PROYECTO DE GRADO

Presentado ante la ilustre Universidad De Los Andes como requisito final para obtener el Título de Ingeniero De Sistemas.

Minería De Datos Aplicada en Series de Tiempo en el Mercado de las Criptomonedas.

Por:

Br. Eris Izairal Ropero Núñez

Tutor: Profesor Rafael E. Borges

Diciembre 2020

Minería De Datos Aplicada en Series de Tiempo en el Mercado de las Criptomonedas.

Br. Eris Izairal Ropero Núñez

Proyecto de Grado — Investigación De Operaciones —69 paginas

Resumen: Hoy en día los avances tecnológicos y las aplicaciones del área de la estadística y la computación, permiten realizar predicciones de eventos a través de la minería de datos. Esta investigación es de naturaleza exploratoria, es decir, busca dar a conocer a través del descubrimiento de conocimiento específicamente mediante las técnicas de minería de datos como lo son (reglas de asociación y de análisis discriminante) aplicadas a datos del mercado de las criptomonedas, dicha exploración se desarrolla con la metodología de CRISP-DM, este tipo de metodología da orden a la investigación ya que se basa en un conjunto de pasos en las cuales en algunos pasos hay retroalimentación. Se genera posibles predicciones acertadas y coherentes, las simulaciones realizadas en rstudio y minitab arrojaron reglas fuertes y grupos clasificados correctamente todos los grupos dieron por encima del 87%, dando a conocer las bondades de la minería de datos aplicada a este mercado.

Palabras clave: Minería de datos, Criptomonedas, Reglas de asociación, Análisis discriminante, Series de tiempo.

ÍNDICE

Índice de tablas

Índice de figuras

Capítulo 1

1.1 Introducción

La minería de datos (MD) conocida también como ciencia de los datos ha ido incursionando en muchos campos de carácter investigativo de acuerdo a los intereses definidos, es por esto que la investigación que se apreciara a continuación está relacionada con las técnicas de (MD) como las reglas de asociaciones y el análisis discriminante en el mercado de las monedas virtuales, cada conjunto de datos representa una serie de tiempo la cual dichos datos son preparados por medio de estadística permitiendo ver el comportamiento de los datos previo a la simulación.

La vinculación que se desea, nos dará patrones o grupos extraídos de una cantidad limitada de datos (base de datos) por medio de simulaciones, estos resultados serán generados como reglas o grupos que dará soporte a las predicciones a través de la toma de decisión, para todos aquellos involucrados de una u otra forma en dicho mercado.

Dicha investigación se realizara bajo la metodología conocida como CRISP-DM que específicamente para la exploración de datos es la más usada y conocida, esto se debe a que se cumple con unas series de etapas en las cuales en algunas hay retroalimentación, es decir, se puede regresar a la etapa anterior para evitar ruidos que perjudique el proceso investigativo.

1.2 Antecedentes

Según [1] en 1996 surgió KDD (Knowledge Discovery in Databases) como un proceso completo de extracción de información, a su vez tiene como objetivos la preparación de datos y el proceso de inferir resultados obtenidos. KDD se puede considerar como un proceso de recuperación de información de grandes cantidades de datos, el aumento de datos se dio como consecuencia de la expansión de los mercados económicos la naturaleza de estos datos son numéricos, de texto ò de otra naturaleza ya sea o no en un mercado industrial, por ende que este crecimiento da interferencia en la toma de decisiones. Es por esto último mencionado que el concepto de KDD se presenta como un proceso no trivial para obtener modelos predictivos e identificar patrones que puedan ser o no considerados potencialmente útiles a la hora de tomar decisiones, considerado también como una herramienta para la predicción basada en la

experiencia, es decir, se realiza una investigación exploratoria dentro de un conjunto finito de datos, los patrones tienen como premisa que son generados a través de técnicas de minería de datos. La MD la definen [2] es una de las etapas del proceso de descubrimiento de conocimiento que consiste en un conjunto de técnicas desarrolladas para el descubrimiento automático de una base de datos previamente preparada de forma que se generen resultados plasmados en forma de patrones, tendencias entre otros que estén acordes a la realidad en los cuales se desea predecir a futuro, además hoy día también se han desarrollado metodologías para llevar una investigación coherente dentro de esta etapa, existen softwares basados en algoritmos diseñados bajo diversos enfoques como lo son las redes neuronales, lógica difusa, algoritmos genéticos y otras técnicas especializadas de análisis de datos no obstante a la hora del uso de algunas de estas herramientas se requiere de bases finitas con datos históricos interno o externo que representan una trama en el área económico o un área de estudio que han sido preparados, dando de este modo eficiencia a la investigación que se desea predecir y así tomar decisiones estratégicas.

1.2.1 Minería de datos

Hoy día se encuentran múltiples investigaciones usando las herramientas tecnológicas de MD, En [3] esta investigación es realizada para obtener reglas que brinden información del patrón que siguen sus consumidores con el objetivo de impulsar la venta de productos de nutrición menos vendidos en el mercado económico de Perú, es por esto que granulan la información de años anteriores obtenidos en el mercado, determinan que la metodología a usar CRISP-DM a su vez la tienda se apoya en otras investigaciones realizadas en Colombia y España, hacen uso de la técnica de reglas de asociación, clustering, redes neuronales, al final de la investigación de las tres técnicas aplicadas observan que las reglas de asociación da resultados precisos a la realidad, determinan que el consumo de suplementos se ve determinado por el conjunto de características del cliente como lo son estado civil del, edad del cliente, sexo del cliente, el número de hijos, el peso del cliente, la estatura del cliente, el ingreso mensual y la actividad física.

En temas como la deserción estudiantil han realizado aplicaciones de minería de datos como muestran en [4] en la presente investigación se apoyan en las técnicas de minería de datos con el objetivo de predecir el comportamiento de la deserción de estudiantes que afecta a la comunidad universitaria en Pasco Perú para realizar un plan estratégico que permita disminuir el índice de deserción, realizan las simulaciones en los programas IBM SPSS Statistics 25 y Weka 3.8.3, los

resultados generados a través de la técnica de clasificación determinan que la variable promedio de notas influye significativamente en la deserción estudiantil, es decir, la mayor cantidad de estudiantes que abandonan es por presentar notas bajas, recomiendan a futuras investigaciones agregar la variable que determine el factor psicológico del estudiante.

En [5] la revista presenta una investigación la aplicación de minería de datos para el análisis de datos climatológicos y brindar una alternativa que permita observar el comportamiento del clima a través de datos climatológicos para guiar a un proceso de toma de decisión, el estudio se basa en los estados Chiapas, Oaxaca, Tabasco y Veracruz de México, se apoyan en la herramienta tecnológica de MD Watson Analytics bajo la metodología de CRISP-DM, la investigación concluye que la herramienta facilita los diversos escenarios de predicción para la toma de decisiones en dicha área de estudio ya mencionada.

1.2.2 Reglas de asociación

Según [6] las reglas de asociación han sido reconocidas como unas de las mejores técnicas de la minería de datos, refleja el estudio de la detección de fraude con tarjetas de crédito por medio del uso de dichas reglas sobre bases de datos transaccionales para predecir futuras acciones fraudulentas y prevenir posibles robos, la investigación da como resultados que entre las principales prioridades a la hora de tomar conocimiento esta conectados a las nuevas exigencias que plantea escenarios competitivos por lo cual resalta el comportamiento de sus clientes, productos, proveedores, vendedores y socios comerciales, entre las dificultades que no permitieron buenos resultados se debe a las representaciones variadas de los datos, recomienda el uso de la lógica difusa como alternativa del estudio.

Entre otras aplicaciones según [6] el objetivo es encontrar asociaciones para impulsar el uso de los recursos de la biblioteca universitaria y a su vez facilitar el trabajo realizados por el personal, la base de datos estuvo constituida por los datos que representa la circulación de los libros y de este modo brindar eficacia de la disposición de un libro en el análisis 2 escenarios por lo que el primer lo etiquetaron es “Espacios Físicos Cerrados” corresponde a el área que solo puede ser accedida por el personal que trabaja en dicha entidad y el segundo lo etiquetaron “Espacios Físicos Abiertos” corresponde al área en la cual concurre en la biblioteca en la búsqueda de un libro, hicieron uso de las características de las asociaciones como lo son el soporte, la confianza y el lift, a partir del análisis los resultados alcanzados asisten al personal

de la biblioteca al momento de planificar y decidir sobre la distribución física de los libros en las estanterías, favorece los tiempos de respuesta de los bibliotecarios ante solicitudes de libros, así como también la ubicación próxima de los libros.

1.2.3 Análisis Discriminante

Según [6] hacen uso del análisis discriminante en la previsión de la insolvencia en las empresas de seguros no vida, tomando como variable explicativa ratios financieros y datos de periodos anteriores permitiendo construir reglas de clasificación para asignar una empresa al grupo de los potencialmente solventes o al grupo de los potencialmente insolventes, los análisis de las clasificaciones de los grupos fueron buenos ya que tuvo un 80% de aciertos de modo que la herramienta ha sido de gran utilidad a la hora de analizar la situación financiera en las empresas de seguros no vida siendo un método rápido y objetivo de evaluar estas empresas.

Algunas aplicaciones buscan en sus estudios demostrar la efectividad de la técnica método análisis discriminante, como muestran en [7] cuyo objetivo es usar dicha herramienta para observar la adicción a los dispositivos móviles conocidos como smartphone en el Centro Universitario Temascaltepec específicamente a los alumnos de la licenciatura en informática administrativa, la base de datos está constituido por una tabla de n individuos y se determinó variables que actúan como perfil de característica de cada una de ella, los resultados de las simulación arrojo que solo 2 variables marcan la diferencia en la representatividad de género, así como también entre las interrogantes como lo es ¿Con que frecuencia haces uso de tu dispositivo móvil? el grupo de las mujeres tuvo un 72.97 % de buena clasificación en comparación al grupo masculino que fue de 54.05% de buena clasificación, en general la investigación encontró el objetivo de demostrar la efectividad del método discriminante ya que determino diferencia en la adicción por género.

Se han realizado aplicaciones en el área de productividad y rentabilidad del petróleo y gas en [8] refleja la investigación entre los años comprendido por los periodos del 2008 y 2010 en Colombia, su metodología consistió en calcular los indicadores de productividad y rentabilidad del petróleo y gas de 116 empresas, la técnica de análisis discriminante les permitió observar la pertenencia y discriminación de las variables de estudio, las simulaciones les permitió concluir que en los periodos estudiados se observa un estancamiento así como también la variable del

indicador margen bruto presenta una diferencia significativa, la función de discriminación presento un 57.3% de efectividad.

1.2.4 Series de tiempo

Según [6] representa un conjunto de observaciones en la cual cada punto constituye un periodo en el tiempo t, permitiendo el pronóstico de distintos eventos de interés en el futuro, permite realizar predicciones y proponer bases de planificación en las áreas de economía, comercio, producción, inventario, control y optimización de industrias. En las aplicaciones en series de temporales según [9] la investigación muestra un modelo para la clasificar series temporales, previamente analizan patrones frecuentes en la secuencia discretizada, es decir, transforman datos numéricos a datos simbólicos de acuerdo al interés del área de estudio que constituyen su base de datos, ilustran con ejemplos lo que significa y los tipos de una serie de tiempo, los datos corresponden pruebas médicas de Potenciales Evocados Auditivos de Tronco Cerebral (PEATCs), el método propuesto por la investigación cuenta con 3 etapas precisas como los son el proceso de transformación, el proceso de descubrimiento y el proceso de clasificación así como también tiene perfecta funcionalidad para el medico como para el principiante esto se debe a que disminuye el tiempo que hay que usar debido a que emplea un cálculo automático.

En aplicaciones a series de tiempo, según [10] La investigación se apoya en las técnicas de las series de tiempo para implementar estrategias que permita disminuir la deserción estudiantil en la universidad de Cartagena de Colombia, plasmaron el número de estudiantes que abandonaron en una línea tendencial a partir de aquí realizaron propuesta con el fin de disminuir dicha cantidad, consideraron las estrategias propuestas por Majzub y Muhammad que establece generar un modelo de financiamiento pensado en el estudiante, detección de señales de alerta temprana entre otros, concluye en implementar como estrategia accionar modelos educativos flexibles, retroalimentación con los estudiantes entre otros.

1.2.5 Criptomonedas

Hoy día los estudios o investigaciones de las aplicaciones de las monedas virtuales abundan en el sector que ha tenido más influencia ha sido en lo económico y financiero, en este caso entre la moneda más conocida esta bitcoin denominada como (BTC) que es valorada del mismo modo que otros activos como lo es el oro, a continuación conoceremos algunas de sus aplicaciones.

Según [11] Representa una tecnología virtual que no requiere de intermediarios, es decir, constituye un sistema no centralizado, funciona bajo un sistema de criptografía denominado blockchain que brinda mayor seguridad en el momento de realizar transacciones generando confianza para el inversor, esta tecnología no está respaldada por ninguna divisa de forma que debe cumplir cierto parámetros para considerarse una criptomoneda, hoy día la criptomoneda más popular a nivel mundial es bitcoin a su vez es reconocida como la primera moneda virtual.

Entre las aplicaciones de criptomonedas en el sector financiero según [12] refleja un estudio de la adopción del blockchain que representa el protocolo del funcionamiento de las monedas virtuales, esto debido a que el mercado de la criptomonedas constituye hoy día unas de las innovaciones tecnológicas más poderosas en la economía, la investigación realiza un análisis a través del modelo de Bass de este modo les permitió tener características de un posible ajuste al ámbito financiero a un conjunto de monedas virtuales, los resultados muestran una adopción exitosa debido a la seguridad de las transacciones sim embargo este proceso aún se encuentra en evolución, el estudio concluye que no existe un registro de las aplicaciones implementadas que permita medir el alcance de esta adopción.

En el área de la educación según [13] este documento se centró en sus posibles aplicaciones educativas y se examinó cómo se puede usar la tecnología blockchain para resolver algunos problemas educativos. Introduce por primera vez las particularidades y ventajas de la tecnología blockchain a continuación explorando algunas de las aplicaciones actuales de blockchain para la educación, el estudio propone algunas aplicaciones innovadoras del uso de la tecnología blockchain, y también se discuten los beneficios y desafíos del uso de la tecnología blockchain para la educación.

1.3 Planteamiento del problema

La Minería de datos (MD) es una ciencia que nace para lograr, extraer y comprender el comportamiento de grandes cantidades de datos usualmente de empresas. Hoy en día es utilizada como soporte investigativo para la predicción en la toma de decisiones.

Según [14] Hoy día es nuevo escuchar opiniones del tipo hay que segmentar a los usuarios a través de la (MD), aplicar la (MD) al historial de compra dará premisas para conocer lo que requieren los usuarios de nuestros servicios o en el mercado se está posicionando por aplicar las nociones de (MD). La apreciación estadística confronta de manera positiva a la minera de datos como un sistema de bajo costo y alto rendimiento empresarial, frente a una sociedad de consumidores en todos sus niveles.

En un mundo globalizado, la MD como ciencia está siendo utilizada en empresas, entidades bancarias y en todas las modalidades empresariales existentes ya sean de manera física o virtual, con la finalidad de asumir decisiones en función de los resultados obtenidos a través de la aplicación de técnicas. En Venezuela no es común el uso de dicha ciencia, sin embargo en [15] se refleja la aplicación de tecnicas de (MD) como lo son las cadenas de Markov, Bayes Ingenuo y Clustering, en el área del Módulo de Departamentos de la página de Control de Estudios y Evaluación de la Universidad del Táchira, con el objetivo de explorar modelos ajustados a distintos grados que ayuden al reconocimiento de patrones de las distintas actuación de los usuarios en un determinar sitio web, permitiendo observar si un usuario es quien dice ser.

En [16] consiste en la muestra de los conceptos, como usarla y el crecimiento en las américas de la moneda virtual bitcoin esto debido a que en el 2009 surge como alternativa a la moneda fiduciaria, estableciendo al bitcoin como un cripto-activo de mayor valor en el mercado de las monedas virtuales en el mundo.

Debido a la escaza aplicación que se le ha dado a la MD en criptomonedas es plantea usar las series de tiempo como herramienta para vincular los datos de las criptomonedas con la ciencia de datos. En investigaciones que se vinculen los mercados de las monedas virtuales con las series de tiempo en [17] muestran un estudio para certificar la relación que pueda o no existir entre los precios del bitcoin con otros mercados como lo son (petroleo Brent y Oro) la exploración se inició con preparación de datos mensuales en un rango de 6 años comprendidos

entre 2012 a 2018, a su vez sobre las series de tiempo del periodo captado aplicaron estimación de un modelo VAR, generando como resultados relaciones claras y pendiente positivas con distintos mercados bursátiles como el caso de Sow Jones sim embargos se observó casos atípicos como la relación del BTC con el índice bursátil de Shanghai que no es clara más sin embargo mostro una pendiente positiva.

La presente investigación que se desea explorar desde el enfoque financiero se quiere aprovechar los patrones con respuesta coherentes logrando generar toma de decisiones especulativas, como expone Ceballos: [18]

> "Un tema interesante al analizar el TIEMPO financiero es la predicción que se relaciona con el TIEMPO de la especulación. El análisis de series temporales, no para su explicación, sino para predecir sus valores futuros y aprovecharse".

Es evidente que el escaso uso de la MD con respecto a las criptomonedas y de allí surge la investigación con respecto al tema, dando origen a una gran interrogante: **¿Sera posible encontrar reglas de asociación fuertes y grupos correctamente clasificados entre monedas con los cuales se puedan tomar decisión para realizar inversión en dicho mercado?**

Con base en esta interrogante se pretende realizar el estudio para observar las bondades de las técnicas conocida como "reglas de asociación" y "Análisis Discriminante", aplicadas a un conjunto de serie de tiempo en el campo de las criptomonedas, dando respuesta a los objetivos de la investigación.

1.1 Alcance

La investigación es de alcance global, es decir, podría generar beneficios estando en cualquier parte del mundo, debido a que los datos que se usaran para ser sometidos en la metodología ya descrita anteriormente, son extraído de forma gratuita de una plataforma, en la que desde cualquier lugar del mundo se puede tener siempre y cuando se tenga el acceso a internet.

1.5 Objetivos

1.5.1 Objetivo general

Interpretar patrones y grupos a través de técnicas de minería de datos, aplicada a una base de datos en la cual contendrá series temporales pasadas de un conjunto de criptomonedas para la toma de decisión.

1.5.2 Objetivos específicos

- Identificar las características de las monedas virtuales que se visualizan en la plataforma CoinMarketCap para la importación de los datos correspondientes a los precios del año 2009 en adelante.
- Usar técnicas estadísticas para la selección de los datos correspondientes a los precios de apertura y cierre de las criptomonedas.
- Transformar los datos correspondientes al rendimiento en atributos discretos por medio de estandarización estadística que permitan tener reglas fuertes y grupos clasificados correctamente.
- Inferir las reglas de asociación provenientes de simulaciones que sirva como soporte para la toma de decisión.
- Inferir grupos provenientes de simulaciones que sirva como soporte para la toma de decisión.

1.6 Metodología

La metodología empleada en la presente investigación es llamada CRISP-DM (del inglés Cross Industry Standard Process for Data Mining). Consiste en un conjunto de etapas o fases en las cuales en algunas de estas etapas hay retroalimentación, está representada por un diagrama como:

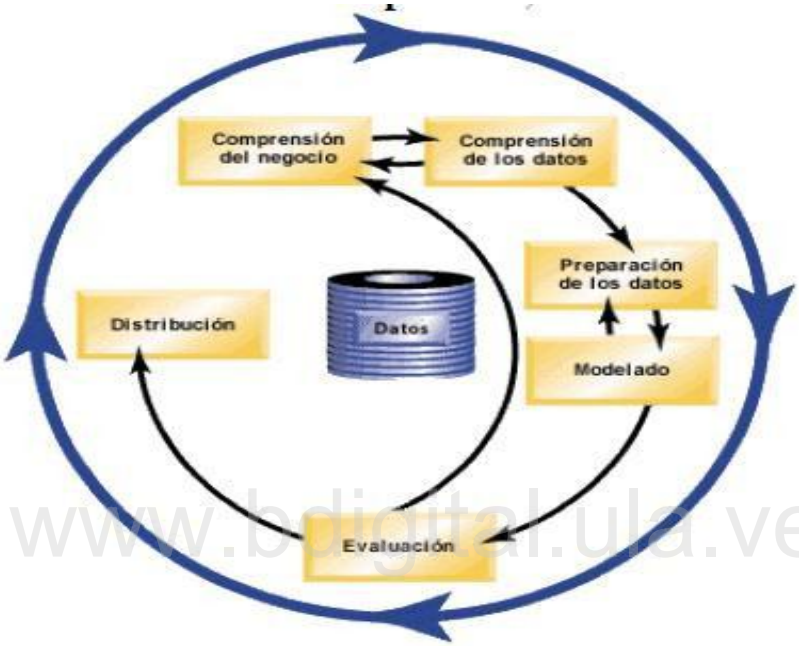


Figura 1 Representación de la metodología CRISP-DM

Comprensión del negocio: En esta etapa se refleja los objetivos de negocio y los de minería de datos, se visualiza el proceso de retroalimentación con la fase de comprensión de los datos, ya que estos objetivos pueden ir variando a medida que en la investigación se concreta los fines.

Comprensión de los datos: Los paramentos a seguir consiste en la recopilación de datos iniciales, descripción de los datos, exploración de datos, verificación de calidad de datos.

Preparación de los datos: Esta comprendida por la selección de datos, limpieza de datos, construcción de nuevos datos, integración de datos, formato de datos.

Modelado: Comprende la selección de técnica de modelado, modelado de supuestos, generación de un diseño de comprobación, generación de modelos, configuración de parámetros, evaluación del modelo.

Evaluación: Comprende la evaluación de los resultados, proceso de revisión, determinación de los pasos siguientes.

Distribución: Comprende la planificación de distribución, planificación del control del mantenimiento, creación de un informe final, revisión final del proyecto.

1.7 Justificación

Este estudio consistirá en interpretar reglas de asociación y grupos aplicadas a series de tiempo de un grupo de criptomonedas en la cual los datos deben estar discretizados, se aplica el algoritmo a priori, la cual se encarga de generar las asociaciones, de igual forma se aplicara cálculos para determinar el porcentaje de clasificación correcto. Cuyo resultado arrojara la tomar decisiones para el riesgo de inversión.

Por lo antes descrito, aportara una investigación más en el área de la ciencia de los datos de manera explícita o implícitamente, el estudio permite visualizar el comportamiento de demanda y oferta del mercado en el que en función de un interés personal o empresarial pueda o no generar una ganancia. Desde el punto de vista estadístico contribuye a la innovación y exploración del conocimiento de la MD como herramienta para toma de decisiones, en tiempo real sin afectar las transacciones La presente línea de investigación aporta las fuentes teóricas en MD aplicadas en series de tiempo en el mercado de criptomonedas, para los futuros trabajos de grados de Ingeniería De Sistemas que quieran indagar en la ciencia de los datos.

Capítulo 2

MARCO TEORICO

El pedestal teórico de esta investigación versa sobre la minería de datos aplicada en series de tiempo en el mercado de las criptomonedas.

Abordando distintas aristas tendentes a visualizar, comprender y realizar una aprehensión de aspectos relativos a la toma de decisiones en mercado competitivo y rentabilidad. Al respecto, se disponen una serie de investigaciones que permiten entender campos del conocimiento de la minería de datos y su importancia en el mercado de las monedas virtuales.

2.1 Minería de datos

Minería de Datos es un término que comprende, técnicas y herramientas para extraer información de grandes bases de datos y obtener los resultados estadísticos que favorezcan la toma de decisiones. Para lo cual se tomarán el enfoque de minería de datos conocida como: MD el cual es un proceso de planteamiento de distintas consultas y extracción de información útil, patrones y tendencias previamente desconocidas desde grandes cantidades de datos posiblemente almacenados en bases de datos" [19].

2.1.1 Proceso de KDD

El proceso de KDD permite conocer la identificación de los datos, el lugar dónde se pueden encontrar y cómo conseguirlos. Este proceso utiliza los datos más apropiados para el objetivo formulado.

En la figura 1.2 se muestra la metodología que debe seguirse para obtener conocimiento a partir de los datos que se encuentran en la base de datos.

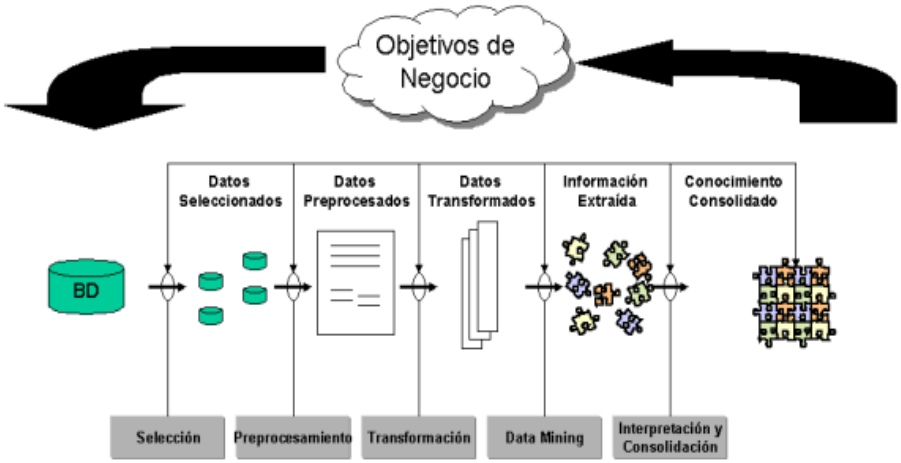


Figura 2 Metodología para el descubrimiento de conocimiento en bases de datos.

El conocimiento de la base de datos está fundamentado en la selección de datos concretos o de características específicas que promuevan un resultado de acuerdo a la intensión del mercado propuesto. [20] Esta data inicial se transforma en datos procesados los cuales arrojaron un resultado de apreciación discreta convirtiendo en datos transformados esta data mining expresa claramente la información extraída la cual es consolidad en resultados consolidados o conocimientos consolidados. Este proceso permite tomar decisiones asertivas.

2.1.2 Almacenamiento de datos

La minería de datos consiste en recopilar y analizar datos para extraer información rentable mediante el tratamiento de los datos con la finalidad de eliminar datos duplicados, datos vacíos, entre otros) permitiendo que estos datos sean manipulables para que los resultados sean consistentes con lo requerido [21]. Es por ello, que la minería de dato es el conjunto de técnicas empleadas para la extracción del conocimiento en grandes volúmenes de información.

2.1.3 Tipos de datos

Como todo proceso informático, requiere de datos de entrada para la ejecución de los procesos posteriores. A cada grupo de datos que representa una característica en especial, llamada atributo; mientras que a la recopilación de atributos de un caso particular se le conoce como instancias y al conjunto de instancias se le llama base de datos.

Entre los tipos de datos en los que se puede representar la información de las organizaciones. Estos pueden ser obtenidos mediante la categorización de los tipos de datos los cuales se subdividen en datos sin dependencia y datos con dependencia. Datos independientes. Estos se refieren a aquellos datos simples recolectados de forma independiente, es decir, sin que se necesite conocer otro valor para ser interpretados. Datos dependientes. Son aquellos datos que dependen de otro factor como el tiempo o el espacio para ser utilizados, estableciendo relaciones implícitas con otros datos.

Los datos Independientes son numéricos categóricos binarios dependientes series de tiempo secuencias discreta. En cuanto a los datos espaciales: Expresan una medida cuantitativa de alguna característica mediante el conjunto de números reales. Conocidos como datos continuos. Mientras que los datos categóricos. Datos cualitativos discretos sin un orden natural entre ellos. Para los datos binarios: es un dato cuantitativo que solo puede tomar uno o dos valores discretos. De igual manera las series de tiempo: es un repertorio de valores, pertenecientes al conjunto de números reales R, recolectados durante un periodo de tiempo específico. [Formalmente, una serie de tiempo se puede expresar como $S = \{S_1, S_2 \cdots, S_t\}$ donde S_t es el valor obtenido en el instante t.

Es por ello, que la minería de datos en series temporales tiene 2 valores que viven en dos espacios diferentes: espacio de valores, que es el conjunto de valores resultante de las mediciones; y espacio temporal, que son los valores de los momentos en que se tomó cada medición. Existe una serie de tiempo especial llamada secuencias discretas o cadenas de caracteres: es una serie de tiempo, pero en lugar de valores continuos o reales, son almacenados valores discretos o categóricos. Otro dato espacia son sobre las mediciones de las características de objetos obtenidos en ubicaciones espaciales almacenadas como coordenadas.

2.1.4 Procesamiento de la información

Dentro del proceso de almacenamiento de los datos, se presentan variables en cuanto al origen de los datos. Estos datos pueden ser: Datos atípicos: Los cuales son valores que sobresalen del modelo general de los datos. Generalmente, tienen un comportamiento aleatorio por lo que son difíciles de detectar y eliminar en los procesos de extracción de conocimiento. En cuanto a los datos incompletos son valores faltantes en las bases de datos debido a fallas en el almacenamiento o simplemente a la falta de medición de sus características en particular. Por su parte, los datos no

estandarizados: Son datos que se presentan en diferentes escalas de referencia de valores, es decir, se pueden presentar características medidas en intervalos pequeños ((0, 1) y otros medidos con magnitudes muy altas [22].

Por lo tanto, la Minería de Datos en series temporales elimina las inconsistencias de los datos, como son: datos faltantes, inconsistentes, con ruido y/o atípicos. Por otro lado, para la detección de datos atípicos, se manejan métodos estadísticos para analizar el comportamiento de los datos e identificar aquellos que no se ajusten al modelo de los mismos.

En cuanto a la transformación de datos, en esta fase, los datos son escalados a un mismo rango de valores usando las medidas de tendencia central y dispersión de la estadística tradicional mediante la reducción de la dimensión. Proceso que consiste en reducir el número de variables aleatorias o atributos, con la finalidad de obtener representaciones más compactas del conjunto original.

La fase de preprocesamiento de la información, no realiza tareas de descubrimiento de conocimiento, debido a que, si los datos de entrada son inconsistentes, con una alta tasa de ruido y con datos faltantes, los resultados obtenidos serán inútiles y fuera de la realidad del problema a resolver. Pero si se presentan inconsistencias, es indispensable realizar este proceso cada vez que se desee aplicar técnicas de minería de datos.

2.1.5 Herramientas para el análisis de la minería de datos

La minería de datos, consiste en descubrir patrones entre los datos y transformándolos en información refinada, lógica y procesable utilizando algoritmos específicos, análisis estadístico, inteligencia artificial y sistemas de bases de datos [23]. Para extraer los datos es necesario un sistema de software especializado para hacer exploración de datos.

2.1.6 Software conocidos en minería

El software Data Mining es el análisis de conjuntos de datos habitualmente extensos en variables para hallar relaciones entre ellos y concretar la información de forma útil .Está ligado a métodos novedosos, el propósito del Data Mining y de la Estadística es el mismo, siendo el matiz del tratamiento de grandes bases de datos lo que le confiere al Data Mining su especificidad.

RapidMiner

Es un software analítico predictivo de código abierto que se utilizado al iniciar cualquier proyecto de minería de datos trabaja con aplicaciones móviles y los chatbots, tienden a depender de esta plataforma de software para el aprendizaje automático, creación rápida de prototipos, desarrollo de aplicaciones, minería de texto y análisis predictivo.

IBM SPSS

Este software te permite generar una gran variedad de algoritmos de minería de datos sin programación. Permite la detección de anomalías, redes bayesianas, Cox y redes neuronales básicas que utilizan perceptores multicapa con aprendizaje retráctil.

R

La R para hacer Data Mining es gratis, tiene código abierto y es fácil de usar sin experiencia en programación. Se ejecuta en casi todos los sistemas operativos y puedes descargarle algoritmos súper avanzados para trabajar con grandes paquetes de información. R te permite manipular datos fácilmente, visualizarlos a través de gráficos interactivos y animados y realizar grandes análisis estadísticos de ellos.

SAS

El SAS Rapid Predictive Modeler, guía a través de un conjunto de quehaceres de minería de datos. Se usa principalmente a nivel empresarial para crear visualizaciones interactivas.

Python

Python es un lenguaje de código abierto gratuito permite aprender a crear toda clase de conjuntos de datos y a realizar análisis de afinidad complejos en cuestión de minutos. Es una herramienta de minería de datos extremadamente efectiva y eficiente.

Orange

Orange es herramientas gratuitas para hacer Data Mining debido a la visualización interactiva por ser un software de aprendizaje automático y de procesos de manipulación de datos y con flujos de trabajo de minería de datos precargados sobre textos, mapas de calor y diagramas de dispersión.

KNIME

Es el software más sencillo de manipular contiene diseño de tablas y gráficos interactivos.

Spark

Excelente en la minería de minería de datos de código abierto de simplicidad, velocidad y compatibilidad con una gran cantidad de lenguajes de programación que incluyen Python, R, Java y Scala.

Apache Mahout

Mahout este software crece continuamente a medida que los algoritmos implementados dentro de Apache Mahout evolucionan, cuenta con una extensa biblioteca en JAVA y su rendimiento al igual que su velocidad es impresionante, permite realizar operaciones matemáticas como el álgebra lineal y las estadísticas.

Weka

Weka admite las principales tareas de minería de datos, incluida la extracción de datos, el procesamiento, la visualización, la regresión, etc. Suponiendo que los datos estén disponibles en forma de un archivo plano. Weka te proporciona acceso a las bases de datos de SQL a través de la conectividad de la base de datos.

2.2 Reglas de asociación

Las reglas de asociación describen la relación entre los elementos de un conjunto de datos. Estas reglas nacen de la investigación de Agrawal, Imielinski y Swami [24] donde consideran la colección de datos que generan las compras en un supermercado, sirviendo como apoyo para saber qué conjunto de productos se compran y generar por medio de las reglas de asociación promociones. En [24] definen las reglas de asociación como:

Dado C como el conjunto de conceptos y T:= $\{t_i \mid i = 1. . . n\}$ como la base de datos de transacciones, donde n es el número total de transacciones y cada transacción t_i consiste en un conjunto de elementos:

$$t_i = \{a_{i,j} \mid j = 1. . . m_i , a_{i,j} \in C\},$$

Y cada elemento $a_{i,j}$ es un elemento del conjunto C y m el número total de elementos en t_i. El algoritmo calcula las reglas de asociación presentadas:

$$X_k \Rightarrow Y_k \, (X_k, Y_k \subset C, X_k \cap Y_k = \{\}).$$

Una regla de asociación es una implicación de la forma $X_k \Rightarrow Y_k$ donde X_k es un conjunto de algunos elementos de C también llamado antecedente y Y_k es un sólo elemento en C que no está presente en X, también llamado consecuente. La regla se satisface si las medidas de soporte, ecuación (2), y confianza, ecuación (3), sean iguales o mayores a las deseadas. El soporte de una regla $X_k \Rightarrow Y_k$ es el porcentaje de transacciones que contiene $X_k \cup Y_k$ como un subconjunto, y la confianza de una regla $X_k \Rightarrow Y_k$ está definida como el porcentaje de transacciones donde Y_k aparece si X_k se encuentra en una transacción:

$$\text{Soporte } (X_k \Rightarrow Y_k) = \frac{|\{t_i \mid X_k \cup Y_k \subseteq t_i\}|}{n}, \qquad \text{Ecuación (1)}$$

$$\text{Confianza } (X_k \Rightarrow Y_k) = \frac{|\{t_i \mid X_k \cup Y_k \subseteq t_i\}|}{|\{t_i \mid X_k \subseteq t_i\}|}, \qquad \text{Ecuación (2)}$$

Las medidas se pueden interpretar como: una regla con bajo soporte indicaría que habrá aparecido por casualidad. Sin embargo, una regla con baja confianza indicaría que no existe

relación entre el antecedente y el consecuente. Además, existe una diferencia entre $X_k \Rightarrow Y_k$ y $Y_k \Rightarrow X_k$, debido a que las reglas comparten el mismo soporte pero su confianza tiende a ser distinta.

Existen casos en que valores altos de confianza se deben a que el producto del lado derecho de la reglas tiene un soporte alto independiente del soporte del producto del lado izquierdo. Tenemos que el lift se representa como la confianza de la regla dividida por el soporte del consecuente.

$$\text{Lift} = \frac{\text{Confianza}(X \rightarrow Y)}{\text{Soporte}(Y)} \qquad \text{Ecuación (3)}$$

Cuando el lift es mayor a uno implica que la probabilidad del consecuente de la regla aumento una vez que se adquiere el antecedente, si el lift es igual a uno significa que la probabilidad no se vio afectada por lo tanto el antecedente no aporta información, si el lift es menor a uno implica que el antecedente tuvo un efecto negativo en la ocurrencia del consecuente de este modo su probabilidad baja.

2.2.1 Algoritmo a priori (Agrawal, 1994)

Busca [25] ítemsets frecuentes usando generación de candidatos. Su nombre se debe a que usa conocimiento a priori para la generación de ítemsets frecuentes. Este algoritmo se resume en dos pasos:

- Generación de todos los ítemsets que contienen un solo elemento, utilización de estos para generar ítemsets que contengan dos elementos, y así sucesivamente. Se toman todos los posibles pares de ítems que cumplen con las medidas mínimas de soporte inicialmente preestablecidas; esto permite ir eliminando posibles combinaciones: aquellas que no cumplan con los requerimientos de soporte no entrarán en el análisis.
- Generación de las reglas revisando que cumplan con el criterio mínimo de confianza. Es interesante observar que si una conjunción de consecuentes de una regla cumple con los niveles mínimos de soporte y confianza, sus subconjuntos (consecuentes) también los cumplen; en el caso contrario, si algún ítem no los cumple no tiene caso considerar sus súper conjuntos.

Así se obtiene un método para construir reglas con un solo consecuente, a partir de ellas construir reglas de dos consecuentes y así sucesivamente; todo se realiza mediante una pasada por

la base de datos para cada conjunto de ítems de diferente tamaño. El esfuerzo computacional depende principalmente de la cobertura mínima requerida, y se lleva prácticamente todo en el primer paso. El proceso de iteración del primer paso se llama level-wise y va considerando los superconjuntos nivel por nivel. De esta manera se tiene una propiedad anti-monótona: si un conjunto de ítems no pasa la prueba de soporte ninguno de sus subconjuntos la pasa; esto se aprovecha en la construcción de candidatos, para no considerar todas las opciones.

Transacción	Listado de productos adquiridos
T1	Computador, impresora
T2	Impresora, DVD
T3	Impresora, cámara de vídeo
T4	Computador, impresora, DVD
T5	Computador, cámara de vídeo
T6	Impresora, cámara de vídeo
T7	Computador, cámara de vídeo
T8	Computador, impresora, cámara de vídeo, scanner
T9	Computador, impresora, cámara de vídeo
T10	Impresora, scanner
T11	Computador, DVD
T12	Computador, impresora, DVD

Tabla 1 Ejemplo de datos de compras de productos

Cada transacción (T1, T2,...) representa una compra realizada por diferentes clientes de DISTCOL; al frente aparecen los productos comprados en cada transacción; por simplicidad, a cada transacción se asigna un identificador. A continuación se explica el algoritmo a priori con base en el contenido de la Tabla 1, que registra 12 transacciones de venta por parte de la empresa.

En la primera iteración del algoritmo, cada ítem es un miembro del conjunto de candidatos 1-ítemsets, C1. El algoritmo explora en orden todas las transacciones para contar el número de ocurrencias de cada ítem.

C_1	
Ítemset	**Contador de soporte (cont_sop)**
{computador}	8
{impresora}	9
{cámara de vídeo}	6
{DVD}	4
{scanner}	2

Tabla 2 Primer paso del algoritmo, contador del soporte de cada producto.

Supóngase que el contador de soporte requerido son dos elementos (min_sop = 2/12 = 16%). El conjunto de los 1-ítemsets frecuentes, L1, puede entonces ser determinado considerando los 1-ítemsets candidatos que satisfacen este mínimo soporte (min_sop).

L_1	
Ítemset	**Contador de soporte (cont_sop)**
{computador}	8
{impresora}	9
{cámara de vídeo}	6
{DVD}	4
{scanner}	2

Tabla 3 Segundo paso del algoritmo, comparación de los candidatos del contador de soporte con el min_sop.

Para descubrir el conjunto de los 2-ítemsets frecuentes L2, el algoritmo a priori usa L1 * L2 para generar un conjunto de candidatos de 2- ítemsets, C_2.

C_2
Ítemset
{computador, impresora}
{computador, cámara de video}
{computador, DVD}
{computador, scanner}
{impresora, cámara de video}
{impresora, DVD}
{impresora, scanner}
{cámara de video, DVD}
{cámara de video, scanner}
{DVD, scanner}

Tabla 4 Tercer paso del algoritmo, generación de C_2 candidatos desde L1.

Las transacciones son exploradas y el contador de soporte de cada ítemset candidato en C2 es acumulado, como se muestra en la Tabla 5.

C_2	
Ítemset	**cont_sop**
{computador, impresora}	5
{computador, cámara de vídeo}	4
{computador, DVD}	3
{computador, scanner}	1
{impresora, cámara de vídeo}	4
{impresora, DVD}	3
{impresora, scanner}	2
{cámara de vídeo, DVD}	0
{cámara de vídeo, scanner}	1
{DVD, scanner}	0

Tabla 5 Cuarto paso del algoritmo, contador de soporte para 2 itemsets.

El conjunto de los 2-ítemsets frecuentes, L2, es entonces determinado por aquellos que cumplan con el mínimo soporte.

L_2	
Ítemset	**cont_sop**
{computador, impresora}	5
{computador, cámara de vídeo}	4
{computador, DVD}	3
{impresora, cámara de vídeo}	4
{impresora, DVD}	3
{impresora, scanner}	2

Tabla 6 Quinto paso del algoritmo, candidatos que cumplen con el min_sop.

La generación del conjunto de candidatos de 3- ítemsets C_3, se muestra en la Tabla 7.

C_3
Ítemset
{computador, impresora, cámara de vídeo}
{computador, impresora, DVD}

Tabla 7 Generar C_3 candidatos desde L2

Las transacciones son exploradas y el contador de soporte de cada ítemset candidato en C3 es acumulado, como se muestra en la Tabla 8; posteriormente, en la Tabla 9 se muestran los ítemsets que cumplen con el min_sop.

C_3	
Ítemset	cont_sop
{computador, impresora, cámara de vídeo}	2
{computador, impresora, DVD}	2

Tabla 8 Octavo paso del algoritmo, contador de soporte para 3 itemsets.

L_3	
Ítemset	cont_sop
{computador, impresora, cámara de vídeo}	2
{computador, impresora, DVD}	2

Tabla 9 Candidatos que cumplen con el min_sop

El algoritmo usa L3 *L3 para generar un conjunto de candidatos de 4-ítemsets, C4; sin embargo, no existen candidatos de 4-ítemsets que cumplan con el min_sop, por lo cual el algoritmo a priori termina. (Fuente tomada y adaptada de Han, 2000).

A continuación se presenta el algoritmo a priori en seudo-código. Encuentra los ítemsets frecuentes. Entradas: Transacciones de una base de datos D; min_sop Salidas: L, ítemsets frecuentes de la base de datos D. Método:

(1) L1 = encontrar_1-ítemsets frecuente(D);
(2) Para (k=2; Lk-1≠∅; k++)
(3) Ck = generar_apriori (Lk-1, min_sop);
(4) Para cada transacción t ∈ D // examinar D para el contador
(5) Ct = subconjuntos (Ck-t) // obtener los subconjuntos t que son candidatos
(6) Para cada candidato c ∈ Ct
(7) c.contador++
(8) fin-para
(9) Lk = {c ε Ck | c.contador≥min_sop}
(10) fin-para
(11) retornar L = Uk Lk ;
Función generar_apriori(Lk-1 : (k-1)-ítemsets frecuentes; min_sop : mínimo soporte)
(1) Para cada ítemset l1 ∈ Lk-1
(2) Para cada ítemset l2 ∈ Lk-1
(3) Si (l1 [1] = l2 [1] ∧ (l1 [2] = l2 [2]) ∧ ... ∧ (l1 [k-2] = l2 [k-2]) ∧ (l1 [k-1] < l2 [k -1]))
Entonces
(4) c = l1 X l2 ; //generar candidatos
(5) Si subconjunto_infrecuente(c, Lk-1) Entonces
(6) eliminar c;
(7) Si No
(8) adicionar c a Ck
(9) fin-para
(10) retornar Ck ;

Función subconjunto_infrecuente(c : k-ítemset candidato; Lk-1 : (k-1)-ítemset frecuente)
(1) Para cada (k-1)-subconjunto s de c
(2) Si (s ∉ Lk-1) Entonces
(3) retornar Verdadero;
(4) retornar Falso;

2.3 Análisis discriminante

Tiene como objetivo analizar la relación entre una variable dependiente categórica con g modalidades, que se corresponden con los grupos analizados, y un conjunto de variables independientes $\mathbf{x_1}$, $\mathbf{x_2}$ … $\mathbf{x_p}$, métricas o cuantitativas, a partir de una serie de funciones discriminantes, [26] que son combinaciones lineales de las variables independientes que mejor discriminan o separan los grupos, y cuya expresión es la siguiente:

$$f_{km} = \mu_0 + \mu_1 x_{1km} + \mu_2 x_{2km} + \ldots\ldots\ldots + \mu_p x_{pkm} \qquad \text{Ecuación (4)}$$

Siendo $\mathbf{f_{km}}$ el valor o puntuación en la función discriminante para el caso **m** en el grupo **k**; $\mathbf{x_{ikm}}$ el valor de la variable discriminante $\mathbf{x_i}$ para el caso **m** en el grupo **k** y μ_i los coeficientes o ponderaciones de las variables $\mathbf{x_i}$.

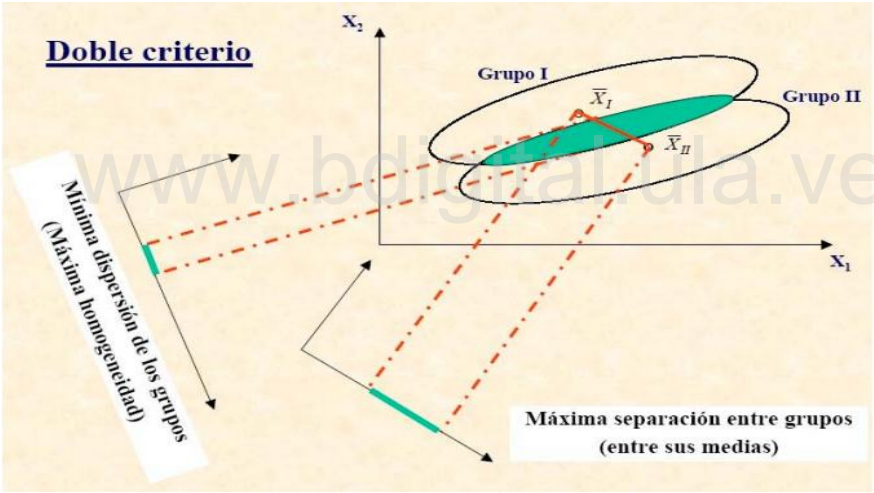


Figura 3 Representación gráfica de la función de fisher.

El análisis discriminante permite estudiar las diferencias entre dos (en el caso del análisis simple) o más (estaríamos ante el análisis discriminante múltiple) grupos de individuos definidos a priori, con respecto a varias variables simultáneamente.

El número de funciones discriminantes a obtener depende, a su vez, del número de grupos definidos por la variable dependiente, ya que se obtienen tantas funciones como grupos menos uno, salvo que el número de variables independientes incluidas en el modelo sea inferior, en cuyo caso el número de funciones discriminantes obtenidas coincide con el de variables. Las funciones discriminantes se obtienen de forma que la primera contiene aquellas variables explicativas cuyos valores más diferencian los distintos grupos, la segunda función es la segunda combinación de

variables que más discrimina entre los grupos, pero con la condición de que los valores obtenidos mediante la primera función no se hallen correlacionados con los de aquélla, y así sucesivamente. Estimadas las funciones discriminantes, su capacidad predictiva se evalúa mediante el establecimiento de una puntuación de corte óptima, que permite asignar los casos a cada uno de los grupos definidos por la variable dependiente, obteniéndose de la puntuación discriminante que corresponde a cada caso, a partir de los valores que presenta el individuo en la combinación de variables explicativas que forman las funciones discriminantes.

2.3.1 Distancia de mahalanobis

La regla de selección en este procedimiento [27] es maximizar la distancia $\mathbf{D^2}$ de Mahalanobis. La distancia multivariante entre los grupos a y b se define como la ecuación (5):

$$D= (n-k)\sum_{i=1}^{p}\sum_{j=1}^{p} w_{ij}^{-1} (x_i^{(a)} - x_i^{(b)})(x_j^{(a)} - x_j^{(b)}) \qquad \text{Ecuación (5)}$$

Donde n es el número de casos válidos, k es el número de grupos, es la media del grupo a en la i-ésima variable independiente, es la media del grupo b en la i-ésima variable independiente, y es un elemento de la inversa de la matriz de varianzas-covarianzas intra-grupos.

Siendo la variabilidad total de la forma de la ecuación (6).

$$\mathbf{T_{ij}} = \mathbf{w_{ij}} + \mathbf{v_{ij}} \qquad \text{Ecuación (6)}$$

La covarianza total es igual a la covarianza dentro de grupos, más la covarianza entre grupos.

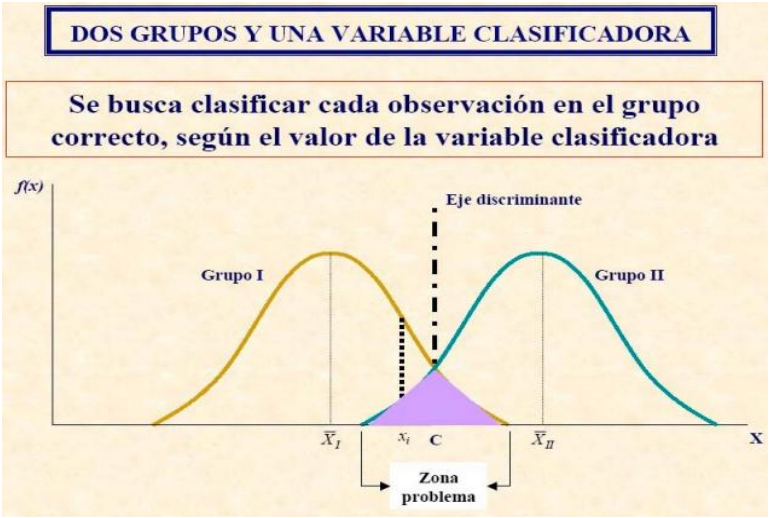


Figura 4 Representación gráfica del análisis de dos grupos y una variable clasificadora.

2.4 Discretizacion de serie de tiempo.

El proceso de discretización consiste en transformar un tipo de dato en valor discreto. Este proceso consiste en transformar la serie continua en secuencias discretas o cadenas de caracteres [28]. Todo proceso de discretización requiere un proceso previo para reducir la dimensionalidad de la serie antes de la transformación.

2.5 Apoyo a la toma de decisiones

Los sistemas de apoyo a la toma de decisiones son las herramientas que usan los directivos para tomar decisiones eficaces, y se basan en la teoría de la decisión. Se puede considerar a las herramientas de Minería de Datos como tipos especiales de herramientas de apoyo a la toma de decisiones. La minería de datos, se convierte en una herramienta estratégica que eleva los niveles de competencia en el cambiante mundo de los negocios. Esto debido a la rapidez con que se identifica y analiza información.

Entre las ventajas de la minería de datos está su facilidad de uso y la aplicabilidad de un conocimiento adecuado de los distintos tipos de algoritmos empleados, y del resultado derivado de la aplicación de la data mining. La cual presenta la tenencia de datos esenciales en una toma de decisión acertada. Con data mining las empresas cuentan con el manejo de los datos fidedignos para la planeación económica, inteligencia empresarial, finanzas, análisis de mercados y análisis de perfiles de clientes entre otros.

Para Molina, (2002). [29] La minería de datos (datamining), es herramienta fundamental de los negocios modernos, ya que es capaz de convertir los datos en inteligencia de negocios -Business Intelligent (BI)- dando así una ventaja de información, sobre prácticas de perfil y comercialización de productos.

2.6 Criptomonedas

Partiendo del origen de criptomoneda derivada del prefijo cripto, proviene de la palabra griega kruptos, que significa oculto, secreto. [30] Criptografía es el estudio de métodos de encriptación de información, principalmente utilizados para enviar un mensaje de manera segura y privada, y para la seguridad y autentificación de datos.

El Banco central europeo (BCE) definió en 2012 la criptomoneda como 'moneda virtual' de "un tipo de dinero no regulado, digital, que se emite y por lo general controlado por sus desarrolladores, y utilizado y aceptado entre los miembros de una comunidad virtual específica. Por lo cual, las criptomonedas son un subconjunto de las monedas digitales basadas en la criptografía.

De igual forma el periódico digital especializado en Bitcoin Coindesk define el término criptomoneda como: "Una forma de moneda basada únicamente en las matemáticas. En lugar de la moneda fiduciaria, que se imprime, una criptomoneda se produce mediante la resolución de problemas matemáticos basados en criptografía."

Por lo tanto, la criptomoneda es una moneda digital diseñada para funcionar como medio de intercambio la cual utiliza la criptografía para asegurar y verificar transacciones, para controlar la creación de nuevas unidades de una criptomoneda particular. De manera generalizada las criptomonedas son entradas limitadas en una base de datos que nadie puede cambiar a menos que se cumplan condiciones específicas.

2.6.1 Tipos de criptomonedas

Con la creación de nuevas criptodivisas, han aparecido nuevas plataformas para enviar, recibir y comprar distintos tipos de criptomonedas como por ejemplo Bitcoin, Bitcoin Cash, Ethereum y Litecoin. Hablamos de Coinbase, una plataforma y monedero digital que permite operar con criptomonedas de forma sencilla y segura. Entre los tipos de monedas virtuales destacan por su fácil manejo y accesibilidad al público:

BITCOIN

El Bitcoin es la criptodivisa o moneda digital pionera, creada 2008 bajo el nombre de Satoshi Nakamoto. Su lanzamiento no tenía el valor que posee ahora y la mayoría tampoco podía llegar a pensar que alcanzaría estos datos. [31] (símbolo: ₿; código: BTC) es un protocolo, proyecto de código abierto y red peer-to-peer que se utiliza como criptomoneda, sistema de pago y mercancía. Bitcoin es un sistema basado en UTXO (siglas en inglés de «Unspent Transaction

Output», comúnmente traducido al español como "monedas no gastadas"). Las cantidades de los UTXO están vinculadas a las direcciones que las pueden gastar por medio del registro de la cadena de bloques. Cuando un usuario (A) desea transferir unidades monetarias a otro usuario (B), construye una transacción especificando en ella la cantidad de bitcoins que cede de los UTXO que desea gastar y la dirección del destinatario (B), la firma con su clave privada y la transmite a la red Bitcoin (BTC). Los nodos que reciben la transacción verifican las firmas criptográficas y la validez de la posesión del UTXO antes de aceptarla y retransmitirla. Este procedimiento propaga la transacción de manera indefinida hasta alcanzar a todos los nodos de la red. La transacción es validada por un minero y minada en un bloque. Una vez que una transacción se encuentra en la cadena de bloques y ha recibido la confirmación de un número razonable de bloques posteriores, la transacción se puede considerar parte permanente de la cadena de bloques.

RIPPLE

Ripple (XRP) moneda digital se trata de un sistema totalmente seguro y encriptado cuya información de las transacciones son públicas pero la información del pago no, es decir, es un sistema confidencial donde el emisor y receptor son los únicos que disponen de la información y el código que la desencripta.

LITECOIN

Litecoin fue lanzado como una alternativa al Bitcoin. Tiene un límite superior al Bitcoin (84 millones, frente a 21 millones) y actualmente existen cerca de 55 millones de Litecoin en circulación. Por este motivo, son muchos los expertos que apuestan por esta criptomoneda en el futuro.

ETHEREUM

Ethereum y el Ether en 2017 las criptodivisas más rentables y es una alternativa basadas en la tecnología como Aragón o Stox. Es la segunda divisa digital en términos de capitalización por eso se le considera una gran alternativa al Bitcoin.

NEO

NEO o el "Ethereum de China» creado en 2014 por Da Hohgfei, NEO permite construir aplicaciones descentralizadas así como contratos inteligentes, además NEO es indivisible, no como ETH.

MONERO

El elemento diferenciador de la criptomoneda Monero es su anonimato. En su filosofía, Monero permite que en cada transacción sea totalmente anónima, incluido remitente, el destinatario y el volumen de la transacción. Para muchos esto supone un problema porque se cree que beneficia a los ciberdelincuentes. Según expertos, Monero tiene perspectiva de vivir su mayor crecimiento y podría llegar a los 50 $.

DASH

Dash es otra criptomoneda peer-to-peer, como el Bitcoin, pero que integra funcionalidades más avanzadas, como: las transacciones instantáneas y las transacciones privadas. Es una moneda digital que vive momentos volátiles y se puede adquirir a muy buen precio.

NEM

NEM, creada en 2015 utiliza la tecnología Blockchain para su gestión: Su elemento más innovador es que permite enviar mensajes, registrar nombres o crear cuentas con varios titulares. A la hora de invertir, es considerada una criptomoneda "low cost"

CARDANO

El sistema de bloques Cardano nació en 2015. Su moneda, Ada, ya cuenta con casi 26 millones de hasta las 45 mil que puede albergar. Esta moneda es su división en dos capas: una la

capa de pagos donde se ejecutan las transacciones y otra donde se llevan a cabo las aplicaciones y contratos, llamada capa de computación

FEDCOIN

Esta criptodivisa creada en Estados Unidos pretende ser el sustituto del dólar. Según sus creadores, es una de las monedas 100% internacional y digital, además, se podría eliminar de golpe los mayores riesgos a la economía que son las corridas bancarias y la hiperinflación.

2.6.2 Uso de las criptomonedas

Inicialmente las criptomomedas, nacieron como una forma de pago anónima y segura en 92 países, de los cuales 6.000 tienen presencia física y hay más de 13 millones de billeteras virtuales creadas. Bajo el espacio financiero las transacciones con criptomonedas facilitan la posibilidad de cambiar bitcoins u otra moneda virtual por monedas de curso legal, entre ellas euros o dólares, o las webs de trading, que permiten comprar y vender bitcoins como si fueran acciones.

Pero el uso de las criptomonedas también cruza al otro lado de la ley. El anonimato que proporcionan las convierte en el medio de pago perfecto para el pago de ilícitos. En la internet profunda (Deep web) los productos que se adquieren en los mercados clandestinos (armas, drogas, pornografía de menores, etc.) se pagan con bitcoin, Otro negativo uso es el pago de rescates tras el ataque de un ramsonware, la actividad ilícita conocida es el criptojackingo minería ilegal. Los ciberdelincuentes toman el control de ordenadores ajenos para utilizarlos para la explotación de criptomonedas como bitcoin.

2.7 Apoyo o beneficio de la minería de datos, la criptomoneda y toma de decisiones.

Los beneficiosos de la minería de datos y la inclusión de las monedas virtuales han consolidado imperios financieros de manera global y con el respaldo del cambio a moneda de curso legal, garantizando inversiones a nivel mundial en tiempo real. Otra forma de beneficios que aporta la minería de datos y la moneda virtual ha sido el crecimiento de grandes plataformas virtuales para el esparcimiento y la recreación. Para que ambas vertientes puedan apoyan la toma de decisiones, se requiere de proceso automático y reutilizables que ayuden a la competencia de los negocios obteniendo de forma rápida la información, descubriendo conocimiento y patrones en base de

datos, como resultado de la aplicación de las técnicas de minería de datos para consolidar empresas, organizaciones y negocios personales con la mayor ventaja de asertiva económica y financiera.

El uso de la minería de datos como soporte a decisiones en los negocios está basada en el analice de la evolución de la empresa, la comparación información en diferentes periodos de tiempo, con el cual se define las medidas cualitativas para los patrones obtenidos como son la precisión, utilidad y beneficio obtenido. Estos procesos de minería de datos a través de la aplicación de técnicas estadísticas avanzadas y nuevos métodos de extracción de conocimiento que permiten al talento humano conocer la proyección e impacto financieros de sus productos a corto, mediano y largo plazo.

Es por ello, que la minería de datos, las monedas virtuales y la toma de decisiones involucran a las organizaciones para definir y establecer un grupo personalizado de fuentes sobre las cuáles poder extraer datos; fijando las frecuencias diarias, semanales y mensuales para establecer un almacén de datos, cuyo principal objetivo está encaminado al descubrimiento de patrones de comportamiento mediante las bases de datos a través de la Inteligencia Artificial con la finalidad de generar nuevas oportunidades de mercado en diferentes sectores. En la banca se utiliza para aminorar los riesgos del mercado, aplicándose habitualmente a la calificación crediticia (rating) y a sistemas inteligentes antifraude para analizar transacciones, movimientos de tarjetas, patrones de compra y datos financieros de los clientes.

En cuanto al comercio, se aplica para explorar bases de datos para optimizar la división del mercado, analizar las relaciones entre cuantificaciones: edad clientes, género, gustos, entre otros. El conocer el comportamiento de los clientes permite dirigir campañas personalizadas de captación; en las zonas de mayor rentabilidad, reduciendo drásticamente el tiempo de búsqueda y evaluación para una inversión exitosa.

Estudios de mercados y valoración permite a los clientes optimizar sus estrategias al anticiparse y conocer el comportamiento del mercado. La herramienta del manejo de minería de datos y las criptomonedas se ha convertido proceso vital para comercio digital, imprescindible en la optimización de los procesos de venta.

Capítulo 3

PREPARACIÓN DE DATOS

En el capitulo presente se trataran los datos mediante herramientas estadisticas que permiten conocer el comportamiento previo de los datos iniciales para luego ser simulados.

3.1 Formato de los datos

Los datos se extrayeron de forma numérica de la plataforma coinmarketcap se tomo la data historica de cada moneda virtual, seleccionando los datos diarios de las columnas de apertura y los de cierre de bitcoin, etherium, riple y litecoin que comprende el periodo octubre del 2015 hasta septiembre del 2020. Estos datos previos permite el calculo de la rentabilidad de 1828 dias en la que cada dia se representa como un punto dentro de una serie de tiempo.

$$R = \frac{\text{cierre} - \text{apertura}}{\text{apertura}} * 100 \qquad \text{Ecuación (7)}$$

3.2 Selección de monedas

Se tomo las primera 4 monedas del ranting de la plataforma web coinmarketcap, es decir, las mas popular y conocidas entre los usuarios que conformar el mercado de compra y venta de dichos activos.

- Bitcoin (BTC): Es la primera criptomoneda creada, popular y de mayor cotizacion a nivel mundial en parte por la seguridad que ofrece a la hora de realizar una transacción, esto debido a su codigo abierto y sin intermediarios.
- Etherium (ETH): Es la segunda criptomoda mas conocida en el mercado, no solo opera como moneda virtual si no que ademas ofrece otros productos a los usuarios ya que permite ser programable, es decir, los usuarios pueden usarla para crear nuevas aplicaciones, al igual que bitcoin ofrece seguridad a la hora de realizar transacciones entre usuarios.
- XRP Ó RIPPLE (XRP): Funciona como criptomoneda y como una plataforma virtual que funciona bajo un protocolo de codigo abierto que esta diseñado para permitir transacciones rapidas y baratas.

- Litecoin (LTC): Funciona como criptomoneda y tambien funciona como software de codigo abierto publicado bajo la licencia MIT, inspirado y casi identico a BTC pero con tarifas de transacciones mucho mas bajas.

3.3 Distribución de los datos

Una de las caracteristicas de la fase de la preparacion esta en observar el comportamiento de los datos, es decir, verificar a que distribucion tienden los valores de estudio, este reconocimiento permite una vision mas clara desde un enfoque estadistico que permita coherencia en la construccion de la base de datos discretizada, esto se debe a que la exploracion que hasta ahora se ha percibido es de carácter continua, pero para llegar a los objetivos planteados al inicio de la investigacion se requiere sacar intervalos para el proceso de etiquetado o discretizacion.

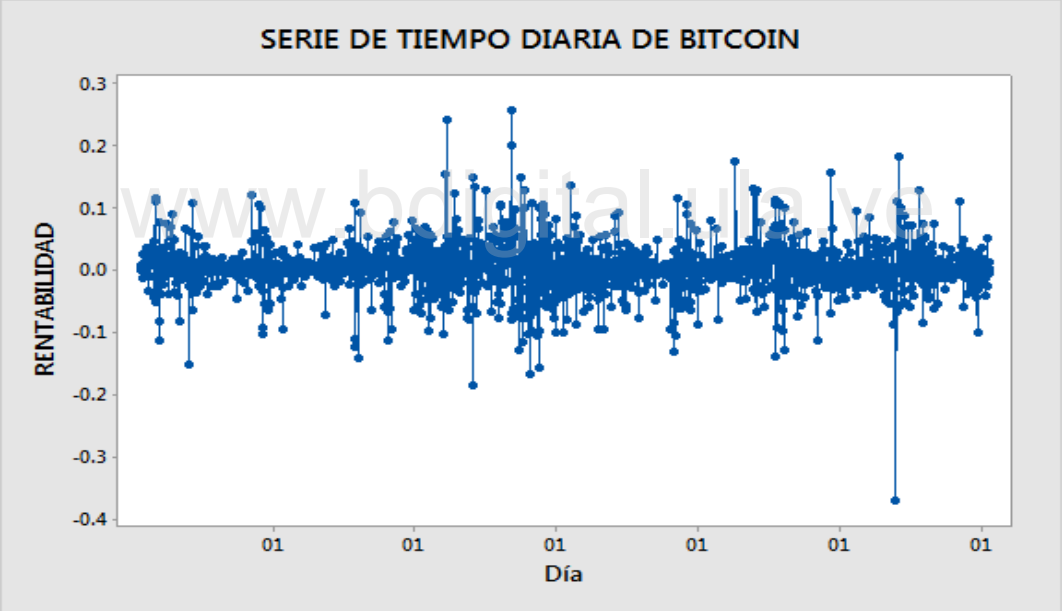
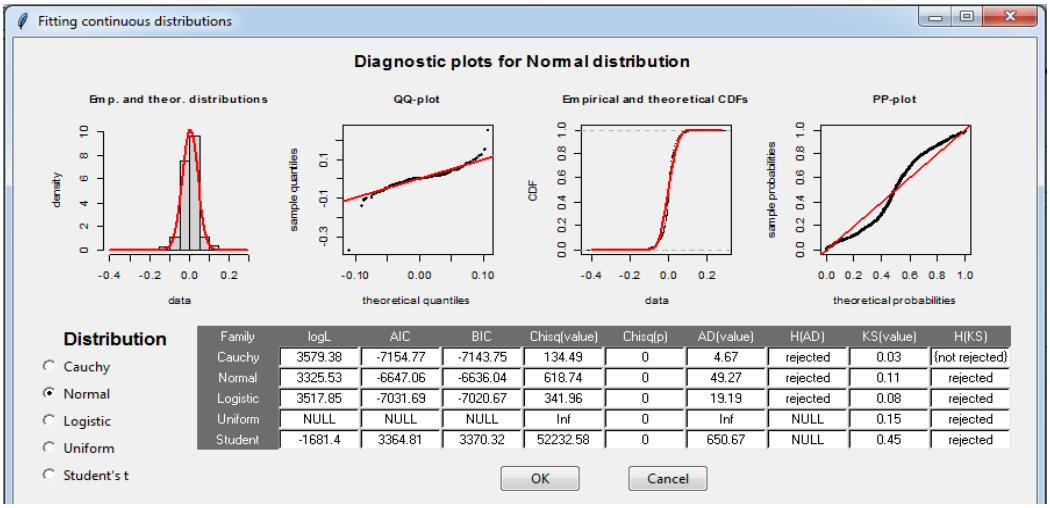


Figura 5 Representación gráfica de la serie de tiempo diaria de bitcoin.

Exploramos a que distribucion se daptan las rentabilidades diarias.



Family	logL	AIC	BIC	Chisq(value)	Chisq(p)	AD(value)	H(AD)	KS(value)	H(KS)
Cauchy	3579.38	-7154.77	-7143.75	134.49	0	4.67	rejected	0.03	{not rejected}
Normal	3325.53	-6647.06	-6636.04	618.74	0	49.27	rejected	0.11	rejected
Logistic	3517.85	-7031.69	-7020.67	341.96	0	19.19	rejected	0.08	rejected
Uniform	NULL	NULL	NULL	Inf	0	Inf	NULL	0.15	rejected
Student	-1681.4	3364.81	3370.32	52232.58	0	650.67	NULL	0.45	rejected

Figura 6 Resultados de las distribuciones a las que pertenecen los datos diarios correspondientes a bitcoin.

En la Figura 6 observamos que los datos no se ajustan a una distribucion normal que es lo que se desea, de este modo procedemos a los datos la transformacion de Jhonson.

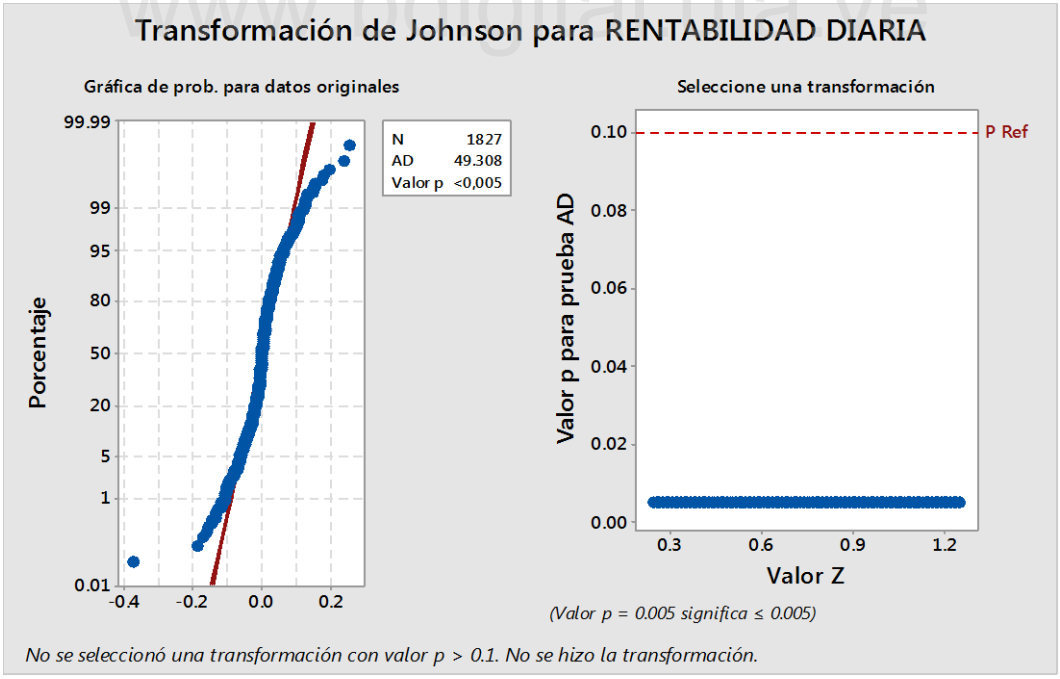


Figura 7 Resultados de la transformación a una distribución normal, se observa que no pudo realizar la transformación esto debido al comportamiento de los datos.

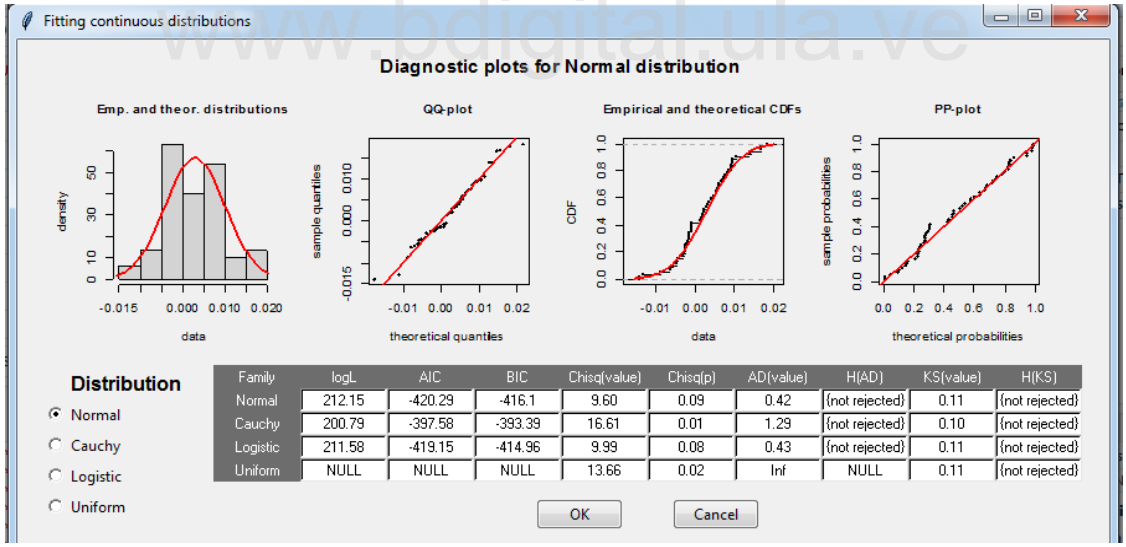
Se observa que las rentabilidades diarias de la moneda bitcoin no brindan informacion que permita realizar la transformacion a un conjunto de datos normales, por lo que se decide explorar su comportamiento mensual con el fin de obtener vertores de datos que se adapten a una distribucion normal permitiendo una preparacion precisa.

3.3.1 Distribución de datos bitcoin

Se representa la rentabilidad mensual de la moneda virtual bitcoin se ajusta a una distribucion normal, cauchy y logistica. Esto implica que los datos no proceden a una transformacion que permita trabajar con datos normales. Para efectos mas practicos se tratara los datos como a una distribucion normal en la cual dicho programa nos da los parametros para su estudio.

Chosen continuous distribution is: Normal (norm)
Fitted parameters are:

mean	sd
0.002832887	0.007050241



Family	logL	AIC	BIC	Chisq(value)	Chisq(p)	AD(value)	H(AD)	KS(value)	H(KS)
Normal	212.15	-420.29	-416.1	9.60	0.09	0.42	{not rejected}	0.11	{not rejected}
Cauchy	200.79	-397.58	-393.39	16.61	0.01	1.29	{not rejected}	0.10	{not rejected}
Logistic	211.58	-419.15	-414.96	9.99	0.08	0.43	{not rejected}	0.11	{not rejected}
Uniform	NULL	NULL	NULL	13.66	0.02	Inf	NULL	0.11	{not rejected}

Figura 8 Resultados de las distribuciones a los datos de rentabilidad mensuales correspondientes a bitcoin.

Dichos parámetros permite obtener los intervalos de confianza del 50 % que permite realizar la discretizacion de todo el vector de datos, por lo tanto 1- $\boldsymbol{\alpha}$ = 0.5 entonces $\boldsymbol{\alpha}/2$ = 0.25 nos queda $z_{\boldsymbol{\alpha/2}}$ = 0.67.

X1= 0.0076

X2= -0.0019

3.3.2 Distribución de datos Ethereum

En la siguiente imagen observamos que los datos que representa la rentabilidad mensual de la moneda virtual Ether nos muestras que no hay certeza de que los datos se ajusten a una normal. Esto implica que los datos proceden a una transformacion que permita trabajar con datos normales.

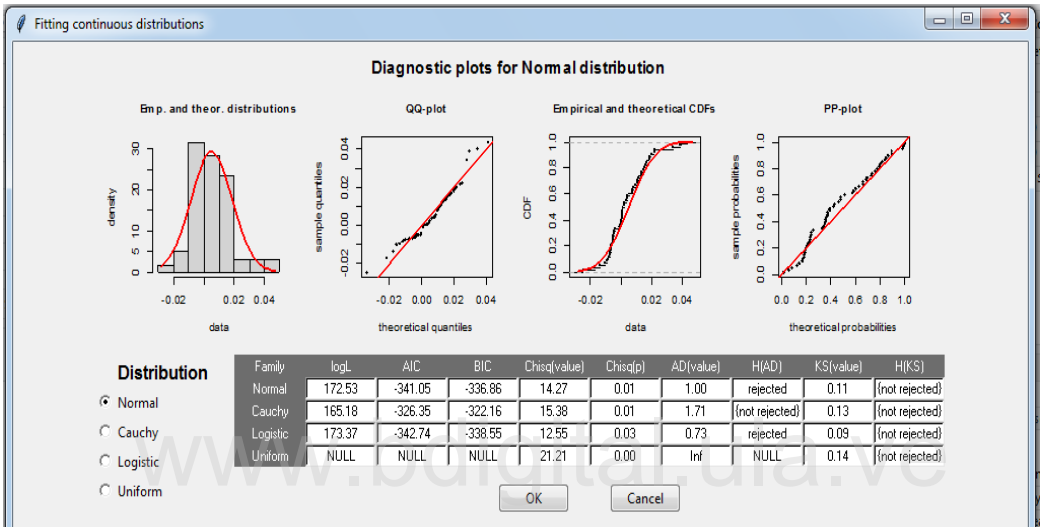


Figura 9 Resultados de las distribuciones a los datos de rentabilidad mensuales correspondientes a ethereum.

En vista que los datos de la rentabilidad poseen negativos y cercanos a cero, se realizara la transformación de Johnson, de modo que el programa Minitab nos arroja el siguiente resultado:

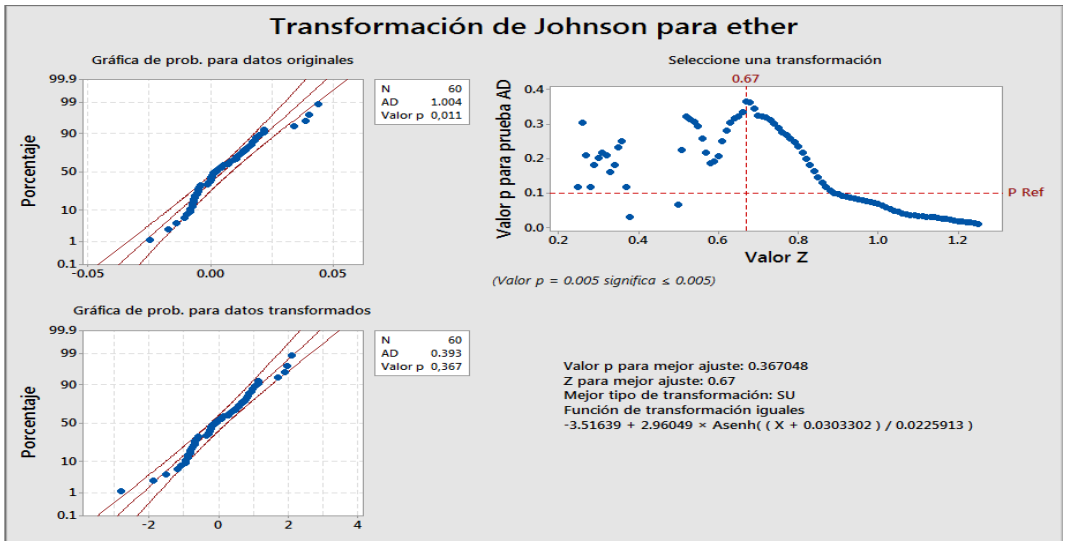
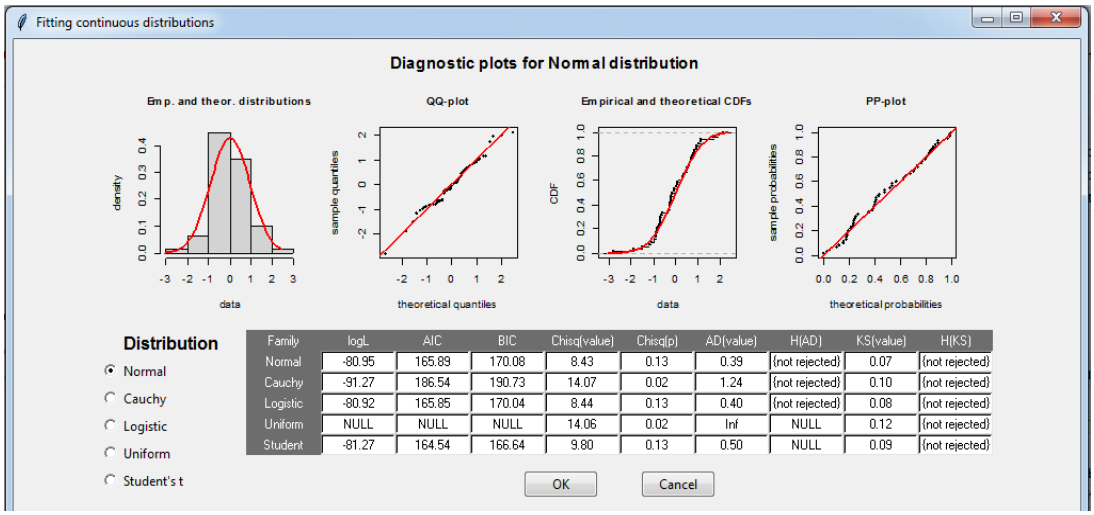


Figura 10 Transformación de las rentabilidades a distribución normal.

Observamos en la figura 10 que el valor requerido para transformar los datos a una distribución normal tiene el ajuste de 0.67, en la imagen inferior llamada "Gráfica de prob. Para datos transformados" se refleja que su significancia tiene un valor p = 0.367 que es mayor a 0.05 generando datos normales siendo este último valor el nivel de referencia para determinar si un conjunto pertenece a una distribución normal.

Al volver a Rstudio verificamos que no hay duda de la normalidad de los nuevos datos. Nos da los siguientes parámetros:

Chosen continuous distribution is: Normal (norm)
Fitted parameters are:
mean sd
0.0053835 0.9325392



Family	logL	AIC	BIC	Chisq(value)	Chisq(p)	AD(value)	H(AD)	KS(value)	H(KS)
Normal	-80.95	165.89	170.08	8.43	0.13	0.39	{not rejected}	0.07	{not rejected}
Cauchy	-91.27	186.54	190.73	14.07	0.02	1.24	{not rejected}	0.10	{not rejected}
Logistic	-80.92	165.85	170.04	8.44	0.13	0.40	{not rejected}	0.08	{not rejected}
Uniform	NULL	NULL	NULL	14.06	0.02	Inf	NULL	0.12	{not rejected}
Student	-81.27	164.54	166.64	9.80	0.13	0.50	NULL	0.09	{not rejected}

Figura 11 Resultados de las distribuciones a los que se adapta el vector de datos transformados.

Dichos parámetros permite obtener los intervalos de confianza del 50 % que permite realizar la discretizacion de todo el vector de datos, por lo tanto 1- α = 0.5 entonces $\alpha/2$ = 0.25 nos queda $z_{\alpha/2}$ = 0.67.

X1= 0.6355

X2= -0.6247

3.3.3 Distribución de datos Ripple

En la siguiente imagen observamos que los datos que representa la rentabilidad mensual de la moneda virtual RIPLE muestra que los datos no se ajusten a una normal. Esto implica que los datos proceden a una transformacion que permita trabajar con datos normales.

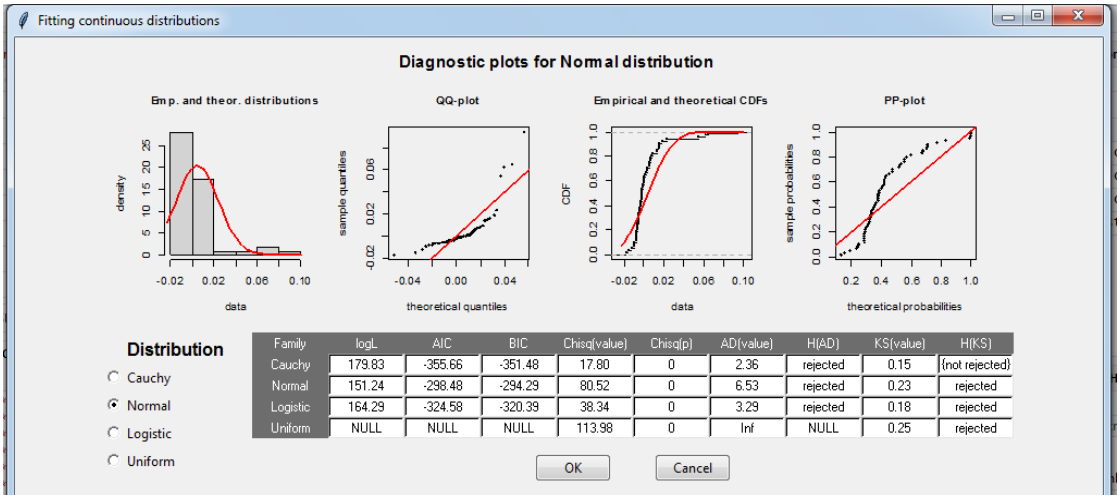


Figura 12 Resultados de las distribuciones a las que se adapta las rentabilidades mensuales.

En vista que los datos de la rentabilidad poseen negativos y cercanos a cero, se realizara la transformación de Johnson, de modo que el programa Minitab nos arroja el siguiente resultado:

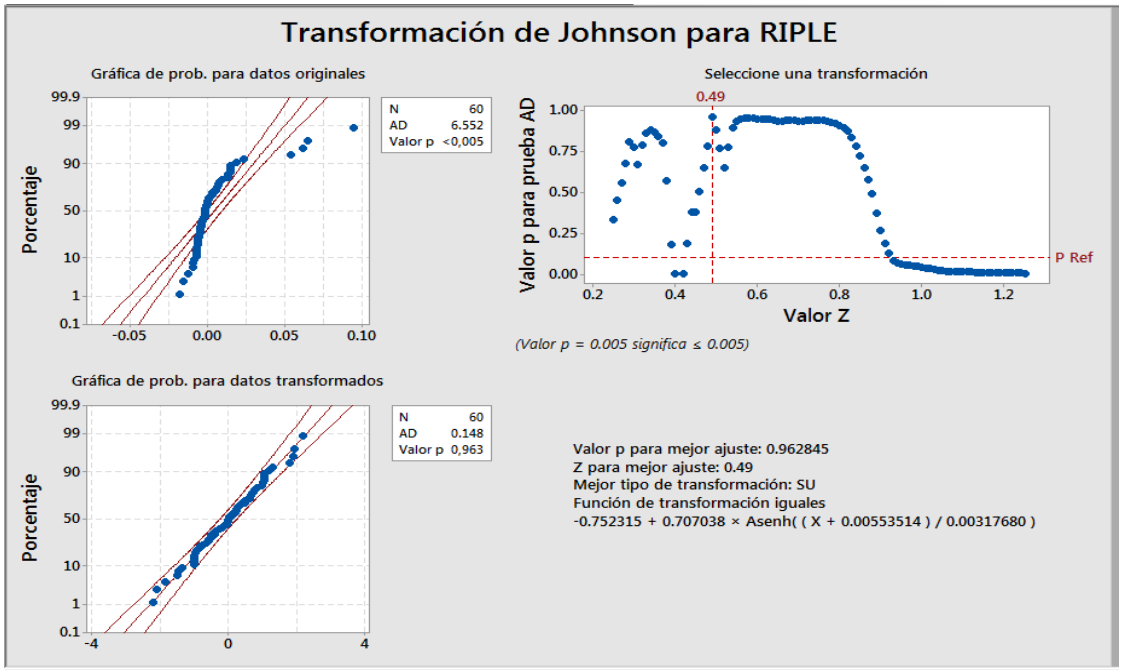


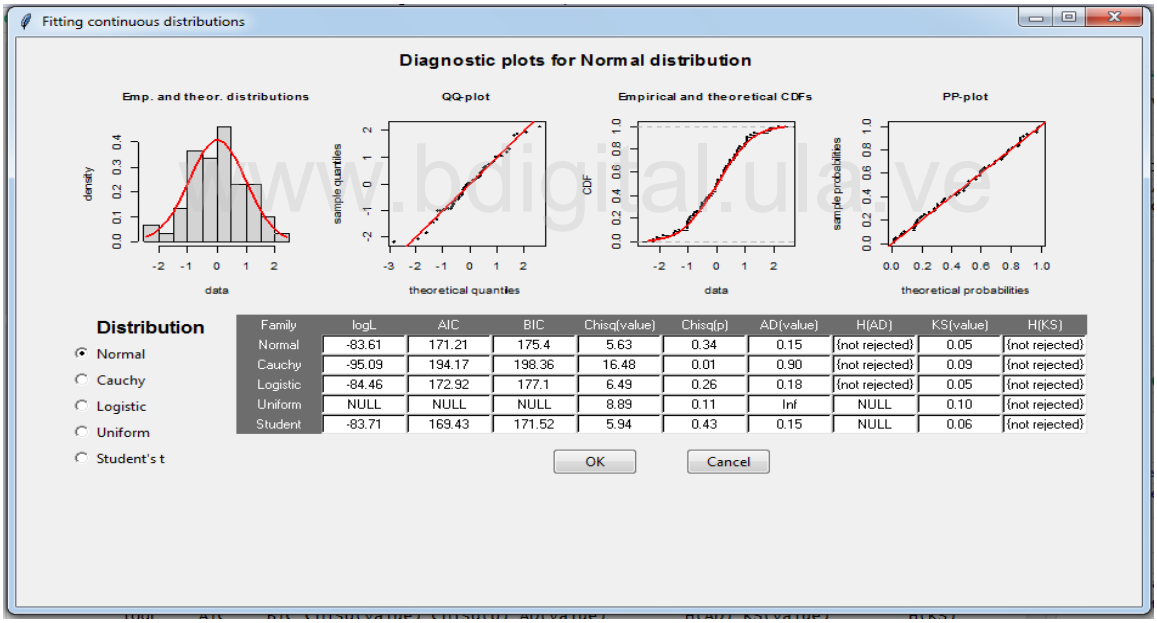
Figura 13 Resultados de la transformación a distribución normal.

Observamos que el valor requerido para transformar los datos a una distribución normal tiene el ajuste de 0.49, en la imagen inferior llamada “Gráfica de prob. Para datos transformados” se refleja que su significancia tiene un valor p = 0.96 que es mayor a 0.05 generando datos normales siendo este último valor el nivel de referencia para determinar si un conjunto pertenece a una distribución normal.

Al volver a Rstudio verificamos que no hay duda de la normalidad de los nuevos datos. Nos da los siguientes parametros:

Chosen continuous distribution is: Normal (norm)
Fitted parameters are:
mean sd

0.007158371 0.974804150



Family	logL	AIC	BIC	Chisq(value)	Chisq(p)	AD(value)	H(AD)	KS(value)	H(KS)
Normal	-83.61	171.21	175.4	5.63	0.34	0.15	{not rejected}	0.05	{not rejected}
Cauchy	-95.09	194.17	198.36	16.48	0.01	0.90	{not rejected}	0.09	{not rejected}
Logistic	-84.46	172.92	177.1	6.49	0.26	0.18	{not rejected}	0.05	{not rejected}
Uniform	NULL	NULL	NULL	8.89	0.11	Inf	NULL	0.10	{not rejected}
Student	-83.71	169.43	171.52	5.94	0.43	0.15	NULL	0.06	{not rejected}

Figura 14 Resultados de las distribuciones a los que se adapta el vector de datos transformados.

Dichos parámetros permite obtener los intervalos de confianza del 50 % que permite realizar la discretizacion de todo el vector de datos, por lo tanto 1- α = 0.5 entonces $\alpha/2$ = 0.25 nos queda $z_{\alpha/2}$ = 0.67.

X1= 0.6658

X2= -0.6515

3.3.4 Distribución de los datos litecoin

En la siguiente imagen observamos que los datos que representa la rentabilidad mensual de la moneda virtual Ether nos muestras que no hay certeza de que los datos se ajusten a una normal. Esto implica que los datos proceden a una transformacion que permita trabajar con datos normales.

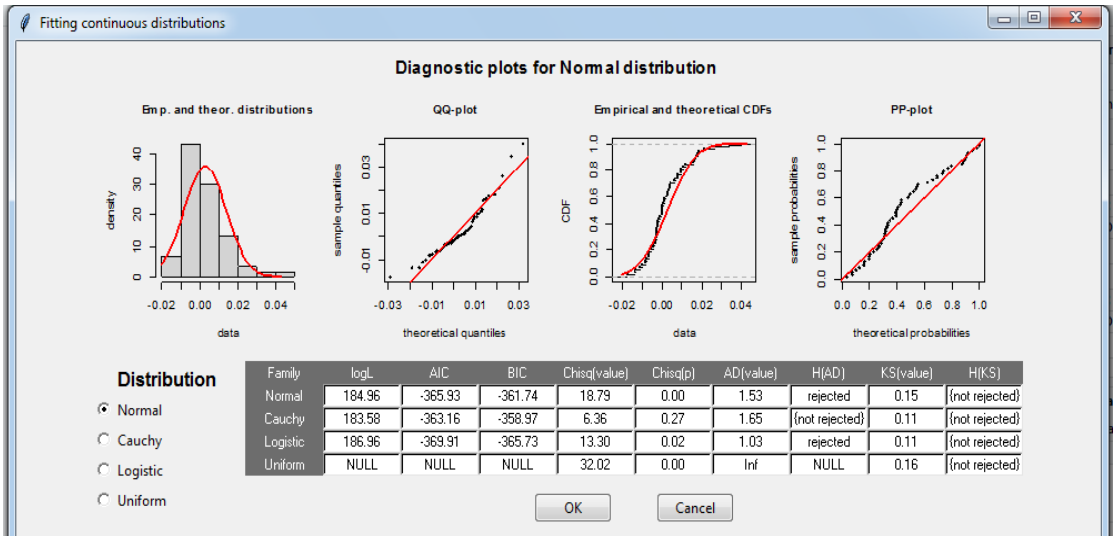


Figura 15 Resultados de las distribuciones a las que se adapta las rentabilidades mensuales.

En vista que los datos de la rentabilidad poseen negativos y cercanos a cero, se realizara la transformación de Johnson, de modo que el programa Minitab nos arroja el siguiente resultado:

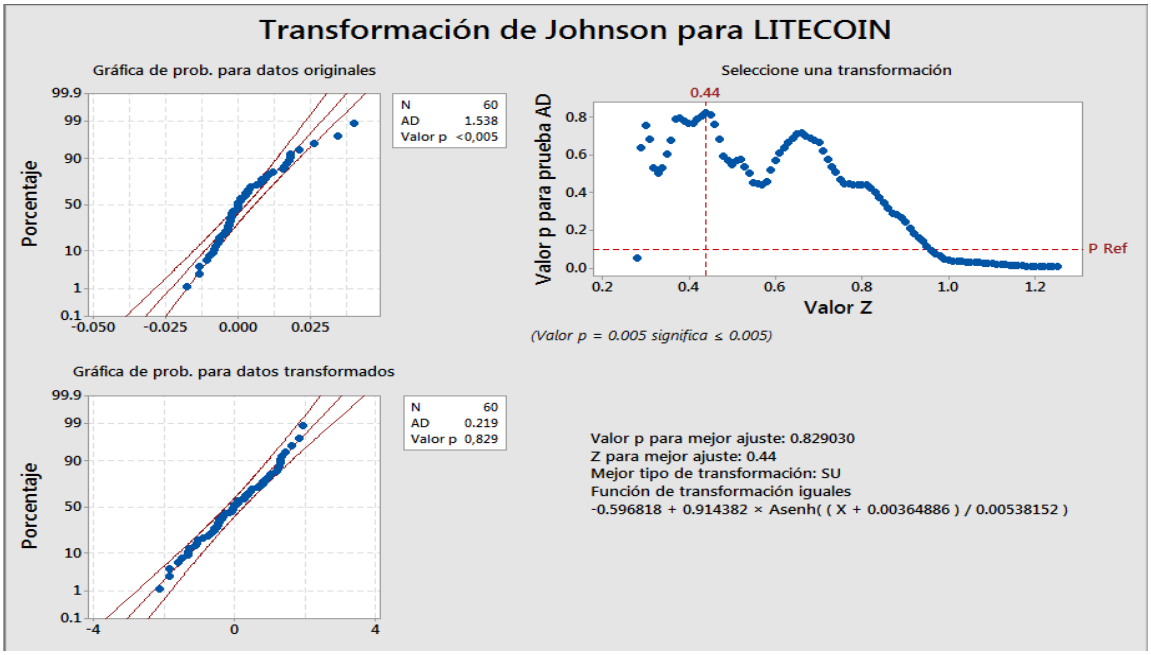
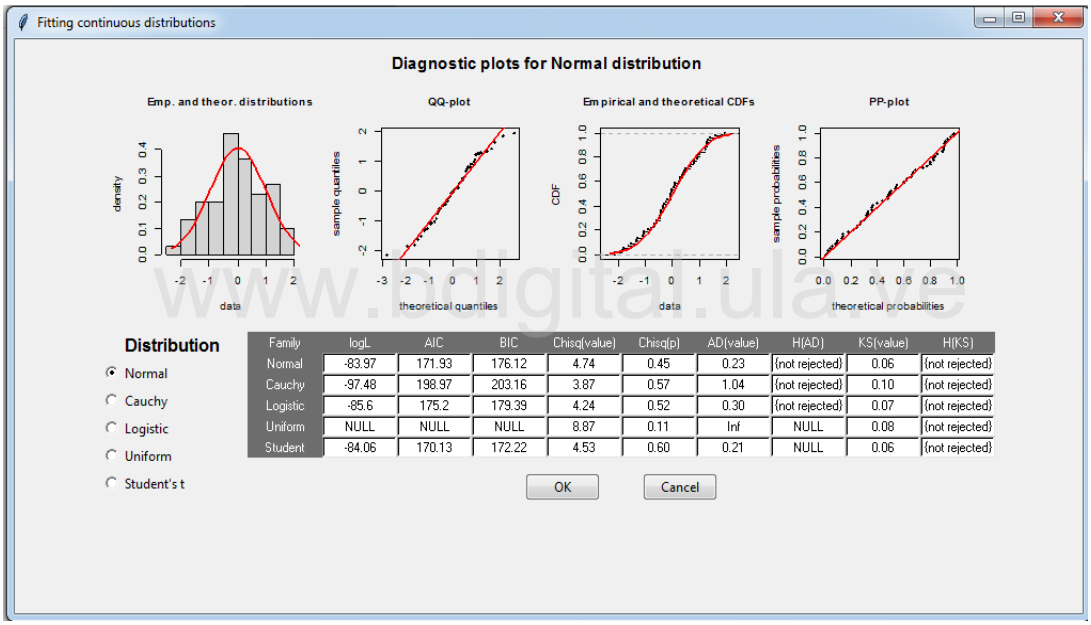


Figura 16 Resultados de la transformación a distribución normal.

Observamos que el valor requerido para transformar los datos a una distribución normal tiene el ajuste de 0.44, en la imagen inferior llamada “Gráfica de prob. Para datos transformados” se refleja que su significancia tiene un valor p = 0.96 que es mayor a 0.05 generando datos normales siendo este último valor el nivel de referencia para determinar si un conjunto pertenece a una distribución normal.

Al volver a Rstudio verificamos que no hay duda de la normalidad de los nuevos datos. Nos da los siguientes parametros:

Chosen continuous distribution is: Normal (norm)
Fitted parameters are:
mean sd
0.007499662 0.980698728



Family	logL	AIC	BIC	Chisq(value)	Chisq(p)	AD(value)	H(AD)	KS(value)	H(KS)
Normal	-83.97	171.93	176.12	4.74	0.45	0.23	{not rejected}	0.06	{not rejected}
Cauchy	-97.48	198.97	203.16	3.87	0.57	1.04	{not rejected}	0.10	{not rejected}
Logistic	-85.6	175.2	179.39	4.24	0.52	0.30	{not rejected}	0.07	{not rejected}
Uniform	NULL	NULL	NULL	8.87	0.11	Inf	NULL	0.08	{not rejected}
Student	-84.06	170.13	172.22	4.53	0.60	0.21	NULL	0.06	{not rejected}

Figura 17 Resultados de las distribuciones a los que se adapta el vector de datos transformados.

Dichos parámetros permite obtener los intervalos de confianza del 50 % que permite realizar la discretizacion de todo el vector de datos, por lo tanto 1- $\boldsymbol{\alpha}$ = 0.5 entonces $\boldsymbol{\alpha}/2$ = 0.25 nos queda $z_{\boldsymbol{\alpha/2}}$ = 0.67.

X1= 0.6701

X2= -0.6551

3.4 Descomposición de la serie de tiempo

La descomposicion da vizualizacion del pronostico y las tendencia de los datos. El analisis estacional refleja patrones anteriores de comportamiento para desarrollar modelos de tendencia utiles en la proyeccion, se puede determinar si la actividad comercial presenta alguna variacion estacional que pueda considerarse para formular planes futuros.

El procedimiento de descomposición analiza los índices estacionales y la variación dentro de cada estación de las series de tiempo.

- Índices estacionales: Los índices estacionales son los efectos estacionales en el tiempo t. Utilice la gráfica para determinar la dirección del efecto estacional.
- Datos con tendencia invertida por estación: Los datos de tendencia invertida son aquellos a los que se les ha eliminado el componente de tendencia. Utilice las gráficas de caja para determinar cuál período estacional tiene la mayor y la menor variación.
- Variación porcentual por estación: La gráfica muestra el porcentaje de variación de cada estación. Utilice la gráfica para cuantificar la variación de cada período estacional.
- Residuos por estación: Los residuos son las diferencias entre los valores observados y los pronosticados. Utilice la gráfica para determinar si existe un efecto estacional en los residuos.

3.4.1 Serie de tiempo bitcoin

Cabe mencionar que la descomposicion es de tipo multiplicativo esto debido a que los datos no poseen media y varianza constante. En la figura que veran a continuacion (MAPE) nos da el error porcentual absoluto medio, (MAD) la desviacion absoluta de la media y (MSD) nos da la desviacion cuadratica media.

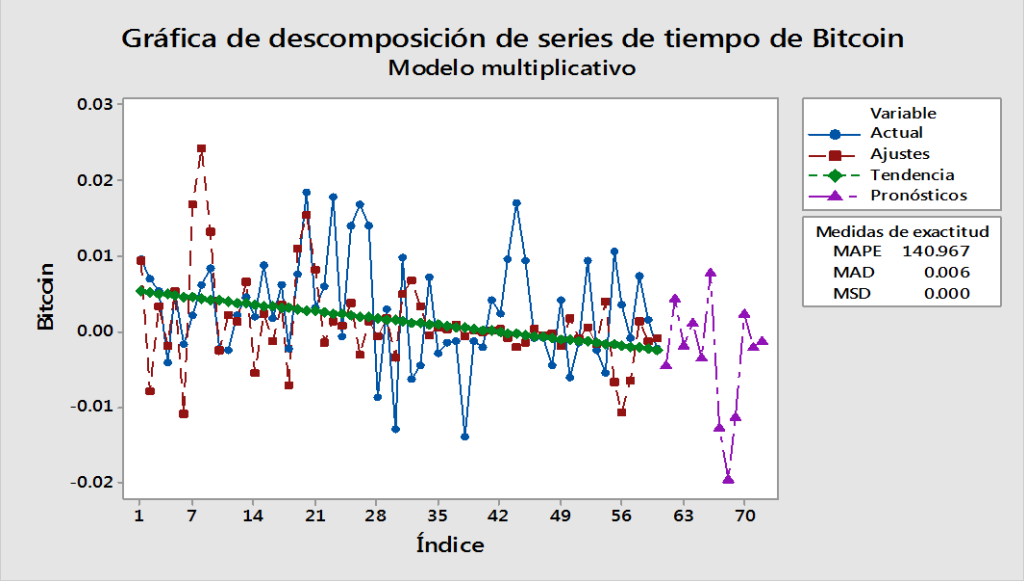


Figura 18 Representación de la tendencia de las rentabilidades de bitcoin.

En la figura 18 el modelo muestra tendencia decreciente la cual viene caracterizada por la siguiente ecuación de tendencia ajustada Yt = 0.00537 - 0.000130×t, del mismo modo predice poca variabilidad para los datos al final de la serie, el valor del MAPE es un porcentaje significativamente alto debido a que gran parte de datos se aproximan o tienden a cero, el MAD y MSD no presentan significancia.

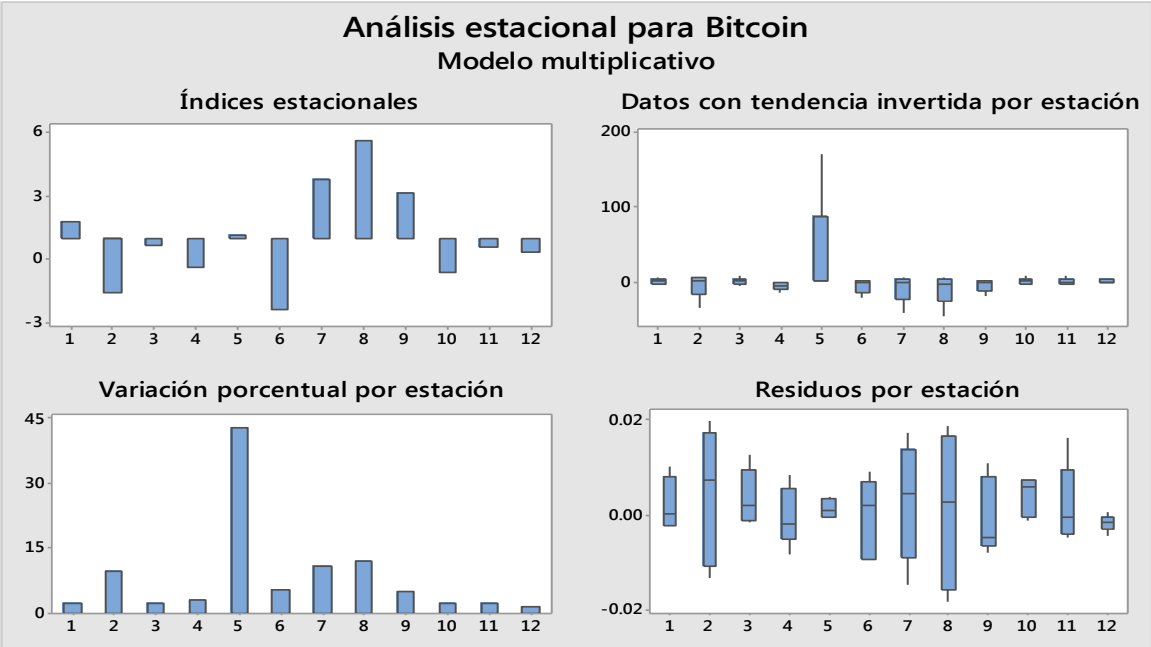


Figura 19 Representación estacional de la serie de tiempo de bitcoin.

En la figura nos proporciona 4 graficas, de modo que en la figura de la parte superior izquierda nos indica que en los meses 2, 4, 6 y 10 son los movimientos descendentes en promedio más significativos de igual modo los meses 3, 5, 11 y 12 tienen un valor descendente más pequeño así como movimientos ascendente en el mes 7 al 9, en la figura de la parte inferior izquierda nos informa que el mes 12 tiene la variación más baja y el mes 5 la más alta, en la parte superior derecha la gráfica de caja de los datos sin tendencia muestra al mes 5 tiene el valor absoluto del efecto estacional más alto esto implica que tienden a tener la más alta variación respecto de los otros meses por último en la gráfica de la parte inferior derecha refleja que no hay efecto moderado entre los residuos.

3.4.2 Serie de tiempo etherium

La descomposicion es de tipo multiplicativo esto debido a que los datos no poseen media y varianza constante. En la figura que veran a continuacion (MAPE) nos da el error porcentual absoluto medio, (MAD) la desviacion absoluta de la media y (MSD) nos da la desviacion cuadratica media.

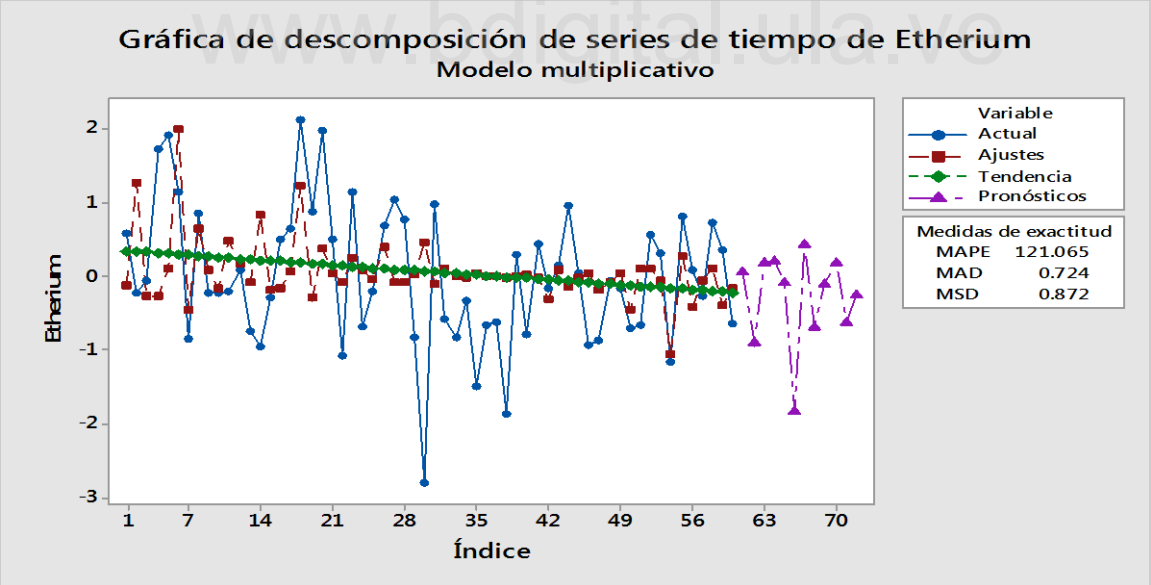


Figura 20 Representación de la tendencia de las rentabilidades.

En esta figura el modelo muestra tendencia decreciente un poco moderada la cual viene caracterizada por la siguiente ecuación de tendencia ajustada Yt = 0.357 - 0.00955×t , la gráfica refleja un pronóstico valores bajos y altos para los datos al final de la serie, el valor del MAPE es un porcentaje significativamente alto debido a que hay gran parte de datos que se aproximan o

tienden a cero, el MSD de 0.872 nos dice que los valores atípicos tienen efecto, el MAD de 0.724 nos dice que el error tiene significancia media.

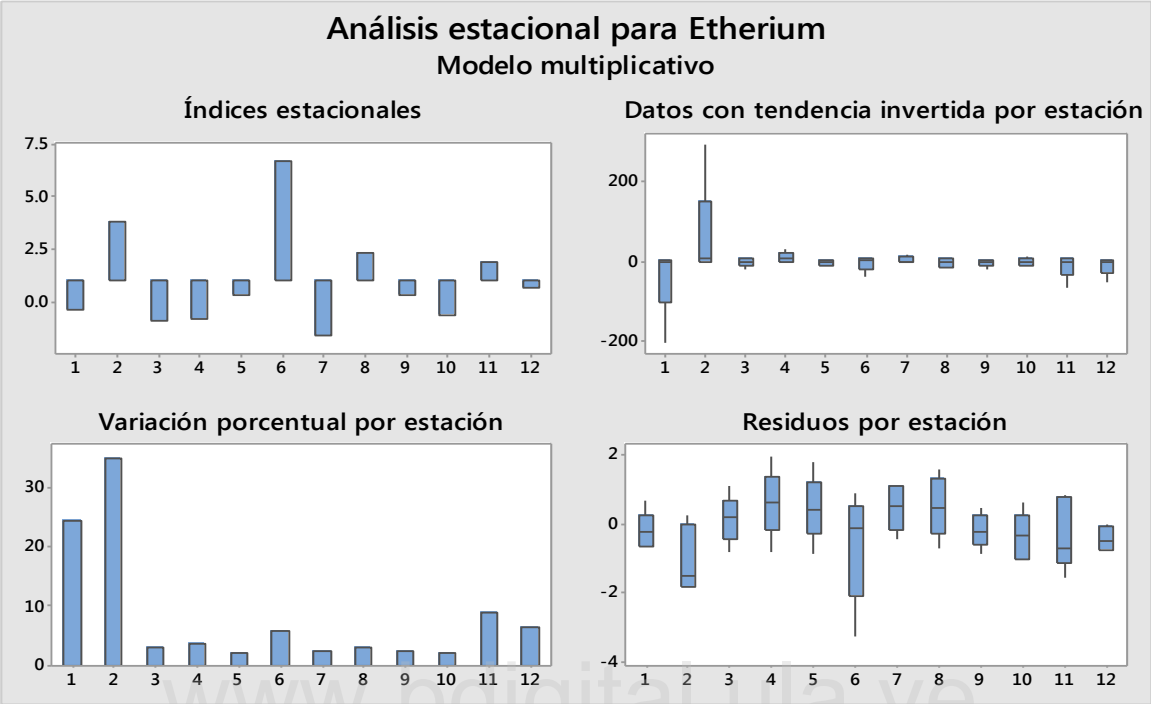


Figura 21 Representación estacional de la serie de tiempo.

En la figura 21 nos proporciona 4 graficas, de modo que en la figura de la parte superior izquierda nos indica que en el mes 7 tiene el movimiento descendente en promedio más significativos de igual modo los meses 1, 3,4,5,9,10 y 12 tiene un valor descendente más pequeño así como movimientos ascendente en el mes 6 y 2, en la figura de la parte inferior izquierda nos informa que los meses 1 y 2 tiene la variación más alta y los meses 3 al 12 la más baja, en la parte superior derecha la gráfica de caja de los datos sin tendencia muestra al mes 1 y 2 que tienen el valor absoluto del efecto estacional más alto esto implica que tienden a tener la variación alta moderada respecto a los otros meses por último en la gráfica de la parte inferior derecha refleja que hay efecto moderado de estación en los residuos.

3.4.3 Serie de tiempo ripple

La descomposicion es de tipo multiplicativo esto debido a que los datos no poseen media y varianza constante. En la figura que veran a continuacion (MAPE) nos da el error porcentual absoluto medio, (MAD) la desviacion absoluta de la media y (MSD) nos da la desviacion cuadratica media.

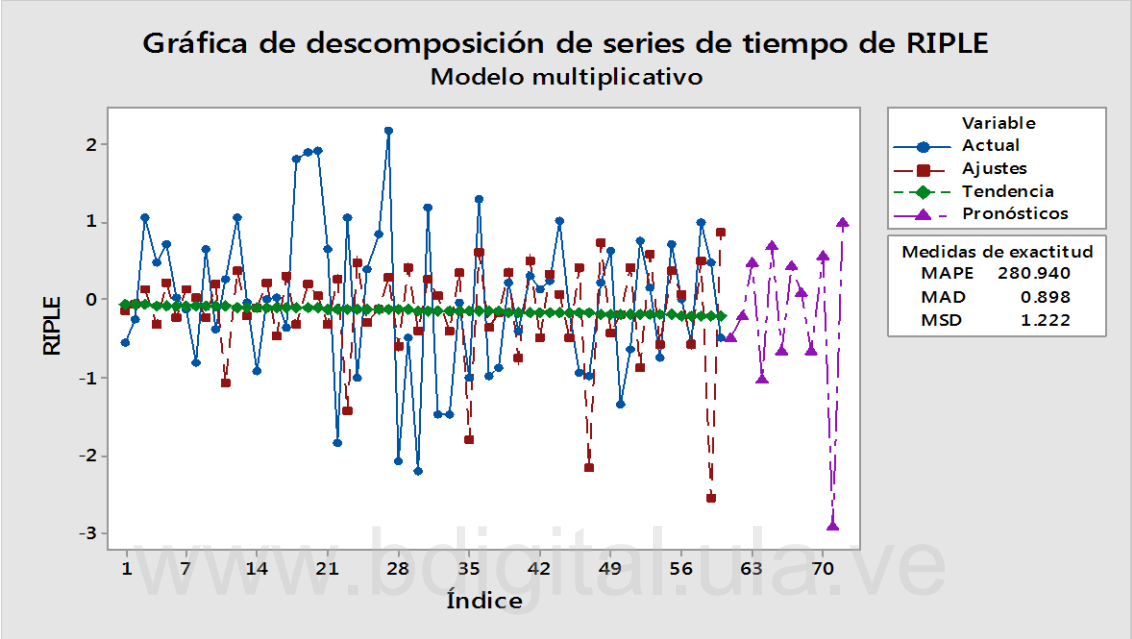


Figura 22 Representación de la tendencia de las rentabilidades

En esta figura el modelo muestra tendencia casi horizontal la cual viene caracterizada por la siguiente ecuación de tendencia ajustada Yt = -0.057 - 0.00245×t la grafica refleja un pronostico poca variabilidad para los datos al final de la serie, el valor del MAPE es un porcentaje significativamente alto debido a que hay gran parte de datos que se aproximan o tienden a cero, el MSD de 0.898 nos dice que los valores atípicos tienen efecto, el MAD 0.898 nos dice que el error es significativo.

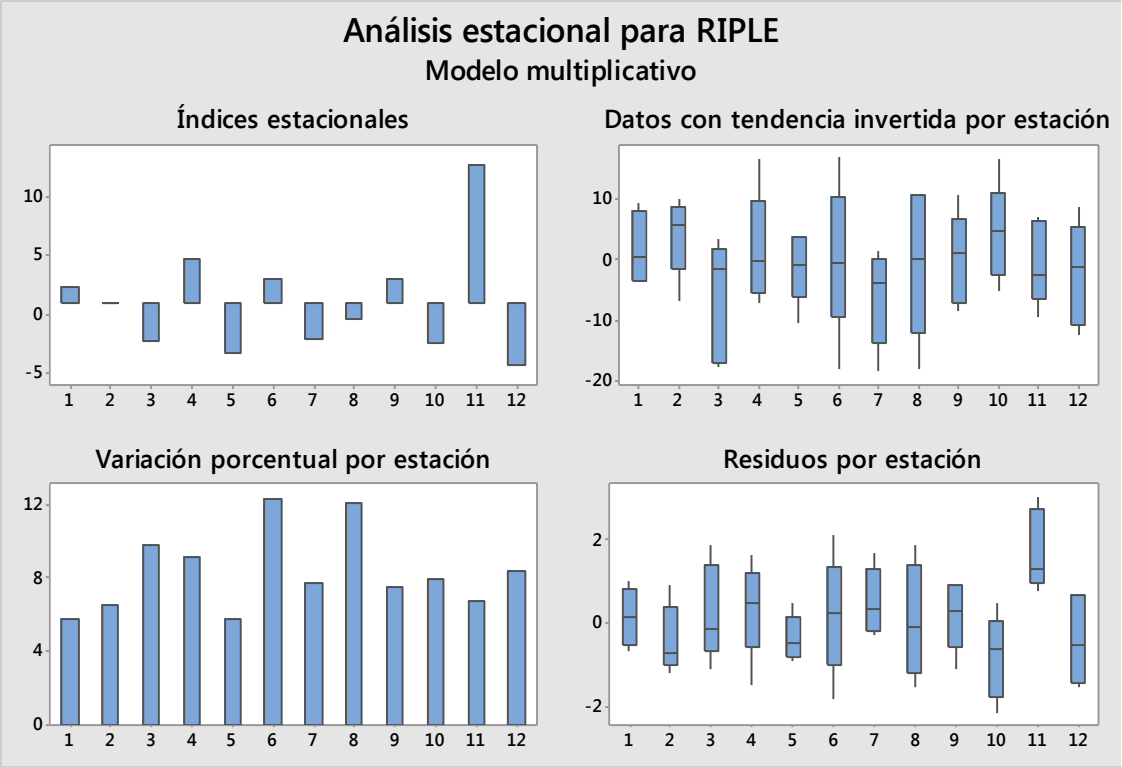


Figura 23 Representacion estacional de la serie de tiempo.

En la figura 23 nos proporciona 4 graficas, de modo que en la figura de la parte superior izquierda nos indica que en el mes 12 tiene el movimiento descendente en promedio más significativos de igual modo los meses 3, 5, 7, 8 y 10 tiene un valores descendente más pequeño así como movimientos ascendente en el mes 4, 6, 9 y 11, en la figura de la parte inferior izquierda nos informa que los meses 6 y 8 tiene la variación más alta y los meses 1, 2 y 5 la más baja, en la parte superior derecha la gráfica de caja de los datos sin tendencia muestra que los 12 meses del año presentan valor absoluto del efecto estacional alto esto implica que tienden a tener variaciones moderadas o casi iguales entre los meses por último en la gráfica de la parte inferior derecha refleja efectos estacionales moderados entre los residuos.

3.4.4 Serie de tiempo litecoin

La descomposicion es de tipo multiplicativo esto debido a que los datos no poseen media y varianza constante. En la figura que veran a continuacion (MAPE) nos da el error porcentual absoluto medio, (MAD) la desviacion absoluta de la media y (MSD) nos da la desviacion cuadratica media.

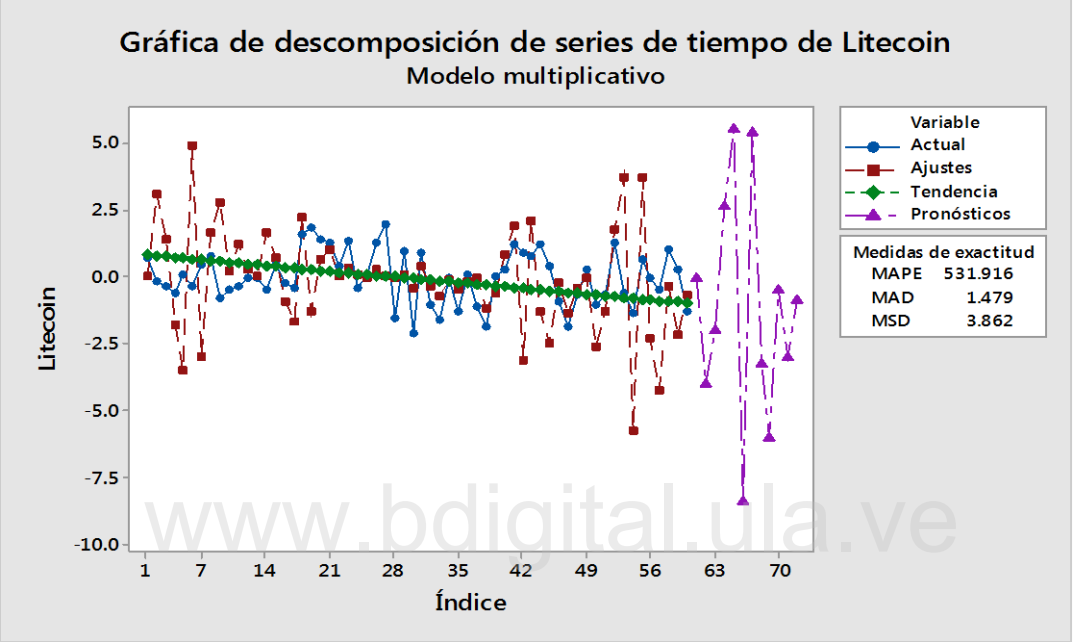


Figura 24 Representación de la tendencia de las rentabilidades.

En la figura 22 el modelo muestra tendencia decreciente minima la cual viene caracterizada por la siguiente ecuación de tendencia ajustada Yt = 0.856 - 0.0304×t la graficas refleja un pronóstico valores altos y bajos para los datos al final de la serie, el valor del MAPE es un porcentaje significativamente alto debido a que hay gran parte de datos que se aproximan o tienden a cero, el MSD de 3.862 nos dice que los valores atípicos tienen efecto, el MAD de 1.479 nos dice que el error es significativo.

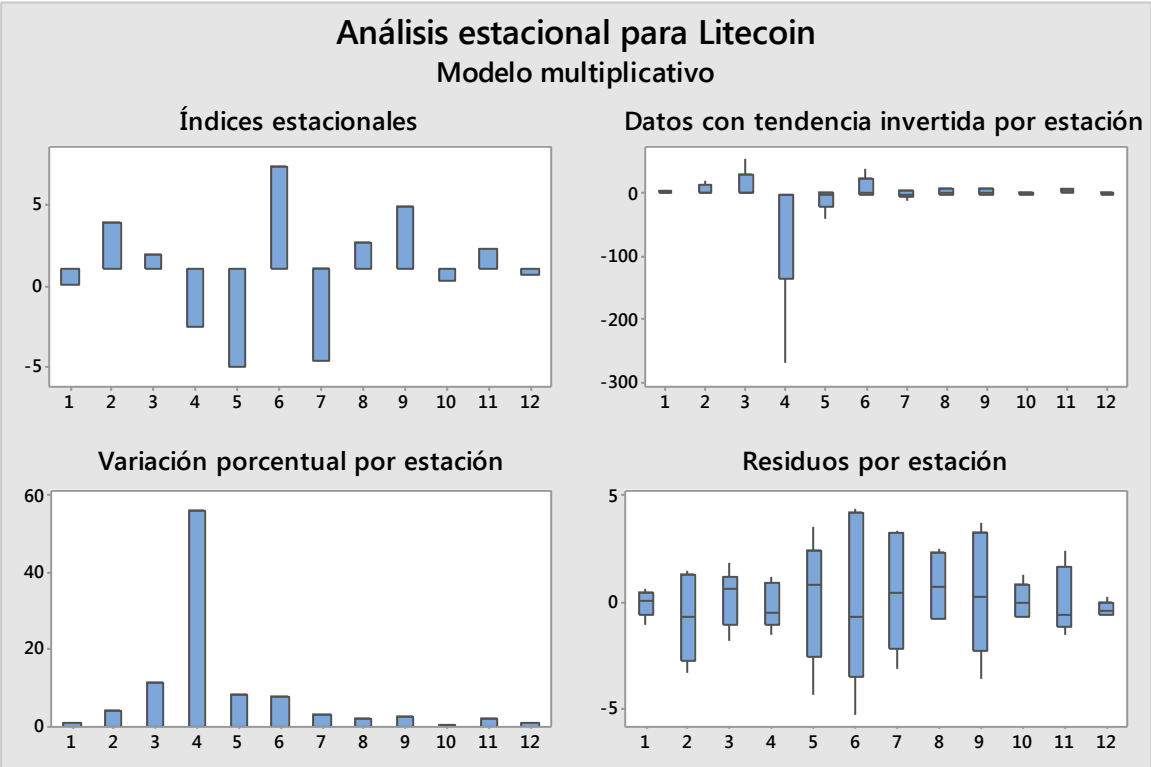


Figura 25 Representacion estacional de la serie de tiempo.

En la figura 23 nos proporciona 4 graficas, de modo que en la figura de la parte superior izquierda nos indica que en los meses 4, 5 y 7 tienen el movimiento descendente en promedio más significativos así como movimientos ascendente en promedio en el mes 2, 6 y 9, en la figura de la parte inferior izquierda nos informa que el mes 4 tiene la variación porcentual más alta y los otros 11 meses tienen la variación porcentual más baja, en la parte superior derecha la gráfica de caja de los datos sin tendencia muestra al mes 4 que tienen el valor absoluto del efecto estacional más alto esto implica que tienden a tener más variación respecto de los otros meses por último en la gráfica de la parte inferior derecha refleja que hay efecto moderado de estación en los residuos.

3.5 Correlación de los datos

A continuación se estudia la correlación o dependencia entre las monedas para observar el comportamiento del conjunto de datos. De modo que los datos están estructurados en forma de rentabilidad por lo cual la correlación indica fuerza y la dirección de una relación lineal y proporcionalidad entre dos variables estadísticas.

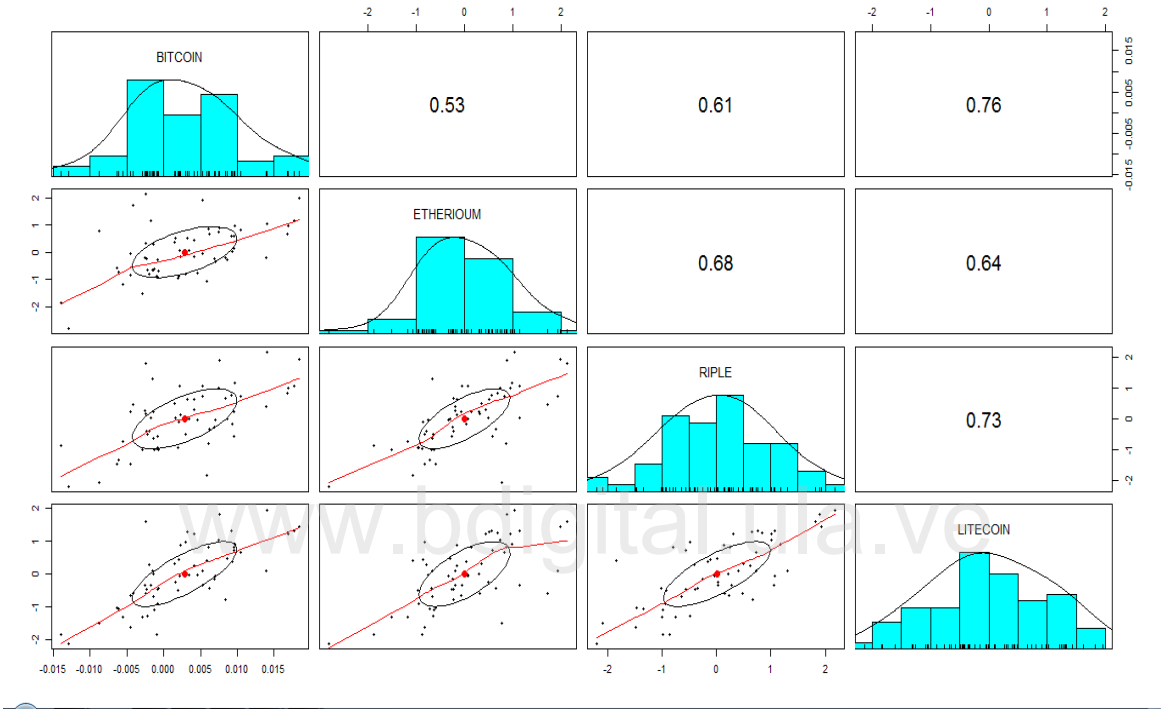


Figura 26 Representación gráfica de la correlación entre las monedas virtuales.

En la figura 26 presentan dos formas de respuestas en la parte superior numérica y en la inferior la representación de dichos valores en forma de gráficas, se observa en la diagonal superior los valores correspondientes a dicha relación por lo que se concluye que en casi todas las combinaciones se determina una dependencia fuerte positiva entre monedas, siendo la relación bitcoin – Litecoin la más fuerte con 0.76 y la relación bitcoin – Etherioum con el valor más pequeño se determina una dependencia moderada. Otra información que nos suministra el grafico es en la diagonal inferior son gráficos en forma de elipses que en algunos son un poco más anchos y en otras un poco más angostas de modo esto debido a la variación numérica de dichas correlaciones. Con esto podemos afirmar que tomando como un ejemplo la relación bitcoin –

Litecoin como ya habiamos mencionada que son positivas y fuerte a medida que la rentabilidad de BTC aumenta la rentabilidad de LTC, pero con una variación de 0.76. Del mismo modo para el resto de las relaciones conforme a su variación numérica.

La información que nos da esta preparación de datos nos brinda premisas para plantear suposiciones antes de ir al capítulo de las simulaciones, es decir, suponer que la preparación está realizada de forma coherente que permita a la hora de aplicar la técnica de reglas de asociación se puedan obtener reglas fuertes y del mismo modo al simular los datos con la técnica de análisis discriminante obtener grupo correctamente clasificados de manera que se puedan predecir futuros comportamientos a través de la toma de decisión.

3.6 Discretización de los datos

Los datos inicialmente son de naturaleza continua y se ajustan a una distribución normal para efecto de lograr los objetivos planteados se requiere realizar una transformación o etiquetado, por lo cual se hará por medio de obtener intervalos de confianza de un 50% a vector de datos de las 4 monedas de estudio, las etiquetas a utilizar son las siguientes:

- SOBRE-COMPRA: Se utiliza para denominar a un activo financiero cuando se incrementa rápidamente su precio. Los valores que constituye este intervalo son los escenarios en los que usuarios obtienen rentabilidad altas significativas.
- COMPRA-VENTA: Constituye el intervalo en el cual el mercado se encuentra en equilibrio, los usuarios comprar y venden sin obtener rentabilidad bajas o altas significativamente.
- SOBRE-VENTA: Se utiliza para denominar a un activo financiero cuando disminuye rápidamente su precio. Los valores que constituye este intervalo son los escenarios en los que usuarios obtienen rentabilidad bajas significativas.

	A	B	C	D	E
1	TRANSACCION	BITCOIN	ETHERIOUM	RIPLE	LITECOIN
2	oct-15	SOBRE-VENTA	COMPRA-VENTA	COMPRA-VENTA	SOBRE-COMPRA
3	nov-15	COMPRA-VENTA	COMPRA-VENTA	COMPRA-VENTA	COMPRA-VENTA
4	dic-15	COMPRA-VENTA	COMPRA-VENTA	SOBRE-COMPRA	COMPRA-VENTA
5	ene-16	SOBRE-VENTA	SOBRE-COMPRA	COMPRA-VENTA	COMPRA-VENTA
6	feb-16	COMPRA-VENTA	SOBRE-COMPRA	SOBRE-COMPRA	COMPRA-VENTA
7	mar-16	COMPRA-VENTA	SOBRE-COMPRA	COMPRA-VENTA	COMPRA-VENTA
8	abr-16	COMPRA-VENTA	SOBRE-VENTA	COMPRA-VENTA	COMPRA-VENTA
9	may-16	COMPRA-VENTA	SOBRE-COMPRA	SOBRE-VENTA	SOBRE-COMPRA
10	jun-16	SOBRE-COMPRA	COMPRA-VENTA	SOBRE-COMPRA	SOBRE-VENTA
11	jul-16	SOBRE-VENTA	COMPRA-VENTA	COMPRA-VENTA	COMPRA-VENTA
12	ago-16	SOBRE-VENTA	COMPRA-VENTA	COMPRA-VENTA	COMPRA-VENTA
13	sep-16	COMPRA-VENTA	COMPRA-VENTA	SOBRE-COMPRA	COMPRA-VENTA
14	oct-16	COMPRA-VENTA	SOBRE-VENTA	COMPRA-VENTA	COMPRA-VENTA
15	nov-16	COMPRA-VENTA	SOBRE-VENTA	SOBRE-VENTA	COMPRA-VENTA
16	dic-16	SOBRE-COMPRA	COMPRA-VENTA	COMPRA-VENTA	COMPRA-VENTA
17	ene-17	COMPRA-VENTA	COMPRA-VENTA	COMPRA-VENTA	COMPRA-VENTA
18	feb-17	COMPRA-VENTA	SOBRE-COMPRA	COMPRA-VENTA	COMPRA-VENTA
19	mar-17	SOBRE-VENTA	SOBRE-COMPRA	SOBRE-COMPRA	SOBRE-COMPRA

Hoja1

Figura 27 Representación de la base de datos discretizada

En la figura se refleja la base de datos discretizada que será simulada en la presente investigación exploratoria. Para efecto de estudio en la parte de la simulación en las reglas de asociación es bueno comprender que es un itemsets en nuestra base de datos un ejemplo (oct- 15/ SOBRE-VENTA, COMPRA-VENTA, COMPRA-VENTA, SOBRE-COMPRA).

CAPITULO 4

SIMULACIÓN

4.1 Reglas de asociación

El algoritmo Apriori hace una búsqueda exhaustiva por niveles de complejidad (de menor a mayor tamaño de itemsets). Para reducir el espacio de búsqueda aplica la norma de "si un itemset no es frecuente, ninguno de sus supersets (itemsets de mayor tamaño que contengan al primero) puede ser frecuente". Visto de otra forma, si un conjunto es infrecuente, entonces, todos los conjuntos donde este último se encuentre, también son infrecuentes.

Para encontrar las reglas de asociación entre las rentabilidades de las monedas de estudio se requiere de las librerías (Matrix) y (arules). La librería arules funciona internamente con el algoritmo apriori.

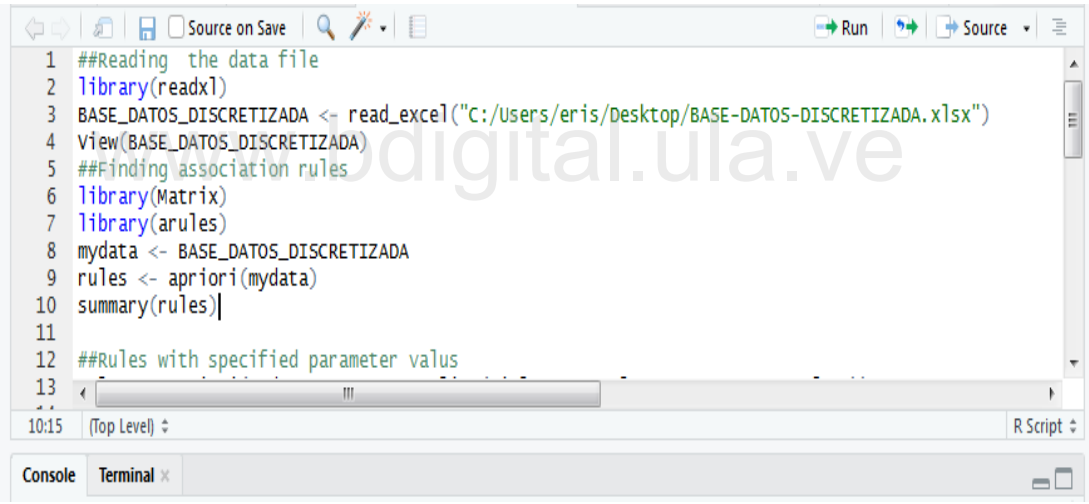
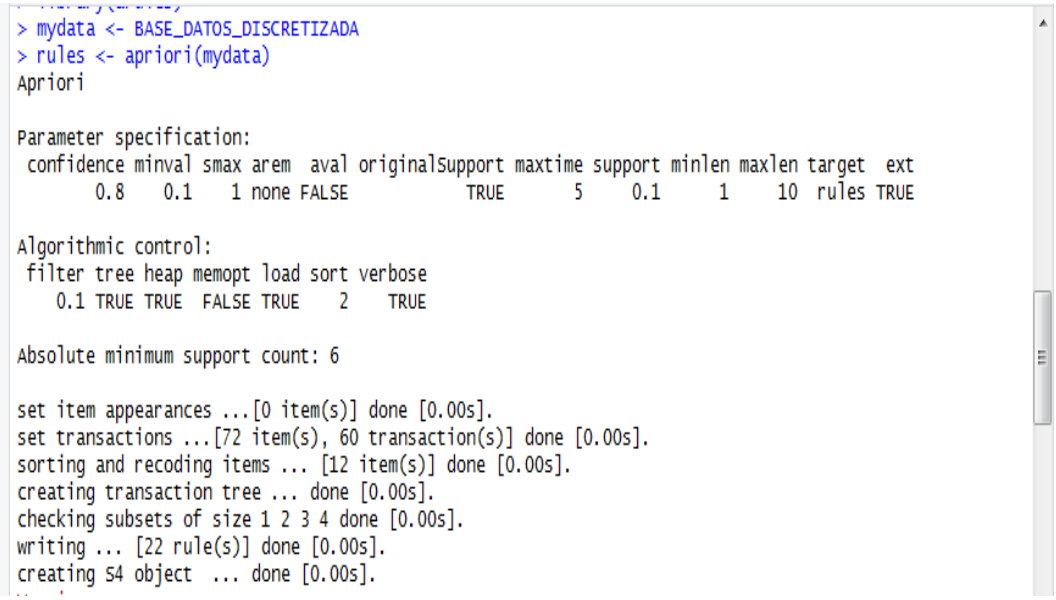


Figura 28 Representación de los comando a usar para la simulación, invocando la función arules.

En la siguiente imagen se refleja la lectura de los datos y el llamado de las librerías Matrix y arules, a la palabra rules le asignamos la invocación del algoritmo a priori sobre la base de datos.

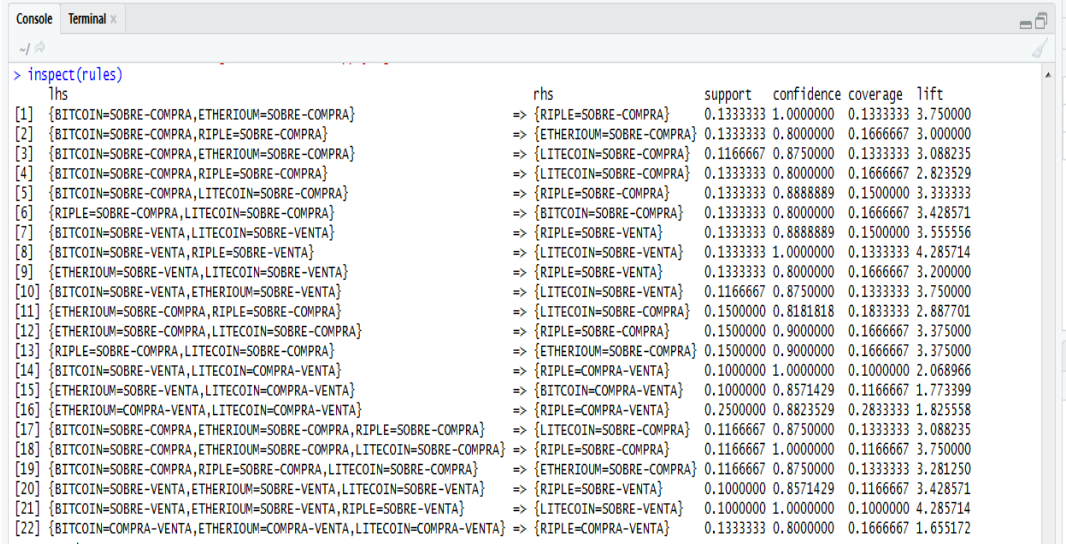
La anterior imagen nos genera las siguientes respuestas:



```
> mydata <- BASE_DATOS_DISCRETIZADA
> rules <- apriori(mydata)
Apriori

Parameter specification:
 confidence minval smax arem  aval originalSupport maxtime support minlen maxlen target  ext
        0.8    0.1    1 none FALSE            TRUE       5     0.1      1     10  rules TRUE

Algorithmic control:
 filter tree heap memopt load sort verbose
    0.1 TRUE TRUE  FALSE TRUE    2    TRUE

Absolute minimum support count: 6

set item appearances ...[0 item(s)] done [0.00s].
set transactions ...[72 item(s), 60 transaction(s)] done [0.00s].
sorting and recoding items ... [12 item(s)] done [0.00s].
creating transaction tree ... done [0.00s].
checking subsets of size 1 2 3 4 done [0.00s].
writing ... [22 rule(s)] done [0.00s].
creating S4 object  ... done [0.00s].
```

Figura 29 Resultados de la simulación aplicando reglas de asociación

En la figura se refleja la simulación de la base de datos con los valores de soporte y confianza por defecto de la librería, es decir, los parámetros no fueron variados por el usuario, con un soporte de 0.8 y una confianza de 0.1, la simulación nos genera 22 reglas.



```
Console  Terminal

> inspect(rules)
     lhs                                                                rhs                      support   confidence coverage  lift
[1]  {BITCOIN=SOBRE-COMPRA,ETHERIOUM=SOBRE-COMPRA}                   => {RIPLE=SOBRE-COMPRA}     0.1333333 1.0000000  0.1333333 3.750000
[2]  {BITCOIN=SOBRE-COMPRA,RIPLE=SOBRE-COMPRA}                       => {ETHERIOUM=SOBRE-COMPRA} 0.1333333 0.8000000  0.1666667 3.000000
[3]  {BITCOIN=SOBRE-COMPRA,ETHERIOUM=SOBRE-COMPRA}                   => {LITECOIN=SOBRE-COMPRA}  0.1166667 0.8750000  0.1333333 3.088235
[4]  {BITCOIN=SOBRE-COMPRA,RIPLE=SOBRE-COMPRA}                       => {LITECOIN=SOBRE-COMPRA}  0.1333333 0.8000000  0.1666667 2.823529
[5]  {BITCOIN=SOBRE-COMPRA,LITECOIN=SOBRE-COMPRA}                    => {RIPLE=SOBRE-COMPRA}     0.1333333 0.8888889  0.1500000 3.333333
[6]  {RIPLE=SOBRE-COMPRA,LITECOIN=SOBRE-COMPRA}                      => {BITCOIN=SOBRE-COMPRA}   0.1333333 0.8000000  0.1666667 3.428571
[7]  {BITCOIN=SOBRE-VENTA,LITECOIN=SOBRE-VENTA}                      => {RIPLE=SOBRE-VENTA}      0.1333333 0.8888889  0.1500000 3.555556
[8]  {BITCOIN=SOBRE-VENTA,RIPLE=SOBRE-VENTA}                         => {LITECOIN=SOBRE-VENTA}   0.1333333 1.0000000  0.1333333 4.285714
[9]  {ETHERIOUM=SOBRE-VENTA,LITECOIN=SOBRE-VENTA}                    => {RIPLE=SOBRE-VENTA}      0.1333333 0.8000000  0.1666667 3.200000
[10] {BITCOIN=SOBRE-VENTA,ETHERIOUM=SOBRE-VENTA}                     => {LITECOIN=SOBRE-VENTA}   0.1166667 0.8750000  0.1333333 3.750000
[11] {ETHERIOUM=SOBRE-COMPRA,RIPLE=SOBRE-COMPRA}                     => {LITECOIN=SOBRE-COMPRA}  0.1500000 0.8181818  0.1833333 2.887701
[12] {ETHERIOUM=SOBRE-COMPRA,LITECOIN=SOBRE-COMPRA}                  => {RIPLE=SOBRE-COMPRA}     0.1500000 0.9000000  0.1666667 3.375000
[13] {RIPLE=SOBRE-COMPRA,LITECOIN=SOBRE-COMPRA}                      => {ETHERIOUM=SOBRE-COMPRA} 0.1500000 0.9000000  0.1666667 3.375000
[14] {BITCOIN=SOBRE-VENTA,LITECOIN=COMPRA-VENTA}                     => {RIPLE=COMPRA-VENTA}     0.1000000 1.0000000  0.1000000 2.068966
[15] {ETHERIOUM=SOBRE-VENTA,LITECOIN=COMPRA-VENTA}                   => {BITCOIN=COMPRA-VENTA}   0.1000000 0.8571429  0.1166667 1.773399
[16] {ETHERIOUM=COMPRA-VENTA,LITECOIN=COMPRA-VENTA}                  => {RIPLE=COMPRA-VENTA}     0.2500000 0.8823529  0.2833333 1.825558
[17] {BITCOIN=SOBRE-COMPRA,ETHERIOUM=SOBRE-COMPRA,RIPLE=SOBRE-COMPRA}    => {LITECOIN=SOBRE-COMPRA}  0.1166667 0.8750000  0.1333333 3.088235
[18] {BITCOIN=SOBRE-COMPRA,ETHERIOUM=SOBRE-COMPRA,LITECOIN=SOBRE-COMPRA} => {RIPLE=SOBRE-COMPRA}     0.1166667 1.0000000  0.1166667 3.750000
[19] {BITCOIN=SOBRE-COMPRA,RIPLE=SOBRE-COMPRA,LITECOIN=SOBRE-COMPRA}     => {ETHERIOUM=SOBRE-COMPRA} 0.1166667 0.8750000  0.1333333 3.281250
[20] {BITCOIN=SOBRE-VENTA,ETHERIOUM=SOBRE-VENTA,LITECOIN=SOBRE-VENTA}    => {RIPLE=SOBRE-VENTA}      0.1000000 0.8571429  0.1166667 3.428571
[21] {BITCOIN=SOBRE-VENTA,ETHERIOUM=SOBRE-VENTA,RIPLE=SOBRE-VENTA}       => {LITECOIN=SOBRE-VENTA}   0.1000000 1.0000000  0.1000000 4.285714
[22] {BITCOIN=COMPRA-VENTA,ETHERIOUM=COMPRA-VENTA,LITECOIN=COMPRA-VENTA} => {RIPLE=COMPRA-VENTA}     0.1333333 0.8000000  0.1666667 1.655172
```

Figura 30 Representación de las 22 reglas generadas en la simulación

En la figura 30 representa la simulación de la aplicación del algoritmo apriori, generando 22 reglas de las cuales 16 reglas son de 3 ítems y 6 de 4 ítems, se observa que el lift en cada una de las reglas es mayor a 1 esto implica que aumenta la probabilidad de cada una de estas relaciones se pueden categorizar como una regla fuerte.


```
##Rules with specified parameter valus
rules <- apriori(mydata,parameter = list(minlen=2, maxlen=3,supp=.1, conf=.8))
inspect(sort(subset(rules, subset=lift > 1), by = "lift"))
inspect(rules)
summary(rules)
```

Figura 31 Representación de las líneas de comando para la variación de parámetros como soporte y confianza en la simulación.

En la figura 31 refleja las líneas de comandos para el caso de estudio en el que se realiza una variación de los parámetros como lo son el soporte que se representa como (supp) y la confianza que se representa como (conf), en maxlen se asigna el número de ítems o tamaño de las reglas a generar.

4.2 Análisis discriminante

En la presente etapa se realizara la simulación del análisis discriminante en el software Minitab 17 de cada una de las monedas en la cual la rentabilidad es la variable que se desea discriminar, esto permitirá observar si los grupos de la base de datos se encuentran correctamente clasificados.

	C1-T	C2
	GRUPOS	RENTABILIDAD BITCOIN
1	SOBRE-VENTA	0.0095735
2	COMPRA-VENTA	0.0069271
3	COMPRA-VENTA	0.0052428
4	SOBRE-VENTA	-0.0041415
5	COMPRA-VENTA	0.0052368
6	COMPRA-VENTA	-0.0016943
7	COMPRA-VENTA	0.0020451
8	COMPRA-VENTA	0.0060661
9	SOBRE-COMPRA	0.0083509
10	SOBRE-VENTA	-0.0022754
11	SOBRE-VENTA	-0.0026100
12	COMPRA-VENTA	0.0021383
13	COMPRA-VENTA	0.0045938
14	COMPRA-VENTA	0.0019784
15	SOBRE-COMPRA	0.0086583
16	COMPRA-VENTA	0.0016470
17	COMPRA-VENTA	0.0062217
18	SOBRE-VENTA	-0.0024014
19	SOBRE-COMPRA	0.0076217
20	SOBRE-COMPRA	0.0184043
21	COMPRA-VENTA	0.0031911
22	COMPRA-VENTA	0.0059022
23	SOBRE-COMPRA	0.0177918
24	COMPRA-VENTA	-0.0006962
25	SOBRE-COMPRA	0.0139831

Figura 32 Representación base de datos correspondiente a bitcoin previo simular en minitab aplicando análisis discriminante.

4.2.1 Resultados bitcoin

Método lineal para respuesta: GRUPOS

Predictores: RENTABILIDAD BITCOIN

Grupo	COMPRA-VENTA	SOBRE-COMPRA	SOBRE-VENTA
Conteo	29	14	17

Resumen de clasificación

Colocar en un grupo	Grupo verdadero COMPRA-VENTA	SOBRE-COMPRA	SOBRE-VENTA
COMPRA-VENTA	22	0	0
SOBRE-COMPRA	1	14	1
SOBRE-VENTA	6	0	16
N Total	29	14	17
N correcta	22	14	16
Proporción	0.759	1.000	0.941

N = 60 N Correcta = 52 Proporción Correcta = 0.867

Distancia cuadrada entre grupos

	COMPRA-VENTA	SOBRE-COMPRA	SOBRE-VENTA
COMPRA-VENTA	0.0000	6.1190	3.1028
SOBRE-COMPRA	6.1190	0.0000	17.9363
SOBRE-VENTA	3.1028	17.9363	0.0000

Función discriminativa lineal para grupos

	COMPRA-VENTA	SOBRE-COMPRA	SOBRE-VENTA
Constante	-0.21	-4.87	-0.62
RENTABILIDAD BITCOIN	165.30	797.90	-285.17

4.2.2 Resultado ETHERIOOM

Predictores: RENTABILIDAD ETHERIOUM

Grupo	COMPRA-VENTA	SOBRE-COMPRA	SOBRE-VENTA
Conteo	26	16	18

Resumen de clasificación

Colocar en un grupo	Grupo verdadero COMPRA-VENTA	SOBRE-COMPRA	SOBRE-VENTA
COMPRA-VENTA	23	0	0
SOBRE-COMPRA	1	16	0
SOBRE-VENTA	2	0	18
N Total	26	16	18
N correcta	23	16	18
Proporción	0.885	1.000	1.000

N = 60 N Correcta = 57 Proporción Correcta = 0.950

Distancia cuadrada entre grupos

	COMPRA-VENTA	SOBRE-COMPRA	SOBRE-VENTA
COMPRA-VENTA	0.0000	6.3365	5.3163
SOBRE-COMPRA	6.3365	0.0000	23.2608
SOBRE-VENTA	5.3163	23.2608	0.0000

Función discriminativa lineal para grupos

	COMPRA-VENTA	SOBRE-COMPRA	SOBRE-VENTA
Constante	-0.0005	-3.2501	-2.5841
RENTABILIDAD ETHERIOUM	0.0713	5.6258	-5.0164

4.2.3 Resultado RIPPLE

Predictores: RENTABILIDAD RIPLE

Grupo	COMPRA-VENTA	SOBRE-COMPRA	SOBRE-VENTA
Conteo	29	16	15

Resumen de clasificación

Colocar en un grupo	Grupo verdadero COMPRA-VENTA	SOBRE-COMPRA	SOBRE-VENTA
COMPRA-VENTA	26	0	0
SOBRE-COMPRA	2	16	0
SOBRE-VENTA	1	0	15
N Total	29	16	15
N correcta	26	16	15
Proporción	0.897	1.000	1.000

N = 60 N Correcta = 57 Proporción Correcta = 0.950

Distancia cuadrada entre grupos

	COMPRA-VENTA	SOBRE-COMPRA	SOBRE-VENTA
COMPRA-VENTA	0.0000	7.9112	8.2999
SOBRE-COMPRA	7.9112	0.0000	32.4177
SOBRE-VENTA	8.2999	32.4177	0.0000

Función discriminativa lineal para grupos

	COMPRA-VENTA	SOBRE-COMPRA	SOBRE-VENTA
Constante	-0.0001	-3.9187	-4.1880
RENTABILIDAD RIPLE	-0.0306	6.5071	-6.7270

4.2.4 Resultados LITECOIN

Predictores: RENTABILIDAD LITECOIN

Grupo	COMPRA-VENTA	SOBRE-COMPRA	SOBRE-VENTA
Conteo	29	17	14

Resumen de clasificación

Colocar en un grupo	Grupo verdadero COMPRA-VENTA	SOBRE-COMPRA	SOBRE-VENTA
COMPRA-VENTA	28	0	1
SOBRE-COMPRA	1	17	0
SOBRE-VENTA	0	0	13
N Total	29	17	14
N correcta	28	17	13
Proporción	0.966	1.000	0.929

N = 60 N Correcta = 58 Proporción Correcta = 0.967

Distancia cuadrada entre grupos

	COMPRA-VENTA	SOBRE-COMPRA	SOBRE-VENTA
COMPRA-VENTA	0.0000	11.3726	10.5952
SOBRE-COMPRA	11.3726	0.0000	43.9218
SOBRE-VENTA	10.5952	43.9218	0.0000

Función discriminativa lineal para grupos

	COMPRA-VENTA	SOBRE-COMPRA	SOBRE-VENTA
Constante	-0.0155	-5.1072	-5.8871
RENTABILIDAD LITECOIN	-0.4621	8.3761	-8.9929

Capítulo 5

Resultados y recomendaciones

En esta sección se dará explicación a los resultados generados en el capítulo anterior así como también recomendaciones a futuros estudios de esta misma índole.

5.1 Reglas de asociación

Las reglas fuertes del conjunto de las 22 reglas tenemos:

1) **SOBRE-COMPRA:**

[2] {BITCOIN=SOBRE-COMPRA, ETHERIOUM=SOBRE-COMPRA}=>{RIPLE=SOBRE-COMPRA}

Soporte Confianza Lift

0.1333333 1.0000000 3.750000

Esta regla nos dice que hay una confianza de 0.1 que implica un 100% de probabilidad empírica de que la suposición hecha por la regla se cumpla, del mismo modo hay un soporte de 0.13 esto implica que el 13% de las transacciones mostraron que las monedas bitcoin, etherium y ripple presentan una sobre-compra, por último el valor del lift de 3.75 incrementa la probabilidad el consecuente, cuando nos enteramos de que ocurrió el antecedente, implica una reglas fuerte y coherente.

2) **SOBRE-VENTA:**

[8] {BITCOIN= SOBRE-VENTA, RIPLE=SOBRE-VENTA} =>{LITECOIN= SOBRE-VENTA }

Soporte Confianza Lift

0.1333333 1.0000000 4.285714

Esta regla nos dice que hay una confianza de 0.1 que implica un 100% de probabilidad empírica de que la suposición hecha por la regla se cumpla, del mismo modo hay un soporte de 0.13 esto implica que el 13% de las transacciones mostraron que las monedas bitcoin, ripple y litecoin presentan una sobre-compra, por último el valor del lift de 4.28 incrementa la probabilidad el

consecuente, cuando nos enteramos de que ocurrió el antecedente, implica una reglas fuerte y coherente.

En general la aplicación de la técnica de regla de asociación a la base de datos que se preparó, recordemos que la fase crucial de la metodología CRISP-DM está en la preparación de datos, sin embargo los resultados de la simulación nos permitió concluir que la técnica es recomendada para este tipo de predicción mientras se mantenga el mismo procedimiento en la preparación de datos.

5.2 Análisis discriminante

A continuación se presenta una tabla resumen de los resultados de la simulación:

Moneda Virtual	Grupo Actual	Número correcto de clasificación	Proporción Correcta de clasificación	Función discriminante
Bitcoin	COMPRA-VENTA SOBRE- COMPRA SOBRE-VENTA	52	0.87	R_{C-V}= 165.30(R_{BTC}) -0.21 R_{S-C}= 797.90(R_{BTC}) – 4.87 R_{S-V}= -285.17(R_{BTC}) - 0.62
Etherium	COMPRA-VENTA SOBRE- COMPRA SOBRE-VENTA	57	0.95	R_{C-V}= 0.0713(R_{ETH}) R_{S-C}= 5.6258(R_{ETH}) – 3.2501 R_{S-V}= -5.0164(R_{ETH}) – 2.5841
Ripple	COMPRA-VENTA SOBRE- COMPRA SOBRE-VENTA	57	0.95	R_{C-V}= -0.0306(R_{XRP}) R_{S-C}= 6.5071(R_{XRP}) – 3.9187 R_{S-V}= -6.7270(R_{XRP}) – 4.1880
Litecoin	COMPRA-VENTA SOBRE- COMPRA SOBRE-VENTA	58	0.96	R_{C-V}= -0.4621(R_{LTC}) – 0.0155 R_{S-C}= 8.3761(R_{LTC}) – 5.1072 R_{S-V}= -8.9929(R_{LTC}) – 5.8871

Tabla 10 Resumen de la simulación en la aplicación de análisis discriminante.

En la presenta tabla se observa que el porcentaje correcto de clasificación de las rentabilidades en cada uno de los grupos es bastante satisfactorio ya que todos están por encima de 87% de este modo podemos trabajar de esta forma la rentabilidad para tomar decisiones, esto nos indica que la fase de preparación de datos específicamente en el proceso de discretización se hizo con el uso de encontrar intervalos de confianza con un 50% de confianza genera buenos porcentajes de clasificación para la toma de decisiones.

5.3 Recomendaciones

- Adaptar el mismo estudio pero con diferentes monedas y estudiar las similitudes y diferencias entre los 2 estudios.
- Hay diferentes metodologías de discretización de datos menciones en diferentes estudios anteriores, se recomienda adaptar un nuevo proceso de etiquetado de datos.
- Para futuros estudios realizar estudios de monedas virtuales con otros activos como el oro, el petróleo entre otros que sea de referencia en la economía mundial.

Bibliografía

[1] Herrero, J. G., & López, J. M. M. (2006). Técnicas de análisis de datos.

[2] Pérez López, C & Santin González, D (2007). Minería de datos, técnicas y herramientas; Editorial Paranienfo.

[3] Grandez Márquez, Ma (2017), Aplicación de minería del datos para determinar patrones de consumo futuro en clientes de una distribuidora de suplementos nutricionales

[4] ALANZA RICALDI, PF (20019) Aplicación de técnicas de minería de datos para predecir la deserción estudiantil de la facultad de Ingeniería de la Universidad Nacional Daniel Alcides Carrión.

[5] Peña, I.A.C., Ávila, A.E.S, A.J.D,& Montelongo, D.L.R (2018). El uso de herramientas tecnológicas de minería de datos en el análisis de datos climáticos/ The use of Tecnonological tool for data mining in the analysis of climatological data. TECI Revista Iberoamericana de las Ciencias Computacionales e Informatica, 7 (13), 1-18

[6] Montgonery. D, Jennings, CH. & Kulahci, M. (2015). INTRODUCCTION TO TIME SERIES ANALAYSIS AND FORECAST ING BY JOHN WILEY & SO

[7] Sánchez, D., Miranda, M., & Cerda, L. (2004). Reglas de asociación aplicadas a la detección de fraude con tarjetas de crédito. In Actas del XII Congreso Español sobre Tecnologías y Lógica Fuzzy (pp. 15-17).

[8] Malberti Riveros, M. A., & Elida Beguerí, G. (2015). Reglas de Asociación con los datos de una biblioteca universitaria. Revista Cubana de Ciencias Informáticas, 9(4), 30-45.

[9] Bustamante, M.E.M (20017) MODELO PARA EL DESCUBRIMIENTO DE PATRONES EN SERIES TEMPORALES SIMBÓLICAS (Doctoral dissertation, Universidad Politécnica de Madrid)

[10] Martelo, R.J, Herrera , K.C.& Villabona, N. (20017). Estrategias para disminuir la deserción universitaria mediante series de tiempo y multipol. Revista Espacios, 38 (45)

[11] Moreno, B. Valencia, N Soto. . & Sánchez, A (2018) Criptomonedas como alternativa de inversión, riesgo regulación y posibilidad de matización en Colombia.

[12] Fuentes, V. (2019). Adopción de criptomonedas y aplicaciones Blockchain en el sistema financiero.

[13] Montealegre, J. I. P. El blockchain y sus posibles aplicaciones para la educación.

[14] Riquelme Santos, J. C., Ruiz, R., & Gilbert, K. (2006). Minería de datos: Conceptos y tendencias. Inteligencia Artificial: Revista Iberoamericana de Inteligencia Artificial, 10 (29), 11-18.

[15] Fortoul Yegues, H. C. (2008). Búsqueda de modelos para el reconocimiento de patrones de uso de un sitio web a través de la minería de datos.

[16] Candelario, B. (2015). Bitcoin. The University of Miami Inter-American Law Review, 47(1), 95-128.

[17] Sánchez Méndez, J. (2018). Relación de largo plazo entre el BitCoin, los indicadores de bolsa y los principales commodities a nivel mundial: un análisis de series de tiempo 2012–2018.

[18] Ceballos Hornero, D. (2004). Análisis del tiempo como variable en economía financiera. Universitat de Barcelona.

[19] Canney, E. (2006). Data mining o minería de datos. Revista Dinero. 9, [Online; accessed 17-Aug-2020]. [Online]. Available: http://www.dinero.com/ columna-del-lector/opinion/articulo/data-mining-omineria-datos/37339

[20] Y.J. & Talavera, R. (2007). Minería de como soporte a la toma de decisiones empresariales, Universidad del Zulia, Revista Opción, 52(23), pp. 104-118.

[21] Pérez, C. & Santín, D. (2007). Minería de datos: técnicas y herramientas. Madrid: Ediciones Paraninfo, S.A

[22] M. Servente. Algoritmos tdidt aplicados a la minería de datos inteligente [tesis de Maestría]. Buenos Aires: Universidad de Buenos Aires, Facultad de Ingeniería. 2002, pp. 14-26.

[23] E. Blasco Ascencio, "Aplicación de técnicas de minería de datos en redes sociales/web," 2015

[24] Rakesh Agrawal, Tomasz Imieli´nski, Arun Swami. Mining Association Rules Between Sets of Items in Large Databases

[25] Amaris, M. E. D. M., & Rodríguez, J. E. R. (2003). La contribución de las reglas de asociación a la minería de datos. Tecnura, 7(13), 94-109.

[26] Quintana, M. J. M., Gallego, A. G., & Pascual, M. E. V. (2005). Aplicación del análisis discriminante y regresión logística en el estudio de la morosidad en las entidades financieras: comparación de resultados. Pecvnia: Revista de la Facultad de Ciencias Económicas y Empresariales, Universidad de León, (1), 175-199.

[27] Herrera, T. F., Gómez, J. M., & de la Hoz Grandaillo, E. (2012). Aplicación de análisis discriminante para evaluar el comportamiento de los indicadores financieros en las empresas del sector carbón en Colombia. *Entramado*, *8*(2), 64-73.

[28] Morgan Kaufmann Publishers. Hernández, J., Ramírez, M.J. & Ferri, C. (2004). Introducción a la minería de datos. Madrid: Pearson Educación. Marcano

[29] Molina, L. C. (2002). Data Mining: torturando a los datos hasta que confiesen. , [Online; accessed 17-Aug-2020]. [Online]. Available http://www.lsi.upc. es/~lcmolina/

[30] Paulo, J Investigación muestra cuales son las 5 criptomonedas más utilizadas para compras. , [Online; accessed 17-Aug-2020]. [Online]. Available https://es.cointelegraph.com/news/the-5-most-used-cryptocurrencies-for-shopping

[31] CoinMarketCap TODAS LAS CRIPTOMONEDAS [Online; accessed 17-Aug-2020]. [Online]. Available https://coinmarketcap.com/es/all/views/all/