

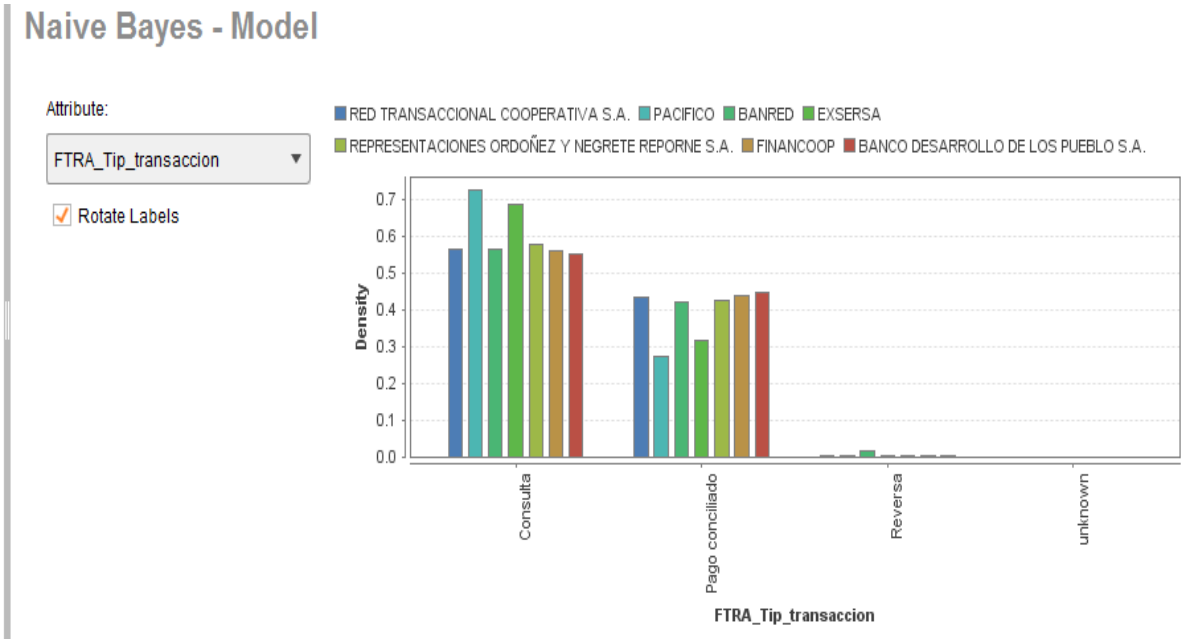


Figura 48. Tipo de transacción por Concentrador

El concentrador Pacifico tiene estandarizado la dirección de sus puntos de pago en la matriz de la Ciudad de Quito Provincia de Pichincha, esto quiere decir que no se especifica las transacciones llevadas a cabo a nivel nacional.

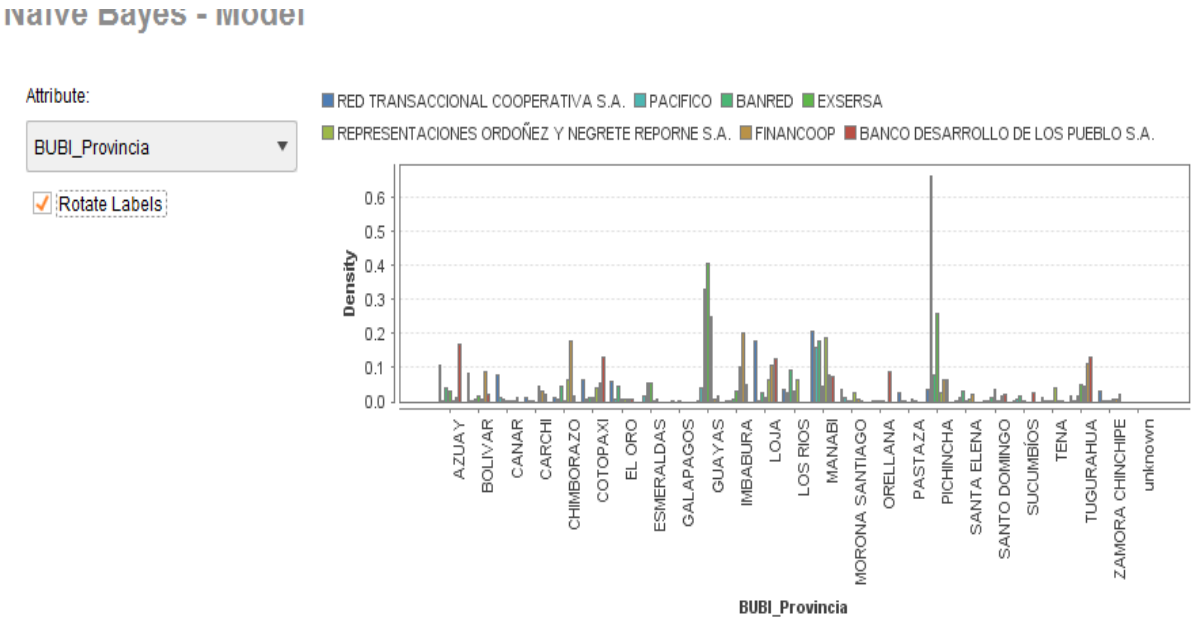


Figura 49. Ubicación por puntos de pago

En el año 2017 el concentrador el mayor número de consultas registra el mayor número de pagos y el concentrador Banred tiene el mayor número de reversos.

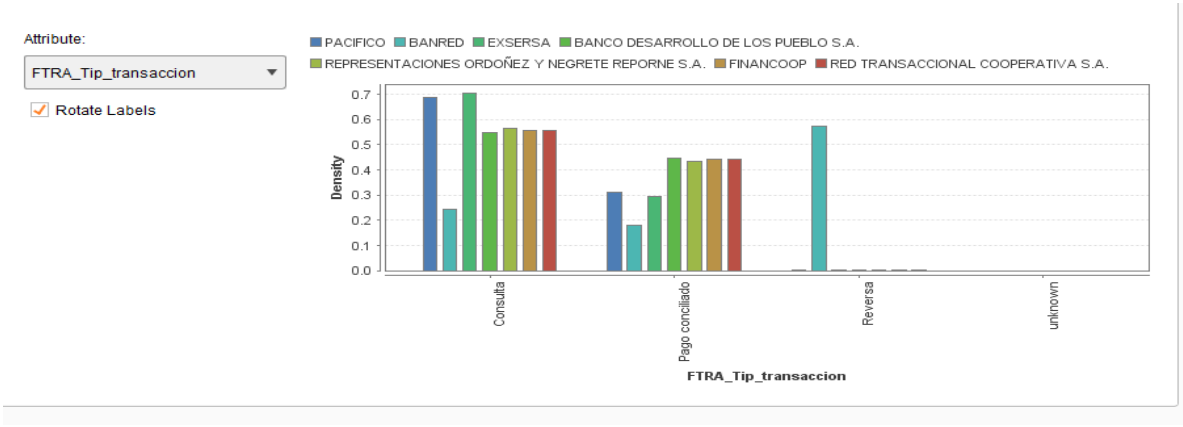


Figura 50. Transacciones realizadas en el año 2017

De acuerdo al gráfico estadístico se registra la mayor transaccionabilidad mediante Pacifico en la Provincia de Pichincha; sin embargo, como se mencionó anteriormente este concentrador no tiene identificado sus punto de pago a nivel nacional. Además el concentrador Banred tiene mayor número de transacciones identificadas a nivel nacional.

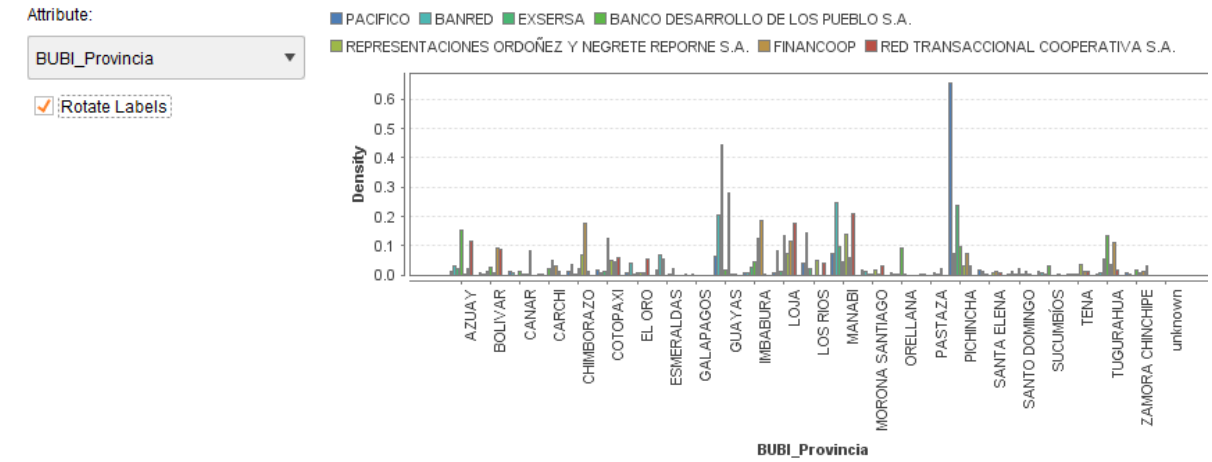


Figura 51. Distribución de los puntos de pago a nivel nacional

A partir del año 2018 el concentrador Banred registra el mayor número de consultas y transacciones exitosas a través de los CNB's,esto se debe a la accesibilidad que tienen los usuarios a estos puntos de pago. Además el concentrador Banred para este año tiene la mayor cantidad de puntos de pago, razón por lo cual registra la mayor transacionalidad a nivel nacional. De acuerdo al análisis realizado acerca a la proyección de cobros esta tendencia se mantendrá para el año 2019.

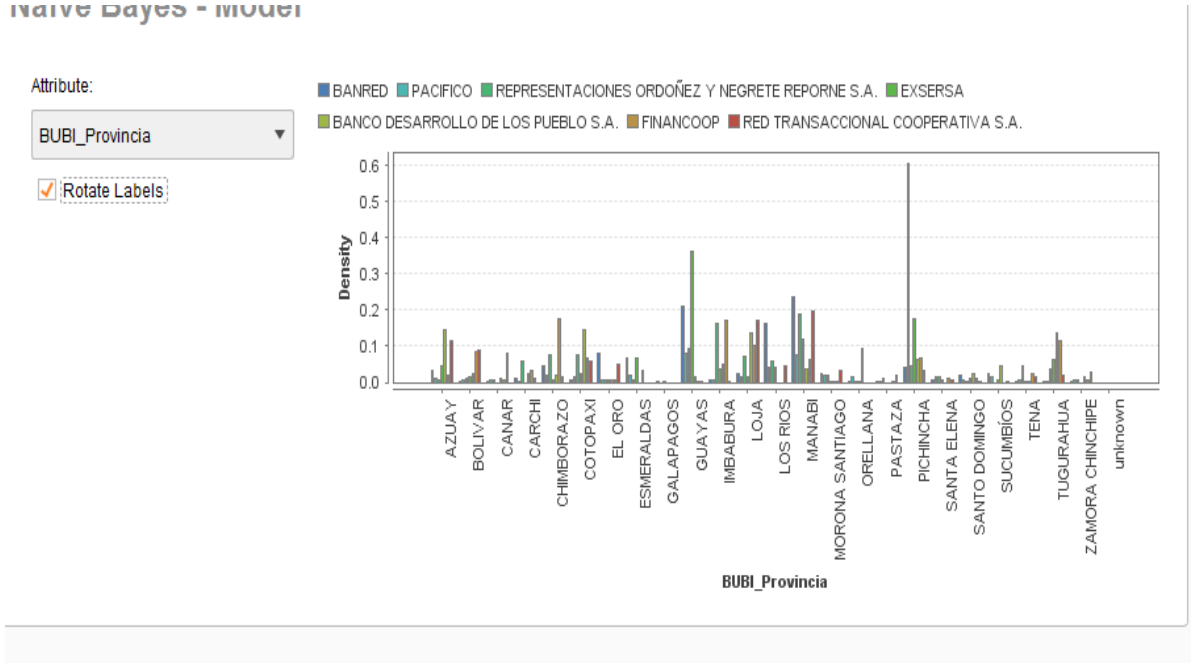


Figura 52. Distribución de puntos de pago en el año 2018

Se abrió el proceso con el cual se visualiza la carga de datos, selección de la variable a predecir DPUN_Comisionista, se seleccionó los factores de entrada como ubicación, tipo de transacción, beneficiarios, periodo, el siguiente paso es el modelo Naive Bayes recomendado por la herramienta y se analiza los resultados obtenidos, como se muestra en la figura 53 y 54:

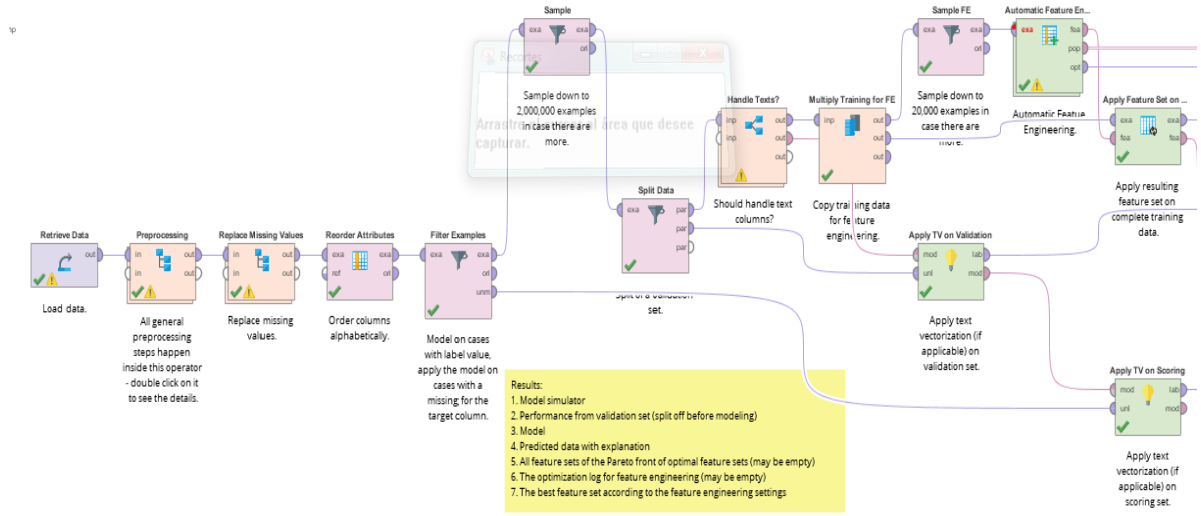


Figura 53. Proceso del modelo Naive Bayes en RapidMiner parte 1

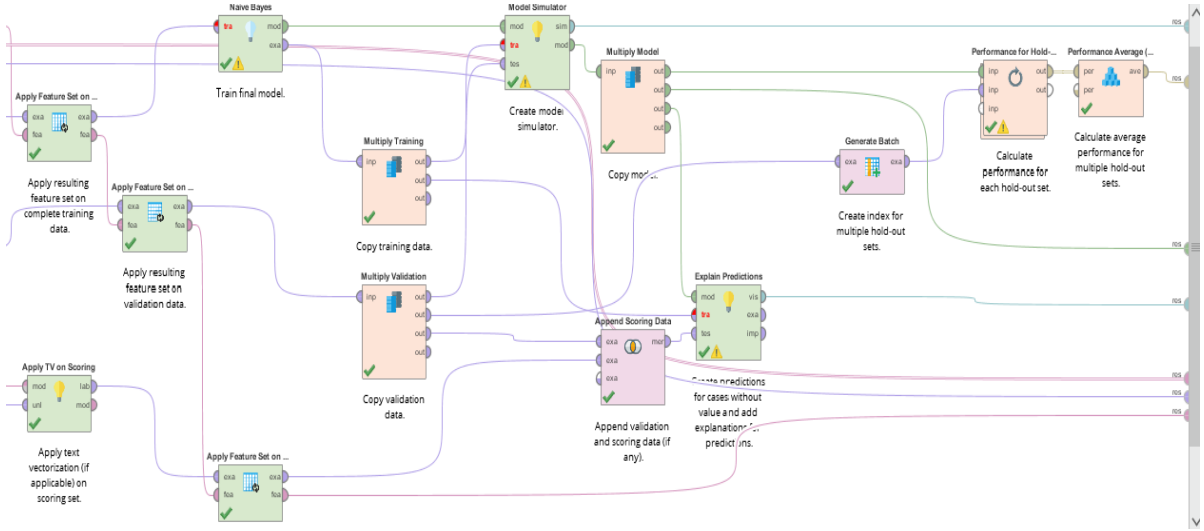


Figura 54. Proceso del modelo Naive Bayes en RapidMiner parte 2

4.1.8. Evaluación del modelo

Se realizó la evaluación de resultados utilizando una matriz de confusión, mediante el operador de validación cruzada que es propio de la herramienta Rapidminer, se separó el set de datos en dos partes, una con un 20% para testeo que sirve para revisar el rendimiento y un segundo del 80% para el grupo de entrenamiento, con el operando validación cruzada, el cual dio como resultado un porcentaje de precisión del modelo implementado.

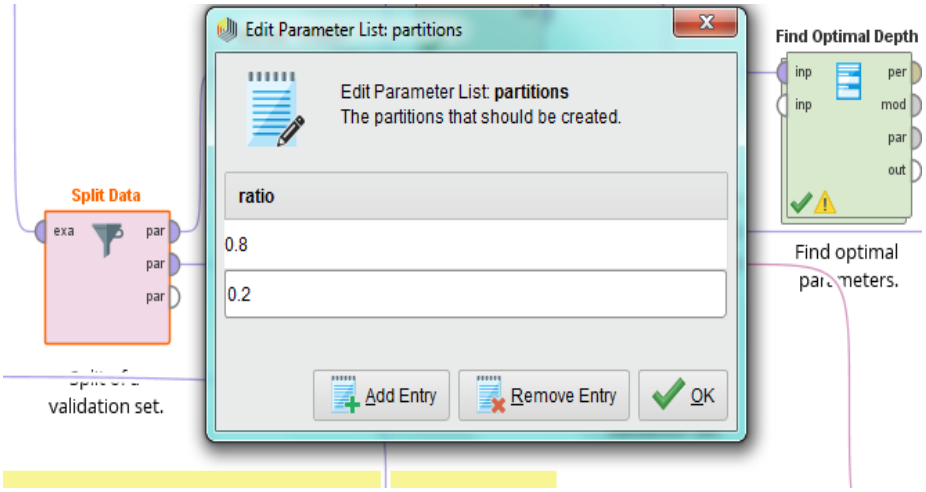


Figura 55. Porcentaje de división de data para el modelo

Con los parámetros de la matriz de confusión se pudo observar el número de aciertos por el factor tipo de bono o pensión y un 75.55% de precisión.

```
Parameter set:

Performance:
PerformanceVector [
-----accuracy: 75.55% +/- 0.26% (micro average: 75.55%)
ConfusionMatrix:
True:   BONO DESARROLLO HUMANO  PENSION ADULTO MAYOR    PERSONAS CON DISCAPACIDAD       MIS MEJORES ANIOS       BONO VARIABLE   Ninguna
BONO DESARROLLO HUMANO: 2071    150     30      14      35      0
PENSION ADULTO MAYOR:   2353    5067    1510    161     14      0
PERSONAS CON DISCAPACIDAD:      3       5       5       5       0       0
MIS MEJORES ANIOS:      111     177     66      1313    54      0
BONO VARIABLE:  62      2       0       7       929     0
Ninguna:        0       0       0       0       0       5318
]
Decision Tree PO.maximal_depth  = 15
```

Figura 56. Parámetros del set de matriz de Confusión

4.2. Resumen comparativo

4.2.1. Denuncias sobre cobros indebidos.

Mediante esta investigación se detectó que las Pensiones Adulto Mayor y Mis Mejores Años, destinadas a personas que tienen 65 años en adelante se encuentran, se encuentran extrema pobreza y Bono Desarrollo Humano tienen tendencia a cobros indebidos como se muestra en la figura 57:

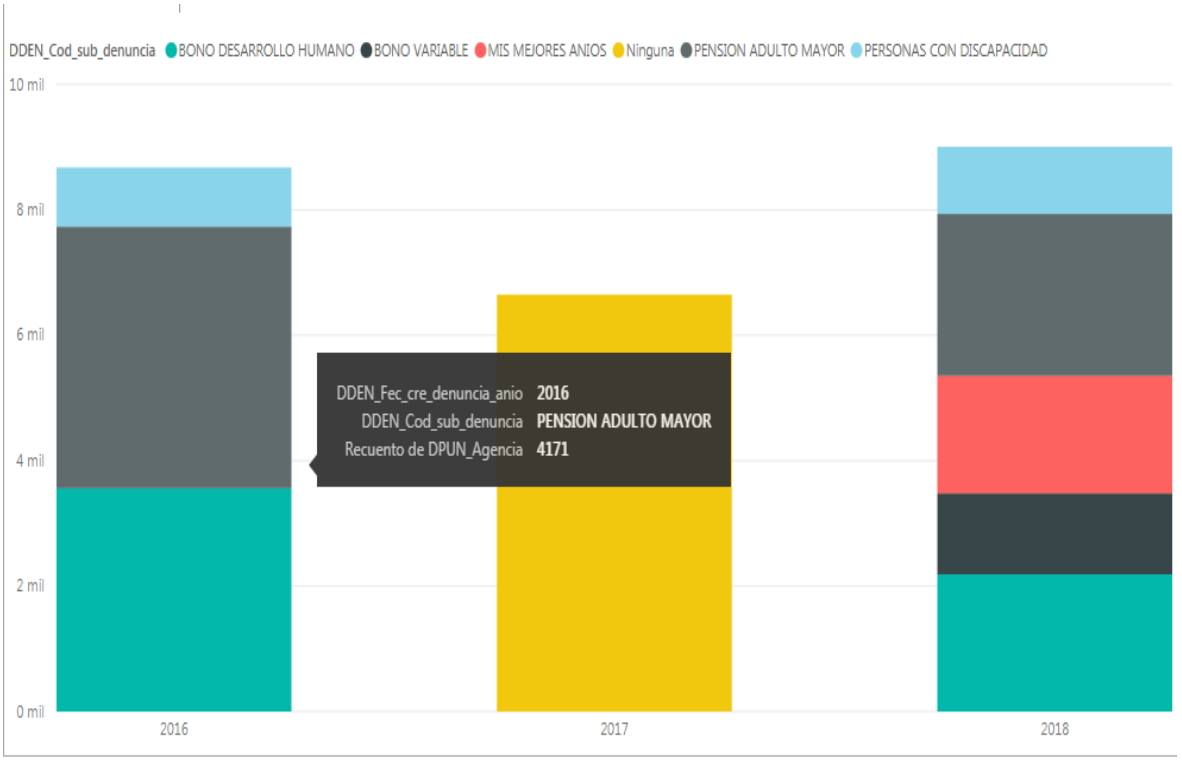


Figura 57. Cuadro comparativo por tipos de bonos y pensiones con denuncias

En el siguiente gráfico se verifica las Provincias que registran mayor número de denuncias a nivel nacional, en base a esta información se sugiere realizar campañas informativas para evitar cobros indebidos. Por lo que, la campaña informativa de manera correctiva se recomienda sea

realizada en las Provincias con mayor índice de denuncias (Pichincha, Guayas, Manabí, Los Ríos, Esmeraldas, Tungurahua) y la campaña preventiva en las demás.

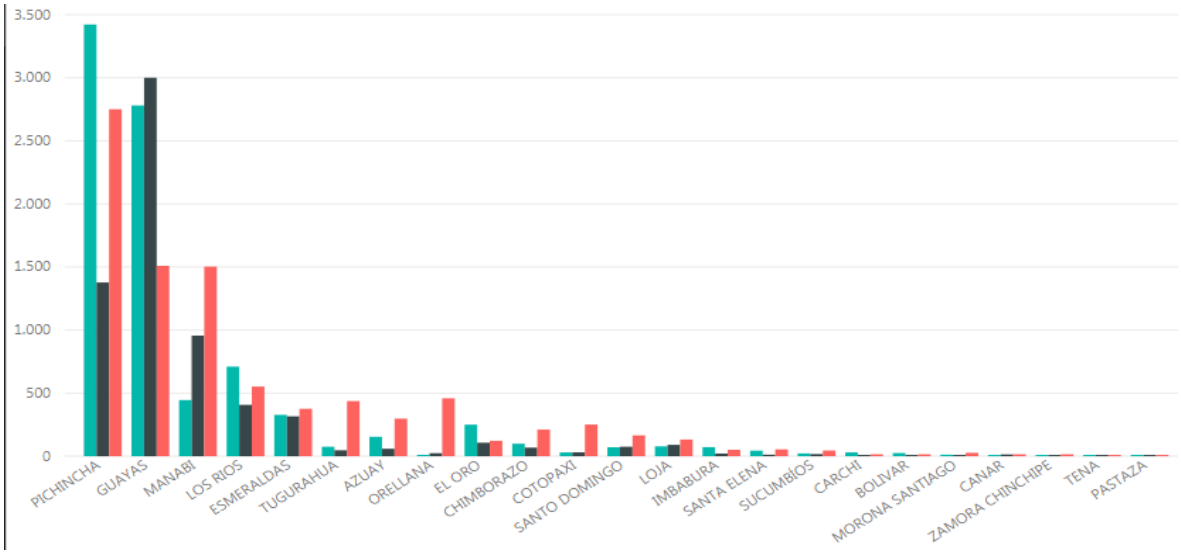


Figura 58. Provincias con mayor índice de denuncias

En el año 2016 se verifica que las Parroquias de Tarqui, Palestina, Puerto Quito, Portoviejo, Iñaquito, Pedro Vicente Maldonado, Bomboli, Febres Cordero, Quevedo, Rocafuerte, Andrés de Vera, El Carmen, Manta es donde más se concentra los puntos inactivos por denuncias.

Figura 59. Puntos inactivos por denuncias en el 2016

En el año 2017 se verifica que las Parroquias de Tarqui, Iñaquito, Portoviejo, El Carmen, Cayambe, Palestina, Manta, Febres Cordero, Lomas de Sargentillo, Puerto Quito, Rocafuerte, Andrés de Vera, Quevedo, La Concordia es donde más se concentra los puntos inactivos por denuncias.

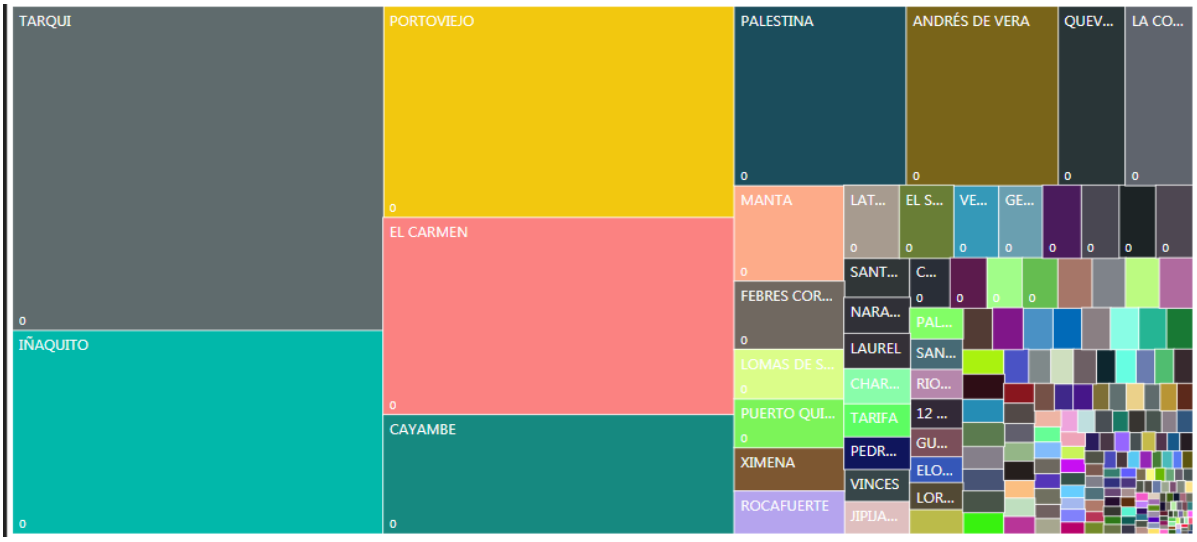


Figura 60. Puntos inactivos por denuncias en el 2017

En el año 2018 se verifica que las Parroquias de Portoviejo, Quevedo, Tarqui, Iñaquito, Andrés de Vera es donde más se concentra los puntos inactivos por denuncias.

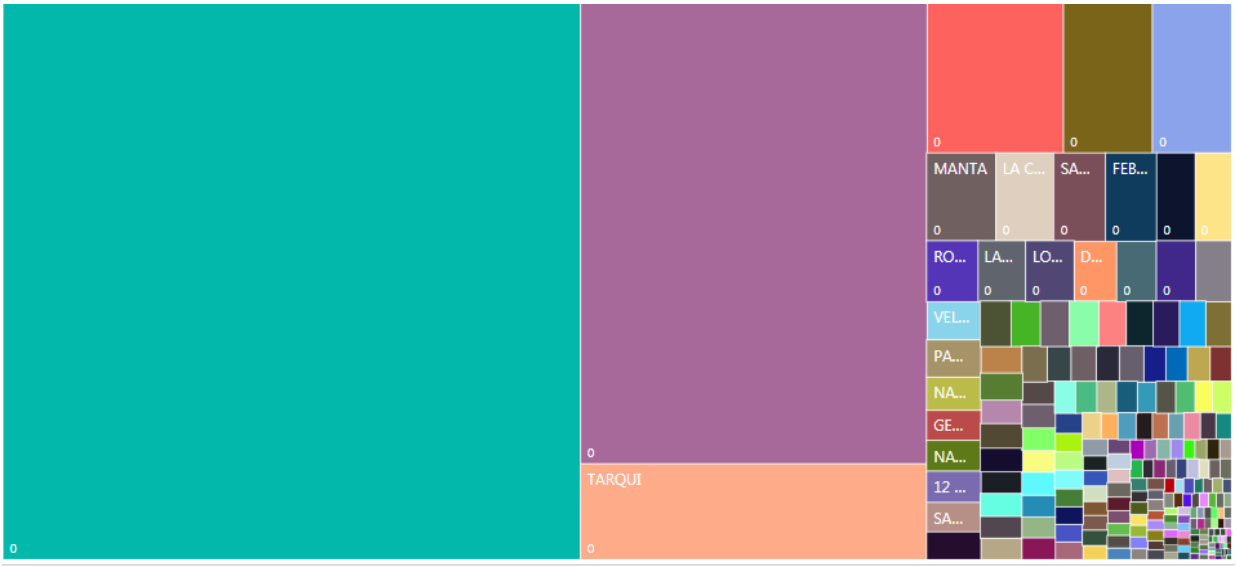


Figura 61. Puntos inactivos por denuncias en el 2018

4.2.2. Distribución de puntos de pago en los años 2016,2017 y 2018.

En el año 2016 se puede visualizar la distribución de puntos de pago, de esta manera los puntos habilitados en estado 1 activo (color rojo) y estado 0 inactivo (color azul), como se demuestra en la figura 62:

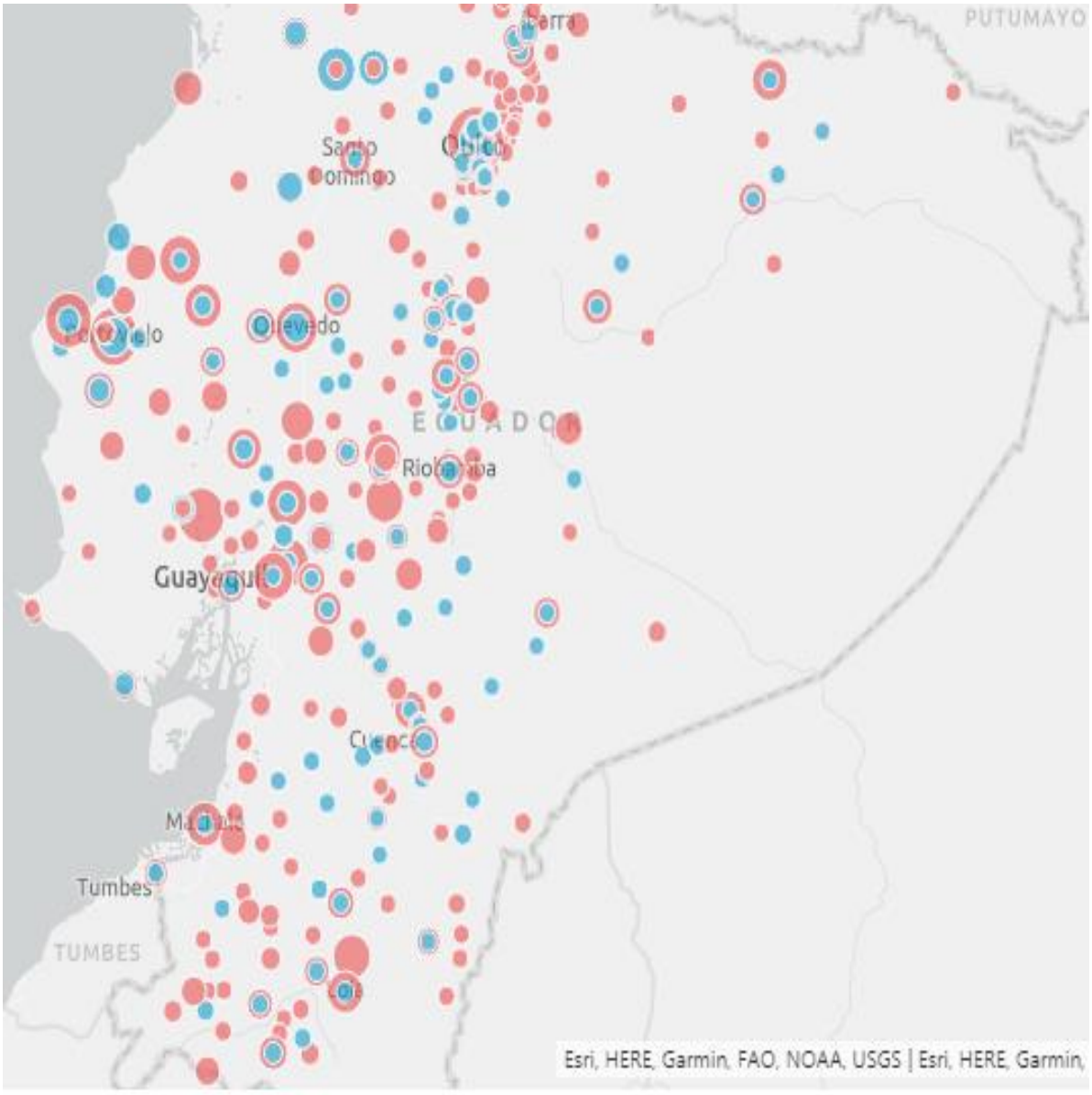


Figura 62. Distribución de puntos de pago en el año 2016

Además, en el análisis realizado a nivel nacional refleja que las Provincias de Pichincha, Guayas, Manabí, Los Ríos es que registra mayor número de puntos pago activos; sin embargo, es Guayas la Provincia con mayor número de puntos pago inactivos.

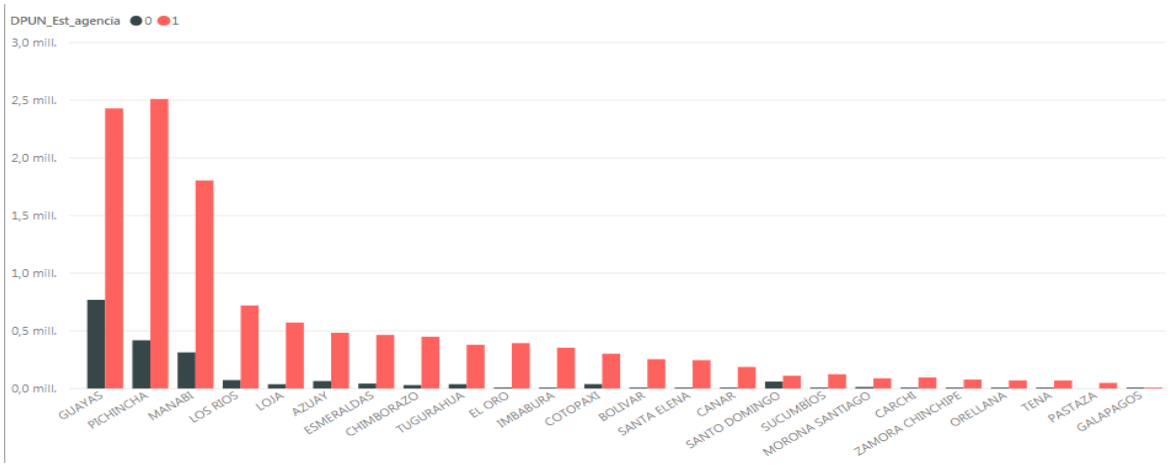


Figura 63. Mayor número de puntos pago activos e inactivos en el 2016

Con respecto a la Ciudad de Guayaquil es la que registra mayor número de puntos inactivos mientras que Colta, Ventanas, Pernales, Nobol, Baños de Agua Santa no tienen puntos inactivos.

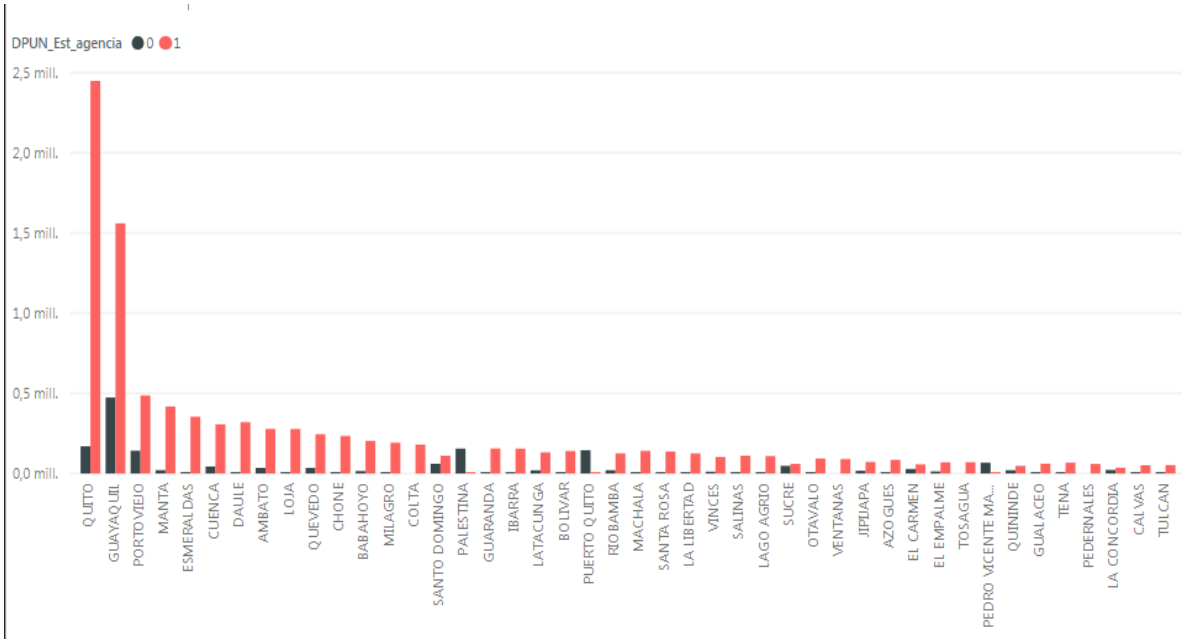


Figura 64. Mayor número de puntos inactivos en el 2016

Por último, las Parroquias que se encuentran con el menor número de puntos pago activos a nivel nacional son: La Merced de Buenos Aires, San José de tambo, Sacapalca, San Francisco, como se muestra en la figura 65:

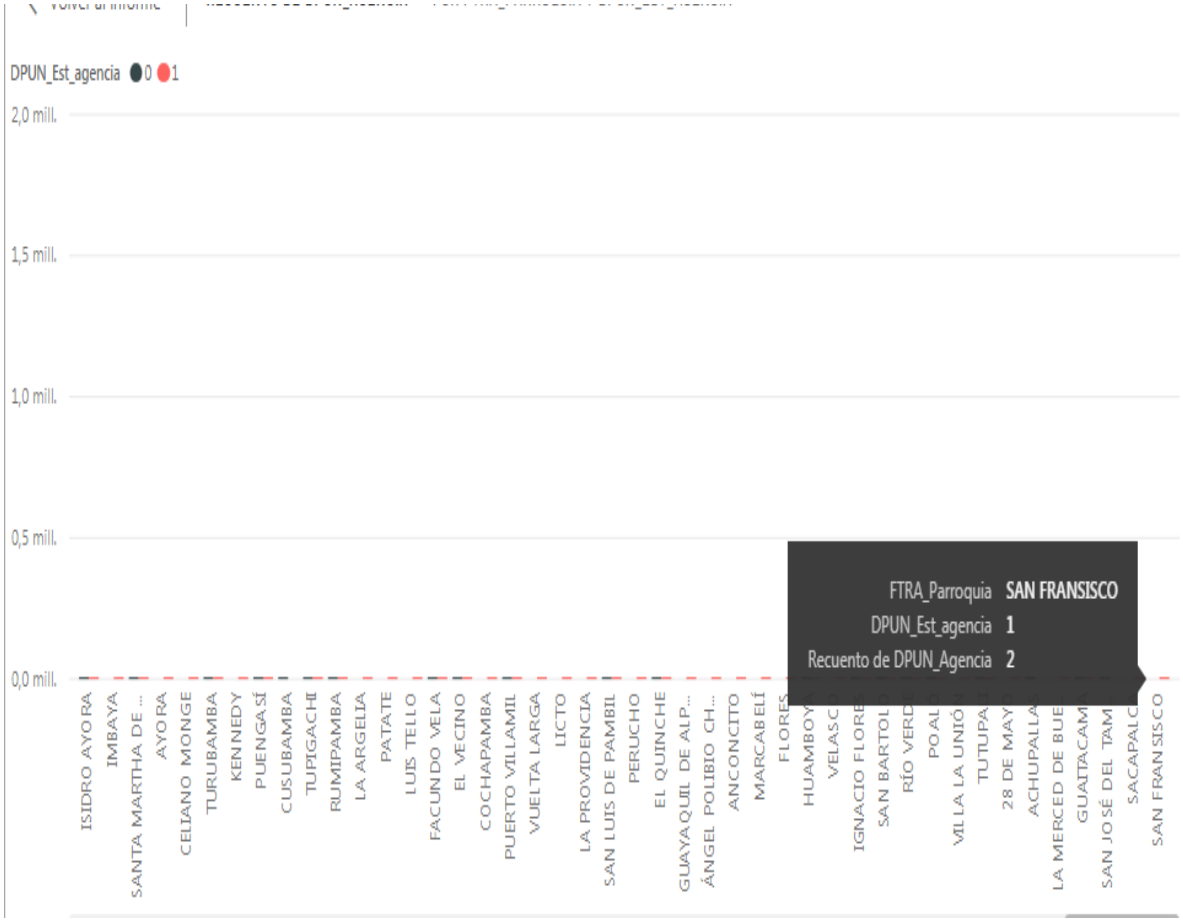


Figura 65. Menor número puntos pago activos en el 2016

En el año 2017 se observa el incremento de los puntos de pago en la región de la Costa, esto se debió al terremoto ocurrido el 16 de abril de 2016, puesto que la Institución Pública entrego un bono de Contingencia por la emergencia Nacional, como se demuestra en la figura 66:

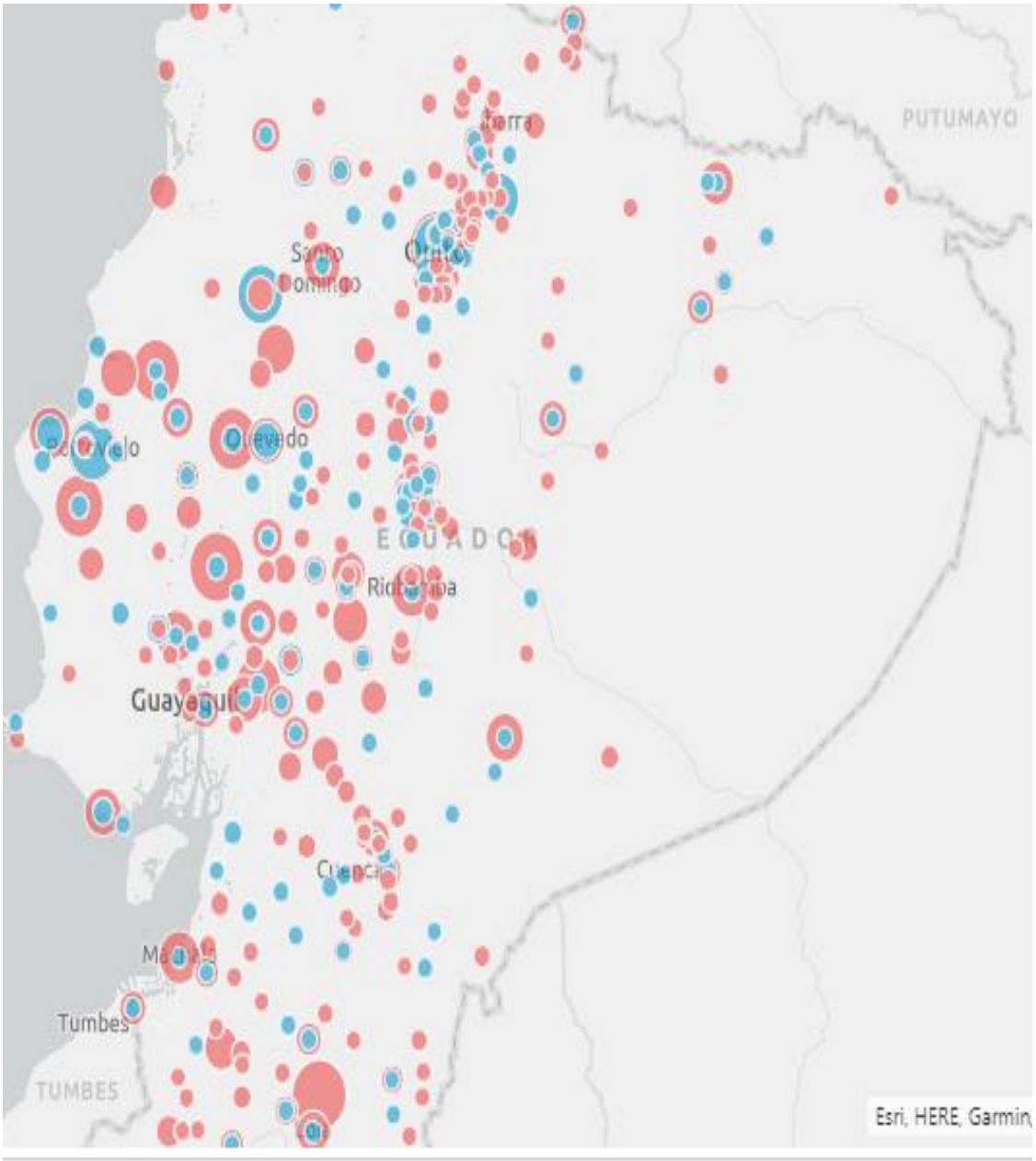


Figura 66. Distribución de puntos de pago en el año 2017

Además, en el análisis realizado a nivel nacional se verificó que Manabí, Guayas, Pichincha, Los Ríos, Loja son las Provincias que registraron una mayor número de puntos pago activos; sin embargo, esta misma Provincia tienen el mayor número de puntos pago inactivos.

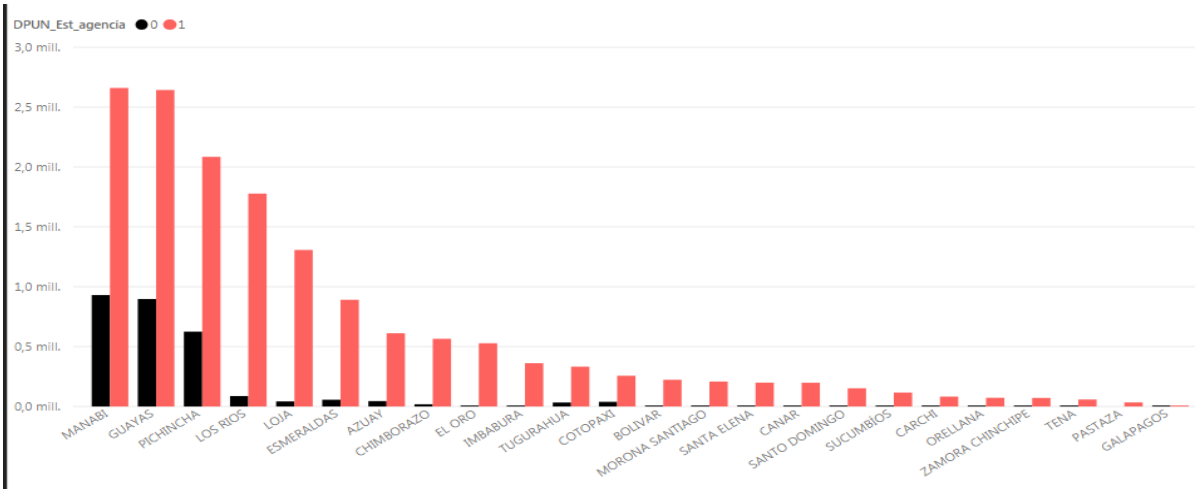


Figura 67. Mayor número de puntos pago activos e inactivos en el 2017

Con respecto a la Ciudad de Guayaquil es la que registra mayor número de puntos pago inactivos, pero se da una fenómeno de puntos inactivos en Portoviejo, El Carmen, Cayambe, Quevedo, Palestina, La Concordia, mientras que Colta, Pedernales, Nobol, Baños de Agua Santa entre otros no tienen puntos inactivos.

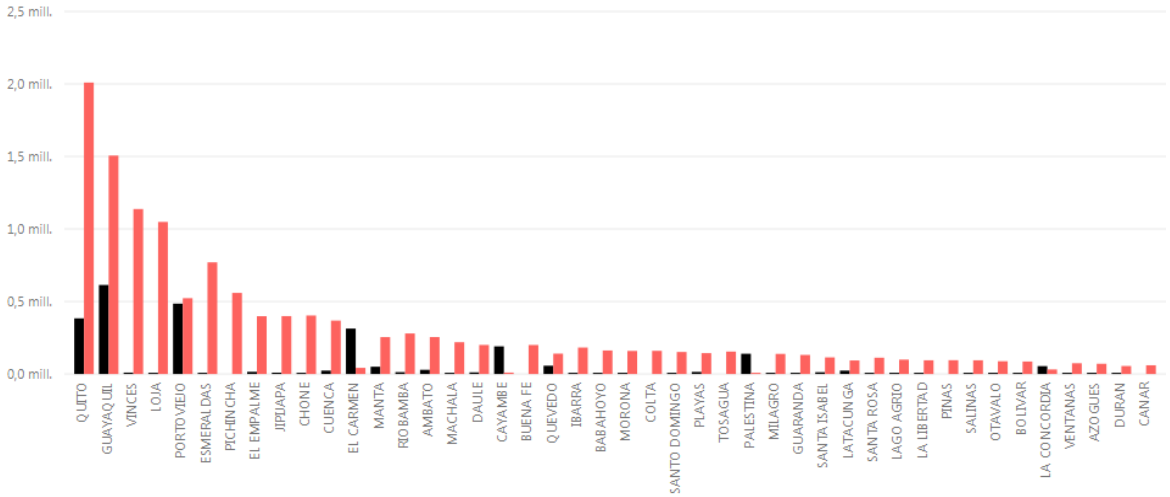


Figura 68. Mayor número de puntos inactivos en el 2017

Por último, en las Parroquias que se encuentran con el menor número de puntos pago activos a nivel nacional son: San Bartolo, Poalo, Huamboya, San Francisco Espejo, Bartolome Ruiz, como se muestra en la figura 69:

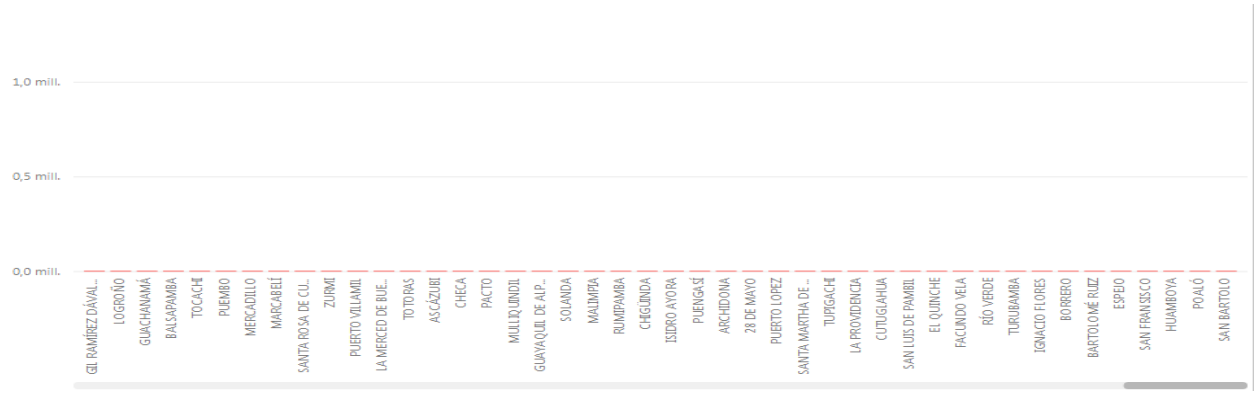


Figura 69. Menor número puntos pago activos en el 2017

A partir del año 2018 se incrementan los Corresponsales No Bancarios (CNB), es por esta razón la diferencia de puntos pago activos en comparación de los inactivos.

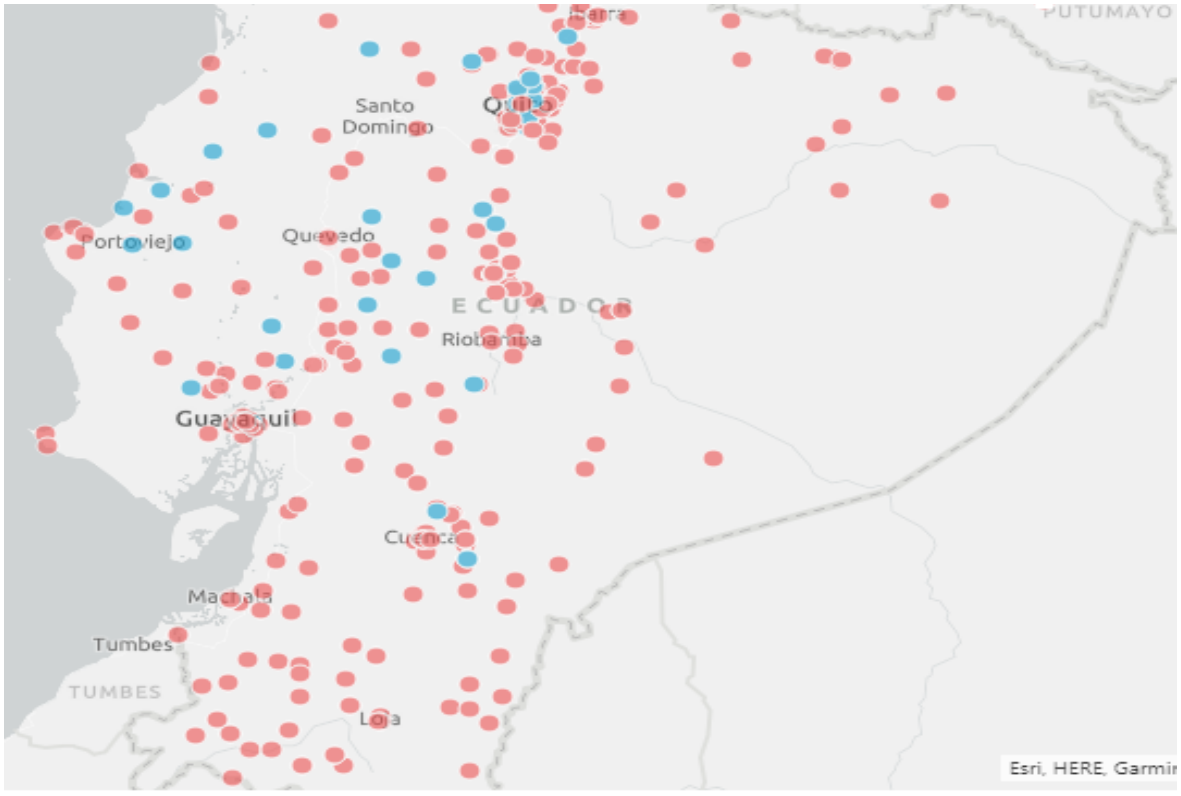


Figura 70. Distribución de puntos de pago en el año 2018

En el análisis realizado en el 2018 a nivel nacional se verificó que Guayas es la Provincia que registra mayor número de puntos pagos activos; sin embargo, es Manabí la Provincia con mayor número de puntos pago inactivo, como se muestra en la figura 70:

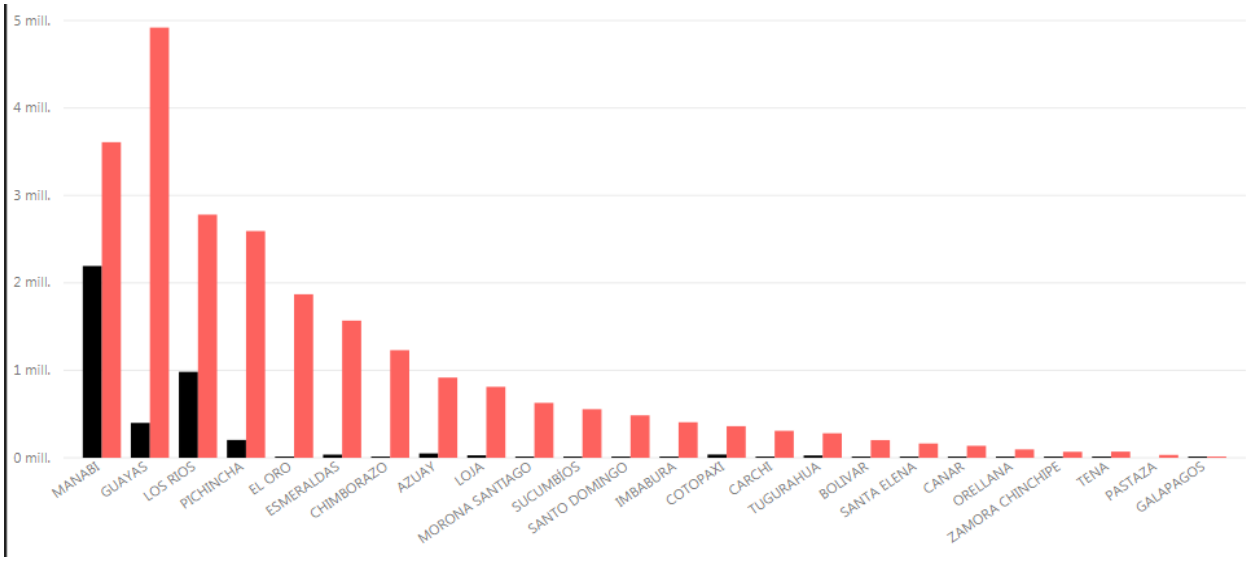


Figura 71. Mayor número de puntos pago activos e inactivos en el 2018

Así también con respecto a la Ciudad Portoviejo, Quevedo, Guayaquil, Quito son las que registran mayor cantidad de puntos inactivos mientras que Espejo, Pallatanga, El Guabo, Guamote, Rio Verde con menor registros de puntos pagos inactivos.

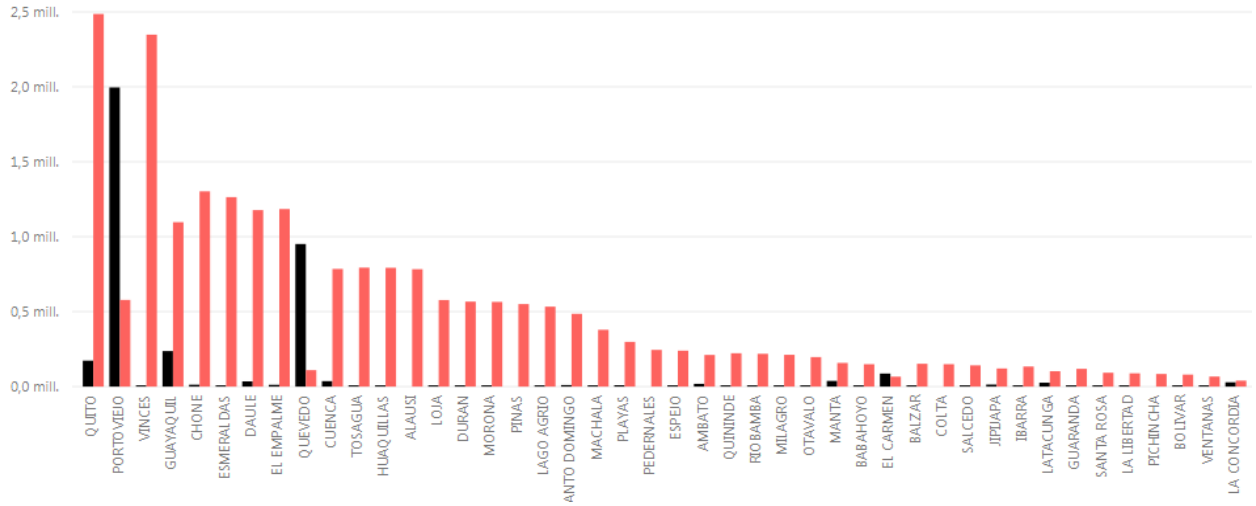


Figura 72. Mayor número de puntos inactivos en el 2018

Por último, en las Parroquias que se encuentran con el menor número de puntos pago activos a nivel nacional son: Uyumbicho, Chiguilpe, Guatiacama, La Providencia, Ignacio Flores, Jijón y Camaño, como se muestra en la figura 72:

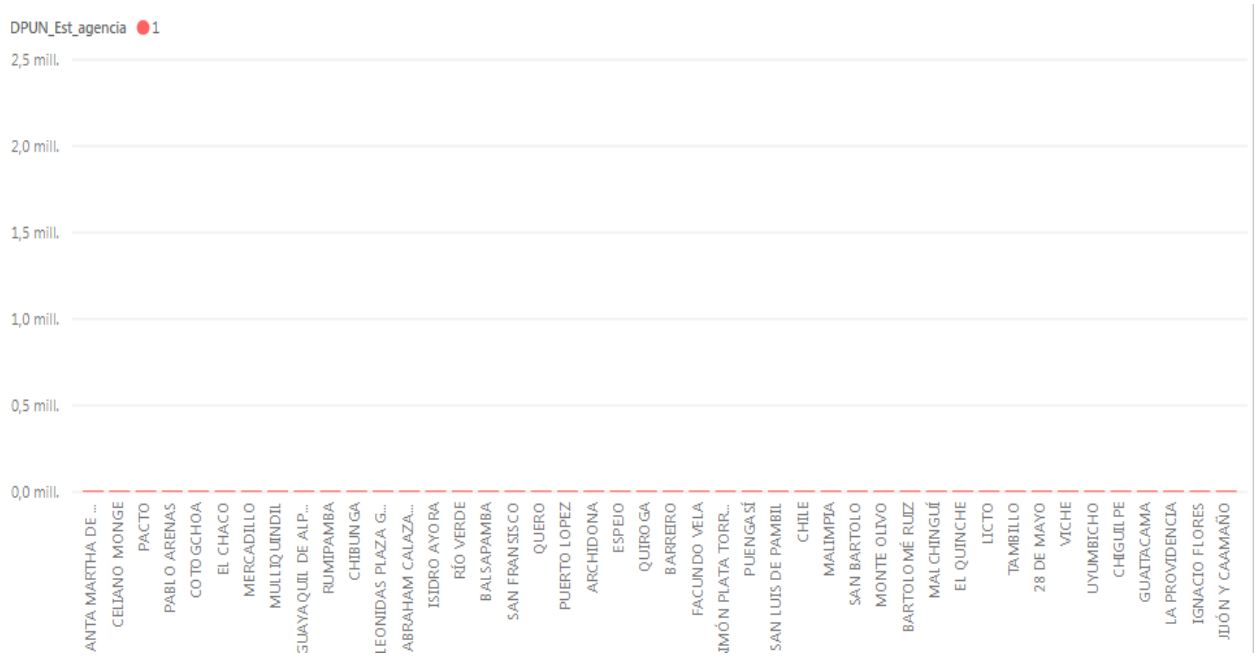


Figura 73. Menor número puntos pago activos en el 2018

CAPÍTULO V

CONCLUSIONES Y RECOMENDACIONES

5.1. Conclusiones

En la institución se fomenta la inclusión económica y social con énfasis en los grupos de extrema pobreza y vulnerabilidad para que puedan desarrollarse y que se dé más igualdad en el ámbito social y económico, todos estos datos son traídos y comparados desde diferentes parte del país para formar grandes volúmenes de datos, los cuales tienen que ser un aporte de información y posterior conocimiento real.

Se analiza todo lo que engloba la Institución frente a los Sistemas Auxiliares de Pago, balcón de servicios con denuncias, entrega de bonos y de pensiones que el Estado otorga a las personas de escasos recursos económicos a nivel nacional, para entender cuáles son sus requerimientos y el resultado que se tiene que obtener. De esta forma se procede a diseñar con la herramienta PowerDesigner el modelo entidad relación y multidimensional, con lo cual se pasa a realizar ETL con la herramienta Rapidminer, la cual brinda las facilidades técnicas para poder obtener los datos validados y ordenados y posterior crear el modelo multidimensional en el motor de base de datos SQL Server. La herramienta Rapidminer cuenta con la extensión auto modelo, la cual acelerar el proceso de creación y validación de diferentes modelos, en los resultados obtenidos se pudo observar que el porcentaje de precisión es del 56% para el modelo Naive Bayes, Generalized Linear Model con un 27,9%, Logistic Regression con un 49,3%, Decision Tree con un 75,5%, con lo cual se escoge el modelo Naive Bayes y Decision Tree que ayudaron a resolver las vulnerabilidades detectadas en el sistema de cobros.

El modelo Naive Bayes deriva la probabilidad de predicción basado en la evidencia subyacente, para lo cual las entradas del modelo son la ubicación, hora de transacción, año, los Sistemas Auxiliares de Pago y el atributo a predecir es el tipo de bono y pensión que otorga por el estado, dando como resultado los cuales fueron sometidos al modelo Naive Bayes y Decission Tree que identificó los patrones de comportamiento de los beneficiarios en los últimos 3 años, se detectó que la Pensión Adulto Mayor y Mis Mejores Años destinadas a personas que tienen 65 años en adelante, se encuentran en extrema pobreza y el Bono Desarrollo Humano tienen tendencia a cobros indebidos, por lo que es necesario establecer campañas informativas en las Parroquias de Tarqui, Palestina, Puerto Quito, Portoviejo, Iñaquito, Pedro Vicente Maldonado, Bomboli, Febres Cordero, Quevedo, Rocafuerte, La Concordia, Andrés de Vera, El Carmen, Cayambe, Manta, Lomas de Sargentillo, Andrés de Vera. Así también, se determinó las zonas geográficas que presenta mayor número de beneficiarios es en la región Costa y Sierra y se identificó que la cantidad de puntos de pago no satisface la demanda de usuarios en el Oriente y Región Insular.

Se realizó la evaluación del modelo analítico predictivo utilizando una matriz de confusión, se tomó un porcentaje de todos los datos para testeo y grupo de entrenamiento con el operando validación cruzada que es propio de la herramienta Rapidminer, el cual dio como resultado un porcentaje de precisión del modelo implementado.

5.2. Recomendaciones

Tener en cuenta los resultados obtenidos de esta investigación como una guía, para que puedan realizar más análisis al gran volumen de datos que se maneja en la Institución diariamente.

Capacitar al personal que maneja los datos de la Institución, para que de esta forma generen conocimiento inmediato a las autoridades y estos puedan tomar las mejores decisiones.

Se recomienda un sistema detector de posibles denuncias por cobros indebidos en base a los datos de denuncia analizados.

En base al gran volumen de datos se recomienda trabajar con servidores robustos y distribuidos para que, al momento de realizar minería de datos, estos puedan procesarse sin novedades.

La herramienta Rapidminer tiene interesantes extensiones (Turbo Pre, Auto Model) entre otras, que ayuda a optimizar tiempo de trabajo y entrega diferentes modelos para poder resolver problemas con mayor precisión.

BIBLIOGRAFÍA

2018 Gartner, Inc. y / o sus Afiliados. (2018). Gartner. Obtenido de https://www.gartner.com/doc/reprints?id=1-65WC0O1&ct=190128&st=sb

Alvarez, L. D. (2005). Seguridad en Informática. Mexico.

Arenas, D. A., & Luna, M. C. (s.f.). Minería de Datos con Búsqueda de Patrones de Comportamiento. 10.

Badilla, I. D. (2010). Seguridad de Linux.

BM Corporation. (2012.). Manual Crisp-Dm de IBM SPSS. Obtenido de ftp://public.dhe.ibm.com/software/analytics/spss/documentation/modeler/15.0/es/CRISP-DM.pdf

Burgos, I. K. (2016). slideshare. Obtenido de https://es.slideshare.net/kreyes1/el-paradigma-dsr-design-science-research

Daimon, H. (2015). Manager Net Services Division. Obtenido de https://www.youtube.com/watch?v=BFfHgwgoD2w&feature=youtu.be

DNTICS. (2000). Manual de Zimbra. Quito.

Evaluando Software. (13 de 07 de 2016). Evaluando Software.com. Obtenido de http://www.evaluandosoftware.com/tecnicas-data-mining/

Fernández, J. &. (2007). Seguridad Informatica.

Fernández, R. M. (2000). Horder 3.1.1.

Flint. (2017). https://roundcube.net/about/.

Gonzales. (2012). Seguridad Informática.

Inostroza, J. C. (2002). Guia de Configuración.

Martínez, B. B. (2013). Benemérita Universidad Autónoma de. Recuperado el 2018, de Minería de Datos: http://bbeltran.cs.buap.mx/NotasMD.pdf

Menendez, L. (2002).

Microsoft . (2018). Obtenido de https://www.microsoft.com/en-us/sql-server/sql-server-2017

Microsoft. (2019). Obtenido de https://powerbi.microsoft.com/en-us/what-is-power-bi/

MIFSUD. (2012).

MySQL. (s.f.). Recuperado el 01 de 2015, de http://www.tutorialspoint.com/mysql/

Omar, G. D. (2007). Software Libre.

PANDORAFMS. (2018 Artica ST). Obtenido de https://blog.pandorafms.org/es/que-son-las-bases-de-datos/

Parraga, V., & Zaldumbide, J. (2018). DATA MINIG MODEL TO IDENTIFY THE FACTORS THAT AFFECT THE ACADEMIC ADVANCEMENT OF HIGHER EDUCATION STUDENT. INTED 2018, 5.

PowerDesigner. (2015). Obtenido de https://www.powerdesigner.biz/ES/powerdesigner/probar-powerdesigner-source_adw847a.html?gclid=EAIaIQobChMIo4KK2emi4gIVAVgNCh3cDwkiEAAYASAAEgLShvD_BwE

RapidMiner. (2019). Obtenido de https://rapidminer.com/resource/gartner-magic-quadrant-data-science-platforms/

Tics. (2012). Políticas de Seguridad. Quito.

Xianlin Zhuo, Z. Y. (2007). Research of Pension Fund Market Risk Model Based on Data Mining.

Yan-hai, L., & lin-yan, S. (2005). Study and applications of data mining to the structure risk analysis of customs declaration cargo.

Yanzapanta. (2013). Implementación de seguridad.