PROYECTO MINERIA DE DATOS  
REAL APLICADA A UN BANCO

GIT HUB PRIMER REPOSITORIO

JOHN ALEXANDER SIERRA YEPES  
ID 366652

CORPORACIÓN UNIVERSITARIA MINUTO DE DIOS

FACULTAD DE INGENIERÍA

PROGRAMA DE INGENIERÍA DE SISTEMAS

SOACHA-CUNDINAMARCA

2019

PROYECTO MINERIA DE DATOS  
REAL APLICADA A UN BANCO

MINERIA DE DATOS

Asesor:

GERARDO ZAMBRANO

CORPORACIÓN UNIVERSITARIA MINUTO DE DIOS

FACULTAD DE INGENIERÍA

PROGRAMA DE INGENIERÍA DE SISTEMAS

SOACHA-CUNDINAMARCA

2019

Contenido

[Introducción 4](#_Toc10045907)

[Objetivos. 4](#_Toc10045908)

[Objetivo general 4](#_Toc10045909)

[Objetivos específicos 4](#_Toc10045910)

[Software. 5](#_Toc10045911)

[SAS Enterprise Miner 5](#_Toc10045912)

[Características: 5](#_Toc10045913)

[PowerBI 6](#_Toc10045914)

[Características 6](#_Toc10045915)

[Proceso de minería de datos. 6](#_Toc10045916)

[Selección de los datos. 6](#_Toc10045917)

[Análisis de los datos. 7](#_Toc10045918)

[Transformación de los datos de entrada. 8](#_Toc10045919)

[Selección y aplicación de las técnicas de minería de datos 8](#_Toc10045920)

**Tabla de Ilustraciones**

[Ilustración 1 Imagen SAS GUIDE 7](#_Toc2451772)

[Ilustración 2 Transformación de datos SAS 8](#_Toc2451773)

[Ilustración 3 Ruta SAS realizada con árboles de decisión 9](#_Toc2451774)

[Ilustración 4 Unión de dos arboles 10](#_Toc2451775)  
[Ilustración 5 Resultado operaciones y tarifas por banca. 11](#_Toc2451775)  
[Ilustración 6 Mapa de calor. 11](#_Toc2451775)  
[Ilustración 7 Simulación. 12](#_Toc2451775)

Introducción

En el siguiente proyecto nos encontraremos con la aplicación real de minería de datos respecto al sector financiero, un banco específicamente. El paso a paso, procesos y algoritmo realizado, para el tratamiento de la información. Ademas el software que se utilizó para el análisis y reporte hecho para llegar a una conclusión que posiblemente ayudara a la toma de decisiones en la inversión de las operaciones realizadas dentro de la renta variable que efectúa el banco.

La Minería de Datos se presenta como una tecnología emergente, con varias ventajas para el sector financiero. Por un lado, resulta un buen punto de encuentro entre los investigadores o analistas financieros y las personas de negocios; por otro, ahorra grandes cantidades de dinero a una empresa; como por ejemplo en: publicidad u ofertas al segmento equivocado y abre nuevas oportunidades de negocios. Además, no hay duda de que trabajar con esta tecnología implica cuidar un sin número de detalles, debido a que el producto final involucra "toma de decisiones".

“No se dice el nombre del Banco, ya que este ejercicio fue una investigación realizada con información real y por motivos de tratamiento de datos.”

Objetivos.

Objetivo general

* Investigar el número de operaciones de renta variable del Banco que se realizaron dentro del año 2018 con sus respectivos fondos.

Objetivos específicos

* Investigar tarifas de cada uno de los fondos de las operaciones realizadas en el año 2018.
* Conocer el valor total de las operaciones por fondo de renta variable del banco.
* Analizar la información recolectada y generar un reporte claro para los usuarios.

Software.

SAS Enterprise Miner

Es un banco de trabajo gráfico completo para la minería de datos. Esta extensa plataforma ampliamente aclamada ofrece capacidades para preparar los datos para la analítica predictiva, identificar las variables más importantes, desarrollar modelos usando los algoritmos más modernos de minería de datos y machine-learning, validar fácilmente la precisión y aptitud del modelo(s), y generar activos que permiten una implementación sencilla de los modelos analíticos en sus aplicaciones operativas para la toma automática de decisiones.

Características:

* Crear series de datos de capacitación y prueba de muestra con un alto valor predictivo.
* Explorar interactivamente las relaciones y anomalías de los datos.
* Crear, transformar y seleccionar las variables más adecuadas para el análisis.
* Aplicar un rango de técnicas de modelado para identificar los patrones en los datos.
* Validar la utilidad y confiabilidad de los hallazgos del proceso de minería de datos.
* Crear todos los activos necesarios para la implementación, monitoreo y gestión de los modelos

PowerBI

Es una solución de análisis empresarial que permite visualizar los datos y compartir información con toda la organización, o insertarla en su aplicación o sitio web. Conéctese a cientos de orígenes de datos y dé vida a sus datos con los paneles e informes dinámicos.

Características

* Transformar los datos en impactantes objetos visuales y compártalos con sus colegas en cualquier dispositivo.
* Explorar y analizar visualmente los datos, en el entorno local y en la nube, todo en una sola vista.
* Colaborar en paneles personalizados e informes interactivos, y compártalos.
* Distribución por la organización con un sistema de gobierno y seguridad integrado

Proceso de minería de datos.

Selección de los datos.

Actualmente el banco tiene un repositorio de datos que esta asociado con el software SAS GUIDE, a través de este se pueden consultar las distintas tablas, claramente con previos permisos y demás. Estas son tablas no relacionales para poder tener una mayor capacidad a la hora de guardar información.

Imagen que contiene captura de pantalla

Descripción generada con confianza muy alta

*Ilustración 1 Imagen SAS GUIDE*

Se realizo una investigación con el area de auditoria, para lograr localizar la tabla de SAS en donde podrían estar las operaciones de renta variable, llegando a la conclusión que nombre de la tabla era operaciones de acciones.

Análisis de los datos.

Para analizar los datos de la tabla de operaciones se realizo una consulta en SQL y a través de un Query para determinar el numero de inserciones que tenia la tabla y no saturar SAS, esta consulta arrojo 15 millones de datos guardados dentro del repositorio. Se tomó la decisión de hacer un filtro especifico utilizando el lenguaje SQL para solo consultar el año 2018, danto como resultado 2 millones de datos.

Transformación de los datos de entrada.

Luego de tener los datos se ingresas a SAS para hacer el análisis de las distintas columnas de la tabla, inicialmente se verifica que la columna de la fecha viene en formato numero y se hace el proceso de cambio a su formato correspondiente, asi como también, examina y el nombre de los fondos no se encuentra solo se encuentra el código, se procede a hacer el cambio utilizando CASE de SQL.

Imagen que contiene captura de pantalla

Descripción generada con confianza muy alta*Ilustración 2 Transformación de datos SAS*

Selección y aplicación de las técnicas de minería de datos

Se selecciono la técnica algoritmo CART o arboles de clasificación y regresión, para esto se identificaron varios aspectos sobre las columnas de la tabla de operaciones, y se debían realizar varios filtros sobre los datos, ademas que se necesitaban datos anteriores con el detalle para realizar los reportes, cabe resaltar que SAS GUIDE es bastante coherente a la hora de hacer agrupaciones y realizar filtros sobre los datos.

Imagen que contiene texto, mapa, interior, cielo

Descripción generada con confianza alta  
*Ilustración 3 Ruta SAS realizada con árboles de decisión*

Ademas de las operaciones realizadas en renta variable se debe cumplir con las tarifas por fondos de la banca, por esta razón se realizó el árbol de decisión de la parte inferior de la ilustración 3, en donde se extrae cada una de las tarifas.

Imagen que contiene texto, mapa, interior

Descripción generada con confianza muy alta

*Ilustración 4 Unión de dos arboles*

Para hacer la unión de los dos árboles se hace a través de SQL, esta se asimila a la unión de dos tablas de base de datos, donde se tiene que tener una llave principal y se indican los campos que va a llevar la tabla resultado.

Finalmente, y con agrupaciones de datos por banca se finaliza la extracción y técnica de minería de datos, dando el resultado presentado en la ilustración 5, un total de 98637 operaciones y están distribuidas por distintas bancas.

Imagen que contiene captura de pantalla

Descripción generada con confianza muy alta

*Ilustración 5 Resultado operaciones y tarifas por banca.*

Presentación.

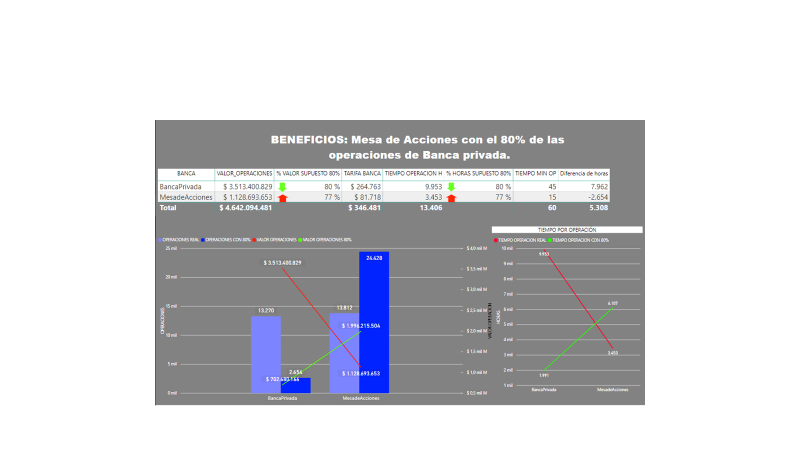
La presentacion de la información se realiza atraves de la herramienta PowerBI, se hace una regresion dentro de el arbol de decision de SAS para ingresar el detalle da las operaciones y asi hacer los graficos.



*Ilustración 6 Mapa de calor.*

Se realiza un mapa de calor para indicar las bancas que tienen el mayor numero de operaciones durente los meses del 2018, se llega a la conclusion que la mayor concentracion de operaciones esta entre Personas y pymes, mesa de acciones y banca privada.

A través del reporte se analizó que se podría hacer una reducción de costos dentro de renta variable en el Banco entre mesa de acciones y banca privada, ya que la cantidad de operaciones durante el año fue muy parecida y se mantuvo un estándar, pero las tarifas eran totalmente distintas una era más económica.

  
*Ilustración 7 Simulación.*

Se simulo un escenario en el cual se traspasaron el 80% de las operaciones al fondo mesa de acciones que tiene una tarifa de operación de $81.718 ahorrando $183.045 por cada operación realizada. En la ilustración 7 se demuestra que puede haber una utilidad de $2.811.006.285 que se puede ahorrar el banco dentro de las operaciones de renta variable.

Actualmente se esta a la espera de la presentación de este análisis de minería de datos ante el presidente del banco y cada una de las gerencias de las bancas involucradas para realizar la toma de decisiones dentro de renta variable para el próximo año.