



UNIVERSIDAD CÉSAR VALLEJO

FACULTAD DE INGENIERÍA Y ARQUITECTURA

ESCUELA PROFESIONAL DE INGENIERÍA DE SISTEMAS

Modelo de predicción para determinar la compra de productos
manteniendo un stock adecuado de artículos de Ferretería. Caso
aplicado a la Ferretería MAYTA.

**TESIS PARA OBTENER EL TÍTULO PROFESIONAL DE:
INGENIERO DE SISTEMAS**

AUTOR:

Flores Zevallos, Max Ignacio (ORCID: 0000-0003-1002-640X)

ASESOR:

Mg. Carranza Barrena, Wilfredo Eduardo. (ORCID: 0000-0003-0845-1984)

LÍNEA DE INVESTIGACIÓN:

Sistema de Información y Comunicaciones.

LIMA – PERÚ

2021

DEDICATORIA

Dedico a mi familia, por apoyarme con mis estudios y permitirme llegar hasta donde estoy.

AGRADECIMIENTO

Agradezco a la Universidad César Vallejo por haberme apoyado en mi formación como profesional.

A mis maestros que me brindaron sus conocimientos y la experiencia en el mundo.

Así mismo a mis compañeros por brindarme su conocimiento y amistad.

ÍNDICE DE CONTENIDO

Carátula.....	i
Dedicatoria.....	ii
Agradecimiento.....	iii
Índice de contenido.....	iv
Índice de tablas.....	v
Índice de figuras.....	vi
Resumen.....	vii
Abstract.....	viii
I. INTRODUCCIÓN.....	1
II. MARCO TEÓRICO.....	4
III. METODOLOGÍA.....	12
3.1 Tipo y Diseño de Investigación.....	12
3.2 Variables y Operaciones.....	12
3.3 Población, muestra, muestreo, unidad de análisis.....	12
3.4 Técnica e Instrumento de recolección de datos.....	14
3.5 Procedimientos para recolección de datos.....	14
3.6 Métodos de análisis de datos.....	14
3.7 Aspectos éticos.....	17
IV. RESULTADOS.....	18
V. DISCUSIÓN.....	30
VI. CONCLUSIONES.....	34
VII. RECOMENDACIONES.....	35
REFERENCIAS.....	36
ANEXOS.....	39

ÍNDICE DE TABLAS

Tabla 1: Productos seleccionamos con mayores ganancias.	18
Tabla 2: Comparación de los diferentes algoritmos aplicados.	30
Tabla 3: Comparación de las diferentes matrices de confusión de árbol.	31
Tabla 4: La matriz de confusión de Rules.	32

ÍNDICE DE FIGURAS

Figura 1: Proceso de CRISP-DM	8
Figura 2: Redes neuronales	10
Figura 3: Cronograma	16
Figura 4: Selección de J48 y el test del 80% para entrenar.	19
Figura 5: Resultado del J48, primera parte información de lo aplicado.....	19
Figura 6: Árbol podado J48, nivel y tamaño del árbol	20
Figura 7: Resumen del J48.	21
Figura 8: Matiz de confusión de J48.....	21
Figura 9: El Grafico del árbol de J48.	22
Figura 10: Clasificador RandomForest, opción del test usando el 80%.	23
Figura 11: Resumen de RandomForest.	23
Figura 12: Matriz de RandomForest.....	24
Figura 13: RandomTree, el tipo de entrenamiento.	24
Figura 14: Lógica aplicada por RandomTree para construir el árbol.....	25
Figura 15: Resumen de RandomTree	25
Figura 16: Matriz de confusión de RandomTree	25
Figura 17: El árbol generado por RandomTree en forma gráfica.	26
Figura 18: Parámetros en lo que se ejecutó el JRip.....	27
Figura 20: Las reglas o relaciones encontradas por JRip	27
Figura 21: Resumen de JRip.....	28
Figura 22: Matriz de confusión de JRip.	28

Resumen

En este proyecto de investigación titulada Modelo de predicción para determina la compra de productos manteniendo un stock adecuado, teniendo como objetivo desarrollar un modelo de predicción para determinar la compra de productos manteniendo un stock adecuado artículos de ferretería. Fue de tipo aplicada y un diseño experimental donde se escogieron 168 productos en base a los datos históricos obtenidos. La cual pasaron un riguroso proceso de recolección y selección de datos y limpieza de las mismas. Los resultados muestran que los modelos de árbol como J48 tienen un 99.85% de correcta clasificación con un error absoluto medio 0.0009, el RandomForest tiene 99.85% de correcta clasificación con un error absoluto medio 0.001, el RandomTree tiene un 99.85% correcto asertividad con un error absoluto medio 0.0009 y el modelo de Rules como JRip tiene un 99.85% con error absoluto medio 0.0052. Se concluye que el modelo J48 es el mejor modelo desarrollado para determinar la compra de productos ya que tiene 99.85% de asertividad y también el TandomTree con un 99.85%, la diferencia con los otros 2 es que tiene un error absoluto medio de 0.0009.

Palabras clave: Modelo de predicción, Predicción, Minería de datos, Aprendizaje automático.

Abstract

In this research project entitled Prediction model to determine the purchase of products while maintaining an adequate stock, with the objective of developing a prediction model to determine the purchase of products while maintaining an adequate stock of hardware items. It was of an applied type and an experimental design where 168 products were chosen based on the historical data obtained. Which underwent a rigorous process of data collection and selection and cleaning of the same. The results show that tree models like J48 have a 99.85% correct classification with a mean absolute error of 0.0009, the RandomForest has a 99.85% correct classification with a mean absolute error 0.001, the RandomTree has a 99.85% correct assertiveness with an error mean absolute 0.0009 and the Rules model like JRip has a 99.85% with mean absolute error 0.0052. It is concluded that the J48 model is the best model developed to determine the purchase of products since it has 99.85% assertiveness and also the TandomTree with 99.85%, the difference with the other 2 is that it has a mean absolute error of 0.0009.

Keywords: Prediction model, prediction, data mining, machine learning.

I. INTRODUCCIÓN.

En la actualidad, los negocio tipo PYMES no realizan con frecuencia una planificación de sus productos que deben comprar en función con ventas, al no haber planificado no va a poder brindar un buen servicio. Al no saber qué productos y cuantos tiene, no va a poder ofrecer con exactitud sus productos, que nuevos productos necesitan sus clientes, etc. Lo cual generaría que el cliente quede insatisfecho, ya sea por la demora de la entrega del producto, el negocio no sepa que producto tiene, no entrega completo los pedidos, contar con productos que no son muy necesarios.

Desde hace años los negocios no toman importancia de los datos o no realizan una planificación de productos que deben comprar. En el proceso de compras lo que realizaban es adquirir sus productos teniendo solo en cuenta, cuanto de stock queda de productos y teniendo referencia si se vende o no el producto, pero como sabemos por nosotros mismos la memoria no siempre es buena en recordar cosas que pasan mucho tiempo. Todo partiendo de la falta contar con un registro de ventas de sus productos y un buen manejo de estos datos, para adquirir conocimiento, al no contar con esto, genera que no podrá tener un juicio más preciso de cuantos productos deben comprar. Con la aparición de la pandemia en el 2019 el problema se agravo ya que un negocio no puede comprar por comprar sus productos, esto debe basarse en datos, que como mencionamos anteriormente estos datos al ser procesados se transforman en información y esta información puede proporcionar conocimientos, y así poder tomar decisiones que tengan menor riesgo, brindando un beneficio al negocio y los clientes. Veamos mediante un gráfico el proceso del negocio para compra del negocio. [\(Anexo 1\)](#)

Lo que se espera es que a al no solucionarse en un futuro esto pueda generar problemas con la atención al cliente y económicos, ya que al comprar productos sin saber si se venderá o no, saber qué productos debe comprar puede provocar que los clientes no vuelvan al negocio o sienta el mismo negocio está gastando más en compras productos y estos no se ve reflejado en las ventas o ganancias. Si el negocio no trata de planificar mejor sus compras no va a poder seguir creciendo o cumplir sus objetivos.

La finalidad de la investigación es mejorar el control de los productos del negocio y también poder predecir cuantos productos debe adquirir para una futura

compra, de lo antes mencionado podemos decir que nosotros buscamos es proponer un modelo de predicción para determinar la compra de productos manteniendo un stock adecuado de artículos de ferretería.

Como consecuencia se formuló el siguiente problema general ¿Cómo determinar la compra de productos manteniendo un stock adecuado de los artículos de ferretería?, y como problemas específicos ¿Cómo identificar los productos con mayor frecuencia de ventas que ayudan a formular el modelo predictivo? ¿Qué algoritmo de clasificación aplica mejor para formular el modelo de predicción? ¿Cómo validar el resultado, el producto de haber aplicado el modelo predictivo?

La justificación de la investigación en lo social es que pretende aportar un conocimiento a los diversos emprendedores o empresarios pequeños y medianas empresas una nueva forma que como van poder manejar el stock, como por ejemplo ya no va ser necesario tener una gran cantidad de productos de diversas marcas lo único que tendrías que hacer las organizaciones es tener un historial de sus ventas o compras y aplicando el modelo predictivo van poder ver que productos son los que necesitan, los que más salen y cuantos en promedio deben tener.

En la justificación institucional es brindar al aliado una herramienta que le brindara una información muy valiosa a la organización como seria los productos que no necesitan comprar en gran cantidad, aquellos que se ven comprar con mayor prioridad, o que nuevos productos debe comprar la organización para poder fortalecer la venta de aquellos productos que se venden bien, con toda esta información la organización poder tomar medidas, ya que una empresa ahora en la actualidad puede estar gastando en recursos innecesarios. En lo personal/ profesional: El trabajo me lleva poder adquirir nuevos conocimientos de lo que en un futuro pueda requerirse ya que lo que el mercado va buscando son profesionales capaces de convertir los datos en información y a la par en conocimientos para la organización.

En la justificación económica: Lo que se estará buscando es que el aliado en su proceso de compra de productos y almacenaje pueda reducir costos, evitando tener productos que no son muy vendidos y generan espacio innecesario y gasto. Tecnológico: Estaremos aplicando los nuevos conocimientos sobre el modelo de predicción, minería de datos y machine learning, y brindar nuevas bases para futuras investigaciones sobre este tema.

Antes esto planteamos nuestro objetivo general es Desarrollar un modelo de predicción para determinar la compra de productos manteniendo un stock adecuado de artículos de ferretería. Y como objetivo específico 1) Identificar datos de productos con mayor rotación de stock y mayor ganancia que faciliten la data de entrada al modelo predictivo. 2) Determinar el mejor algoritmo de clasificación para la formulación del modelo. 3) Aplicar un 80% para procesar data y un 20% como elemento de control para validar los resultados del modelo.

Terminando esta parte se planteó la hipótesis general como: El modelo de predicción ayuda a determinar que productos comprar para mantener un stock adecuado de artículos de ferretería, y Las hipótesis específicas 1) Los datos de productos identificados ayudaron a los datos de entrada al modelo predictivo. 2) El algoritmo de clasificación identificado ayuda a la formulación del modelo. 3) Con el procesamiento del 80% se formula el modelo y el 20 % se valida el modelo resultante.

II. MARCO TEÓRICO

En base a la búsqueda realizada en el artículo de revisión, de diferentes investigaciones y artículos sobre el modelo de predicción se encontró con los siguientes autores que trataron con un tema parecido del *modelo de predicción*.

2.1 Antecedentes Internaciones.

Rodríguez (2017), en su investigación titulada "*Creación de modelo de predictivo para la demanda de clientes de una empresa del rubro de TI*". Tuvo como objetivo creación de un modelo que prediga la demanda de cada cliente en el futuro con el fin de adelantarse a las necesidades de éstos y de esta manera generar nuevas oportunidades. Fue un estudio experimental, La población fue la empresa de TI, la muestra fueron los datos de los clientes y muestreo probabilístico; la obtención de la información fue de la base de datos del negocio. Los resultados obtenidos fueron que el aspecto a resaltar es la gran exactitud del pronóstico que alcanza el modelo, teniendo un error absoluto medio de 0.7115, indicando estos valores que se pronosticaron se diferencias en promedio de dicha cantidad. Lo que uno busca generalmente es que el valor será lo mínimo posible. En consecuencia, se decide usar el modelo realizado con la técnica Decision Tree Regressor, teniendo en cuenta que posee un tiempo de procesamiento menor que el empleado por Gradient Boosted Tree. Se concluyo que, al estudiar varias metodologías, se estima que CRISP-DM fue la mejor decisión para elaborar el trabajo, porque para el proceso era necesario un levantamiento de requerimientos y entender el negocio que otras técnicas no otorgan. Sabiendo que si se usaba la metodología SEMMA hubiéramos agilizado el proceso, no se hubiéramos tenido un desarrollo organizado y eficaz, ya que esto se logra teniendo buenas etapas que propone la metodología usada. Una de las principales ventajas es la capacidad de combinar los aspectos del negocio con los aspectos técnicos, porque se usará el plan de trabajo teniendo en cuenta las necesidades de minería de datos ofreciendo una planificación bastante útil a la hora de reconocer errores y realizar evaluaciones de forma interactiva.

Moreno y Michel (2019), en su investigación titulada "*Análisis de comportamiento de la predicción de ventas con diferentes técnicas de minería de*

datos". Tuvo como objetivo tomar una muestra de la información de las ventas de 6 productos, de línea de artículos de mayor movimiento de una empresa, tomando para el análisis dos variables de tipo cuantitativa, las ventas y el tiempo, para aplicarlas a un modelo de predicción. Fue un estudio de tipo no experimental, la población de estudio fueron los productos de venta, la muestra fueron 6 productos con mayor movimiento y muestreo fue probabilístico; La información fue recogida de la base de datos del negocio. Los principales resultados mencionan que al aplicar la ecuación (2) descrita en el fundamento teórico, la cual corresponde al modelo de promedio móvil, que al analizar la predicción generada (línea roja) con respecto a la venta real que representa su comportamiento histórico (línea azul), muestra un desfase, así la cantidad de meses que se tomaron para promedio móvil fue de tres, ya que se hicieron pruebas que demostraron que esta cantidad fue la más idónea para continuar con la prueba de los demás productos, el otro resultado menciona que el producto 1 tiene un porcentaje más bajo con el modelo de regresión lineal; el producto 2 tiene un porcentaje de 0.90% de error en la técnica de árboles aleatorios, siendo la mejor para éste. El producto 3 muestra similitud con el producto 1 ya que el porcentaje más bajo es con la técnica de regresión lineal; para el producto 4 se puede notar una diferencia en el modelo con el porcentaje de error más bajo en la técnica de suavización exponencial. Por último, para los productos 5 y 6 se mostraron porcentajes más bajos con el modelo de regresión lineal. Se concluyó la predicción de ventas es algo vital hoy en día y el saber qué rumbo tendrá un negocio puede ser de gran ayuda para los empresarios. Al desarrollar esta investigación se intentó elaborar una herramienta para apoyar a la toma de decisiones dentro de una abarrotera ubicada en Cd. Guzmán, tras aplicar método de minería de datos, así como regresión lineal, se concluyó que al aplicarlas no se pudo obtener un resultado óptimo en comparación con las ventas reales de la empresa.

En no realizar muy largo, se movieron algunos antecedentes al [anexo 2](#).

2.2 Antecedentes Nacionales.

Castro y Hernández (2016), en su investigación titulada "*Implementación de un modelo de predictivo basado en data mining y soportado por SAP predictive analytics en retails*". Tuvo como objetivo implementar un modelo predictivo basado

data mining y soportado por SAP en una empresa Retails. Fue un estudio de tipo experimental, se analizó la empresa Rattail; Se realizó una recolección de la base de datos del área de planeamiento Comercial. Los resultados obtenidos fueron que cuando se compró el algoritmo Triple Exponencial Smoothing con el pronóstico que realiza el área de planeamiento comercial de Topitop, se pudo identificar que se mejoró en la predicción del alcance, minimizando el Error Porcentual Absoluto Medio (MAPE) en 28.49%. Se pudo concluir que La predicción de las ventas juega un papel imprescindible en el cumplimiento del proceso para contar con productos en la empresa, evitando la ausencia o el sobre abastecimiento de stock, esto afin evita tener pérdidas monetarias, contando con una gran cantidad de datos se puede lograr mejorar resultados. Contando con los datos de los dos últimos años de ventas, los algoritmos utilizados nos proporcionaron resultados bastante favorable e incluso mucho más preciso a una estimación matemática.

De la Oliva (2019), en su investigación titulada “*Sistema de pronóstico de compra utilizando un modelo de predicción para mejorar el monitoreo comercial de un supermercado de la ciudad de Chiclayo*”. Tuvo como objetivo Mejorar el monitoreo comercial mediante la implementación de un sistema de pronóstico de compra utilizando un modelo de predicción. Fue un estudio de causi-experimental, la población de estudio fue el supermercado de la ciudad de Chiclayo, la muestra los clientes internos, externos y documentos, y muestreo fue probabilístico; el instrumento usado para recoger la información fueron las hojas de apuntes, alguna grabadora y preguntas preparadas. Los principales resultados fueron que, Se incrementó la satisfacción de clientes tanto internos como externos en un 48 y 53%. Se ampliaron los reportes de predicción en un 67% a través de tener más datos precisos y de importancia para la empresa, y se disminuyó el tiempo en que se realizan las predicciones de compra en un 95%. Se concluyó que con la implementación del sistema para predicción y el uso de modelo se puede lograr satisfacer a los clientes tanto internos como externos. Con lo observado antes se pudo ver que los inconvenientes presentados en la organización no permitían un desempeño eficiente a fin de pronosticar la compra generando insatisfacción por parte del trabajador. Con la elaboración del sistema se aumentó la satisfacción en un 45% y 53% de los clientes internos y externo respectivamente.

Para más antecedentes revisar [anexo 3](#).

2.3 Definición de Términos.

2.3.1 Minería de datos.

Cuando escuchamos por primera vez la palabra minería podemos confundirnos con minería normal que se conoce, pero si esta palabra los separamos podemos ver que una trata algo de escarba o buscar y datos son un registro o lista, al tratarlo de relacionar podemos decir que se trata buscar información relevante en la lista de datos que tenemos, bueno para tener más claro y no confundirnos veamos que nos dice el siguiente autor. Daza (2016) menciona que:

La minería de datos hace referencia a la integración de datos que puede ser de diferentes fuentes como SQL server, Oracle, Excel, etc., para después objeten un importante conocimiento, en otras palabras, identificar información trascendental, importante y útil, de lo cual las organizaciones podrán tomar decisiones. (p. 16)

En conclusión, la minería de datos es donde se extraen datos de diferentes fuentes, estos datos procesarlas y extraer o identificar información importante, y esta información va a poder ayudar a la organización a tomar medias en base a datos.

Para esta investigación vamos a aplicar una de las metodologías de minería de datos que es CRISP-DM.

a) CRISP – DM (Cross-Industry Standard Process for Data Mining)

Parecido a los anteriores, esta es una metodología, pero también podemos decir que es modelo de procesos para descubrir la manera en la que los expertos en la materia abordan el problema. (Galán, 2015, p. 21). En caso de querer saber más a profundidad cada fase de la metodología ir al [anexo 4](#).

Proceso de la metodología CRISP-DM.

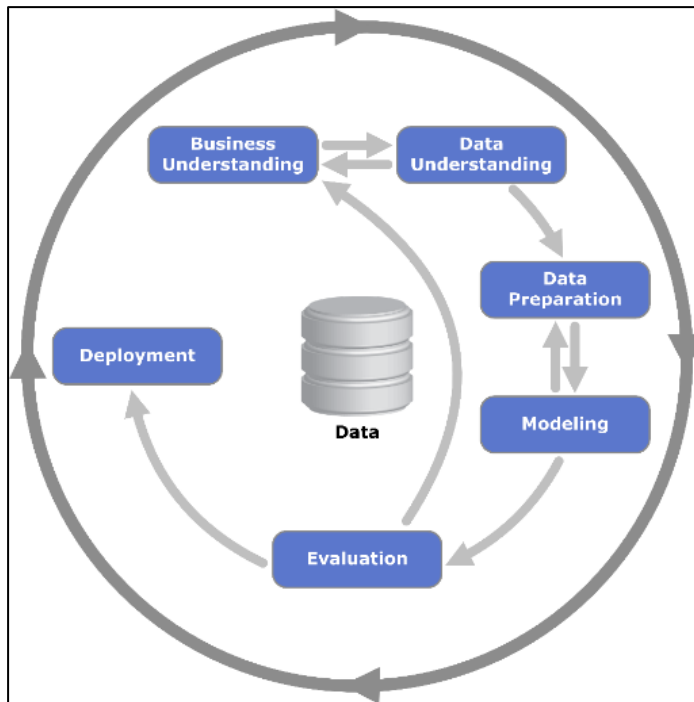


Figura 1: Proceso de CRISP-DM

Fuente:

https://disi.unal.edu.co/~eleonguz/cursos/md/presentaciones/Sesion5_Metodologias.pdf

2.3.2 Machine Learning

Un autor nos menciona que “Machine learning o inteligencia artificial es la capacidad de un dispositivo tecnológico para realizar tareas específicas y tomando decisiones como una persona en base al entrenamiento que han tenido”. (Maisueche, 2019, p. 13)

Para aprendizaje existen dos tipos, según nos menciona Sandoval (2018):

a) Aprendizaje Supervisado.

Es cuando le proporcionamos las preguntas (Características) y respuestas (etiquetas) de los datos tratando de entrenarlo. Entre estos existen 2 tipos de algoritmos.

- Algoritmo de Clasificación: Lo que se espera es que el algoritmo nos indique a que grupo pertenece el elemento de estudio.
- Algoritmo de regresión: Lo que se espera es la respuesta un número y no se ubica en un grupo.

b) Aprendizaje no supervisado.

En esta parte a diferencia del anterior solo le damos las preguntas(característica) al algoritmo. Lo que se realizara el algoritmo es agrupar a los elementos según su característica y brindar la respuesta según las características y decir a que grupo pertenece.

2.3.3 Modelos de Machin Learning

Podemos agrupar en 3 modelos.

- a) **Modelos lineales:** Es encontrar una línea entre una línea que se ajuste a la nube de puntos. Los más conocidos y usados son:
 - Regresión lineal.
 - La logística.
- b) **Modelo de árbol:** Es el más preciso, estable y más fácil de interpretar, ya que construye reglas de decisiones que se puedan representar como árbol. A diferencia del anterior puede representarse relaciones no lineales. Los que destacan son:
 - Los árboles de decisión.
 - Los random forest.
- c) **Redes neuronales:** Caparrós (1994) menciona que es un procedimiento de cómputo que trata de simular el comportamiento de neuronas durante el aprendizaje, está conformado por elementos interconectados (neuronas) que reciben información y generan señales que se trasmite hasta generar una respuesta (p. 221)

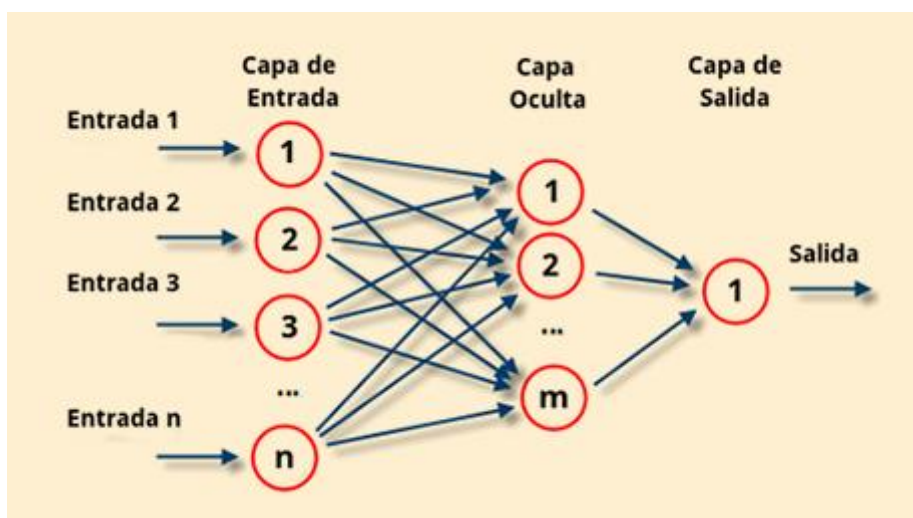


Figura 2: Redes neuronales

Fuente: <https://monitor.iiiipe.edu.mx/notas/%C2%BFpor-qu%C3%A9-las-redes-neuronales-artificiales-son-el-futuro>

2.3.4 Modelo de Predicción.

En esta parte mencionare algunas definiciones que mencionan algunos autores.

Uno de esos términos será el modelo de predicción en donde Redacción España (2020) mencionan que “son procesos mediante técnicas computacionales de estudio de datos que ayudan a deducir la probabilidad de que ocurra ciertas situaciones previas, también podemos poner descubrir oportunidades de negocio” (párr. 2). De lo que se menciona podemos decir que un modelo de predicción es conjuntos de técnicas computacionales todo a partir de datos sobre lo que puede suceder a un futuro, así el negocio puede tomar decisiones de detectar las oportunidades.

Para ampliar un poco más veamos lo que menciona el sitio web NeoAttack que también define el modelo predictivo como “proceso de estadística y análisis d datos a fin de definir las hipótesis o descubrir resultados o sucesos futuros.” (párr.3).

Como ultima mención es que menciona que el modelo predictivo puede servir para hallar oportunidades y detectar situaciones adversas.

2.3.5 Stock

Continuando, ahora veamos ahora lo que es el Stock. En primera impresión y de lo que nosotros escuchamos ya que es un término muy importante en una empresa, de la más pequeña a la más grande. Stock el almacenamiento de productos en un lugar específico para que la empresa pueda hacer uso de ellos cuando sean necesario, para tener más claro veamos lo que menciona Páez (2020) donde menciona que “es el conjunto de mercancías almacenas por una empresa. Esta mercancía se refiere a materiales primarias y productos terminados y distribuidos para ser entregados a clientes, y estos

tiene que estar asociados al proceso productivo del negocio”. (párr.1). De esta cita podemos concluir que el stock son productos que se relacionan con la actividad del negocio y estos están almacenado para su futuro uso que la empresa requiera.

Por otra parte, InnovaCommerce menciona que “La palabra stock es una palabra anglosajona, que, si la traducimos, hace referencia a la cantidad de productos, artículos, o bienes de los que tiene una organización determina o una persona en concreto”. (párr. 1). De la cita podemos resaltar que hace menciona a cantidad de productos que cuenta una organización, y cuando hace mención a anglosajona se refiere que la palabra proviene hace mucho tiempo aproximadamente del siglo V, de pueblos germanos que invaden Inglaterra.

2.3.6 WEKA

Weka es un software que nos ayudara en el proyecto con el manejo de los datos. Para tener más claro veamos lo que nos indica la misma página de weka 3 la cual indica que es una serie de algoritmos de formación automático en data mining. Conteniendo herramientas para preparar los datos, de ordenar, de análisis de regresión, clustering, reglas de asociación y visualización. (párr. 1). De lo que podemos concluir es que weka es una plataforma que tiene varios algoritmos y herramientas para la minería de datos.

III. METODOLOGÍA

3.1 Tipo y Diseño de Investigación.

3.1.1 Tipo de investigación: Aplicada.

Según Rodríguez (2020) menciona que “Es aquella investigación en el cual el problema ya se determinó y es de conocimiento del investigador, por ello se utiliza la investigación para dar la respuesta a preguntas específicas” (párr. 1). Otro autor que también menciona sobre el tema es Lozada (2014) quien menciona que:

Busca generar conocimiento utilizando directa y en mediano plazo en la sociedad a o en el ámbito productivo. [...] también se utiliza conocimiento que proviene de la investigación básica y afectando indirectamente cómo incremento del nivel de vida de la población y en la creación de puesto de trabajo. (p. 35)

De las dos citas anteriores podemos decir que la investigación de tipo aplicada es una investigación que nace a partir de un problema y buscaremos una solución a esta partiendo de conocimiento básico sobre el tema.

3.1.2 Diseño de investigación: Experimental.

En cuanto a la investigación experimental Arias (2012) menciona que “la investigación experimental es la manipulación de la variable independiente a determinadas condiciones, estímulos o tratamientos, para así ver los efectos o reacciones que producen en la variable dependiente.” (p.34).

3.2 Variables y Operaciones.

3.2.1 Variable independiente: Modelo de predicción.

3.2.2 Variable dependiente: Stock adecuado.

Para más detalles puede revisar en [anexo 5](#) (Matriz Operacional). Y [anexo 6](#) (Matriz de consistencia).

3.3 Población, muestra, muestreo, unidad de análisis.

3.3.1 Población: Los 298 productos que se vendieron entre los años del 2020-2021.

3.3.2 Muestra: Los 168 productos que se vendieron entre los años 2020 al agosto del 2021.

$$n = \frac{N * Z^2 * p * q}{d^2 * (N - 1) + Z^2 * p * q}$$

En donde:
 N= tamaño de la población.
 Z = nivel de confianza.
 P = probabilidad de éxito o proporción esperada.
 q = probabilidad de fracaso.
 d² = precisión (error máximo admisible en términos de proporción)

Figura 23: Formula para calcular muestra.

CALCULO DE TAMAÑO DE MUESTRA FINA	
PARAMETRO	INSERTAR VALOR
N	298
Z	1.96
P	0.5
Q	0.5
e	0.05
Tamaño de muestra	
	168
Numerador	286.1992
Denominador	1.7029

Figura 24: Calculo de la muestra de los productos.

3.3.3 Muestreo: Estratificado.

Pero el muestro estratificado es dividir la población en grupos y de esos grupos seleccionar aleatoriamente la muestra. Pero para tener más claro veamos que nos dice Porras quien menciona que esto “implica dividir a la población en clases y o grupos, denominados estratos (subgrupo). Las unidades incluidas en cada subgrupo deben ser relativamente homogéneas con respecto a las características a estudiar” (s.f., p.5).

Otro autor también que nos habla es Roldán (2019) quien menciona que “es una técnica o procedimiento en la que se divide la población bajo estudio en distintos subgrupos o estratos. La característica es que cada subgrupo debe tener elementos únicos y estos elementos no pueden estar en más de uno” (parr. 1).

3.3.4 Unidad de análisis: Productos con mayor demanda.

3.4 Técnica e Instrumento de recolección de datos.

No aplica para esta investigación.

Pero podemos mencionar que se extraerá los datos del Excel del negocio (BD), que estén entre los años 2020-2021.

3.5 Procedimientos para recolección de datos.

Los datos estén en físico se pasará a Excel para más comodidad y manejo de los datos. Cuando lo tengamos en Excel tendremos que limpiar la información, llenando los campos vacíos o eliminar los repetidos, en otros casos tener que cambiar los valores de texto por números. Y aplicamos un modelo de predicción, analizamos los resultados y brindamos esa información al aliado para que tome sus medidas en base a datos reales.

- El procesamiento de los datos se realizará a través de la plataforma weka que ayuda realizar modelos predictivos aplicando algoritmos de clasificación.
- Para realizar el uso de la plataforma weka primero debemos tener los datos en un archivo con un formato csv.
- Después recién aplicamos los algoritmos de clasificación árbol, pero esto puede variar dependiendo la investigación.
- Después de aplicado algún algoritmo y haber obtenido una buena confianza de los algoritmos, podemos cargar un archivo para probar si funciona el algoritmo para predecir. Tener en cuenta que estos archivos también deben estar en un formato .csv o .arff.
- Después de ver los resultados debemos documentar los resultados

3.6 Métodos de análisis de datos.

El método que se aplicara es en método CRISP-DM para recolección, procesamiento y análisis de datos. Para mayor entendimiento se recomienda revisar el marco teórico, y si tienes mayor interés puede investigar por su cuenta nada nos detiene en un mundo cada vez más accesible a la información.

También se aplicará la herramienta WEKA la cual es usada para minería de datos y machine learning, la cual nos ayuda a con el modelo de predicción.

Después de haber recogido toda la información se procede a usar WEKA para aplicar los diferentes algoritmos de agrupación y clasificación. Podremos visualizar y al resultado de aplicar los algoritmos poder comprobar si está bien el modelo.

Figura 3: Cronograma

	📌	Nombre	Duración	Inicio	Terminado
1		☐ Modelo de Predicción - CRISP-DM	51 days?	13/09/21 08:00 AM	22/11/21 05:00 PM
2		Aprobación del Plan	1 day?	13/09/21 08:00 AM	13/09/21 05:00 PM
3		☐ Comprensión del negocio	5 days?	13/09/21 08:00 AM	17/09/21 05:00 PM
4		☐ Determinar objetivos del negocio	5 days?	13/09/21 08:00 AM	17/09/21 05:00 PM
5		Objetivos del negocio	5 days?	13/09/21 08:00 AM	17/09/21 05:00 PM
6		☐ Valoración de la situación	5 days?	13/09/21 08:00 AM	17/09/21 05:00 PM
7		Requisitos, supuestos y restricciones	5 days?	13/09/21 08:00 AM	17/09/21 05:00 PM
8		☐ Determinar los objetivos de DM	5 days?	13/09/21 08:00 AM	17/09/21 05:00 PM
9	📌	Metas de Data Mining	5 days?	13/09/21 08:00 AM	17/09/21 05:00 PM
10		☐ Realizar el plan del proyecto	5 days?	13/09/21 08:00 AM	17/09/21 05:00 PM
11		Plan de proyecto	5 days?	13/09/21 08:00 AM	17/09/21 05:00 PM
12		☐ Comprensión de los datos	7 days?	25/09/21 08:00 AM	05/10/21 05:00 PM
13		☐ Recolectar los datos iniciales	5 days?	25/09/21 08:00 AM	01/10/21 05:00 PM
14	📌	Reporte de recolección de datos	5 days?	25/09/21 08:00 AM	01/10/21 05:00 PM
15		☐ Descripción de los datos	5 days?	25/09/21 08:00 AM	01/10/21 05:00 PM
16	📌	Reporte de descripción	5 days?	25/09/21 08:00 AM	01/10/21 05:00 PM
17		☐ Exploración de los datos	5 days?	29/09/21 08:00 AM	05/10/21 05:00 PM
18	📌	Reporte de exploración de los datos	5 days?	29/09/21 08:00 AM	05/10/21 05:00 PM
19		☐ Verificar la calidad de los datos	5 days?	29/09/21 08:00 AM	05/10/21 05:00 PM
20	📌	Reporte de calidad de datos	5 days?	29/09/21 08:00 AM	05/10/21 05:00 PM
21		☐ Preparación de los datos	9 days?	05/10/21 08:00 AM	15/10/21 05:00 PM
22	📌	Descripción del dataset	1 day?	05/10/21 08:00 AM	05/10/21 05:00 PM
23	📌	dataset o conjunto de datos	1 day?	05/10/21 08:00 AM	05/10/21 05:00 PM
24		☐ Seleccionar los datos	7 days?	05/10/21 08:00 AM	13/10/21 05:00 PM
25	📌	Inclusión/exclusión de datos	7 days?	05/10/21 08:00 AM	13/10/21 05:00 PM
26		☐ Limpiar los datos	7 days?	07/10/21 08:00 AM	15/10/21 05:00 PM
27	📌	Reporte de calidad de los Datos	7 days?	07/10/21 08:00 AM	15/10/21 05:00 PM
28		☐ Estructurar los datos	7 days?	07/10/21 08:00 AM	15/10/21 05:00 PM
29	📌	Derivación de atributos	7 days?	07/10/21 08:00 AM	15/10/21 05:00 PM
30	📌	Generación de registros	7 days?	07/10/21 08:00 AM	15/10/21 05:00 PM
31		☐ Integrar los datos	2 days?	14/10/21 08:00 AM	15/10/21 05:00 PM
32	📌	Unificación de datos	2 days?	14/10/21 08:00 AM	15/10/21 05:00 PM
33		☐ Formateo de los datos	2 days?	14/10/21 08:00 AM	15/10/21 05:00 PM
34	📌	Reporte de calidad de datos	2 days?	14/10/21 08:00 AM	15/10/21 05:00 PM
35		☐ Modelado	7 days?	15/10/21 08:00 AM	25/10/21 05:00 PM
36		☐ Seleccionar técnica de modelado	1 day?	15/10/21 08:00 AM	15/10/21 05:00 PM
37	📌	Técnica seleccionada	1 day?	15/10/21 08:00 AM	15/10/21 05:00 PM
38	📌	Supuesto del modelo	1 day?	15/10/21 08:00 AM	15/10/21 05:00 PM
39		☐ Generar el plan de prueba	1 day?	20/10/21 08:00 AM	20/10/21 05:00 PM
40	📌	Plan de pruebas	1 day?	20/10/21 08:00 AM	20/10/21 05:00 PM
41		☐ Construir el modelo	1 day?	20/10/21 08:00 AM	20/10/21 05:00 PM
42	📌	Modelo	1 day?	20/10/21 08:00 AM	20/10/21 05:00 PM
43	📌	Descripción del modelo	1 day?	20/10/21 08:00 AM	20/10/21 05:00 PM
44		☐ Evaluar el modelo	1 day?	23/10/21 08:00 AM	25/10/21 05:00 PM
45	📌	Evaluar el modelo	1 day?	23/10/21 08:00 AM	25/10/21 05:00 PM
46	📌	Revisión de los parámetros	1 day?	23/10/21 08:00 AM	25/10/21 05:00 PM
48		☐ Evaluar los resultados	5 days?	25/10/21 08:00 AM	29/10/21 05:00 PM
49	📌	Valoración de los resultados	5 days?	25/10/21 08:00 AM	29/10/21 05:00 PM
50	📌	Modelos aprobados	5 days?	25/10/21 08:00 AM	29/10/21 05:00 PM
51		☐ Revisión del proceso	3 days?	30/10/21 08:00 AM	03/11/21 05:00 PM
52	📌	Revisión del proceso	3 days?	30/10/21 08:00 AM	03/11/21 05:00 PM
53		☐ Determinar próximos pasos	5 days?	30/10/21 08:00 AM	05/11/21 05:00 PM
54	📌	Técnica modelada	5 days?	30/10/21 08:00 AM	05/11/21 05:00 PM
55	📌	Lista de posibles acciones	5 days?	30/10/21 08:00 AM	05/11/21 05:00 PM
56		☐ Implantación	11 days?	06/11/21 08:00 AM	22/11/21 05:00 PM
57		☐ Plan de implantación	1 day?	06/11/21 08:00 AM	08/11/21 05:00 PM
58	📌	plan de implementación	1 day?	06/11/21 08:00 AM	08/11/21 05:00 PM
59		☐ Plan de monitoreo y mantenimiento	1 day?	09/11/21 08:00 AM	09/11/21 05:00 PM
60	📌	Plan de monitorización y mantenimiento	1 day?	09/11/21 08:00 AM	09/11/21 05:00 PM
61		☐ Informe final	5 days?	13/11/21 08:00 AM	19/11/21 05:00 PM
62	📌	Informe final	5 days?	13/11/21 08:00 AM	19/11/21 05:00 PM
63	📌	Modelos aprobados	5 days?	13/11/21 08:00 AM	19/11/21 05:00 PM
64		☐ Revisión del proyecto	1 day?	20/11/21 08:00 AM	22/11/21 05:00 PM
65	📌	Documentación de experiencias	1 day?	20/11/21 08:00 AM	22/11/21 05:00 PM

3.7 Aspectos éticos.

En esta parte es donde delimitare lo que se debe cumplir para que el trabajo sea de alta calidad y no tenga algún objeción u observación del contenido:

- Respetar la autoría de otros autores.
- Siempre solicitar permiso para usar datos de una empresa en tu investigación.
- Los datos deben tener un uso exclusivo para la investigación y no otros fines.

Concytec “menciona que los principios para las fases de una investigación científica son las siguientes:

- a) **Integridad** en las actividades de investigaciones científicas y gestión.
- b) **Honestidad intelectual en todos los aspectos de la investigación científica.**
- c) **Objetividad, justicia y responsabilidad** en la ejecución y difusión de los resultados de la investigación científica.
- d) **Transparencia**, actuando sin conflicto de interés, declarando y manejando el conflicto, sea este económico o de otra índole.”

Bases legales.

- Ley 28303, Ley marco de Ciencia, Tecnología e Innovación Tecnológica.
- Ley N° 28613, Ley del Consejo Nacional de Ciencia, Tecnología e Innovación Tecnológica.
- Ley N° 30035, Ley que regula el Repositorio Nacional Digital de Ciencia, Tecnología e Innovación de acceso abierto.
- Ley N° 30220, Ley Universitaria.
- Ley N° 30806, Ley que modifica diversos artículos de la Ley 28303, Ley marco de Ciencia, Tecnología e Innovación Tecnológica; y de la Ley 28613, Ley del Consejo Nacional de Ciencia, Tecnología e Innovación Tecnológica.
- Ley N° 30407, Ley de Protección y Bienestar Animal.

Fuente: <https://portal.concytec.gob.pe/images/publicaciones/Codigo-integridad-cientifica.pdf>

IV. RESULTADOS

- Identificar datos de productos con mayor rotación de stock y mayor ganancia que faciliten la data de entrada al modelo predictivo.

Se lograron identificar todos los productos y realizar un rango de los que más ganancia generaban a los que menos ganancias generaban de todos esto se selecciona a 168 productos y cada uno contando un 20 registro por los 20 meses que usamos como muestra. Cada uno se le asigno un código. Para ver la lista completa podrá realizarlo en el anexo 9.

Productos	Código
<i>ABRAZADERA</i>	<i>1</i>
<i>ADAPTADOR</i>	<i>2</i>
<i>ALAMBRE</i>	<i>3</i>
<i>ARENA FINA</i>	<i>4</i>
<i>ARENA GRUESA</i>	<i>5</i>
<i>BALDE SIKA IMPERMEABLE</i>	<i>6</i>
<i>BALDE VACIOS</i>	<i>7</i>
<i>BATEAS</i>	<i>8</i>
<i>BLOQUETA</i>	<i>9</i>
<i>BOLSA DE AGUA</i>	<i>10</i>

Tabla 1: Productos seleccionamos con mayores ganancias.

- Determinar el mejor algoritmo de clasificación para la formulación del modelo.
- Aplicar un 80% para procesar data y un 20% como elemento de control para validar los resultados del modelo

En este caso los dos objetivos vamos cumplirlos en un solo proceso ya que WEKA permite realizar la creación de modelo usando un porcentaje y el resto para entrenar. Al final comprar cual es el porcentaje mejor con mayor asertividad y así determinar el mejor modelo.

Modelos de árbol que aplicamos a nuestros datos como el J48, RandomForest y RandomTree y uno del modelo de reglas el algoritmo JRip tambien conocido como RIPPER.

- Primero estaremos probando con el J48 y también conocido como C4.5 que es el más usado para este tipo de predicciones. Y en Test Opción seleccionaremos la opción Porcentaje Split, donde asignaremos un 80% para entrenar el modelo y el restante para el test.

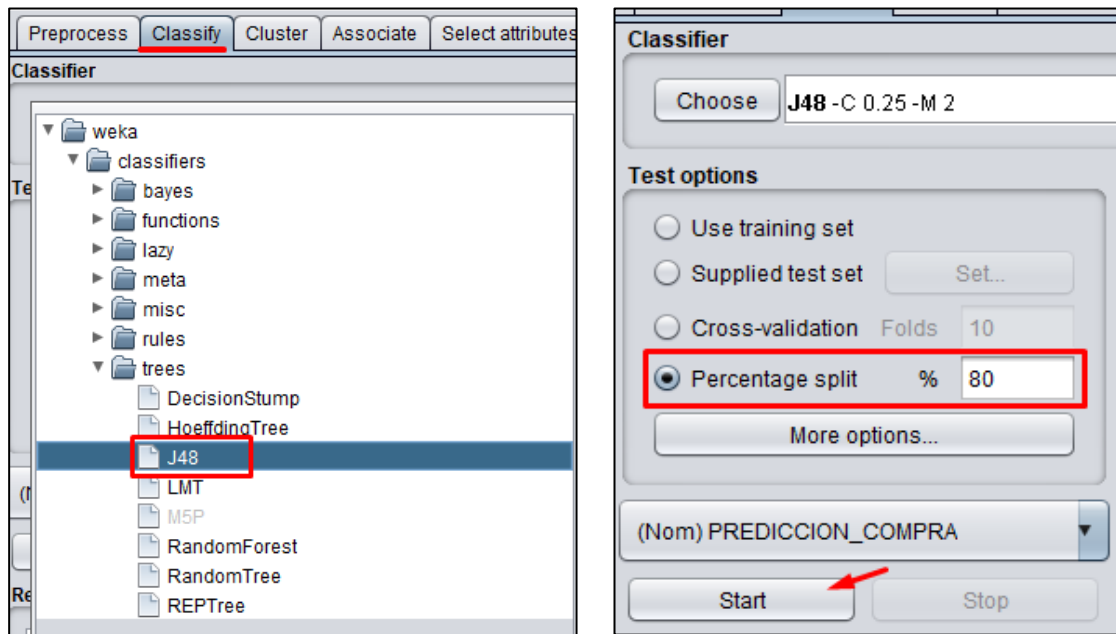


Figura 4: Selección de J48 y el test del 80% para entrenar.

Después de realizado nos mostrar el resultado, en la parte superior nos muestra, nombre del algoritmo aplicado, base de datos, instancias, atributos, y el modo de tes aplicado.

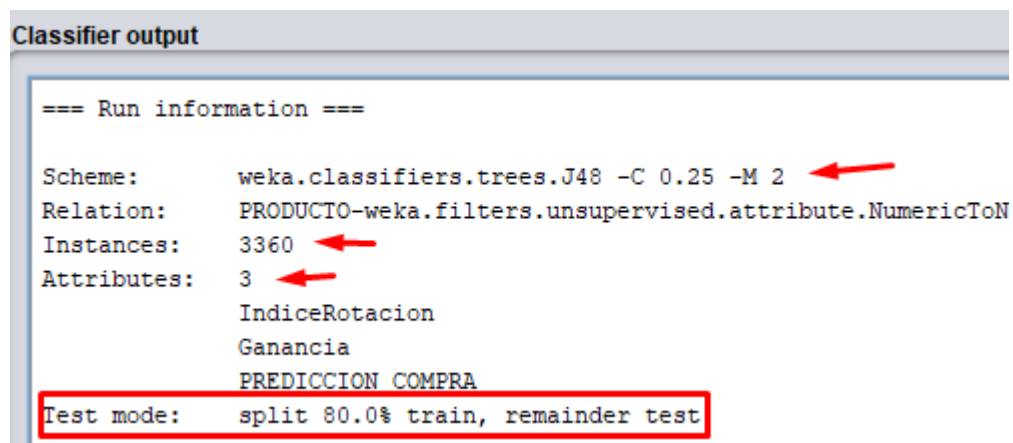


Figura 5: Resultado del J48, primera parte información de lo aplicado.

Continuando veremos las interacciones lógicas que siguió el modelo, también el tamaño y nivel del árbol.

```

J48 pruned tree
-----

Ganancia = NULA: NO_COMPRAR (2513.0)
Ganancia = BAJA
|  IndiceRotacion = NULA: NO_COMPRAR (10.0)
|  IndiceRotacion = BAJA: EVALUAR (411.0)
|  IndiceRotacion = MEDIA: EVALUAR (25.0)
|  IndiceRotacion = ALTA: EVALUAR (40.0)
|  IndiceRotacion = MUY_ALTA: EVALUAR (0.0)
Ganancia = MEDIA
|  IndiceRotacion = NULA: EVALUAR (0.0)
|  IndiceRotacion = BAJA: EVALUAR (43.0)
|  IndiceRotacion = MEDIA: COMPRAR (5.0)
|  IndiceRotacion = ALTA: COMPRAR (4.0)
|  IndiceRotacion = MUY_ALTA: EVALUAR (0.0)
Ganancia = ALTA
|  IndiceRotacion = NULA: EVALUAR (15.0)
|  IndiceRotacion = BAJA: COMPRAR (167.0)
|  IndiceRotacion = MEDIA: COMPRAR (18.0)
|  IndiceRotacion = ALTA: COMPRAR (107.0)
|  IndiceRotacion = MUY_ALTA: COMPRAR (1.0)
Ganancia = MUY_ALTA: COMPRAR (1.0)

Number of Leaves :      17

Size of the tree :      21

```

Figura 6: Árbol podado J48, nivel y tamaño del árbol

Lo más importante de ver la correcta clasificación de las instancias se recomienda que mientras más cerca de 100% mejor va ser el modelo. En este caso observamos que obtuvimos un 99.85% de asertividad. Obtenemos también Error absoluto medio, error cuadrático medio, error absoluto relativo y error cuadrático medio de la raíz.

```

=== Summary ===

Correctly Classified Instances      671          99.8512 %
Incorrectly Classified Instances     1           0.1488 %
Kappa statistic                    0.9962
Mean absolute error                 0.0009
Root mean squared error             0.0265
Relative absolute error             0.357 %
Root relative squared error        7.3557 %
Total Number of Instances          672

=== Detailed Accuracy By Class ===

      TP Rate  FP Rate  Precision  Recall  F-Measure  MCC   ROC Area  PRC Area  Class
1.000  0.006  0.998    1.000  0.999    0.996  1.000    1.000    NO_COMPRAR
1.000  0.000  1.000    1.000  1.000    1.000  1.000    1.000    EVALUAR
0.981  0.000  1.000    0.981  0.990    0.990  1.000    1.000    COMPRAR
Weighted Avg.  0.999  0.005  0.999    0.999  0.999    0.996  1.000    1.000

```

Figura 7: Resumen del J48.

Otro muy importante también es nuestra matriz de confusión, donde la diagonal comenzando de la a, la cual son las acertadas y los demás no.

```

=== Confusion Matrix ===

  a   b   c  <-- classified as
512  0   0 |  a = NO_COMPRAR
  0 107  0 |  b = EVALUAR
  1   0 52 |  c = COMPRAR

```

Figura 8: Matiz de confusión de J48.

Una opción más que podemos encontrar en WEKA es y en árbol, es poder visualizar en forma gráfica. En donde en una forma más grafica podemos observar que comienza con ganancia y según el valor no indica que resultado tomar.

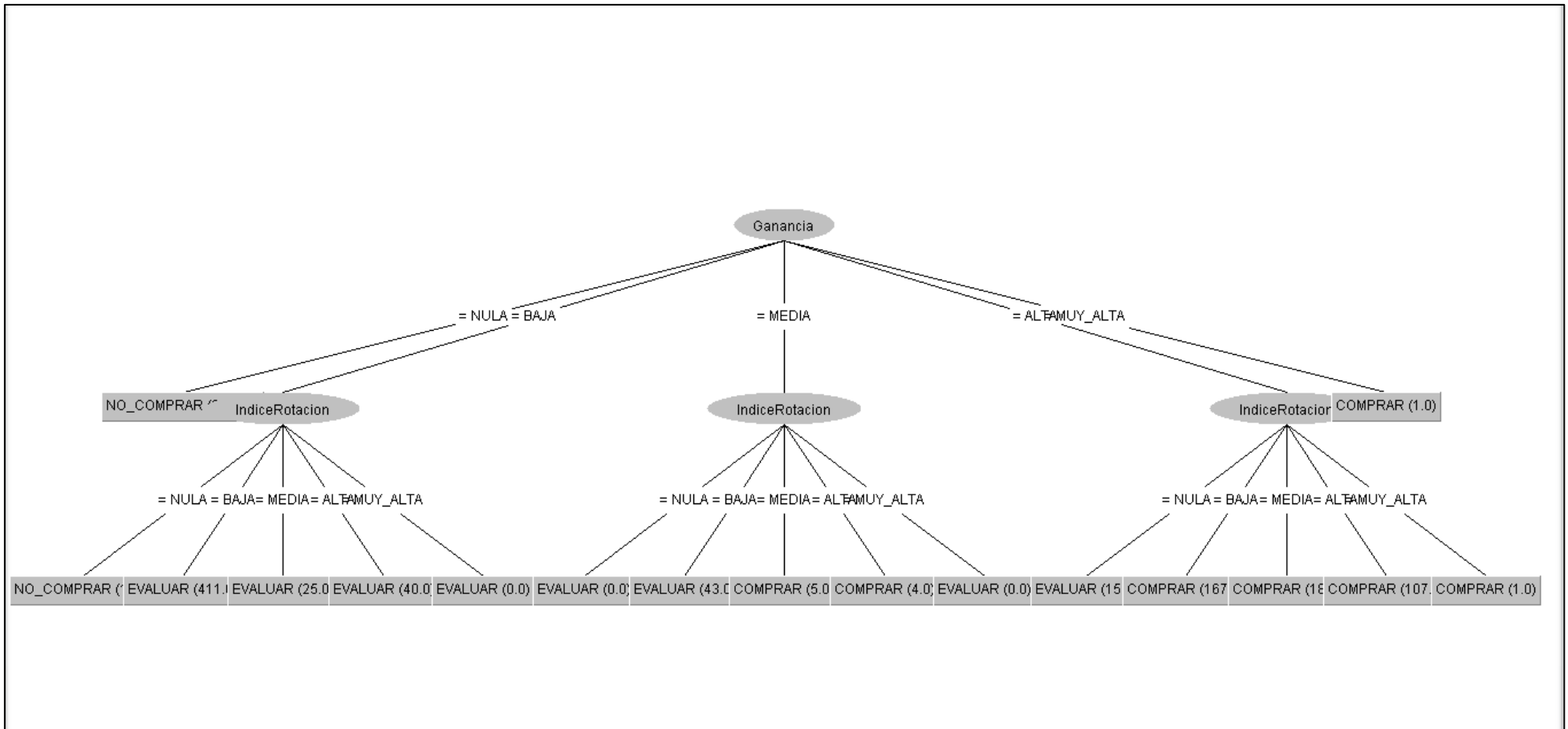


Figura 9: El Grafico del árbol de J48.

- Ahora veamos el algoritmo RandomForest, aplicaremos el mismo Test Options que el anterior.

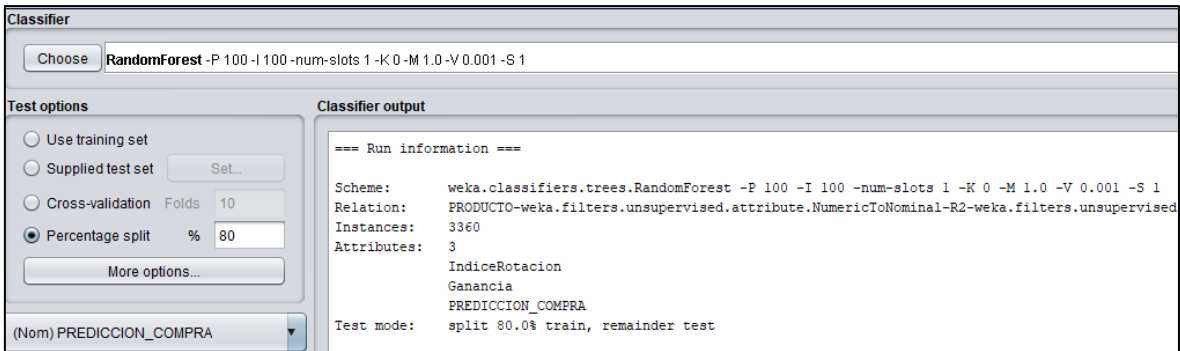


Figura 10: Clasificador RandomForest, opción del test usando el 80%.

En la correcta clasificación obtenemos un 99.85%. Obtenemos también Error absoluto medio, error cuadrático medio, error absoluto relativo y error cuadrático medio de la raíz.

```

=== Summary ===

Correctly Classified Instances      671          99.8512 %
Incorrectly Classified Instances     1           0.1488 %
Kappa statistic                    0.9962
Mean absolute error                 0.001
Root mean squared error             0.0266
Relative absolute error             0.3933 %
Root relative squared error         7.3804 %
Total Number of Instances          672

=== Detailed Accuracy By Class ===

              TP Rate  FP Rate  Precision  Recall  F-Measure  MCC      ROC Area  PRC Area  Class
              1.000   0.006   0.998     1.000   0.999     0.996   1.000    1.000    NO_COMPRAR
              1.000   0.000   1.000     1.000   1.000     1.000   1.000    1.000    EVALUAR
              0.981   0.000   1.000     0.981   0.990     0.990   1.000    1.000    COMPRAR
Weighted Avg.  0.999   0.005   0.999     0.999   0.999     0.996   1.000    1.000
  
```

Figura 11: Resumen de RandomForest.

La matriz donde muestra de 512 NO_COMPRAR se clasificó correctamente, pero 1 no, 107 Evaluar se clasificaron correctamente, 52 Comprar están bien,

```

=== Confusion Matrix ===

  a  b  c  <-- classified as
512  0  0 |  a = NO_COMPRAR
  0 107  0 |  b = EVALUAR
  1  0 52 |  c = COMPRAR
  
```

Figura 12: Matriz de RandomForest

- Como tercero que veremos de árbol será el RandomTree.

```
=== Run information ===  
  
Scheme:      weka.classifiers.trees.RandomTree -K 0 -M 1.0 -V 0.001 -S 1  
Relation:    PRODUCTO-weka.filters.unsupervised.attribute.Remove-R1-2  
Instances:   3360  
Attributes:  3  
             IndiceRotacion  
             Ganancia  
             PREDICCION_COMPRA  
Test mode:   split 80.0% train, remainder test
```

Figura 13: RandomTree, el tipo de entrenamiento.

También podemos observar las diferentes relaciones e interacciones que a realizado el algoritmo y el tamaño del árbol es de 21.

```
=== Classifier model (full training set) ===  
  
RandomTree  
=====  
  
Ganancia = NULA : NO_COMPRAR (2513/0)  
Ganancia = BAJA  
|  IndiceRotacion = NULA : NO_COMPRAR (10/0)  
|  IndiceRotacion = BAJA : EVALUAR (411/0)  
|  IndiceRotacion = MEDIA : EVALUAR (25/0)  
|  IndiceRotacion = ALTA : EVALUAR (40/0)  
|  IndiceRotacion = MUY_ALTA : NO_COMPRAR (0/0)  
Ganancia = MEDIA  
|  IndiceRotacion = NULA : NO_COMPRAR (0/0)  
|  IndiceRotacion = BAJA : EVALUAR (43/0)  
|  IndiceRotacion = MEDIA : COMPRAR (5/0)  
|  IndiceRotacion = ALTA : COMPRAR (4/0)  
|  IndiceRotacion = MUY_ALTA : NO_COMPRAR (0/0)  
Ganancia = ALTA  
|  IndiceRotacion = NULA : EVALUAR (15/0)  
|  IndiceRotacion = BAJA : COMPRAR (167/0)  
|  IndiceRotacion = MEDIA : COMPRAR (18/0)  
|  IndiceRotacion = ALTA : COMPRAR (107/0)  
|  IndiceRotacion = MUY_ALTA : COMPRAR (1/0)  
Ganancia = MUY_ALTA : COMPRAR (1/0)  
  
Size of the tree : 21
```


Figura 14: Lógica aplicada por RandomTree para construir el árbol.

Continuando igual a los anteriores vemos que la correcta clasificación es de un 99.85%. Obtenemos también Error absoluto medio, error cuadrático medio, error absoluto relativo y error cuadrático medio de la raíz.

```
=== Summary ===
Correctly Classified Instances      671          99.8512 %
Incorrectly Classified Instances      1          0.1488 %
Kappa statistic                    0.9962
Mean absolute error                  0.0009
Root mean squared error              0.0265
Relative absolute error              0.357 %
Root relative squared error          7.3557 %
Total Number of Instances           672

=== Detailed Accuracy By Class ===

```

	TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area	Class
	1.000	0.006	0.998	1.000	0.999	0.996	1.000	1.000	NO_COMPRAR
	1.000	0.000	1.000	1.000	1.000	1.000	1.000	1.000	EVALUAR
	0.981	0.000	1.000	0.981	0.990	0.990	1.000	1.000	COMPRAR
Weighted Avg.	0.999	0.005	0.999	0.999	0.999	0.996	1.000	1.000	

Figura 15: Resumen de RandomTree

La matriz de confusión clasificador usado.

```
=== Confusion Matrix ===

```

	a	b	c	<-- classified as
	512	0	0	a = NO_COMPRAR
	0	107	0	b = EVALUAR
	1	0	52	c = COMPRAR

Figura 16: Matriz de confusión de RandomTree

Este algoritmo nos permite también visualizar el modelo en una forma gráfica donde comienza con ganancia y de va dividiendo según el valor que toman.

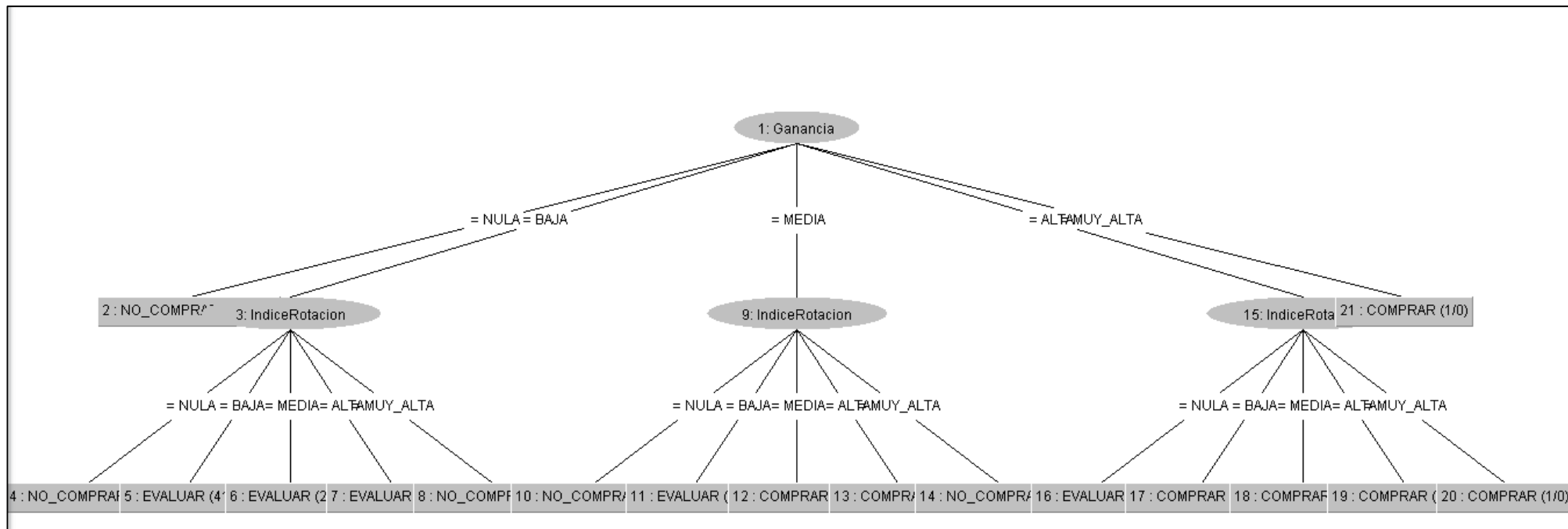


Figura 17: El árbol generado por RandomTree en forma gráfica.

- Un último que probaremos será un clasificador *rules* y el algoritmo JRip pero normalmente conocido como RIPPER. Dejamos las opciones los mismo que los anteriores, teniendo 3360 instancias y 3 atributos. 80% para entrenar y el restante para entrenar.

```

=== Run information ===

Scheme:      weka.classifiers.rules.JRip -F 3 -N 2.0 -O 2 -S 1
Relation:    PRODUCTO-weka.filters.unsupervised.attribute.Remove-R1-2
Instances:   3360
Attributes:  3
              IndiceRotacion
              Ganancia
              PREDICCIÓN_COMPRA
Test mode:   split 80.0% train, remainder test

```

Figura 18: Parámetros en lo que se ejecutó el JRip.

En JRIP rules nos muestra las relaciones que existen, como ejemplo podemos ver que si la ganancia es ALTA y el índice de rotación es BAJA la predicción es que se compraron todos (167) y así los demás. También nos indica que son 11 reglas

```

JRIP rules:
=====

(Ganancia = ALTA) and (IndiceRotacion = BAJA) => PREDICCIÓN_COMPRA=COMPRAR (167.0/0.0)
(Ganancia = ALTA) and (IndiceRotacion = ALTA) => PREDICCIÓN_COMPRA=COMPRAR (107.0/0.0)
(IndiceRotacion = MEDIA) and (Ganancia = ALTA) => PREDICCIÓN_COMPRA=COMPRAR (18.0/0.0)
(Ganancia = MEDIA) and (IndiceRotacion = MEDIA) => PREDICCIÓN_COMPRA=COMPRAR (5.0/0.0)
(IndiceRotacion = ALTA) and (Ganancia = MEDIA) => PREDICCIÓN_COMPRA=COMPRAR (4.0/0.0)
(Ganancia = BAJA) and (IndiceRotacion = BAJA) => PREDICCIÓN_COMPRA=EVALUAR (411.0/0.0)
(Ganancia = BAJA) and (IndiceRotacion = ALTA) => PREDICCIÓN_COMPRA=EVALUAR (40.0/0.0)
(IndiceRotacion = BAJA) => PREDICCIÓN_COMPRA=EVALUAR (43.0/0.0)
(IndiceRotacion = MEDIA) => PREDICCIÓN_COMPRA=EVALUAR (25.0/0.0)
(Ganancia = ALTA) => PREDICCIÓN_COMPRA=EVALUAR (16.0/1.0)
=> PREDICCIÓN_COMPRA=NO_COMPRAR (2524.0/1.0)

Number of Rules : 11

Time taken to build model: 0.2 seconds

=== Evaluation on test split ===

Time taken to test model on test split: 0 seconds

```

Figura 20: Las reglas o relaciones encontradas por JRip

Ahora observaremos que tenemos un 99.55 de asertividad con el algoritmo JRip. Que sería 669 de aciertos y 3 incorrectos. Obtenemos también Error absoluto

medio, error cuadrático medio, error absoluto relativo y error cuadrático medio de la raíz.

```

=== Summary ===
Correctly Classified Instances      669          99.5536 %
Incorrectly Classified Instances     3           0.4464 %
Kappa statistic                    0.9885
Mean absolute error                 0.0052
Root mean squared error             0.0546
Relative absolute error              1.9559 %
Root relative squared error         15.179 %
Total Number of Instances          672

=== Detailed Accuracy By Class ===

```

	TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area	Class
	0.998	0.006	0.998	0.998	0.998	0.992	0.996	0.998	NO_COMPRAR
	1.000	0.004	0.982	1.000	0.991	0.989	0.998	0.975	EVALUAR
	0.962	0.000	1.000	0.962	0.981	0.979	0.981	0.965	COMPRAR
Weighted Avg.	0.996	0.005	0.996	0.996	0.996	0.990	0.995	0.992	

Figura 21: Resumen de JRip.

Y por último la matriz, observamos que en NO_COMPRAR 511 se clasificaron correctamente, pero 1 no, en EVALUAR 107 correctas y 2 no y en COMPRAR 51 correctos.

```

=== Confusion Matrix ===

```

	a	b	c	<-- classified as
511	1	0	0	a = NO_COMPRAR
0	107	0	0	b = EVALUAR
1	1	51	0	c = COMPRAR

Figura 22: Matriz de confusión de JRip.

En anexo 9, la fase de Modelado: Evaluación del modelo podrá visualizar una prueba test.

Verificando si nuestras hipótesis están bien o las descartamos por nueva que se genere.

En nuestra Hipótesis general donde mencionábamos que el modelo de predicción ayuda a determinar que productos comprar para mantener un stock adecuado de artículos de ferretería. Podemos mencionar que, si ya que con el modelo crado al realizar el test que puede verse en anexos, comprobamos que de los veinte registros a precedir todos arrojaron un 100% de asertividad con el modelo J48.

La hipótesis específica 1: Los datos de productos identificados ayudaron a los datos de entrada al modelo de predictivo. Esta hipótesis es correcta ya que los datos identificados ayudaron a entrenar al modelo e identificar cual de todos es el mejor y casi todos arrojaron un 99.85% de correcta clasificación.

La hipótesis específica 2: El algoritmo de clasificación identificado ayudo a la formulación del modelo. Si ya que arrojó una correcta clasificación de 99.85%.

La hipótesis específica 3: Con el procesamiento del 80% se formula el modelo y el 20% se valida el modelo resultante. Si, ya que ayuda al modelo a entrenar y al momento de realizar la prueba con datos este mostro una predicción de 100%.

V. DISCUSIÓN.

Ahora compararemos los diferentes algoritmos usados y veremos cual es la mejor para nuestro modelo que estamos planteando. Para ellos aremos una tabla y compraremos los porcentajes de asertividad de cada algoritmo o modelo

N°	CLASIFICADOR	ALGORITMO	PORCENTAJE DE ASERTIVIDAD	ERROR ABSOLUTO MEDIO
1	Árbol	J48	99.85%	0.0009
2	Árbol	RANDOMFOREST	99.85%	0.001
3	Árbol	RANDOMTREE	99.85%	0.0009
4	Rules	JRIP	99.85%	0.0052

Tabla 2: Comparación de los diferentes algoritmos aplicados.

De esta tabla de comparación que el clasificador árbol: J48 tiene un asertividad de 99.85%, el árbol: RandomForest tiene un asertividad de 99.85%, el árbol RandomTree tiene un 99.85% de asertividad y por último un Rules: JRIP tiene un 99.85% de asertividad.

Rodríguez (2017), en su investigación titulada “Creación de modelo de predictivo para la demanda de clientes de una empresa del rubro de TI. Los resultados obtenidos fueron que el aspecto a resaltar es la gran exactitud del pronóstico que alcanza el modelo, teniendo un error absoluto medio de 0.7115, indicando estos valores que se pronosticaron se diferencias en promedio de dicha cantidad.

En cuanto a lo que menciona de error absoluto medio en nuestro caso tenemos en nuestro modelo J48 y RandomTree el error absoluto medio de 0.0009, en randomForest de 0.001 y JRIP tiene un 0.0052.

El presente estudio muestra la lo que se puede hacer con datos que la empresa puede tener si se realizar un buen proceso de clasificación y manejo, y de esta importancia nos mencionan Moreno y Michel (2019), en su investigación titulada “Análisis de comportamiento de la predicción de ventas con diferentes técnicas de minera de datos”. Donde se concluye la predicción de ventas es algo vital hoy en día y el saber qué rumbo tendrá un negocio puede ser de gran ayuda para los empresarios

Castro y Hernández (2016), en su investigación titulada “Implementación de un modelo de predictivo basado en data mining y soportado por SAP predictive analytics en retails”. Se pudo concluir que la predicción de las ventas juega un papel imprescindible en el cumplimiento del proceso para contar con productos en la empresa, evitando la ausencia o el sobre abastecimiento de stock, esto afín evite tener pérdidas monetarias, contando con una gran cantidad de datos se puede lograr mejorar resultados.

Otra comparativa que tenemos son las matrices de confusiones de cada modelado. Donde nos indica que en forma diagonal la correcta clasificación. Podemos Decir que las predicciones son las verticales y las horizontales son clase real.

J48: La columna ‘a’ que es No comprar indica que 512 instancias se han clasificado correctamente en clase No comprar. La columna ‘b’ 107 instancias de la clase Evaluar están correctamente clasificadas en clase Evaluar y ninguna incorrecta, y la coliman ‘c’ indica que 52 instancias están clasificaciones correctas, pero 1 instancia de la clase Comprar está clasificada incorrectamente como No comprar. Esta indicación se repite en RandomForest y RandomTree.

Nota: La fila *a* indica NO_COMPRAR, la *b* indica EVALULAR y la *c* es COMPRAR

	J48			RandomForest			RandomTree		
	a	b	c	a	b	c	a	b	c
a	512	0	0	512	0	0	512	0	0
b	0	107	0	0	107	0	0	107	0
c	1	0	52	1	0	52	1	0	52

Tabla 3: Comparación de las diferentes matrices de confusión de árbol.

JRip: La columna ‘a’ que es no comprar indica que 511 instancias de la clase No comprar se han clasificado correctamente en clase No comprar, pero hay uno donde no se clasificó correctamente y ha sido asignado a clase Evaluar. La columna ‘b’ 107 instancias de la clase Evaluar han sido asignas correctamente en la clase Evaluar, y la columna ‘c’ indica que 52 instancias de la clase comprar han sido clasificaciones correctas en la clase Comprar, mientras 2 instancias de clase

Comprar han sido incorrectamente asignadas una en No comprar y la otra en evaluar.

Nota: La fila a indica NO_COMPRAR, la b indica EVALULAR y la c es COMPRAR

	JRip		
	a	b	c
a	511	1	0
b	0	107	0
c	1	1	51

Tabla 4: La matriz de confusión de Rules.

Las fortalezas que podemos mencionar respecto a la metodología CRISP-DM es que nos permite comenzar desde el análisis del negocio, nos permite usar varias herramientas y no nos delimita o impone una, también permite revisar los requerimientos, y es una de las más usadas para el análisis de datos, minería de datos, etc.

La investigación nos permite ver un nuevo campo que está tomando mayor fuerza en las organizaciones ya que dependiendo de los datos que maneje una organización esta va poder tomar medidas que mejoren su negocio, pero a partir de los datos que cuenta la organización y que tan seguro los tenga. Teniendo los datos un realizándolos un proceso para transformarlo en información y estos den conocimiento al negocio para tomar decisiones.

En este caso la el negocio necesitaba ver que productos comprar. Para ello se planteó tener en cuenta 2 variables que son el índice de rotación y la ganancia de los productos. A partir de esto comenzar a transformar los datos en información y estos en conocimientos. El negocio no necesitaba tener un producto en su almacén por mucho tiempo solo generando gasto en el mantenimiento, pero luego se analizó y se vio que no necesariamente el producto que salía mucho generaba ganancias altas, ya que puede salir poco, pero genera ganancias altas. Para ellos planteamos un modelo que a partir del índice y la ganancia puede determinar si compramos el producto no se compra.

En lo último menciona es donde podemos mencionar que se ve la relevancia de la investigación, ya que se estaría viendo un tema antiguo y muy poco visto pero que en los últimos años o meses está tomando fuerza.

VI. CONCLUSIONES

1. El modelo Árbol: J48 es el mejor modelo para predecir la compra de productos manteniendo un stock adecuado de artículos de ferretería teniendo un 99.85% de asertividad. Como segunda opción el árbol: RandomTree también teniendo un 99.85% ya que los 2 tienen un 0.0009 de error absoluto medio.
2. Se logro identificar un total de 3360 registros de los productos que generan mayores ganancias y estos faciliten la data de entrada ya que no necesariamente el producto que tiene un alto índice de rotación genera mayor ganancia.
3. El mejor algoritmo de clasificación es el del árbol y uno de rules, que las 4 elegidas tuvieron un porcentaje superior al 99% de asertividad en la clasificación de nuestros datos.
4. Se logro validar el modelo ya que se usó el 80% de los datos para el entrenamiento y el restante para validar el dato, al realizar la predicción el resultado fue que todos arrojaron un 100% de precisión.

VII. RECOMENDACIONES

Una recomendación importante resaltado por el jurado es que en futuras investigaciones estudiar variables que tengan mayor incidencia en la ganancia del negocio y así se mejore el grado de asertividad.

Se recomienda contar con datos reales de la organización o lugar de estudio donde se aplicará la investigación con lo cual se logrará obtener resultados más fieles a datos reales.

Se recomienda realizar este tipo de investigaciones ya que son temas que están tomando mayor relevancia y sin mencionar que los datos son el poder, ya que los datos transformados o procesados se convierten en información y esto en conocimientos permiten que el negocio tome buenas decisiones.

Tratar de revisar el tiempo y ver si los datos que nos proporcionaran serán en físico o ya digital, ya que si es lo segundo debemos tomar un poco más de tiempo en la fase de recolección de datos de la metodología CRISP-DM.

Se recomienda usar WEKA ya que permite usar varios algoritmos de clasificación y agrupamiento, también según los resultados de asertividad en la clasificación compararlos y escoger cual es mejor para nuestra investigación, en caso de querer programación ya identificado nuestro algoritmo o modelo es Python es una buena opción.

REFERENCIAS

CÁCERES, Víctor. Predicción de precios de acciones de bolsa de valores utilizando support vector regression. (tesis de pregrado). Universidad de Lima. Lima. 2018.

Disponible en:
https://repositorio.ulima.edu.pe/bitstream/handle/20.500.12724/6973/C%c3%a1ceres_Chian_V%c3%adctor_Andr%c3%a9s_Edgard.pdf?sequence=1&isAllowed=y

CALI, Silva. Diseño de un modelo de control de inventarios para la empresa ferretería "Andina" en la ciudad de Riobamba, provincia de Chimbrazo. (Tesis de pregrado). Escuela superior politécnica de Chimborazo. Riobamba. 2019.

Disponible en:
<http://dspace.esPOCH.edu.ec/bitstream/123456789/11304/1/22T0504.pdf>

CAPARRÓS, Egea. Redes neuronales: concepto, fundamentos y aplicaciones en el laboratorio clínico. Universidad virgen de la Arrixaca. pp. 221-228. 1994.

Disponible en:
[https://www.seqc.es/download/revista/388/1224/5818427/1024/cms/Qu%C3%ADmica%20I%C3%ADnica%201994;13%20\(5\)%20221-228.pdf/](https://www.seqc.es/download/revista/388/1224/5818427/1024/cms/Qu%C3%ADmica%20I%C3%ADnica%201994;13%20(5)%20221-228.pdf/)

CÁRDENAS, Joaquín. Predicción de potenciales clientes de inmuebles para aitué, basado en datos históricos de sus clientes. (Tesis de pregrado). Universidad de Concepción. Concepción. 2019.

Disponible en:
http://repositorio.udec.cl/bitstream/11594/400/1/Tesis_Prediccion_de_potenciales_clientes.pdf

CARRIZO, Claudio. Modelo predictivo para determinar la graduación de alumnos de carreras de ingeniería aplicando técnicas de minería de datos. Universidad Tecnológica Nacional San Francisco. San Francisco. 2018.

Disponible en:
https://repositorio.ulima.edu.pe/bitstream/handle/20.500.12724/8724/Carrizo_Claudio_Jos%c3%a9.pdf?sequence=1&isAllowed=y

CASTRO, Alexandra y HERNÁNDEZ, Juan. Implementación de un modelo predictivo basado en data mining y soportado por sap predictive analytics en ratails. (Tesis de pregrado), Universidad Peruana de Ciencias Aplicadas (UPC). Lima. 2016.

Disponible en:
<https://repositorioacademico.upc.edu.pe/bitstream/handle/10757/620850/SAPPRE-D-Memoria%20Final%20de%20Proyecto%20v1.7.pdf?sequence=1&isAllowed=y>

DAZA, Alfredo. Data mining (minería de datos). Macro EIRL. Lima, Perú. 2016.
Disponible en: <https://editorialmacro.com/wp-content/uploads/2021/02/9786123044176.pdf>

GALÁN, Víctor. Aplicación de la metodología CRISP-DM a un proyecto de minería de datos en el entorno universitario. Universidad Carlos III de Madrid. 2015.
Disponible en: https://e-archivo.uc3m.es/bitstream/handle/10016/22198/PFC_Victor_Galan_Cortina.pdf

GRÁNDEZ, Miguel. Aplicación de minería de datos para determinar patrones de consumo futuro en clientes de una distribuidora de suplementos nutricionales. (Tesis de pregrado). Universidad San Ignacio de Loyola. Lima. 2017.
Disponible en: http://repositorio.usil.edu.pe/bitstream/USIL/2763/1/2017_Granda_Aplicacion-de-mineria-datos.pdf

INNOVA COMMERCE. ¿Qué es stock? [en línea]. Recuperado desde: <https://www.innova-commerce.com/que-es-stock/>

LOZADA, José. Investigación aplicada: definición, Propiedad intelectual e industria. Vol. 3. pp. 47-50. 2014.
Disponible en: <https://dialnet.unirioja.es/servlet/articulo?codigo=6163749>
ISSN-e 1390-9592

MAISUECHE, Aalberto. Utilización del machine learning en la industria 4.0. Universidad de Valladolid. Valladolid. 2019.
Disponible en: <https://core.ac.uk/download/pdf/228074134.pdf>

MORENO, Teresita y MICHEL, Rosa. Análisis del comportamiento de la predicción de ventas con diferentes técnicas de minería de datos. Maestría ciencia de la computación. Instituto Tecnología de Cd. Guzmán. Guzmán. 2019.
Disponible en: <http://reibci.org/publicados/2019/oct/3600106.pdf>

NEOATTACK. Modelo predictivo. España. Recuperado desde: <https://neoattack.com/neowiki/modelo-predictivo/>

DE LA OLIVA, Lesly. Sistema de pronóstico de compra utilizando un modelo de predicción para mejorar el monitoreo comercial de un supermercado de la ciudad de Chiclayo. Tesis de pregrado, Universidad católica Santo Toribio de Mogrovejo. 2019.
Disponible en: <http://tesis.usat.edu.pe/handle/20.500.12423/2231>

PÁEZ, Gabriel. Stock (estocaje). Economipedia.com. 2020.
Disponible en: <https://economipedia.com/definiciones/stock-estocaje.html>

REDACCIÓN ESPAÑA. Qué es un modelo predictivo y cómo se aplica al negocio. 2020.

Disponible en: <https://agenciab12.com/noticia/que-es-modelo-predictivo-como-aplica-negocio>

RODRÍGUEZ, Ignacio. Creación de modelos predictivos para la demanda de clientes de una empresa del rubro TI. (Tesis de pregrado). Universidad Técnica Federico Santo María. Santiago – Chile. 2017.

Disponible en: <https://repositorio.usm.cl/bitstream/handle/11673/15564/3560902038239UTFSM.pdf?sequence=1&isAllowed=y>

RODRÍGUEZ, Daniela. Investigación aplicada: Características, definición, ejemplos. Lifeder. 2020.

Disponible en: <https://www.lifeder.com/investigacion-aplicada/>

WEKA 3. Weka 3: Machine Learning Software in Java. s.f.

Disponible en: <https://www.cs.waikato.ac.nz/ml/weka/>

ANEXOS

ANEXO 1: DIAGRAMA DE COMPRA DE PRODUCTO ACTUAL

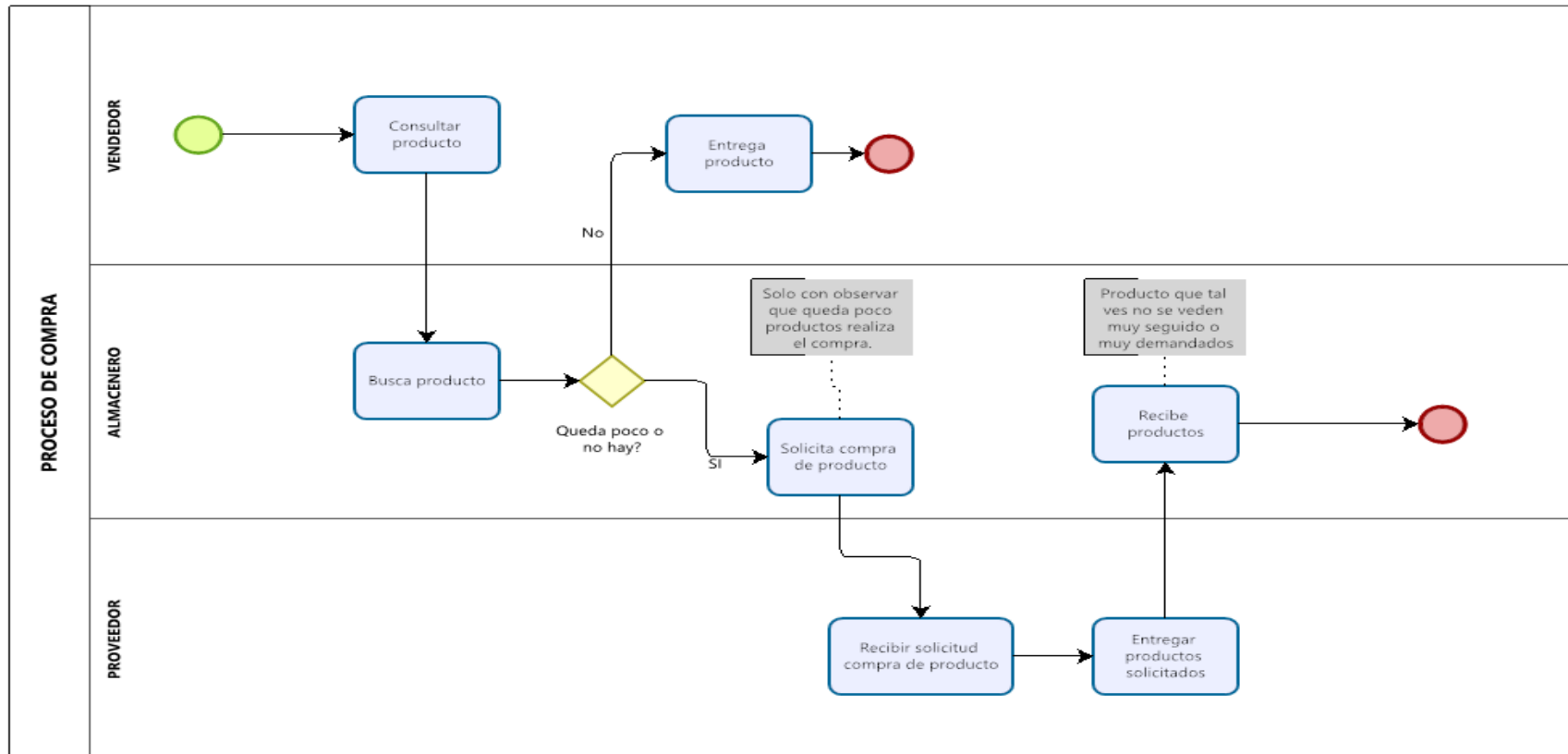


Figura 1: Diagrama de compra de producto

ANEXO 2: ANTECEDENTES INTERNACIONALES.

Carrizo (2018), en su investigación titulada "*Modelo predictivo para determinar la graduación de alumnos de carreras de ingenierías aplicando técnicas de minería de datos*". Tuvo como objetivo construir un modelo predictivo a través de técnicas de minería de datos que permita, determinar la cantidad de alumnos que puedan graduarse en carreras de ingenierías y patrones que puedan incidir en la graduación. Fue un estudio de tipo experimental, la población de estudio son los estudiantes de la universidad UTN, la muestra son los estudiantes de ingeniería y muestreo fue probabilístico; el instrumento de donde se obtendrán son los datos es del Sistema SysAcad. Los resultados muestran que, la matriz de confusión al haber aplicado la técnica del árbol de decisión, en la que se pudo visualizar que el entrenamiento arroja un acierto global de 97.53% de lo que podemos decir que el modelo entregado tiene un rendimiento alto y es eficiente. Continuando, al analizar los resultados arrojados por el operador `apply model` en el modelo predictivo, evidenciamos que la variable que predcimos denominada "Graduado", se observa que el periodo 2012-2018 se graduaron 8 alumnos de carrera de Ingeniería. En último, en el resultado de aplicar el árbol de decisiones observamos patrones relación-nodos en aspectos académicos y aspectos personales. Se concluyo que se puedo construir un modelo satisfactorio, teniendo en cuenta que tuvo un nivel de aceptación en cantidad de aciertos, el cual permitió determinar la graduación de alumnos de carreras de ingeniería a través de la aplicación de la técnica de árbol de decisión.

Cárdenas (2019), en su investigación titulada "*Predicción de potenciales clientes de inmuebles para Aitúé, basado en datos históricos de sus clientes*". Tuvo como objetivo generar un modelo de clasificación binaria de cotizantes a clientes que además estime la probabilidad de que un cotizante se convierta en cliente. Fue un estudio de tipo experimental, la población son los clientes de inmueble, la muestra son los clientes de los 5 proyectos en venta, y el muestreo fue probabilístico; el instrumento usado fue ficha de registro de datos. Los principales resultados fueron 1) Para la tarea 1 y la clasificación de la variable objetivo negocio, los valores obtenidos de la validación cruzada muestran que el rendimiento más alto fue obtenido por XGB con una exactitud de 0.864, seguido de SVM y RF. Desde

la perspectiva de las métricas de f1-score y AUC, el clasificador que obtuvo los puntajes más altos en ambas métricas fue también XGB con un f1-score de 0.652 y un AUC de 0.827 seguido por SVM. Los detalles se pueden ver en la tabla 13, la figura 16 y figura 17. 2) Para la tarea 1 y la clasificación de la variable objetivo “compra”, los resultados muestran que el rendimiento más alto fue obtenido por RF con una exactitud de 0.891, seguido del SVM con un 0.821. Las métricas de f1-score y AUC, el clasificador que obtuvo los puntajes más altos en ambos fue RF con un f1-score de 0.492 y un AUC de 0.926 seguido por SVM con un f1-score de 0.478 y un auc de 0.905 en el test de validación cruzada. Coronando a RF como el mejor clasificador para esta tarea. Los detalles se pueden ver en la tabla 14, la figura 18 y figura 19. Se concluyo que los resultados de los mejores modelos para las tareas son: XGB con una exactitud de 0.86, un F1_score de 0.65 y un Area Under the Curve (AUC) del grafico Reciever Operating Curve de 0.83 para la tarea 1 y v.o. “negocio”, RF con una exactitud de 0.89, un F1_score de 0.5 y un AUC de 0.93 para la tarea 1 y v.o. “compra”, LR con una exactitud de 0.80, un F1_score de 0.6 y un AUC de 0.83 para la tarea 2 y v.o. negocio y, RF con una exactitud de 0.82, un F1_score de 0.3 y un AUC de 0.81 para la tarea 2 y v.o. compra.

Cali (2019), en si investigación titulada “*Diseño de un modelo de control de inventario para la empresa ferretería Andina en la ciudad de Riobamba, provincia de Chimborazo*”. Tuvo como objetivo Diseñar un modelo de control de inventario para la empresa ferretería Andina en la ciudad de Riobamba, provincia de Chimborazo, con la finalidad de optimizar los recursos y controlar los inventarios, evitando perdida de productos. Fue un estudio de tipo descriptiva, la población de estudio fue la empresa Ferretería “Andina”, muestra personal de área contable y los vendedores, y muestreo probabilístico; el instrumento usado fue un cuestionario online dirigida al personal de área contable y entrevista a la propietaria de la empresa. Los principales resultados fueron que el negocio no tiene adecuado control de inventario, no tienen índice de mercancía duplicados o erróneos, se evidencia diferencia entre lo físico en bodega o almacén, con lo que se ve en el sistema contable, no cuenta con la persona que se encarga del control de la mercancía y las operaciones que realizan los trabajadores de la empresa no tienen firmas de responsabilidad. Se concluye que la ferretería Andina no presento una relación entre los productos en físicas y contables, se presentó duplicidad de

códigos (índice), también productos que no se venden, productos se van deteriorando y no se evidencio un modelo de control de inventarios.

ANEXO 3: ANTECEDENTES NACIONALES

Grández (2017), en su investigación titulada “*Aplicación de minería de datos para determinar patrones de consumo futuro de clientes de una distribuidora de suplemento nutricionales*”. Tuvo como objetivo aplicación de minería de datos a través de un software informático para encontrar reglas de consumo que determine patrones de consumo futuro de una distribuidora de suplementos nutricionales con el fin implementar políticas que incrementen el nivel de venta. Fue un estudio de tipo no-experimentar, la población son clientes de la distribuidora de suplementos nutricionales-local Miraflores, la muestra son 611 clientes que realizaron compra entre enero del 2016 a julio del 2016, y el muestro fue probabilístico; el instrumento fue ficha de registro de datos. Los resultados fueron el total de nuestros clientes de entre el 25-32 años que realizan activadas físicas Pilates van comprar el producto Harbinger Fitness correa big grip pro lifting straps, y el total de los clientes que realizan Kickboxing y tienen menos de 2 hijo compren el producto Solgar glucosamine chondroitin complex. El total probable que los clientes que tienen 3 o más hijos y pesan entre 80,4 a 91.5 compren el producto Syntrax néctar medical. Se concluyo que Al realizar el data mining, es sumamente importante la metodología que elijamos, ya que el proceso de análisis de la información y el mismo procesamiento son importantes, tener en cuenta una metodología que aplica la difusión, que cuente con respaldo de instituciones lideres el mercado, pero no solo es la metodología sino la herramienta precisa la cual debemos escoger, esta debe contar con documentación, recursos en línea, foros activos permitiendo y facilitando el intercambio de ideas.

Cáceres (2018), en su investigación titulada “*Predicción de precios de acciones de bolsa de valores utilizando support vector regresión*”. Tuvo como objetivo implementar y evaluar la factibilidad de un modelo de Support Vector Machine para la predicción de precios de acciones a mediano plazo (un trimestre) de la bolsa de valores de S&P. El tipo de estudio fue experimental, la población de estudio son las empresas de la bolsa de valores de NASDAQ o NYSE, la muestra son 40 empresa en el índice de Standard & Poor 500, y el muestreo es probabilístico. Se recogió los datos históricos de estado financiero y cotizaciones de acciones de los años 2011 a 2015. Los principales resultados mencionan que

las 320 acciones de datos de entrenamiento tuvieron un rendimiento de 3.994753%. Los esquemas de predicción mostrados en la tabla permiten elegir las acciones que tuvieron mejor rendimiento después de aplicar las predicciones. Support Vector Machine, cuando clasificamos las acciones, al prever la dirección de los precios, no se diferenciará si el incremento es encima o debajo del mercado. Por ello, no podemos diferenciar que acciones seleccionadas por el modelo presentaran un rendimiento superior al promedio del mercado. En consecuencia, no se sabe que acciones se escogieron de todas sugeridas, ya que logran incrementar, pero no en línea a las expectativas. Se concluye que los valores obtenidos del modelo Support Vector Regression se evalúan aplicando indicadores estadísticos como DA, MAE, MAPE y MSE. Los 4 elegidos, el primero de análisis técnico, el segundo de análisis fundamental, el tercero de combinado de las 2 metodologías y el último de un subgrupo de indicadores de último grupo, tienen resultados similares entre sí. Muestran resultados promedio en dirección correcta de 60%, error que va alrededor de 5 dólares, error porcentual de 10% y por último error promedio cuadrado de 120 dólares.

ANEXO 4: DESCRIPCIÓN DE LAS FASES DE CRISP-DM

Esta metodología igual que otras consta de fases o procesos y en este caso son 6. Rodríguez (2010) referenciado por Galán los menciona:

Comprensión del negocio: “En esta etapa es una de las más importantes, comprende un conjunto de actividades con el fin de entender los objetivos y requisitos del proyecto, pero siempre teniendo en cuenta el negocio, a fin de convertir en objetivos técnicos y en la elaboración de plan de proyecto. En consecuencia de aprovechar la minería de datos, es muy importante entender de manera clara y precisa el problema que vamos a resolver, esto ayuda a recolectar datos correctos y su interpretación de los resultados”. (Rodríguez, 2010)

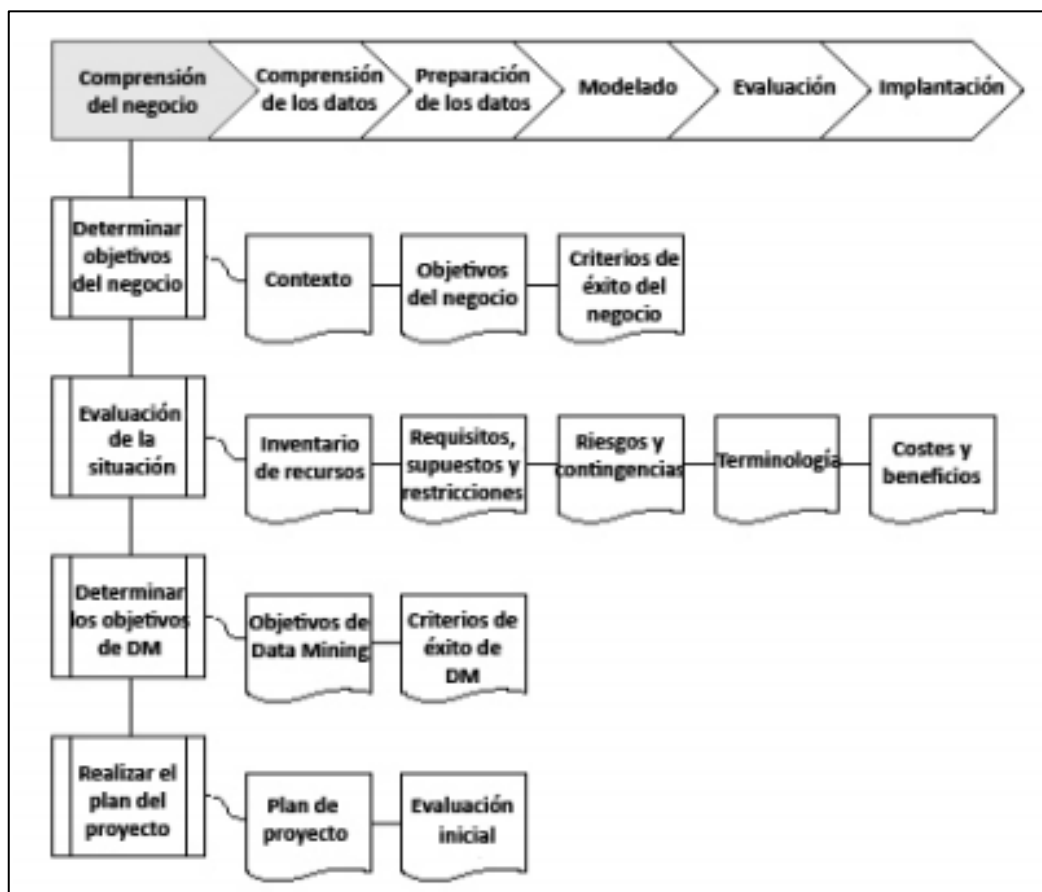


Imagen 1: Fase de comprensión de negocio.

Comprensión de los datos: “Recogemos los datos iniciales a fin de conocer y familiarizarse con ellos, identificar su calidad, establecer el vínculo al cual permite definir las primeras hipótesis. Las que más demanda esfuerzo y tiempo en un proyecto de minería de datos requieren son las 2, 3 y 4 fase de esta metodología.

En la mayoría de los casos en que la organización cuenta con una BD corporativa, se recomienda crear una nueva BD específica para el proyecto, ya que se puede producir acceso a la BD con el fin de realizar consultas y modificaciones, esto generando problemas”. (Rodríguez, 2010). Vamos las tareas en la imagen 2 de la fase:

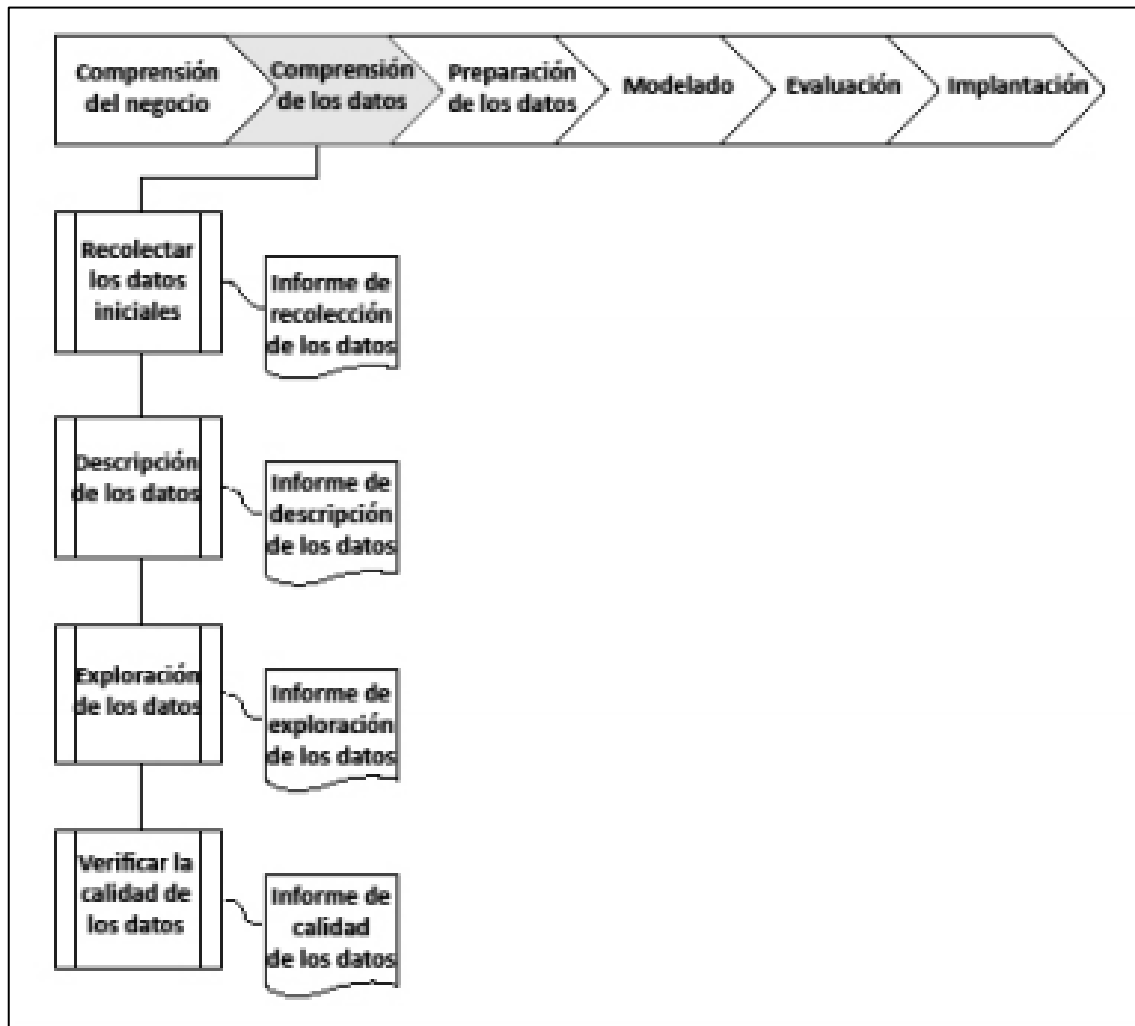


Imagen 2: Fase de comprensión de datos y tareas.

Preparación de datos: “Después de ya haber recogido los datos iniciales en la anterior fase, procedemos a prepararlos a fin de aplicar técnicas de minería de datos, estas pueden ser de visualización de datos, búsqueda de relación entre variables u otras medidas de explotación de los datos. Cuando hablamos de prepararlos nos referimos a actividades generales de selección de datos a los cuales aplicaremos una determinada técnica de modelado, la limpieza de datos, la generación de variables adicionales, integración de diferentes orígenes de datos y los

cambios de formato” (Rodríguez, 2010). En la Imagen 3 se podrá ver las tareas de esta fase.

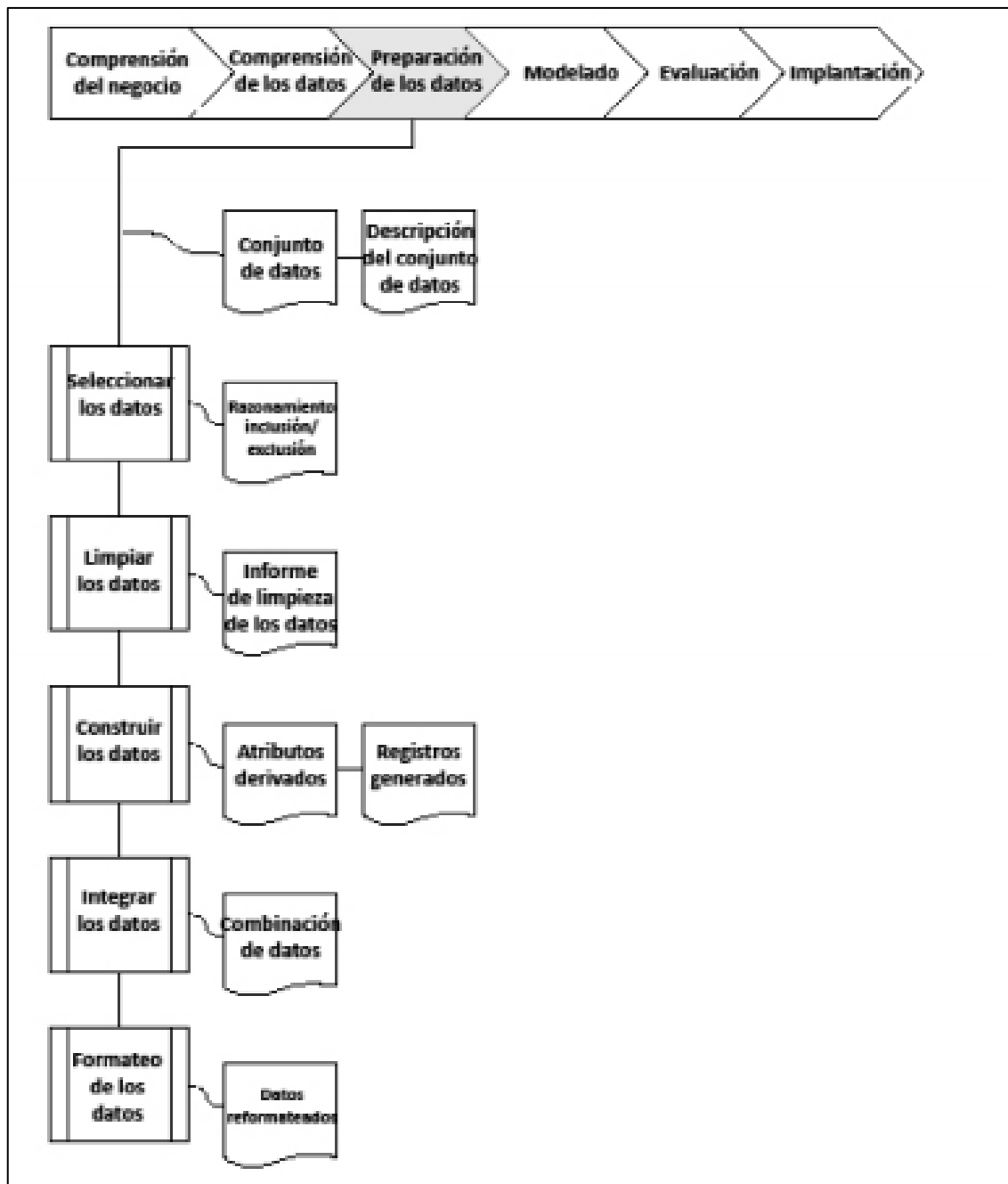


Imagen 3: Fase de preparación de los datos y tareas.

Modelamiento: “Aquí es donde escogeremos la técnica de modelado apropiado para el proyecto. La técnica a usar va en función a los siguientes criterios: Tiene que ser adecuado para el problema, disponer de los datos apropiados, cumple los

requisitos del problema, el tiempo para obtener el modelo y conocer la técnica”. (Rodríguez, 2010). La Imagen 4 muestra las tareas y salidas que se obtienen.

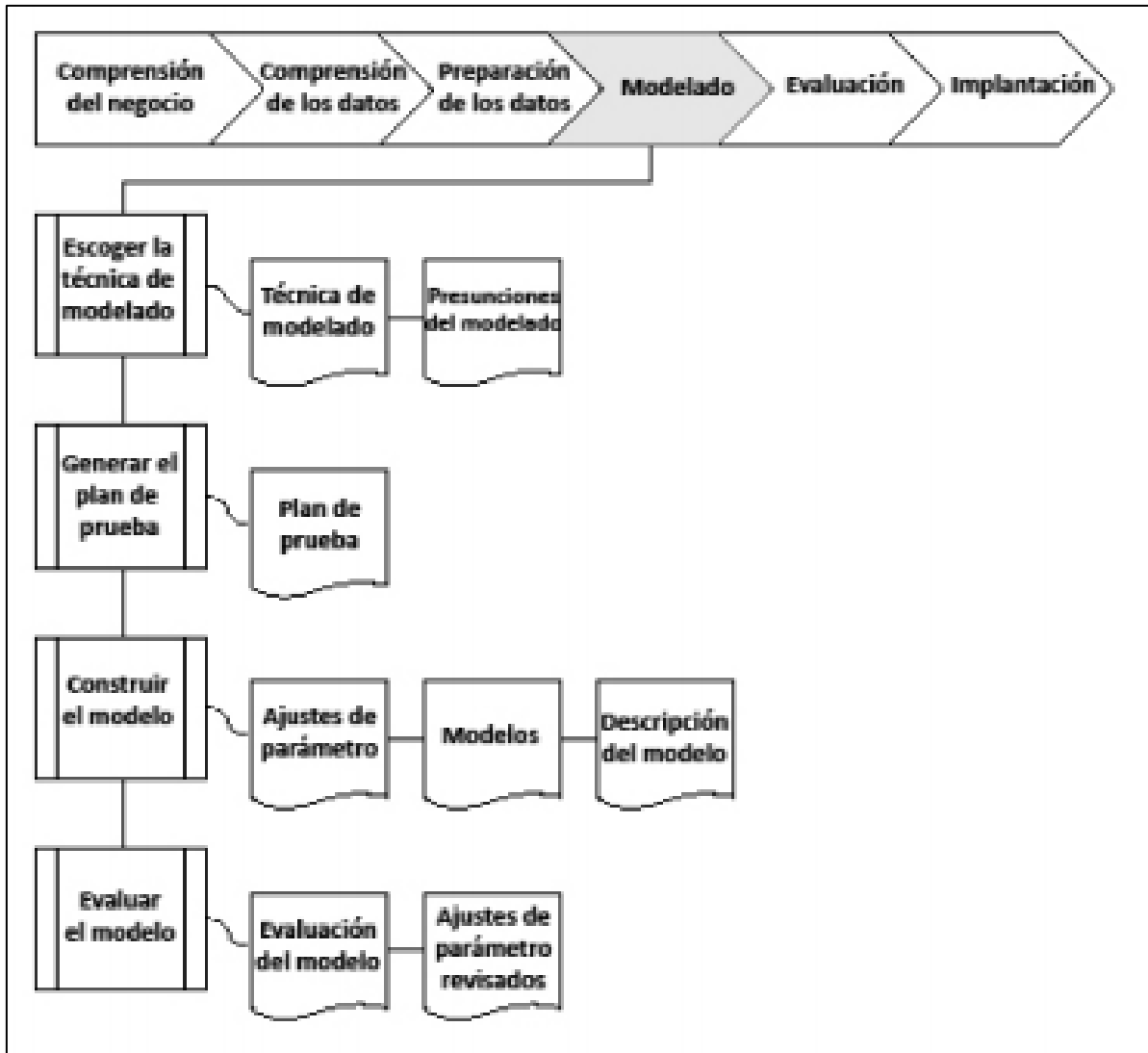


Imagen 4: Fase de modelado y tareas.

Evaluación: “Veamos el cumplimiento de los criterios de éxito del problema a fin de evaluar el modelo. Otro factor en tener en cuenta es la finalidad calculada del modelo, esto solo aplicaremos a los datos que se realizaron el análisis. Es necesario revisar el desarrollo, partiendo de los resultados obtenidos, y así, si es el caso repetir la anterior actividad, en caso de haber cometido error. Cuando el modelo es válido procedemos con la explotación del modelo” (Rodríguez, 2010). En la imagen 5 observaremos las tareas que componen esta fase y los resultados que deben obtener.

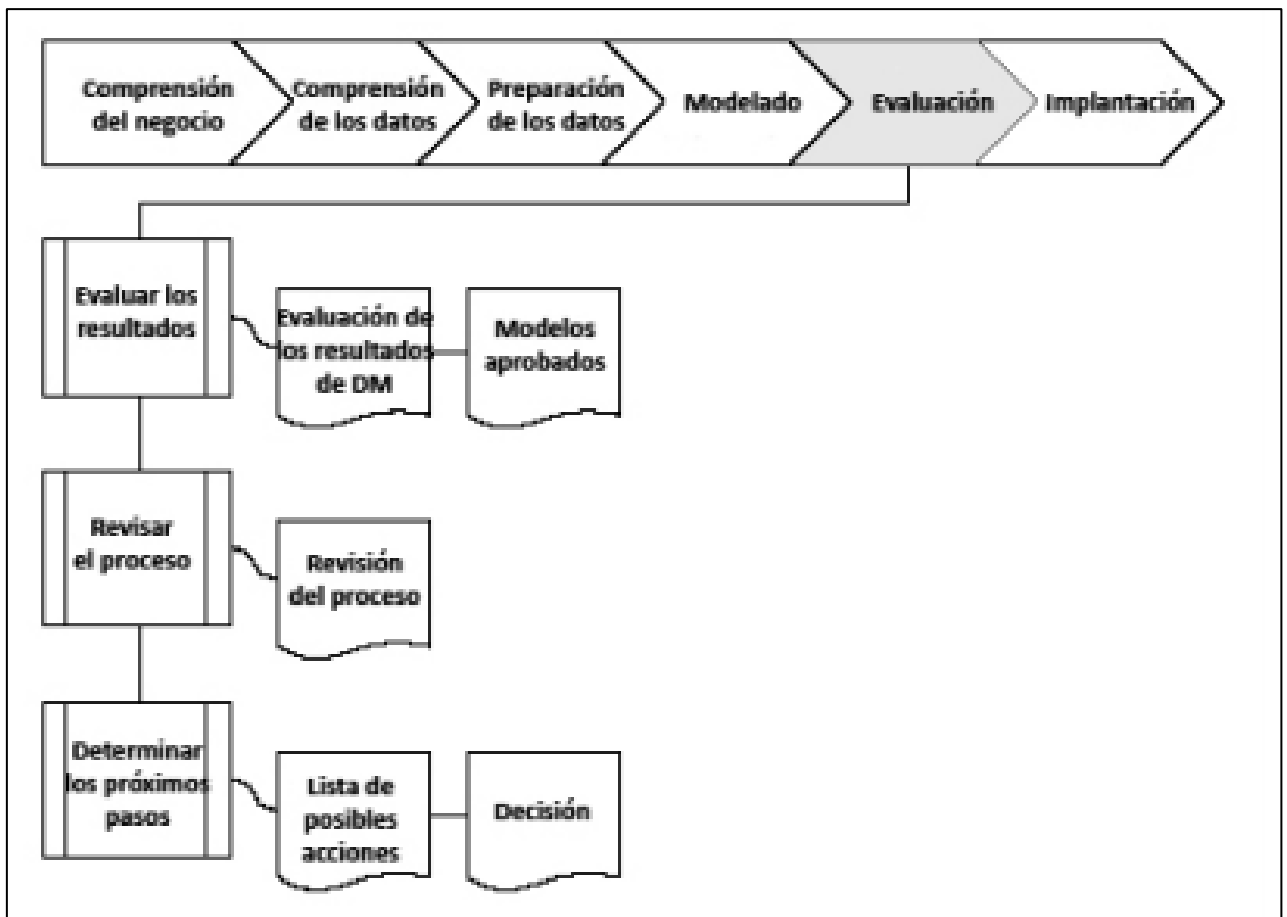


Imagen 5: Fase de evaluación y sus tareas.

Despliegue o implantación: “Cuando el modelo está hecho y validado, debemos aplicar la inteligencia obtenida en acciones dentro del proceso de la empresa, esto puede hacerse, por ejemplo, el análisis recomienda acciones en base a la exploración del modelo y sus resultados. Por lo regular el proyecto de data mining no termina cuando se implanta el modelo, ya que debemos redactar y presentar los resultados de manera clara para el usuario entienda y lograr incrementar el conocimiento”. (Rodríguez, 2010). En la Imagen 6 se muestran las tareas que componen la fase:

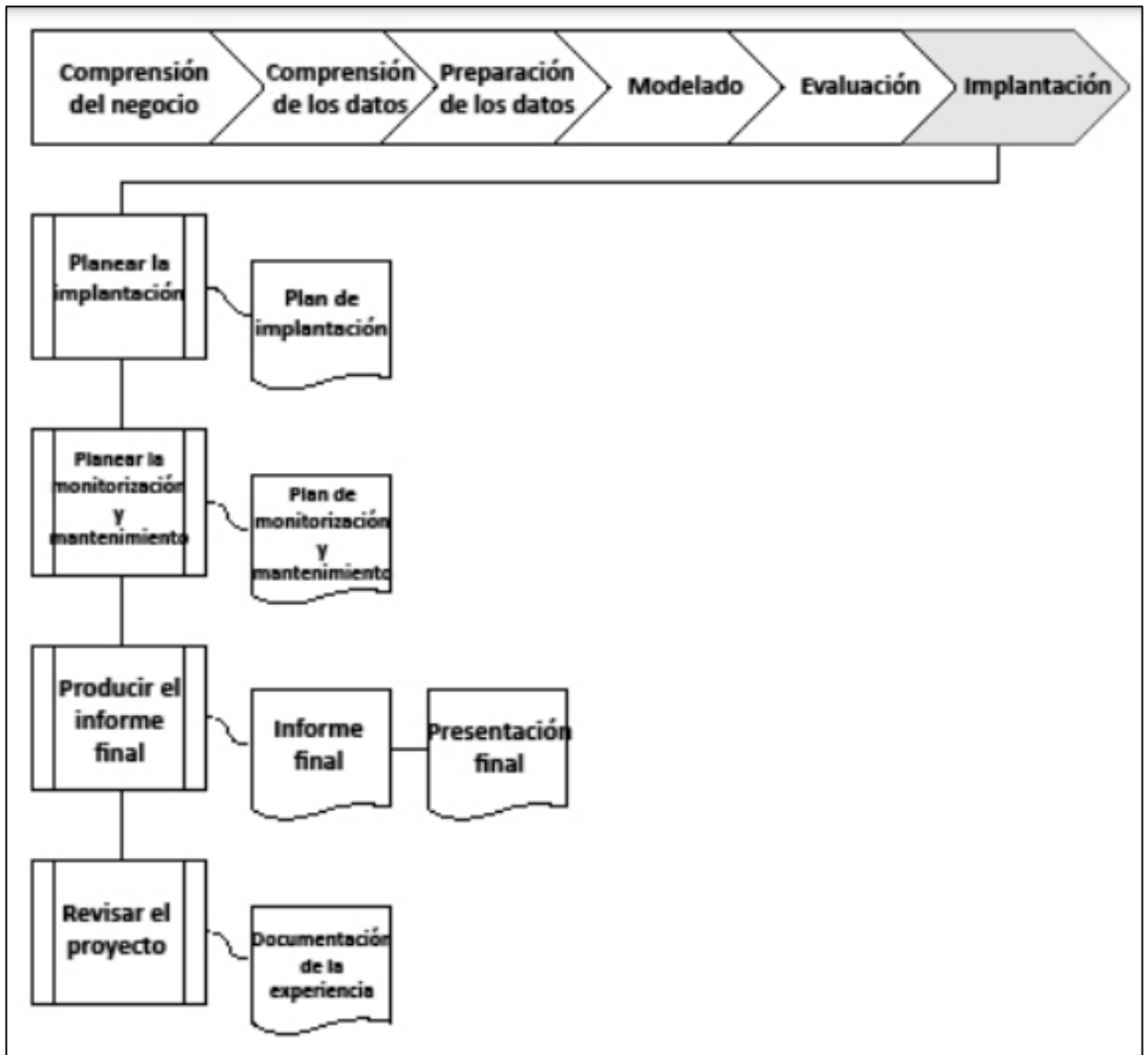


Figura 6: Fases de implementación y tareas.

ANEXO 5: MATRIZ DE OPERACIONALIZACIÓN DE VARIABLES

VARIABLES	DEFINICIÓN CONCEPTUAL	DEFINICIÓN OPERACIONAL	DIMENSIONES	INDICADORES	INSTRUMENTO	FÓRMULA	ESCALA DE MEDICIÓN
Modelo de predicción (VI)	Espino (2017) menciona que “Son modelos de la relación entre el rendimiento específico de una unidad en una muestra y uno o más atributos o características conocido de la unidad” (p.13)	Modelo que estima valores adecuados para regular un proceso de compra.	Data de entrada Algoritmo de clasificación Reporte de resultado		Plataforma Software WEKA		
Stock adecuado (VD)	Páez (2020) menciona que stock “es el conjunto de mercancías almacenadas por una empresa” (par. 1) Carro y González “Se crea cuando el volumen de materiales, partes o bienes terminados que se recibe es mayor que el volumen de los	Cantidad de artículos que se mantienen en almacén garantizando un abastecimiento que posibilite rentabilidad del negocio.	Rotación de productos.	Índice de rotación de stock.	Registro digital	$IR = \frac{VT}{SP}$ IR=Índice Rotación VT= Ventas totales del periodo (mensual) SP=Stock Inicial (mensual)	Razón

	mismos que se distribuye. El stock se agota cuando la distribución es mayor que la recepción de materiales.” (p. 1)		Rentabilidad.	Ganancia de cada producto. (productos que generan mayor ganancia)		$G = (V - CCP) * C$ <p>G=Ganancia V= Costo de venta. CCP= costo de compra del producto. C=Cantidad de venta (mensual)</p>	Razón
--	---	--	---------------	---	--	--	-------

ANEXO 6: MATRIZ DE CONSISTENCIA.

PROBLEMA	OBJETIVOS	HIPÓTESIS	VARIABLES E INDICADORES	METODOLOGÍA
<ul style="list-style-type: none"> • Problema General. • ¿Cómo determinar la compra de productos manteniendo un stock adecuado de los artículos de ferretería? • Problema específico • ¿Cómo identificar los productos con mayor frecuencia de ventas que ayuden a formular el modelo predictivo? • ¿Qué algoritmo de clasificación aplica mejor para formular el modelo de predicción? 	<ul style="list-style-type: none"> • Objetivo general. Desarrollar un modelo de predicción para determinar la compra de productos manteniendo un stock adecuado de artículos de ferretería. • Objetivos Específico. • Identificar datos de productos con mayor rotación de stock y mayor ganancia que faciliten la data de entrada al modelo predictivo. • Determinar el mejor algoritmo de clasificación para la formulación del modelo. • Aplicar un 80% para procesar data y un 20% como 	<ul style="list-style-type: none"> • Hipótesis general El modelo de predicción ayuda a determinar que productos comprar para mantener un stock adecuado de artículos de ferretería. • Hipótesis específico • Los datos de productos identificados ayudaron a los datos de entrada al modelo de predictivo. • El algoritmo de clasificación identificado ayudo a la formulación del modelo • Con el procesamiento del 80% se formula el modelo y 	<p>VI: Modelo de predicción</p> <p>VD: Stock adecuado</p> <p>Índice de rotación de stock.</p> <p>Productos que generan mayor ganancia</p>	<p>Tipo de investigación: aplicada.</p> <p>Diseño de investigación: Experimental.</p> <p>Población. Los 298 productos que se vendieron entre los años del 2020-2021.</p> <p>Muestra: Los 168 productos que se vendieron entre los años 2020 al agosto del 2021.</p>

<ul style="list-style-type: none">• ¿Cómo validar el resultado, el producto de haber aplicación del modelo predictivo?	elemento de control para validar los resultados del modelo.	el 20% se valida el modelo resultante.		Muestreo: Estratificado. Unidad de análisis: productos con mayor demanda.
--	---	--	--	--

ANEXO 07: CARTA DE ACEPTACIÓN.

CARTA DE PRESENTACIÓN

Fecha: 28 / 06 / 2021

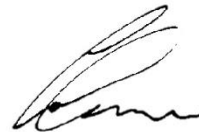
JHONATAN MAYTA.

Tenga mi cordial saludo, Yo **Max Ignacio Flores Zevallos** con DNI **72512652** estudiante de la carrera de Ingeniería de Sistemas del IX ciclo de la Universidad César Vallejo.

Le hago presente esta carta de presentación, como una muestra de compromiso para poder colaborar en el proyecto titulado "**Modelo de predicción para determinar la compra de productos manteniendo un adecuado stock**". El cual buscara crear un modelo de predicción que proporcione pronósticos sobre el stock de artículos de ferretería, como la cantidad de productos necesitaría comprar. **Todos los datos serán usados con fines académicos.**

Gracias por su atención.

Max Flores Zevallos



Jhonatan Omar Mayta Solorzano
(Ferretería Mayta)

ANEXO 9: DESARROLLO DEL MODELO APLICANDO LA METODOLIA CRISP-DM

1. COMPRENSIÓN DEL NEGOCIO

En esta etapa trataremos de conocer el negocio y si esto esta alineado con lo que vamos hacer.

1.1. Objetivos del negocio.

La ferrería Mayta es una empresa encargada de la compra, venta y distribución de materiales y herramientas de construcción, carpintería y afines. Teniendo en cuenta la compra de estos materiales y herramientas el negocio necesita tener un buen plan de compra y no comprar producto que no se estarán vendiendo o comprar en menos cantidad algunos productos.

El objetivo final contar con productos que se vendan y generen mayor ganancia, tener un buen gasto en la compra de los productos que satisfagan a sus clientes.

1.2. Valoración de la situación actual del negocio

EL negocio cuenta no cuanta con un registro de los productos, cuando, cantidad, fecha, etc., que necesitarías para la investigación, aun no cuenta con un registro digital, solo físico, para ellos tendremos que realizar la recolección de los datos en un Excel para su manejo más fácil. En cuando a la cantidad de los datos contamos con las boletas de los años 2020- setiembre de 2021.

1.3. Requisitos, supuestos y restricciones.

Los datos a usar son desde el año 2020 a setiembre del 2021, y esto contengan criterios como, el nombre del producto, fecha, cantidad, y los datos basicos que se iran viendo los datos mas relevantes y lo que no.

1.4. Objetivos de minería de datos.

El objetivo es encontrar un modelo que sea capaz de predecir la compra de productos a partir de las variables rotación, ganancia.

Criterio de éxito: se considerará exitoso el modelo si es más del 70%-80% confiable o exacto.

1.5. Plan de proyecto.

Para la elaboración de la investigación se plantean las siguientes etapas generales:

- 1) Recolección de los datos desde las boletas físicas.
- 2) Explorar y verificar la calidad de los datos.
- 3) Preparar los datos para el análisis, lo que implica eliminar campos básicos o datos duplicados.
- 4) Modelado, lo que sería el uso de herramientas para análisis encontrando relaciones y patrones.
- 5) Evaluación de los resultados, comparando los resultados con los datos reales para ver si funciona.
- 6) Implementación. Se estarán dando nuevos consejos en base a los resultados del análisis, de igual forma servirá como base para nuevos proyectos de minería de datos u otros de IT.

2. COMPRENSIÓN DE LOS DATOS

2.1. Recolección de los datos iniciales

Los datos que vamos a usar son datos de las boletas de ventas que nos proporcionó el negocio, de estas para su mejor manejo y tratamiento debemos pasarlo a un formato digital que en este caso será en un Excel. Los campos encontrados en las boletas son Id, el cliente, teléfono, fecha, dirección, cantidad, etc.

Por otro lado, los datos de compra, en este caso no se lograron conseguir, por ello en este caso vamos a generar datos aleatorios, a partir de la cantidad de productos que nos salga de la boleta de ventas.

Lista de datos adquiridos de las boletas:

- **Numero de boleta**

Es el que identifica cada boleta de venta

- **Nombre Cliente**

Nombre del cliente que ha adquirido el producto o productos

- **Teléfono**

El número de teléfono del cliente

- **Fecha**

La fecha de la compra realizada por el cliente

- **Dirección**

La dirección de entrega del producto

- **Cantidad**

Que cantidad de un producto se está llevando según su unidad de medida

- **Descripción (producto)**

Es el nombre del producto

- **Precio por producto**

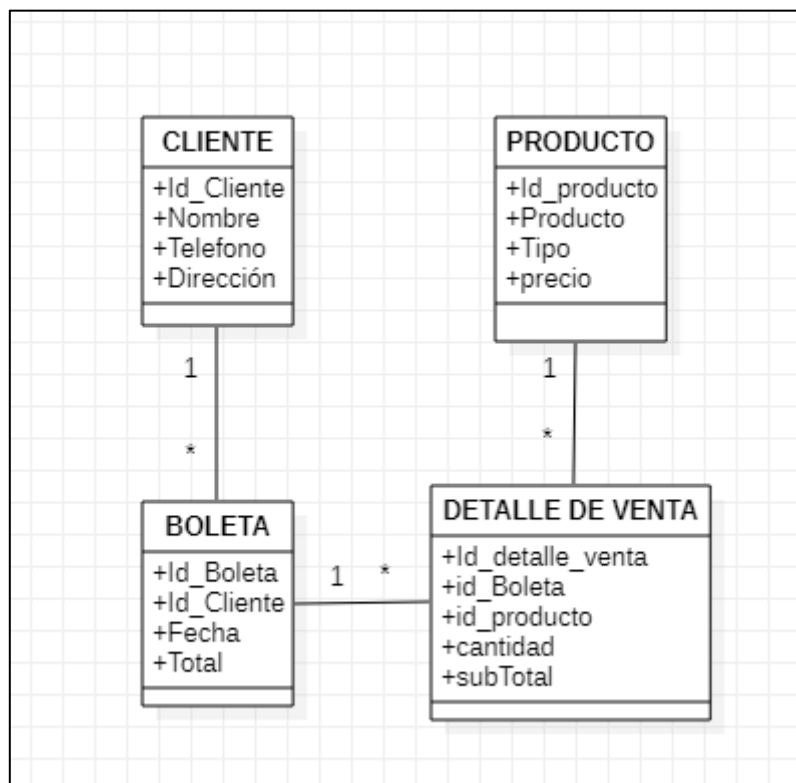
Cuánto cuesta cada producto por unidad o según su unidad.

Para la elaboración de la minería de datos tendremos que realizar algunos cálculos y limpiezas de los datos de lo que saldrán.

- Producto
- Índice de rotación
- Ganancia
- Stock (lo que va a predecir)

2.2. Descripción de los datos

Los datos recogidos si los representamos en una forma de una base de datos sería parecido de la figura, para generar esta figura se ha usado la herramienta *StarUML* que sirve para generar diversos modelos UML. En la figura siguiente se ve la Propuesta de Diagrama clase proceso de venta



En esta figura podemos observar que las ventas constan de 4 tablas básicas: CLIENTE, BOLETA, DETALLE VENTA y PRODUCTO.

Una modificación a adicional que vamos hacer es crear un único *Numero de boleta*, cada boleta consta de 100 hojas. A los número repetidos se les agregara un extensión “.*[numero]*”

N° BOLETAS	NÚMERO DE BOLETAS	MESES
1	401-500	Noviembre, diciembre 2019 y enero 2020
2	1701 – 1800	Enero, febrero, marzo 2020
3	251.2 – 350.2	Agosto 2021
4	201.3 – 250.3	Agosto 2021
5	451.2 – 500.2	Agosto, Setiembre 2021
6	1 – 100	Marzo, mayo, junio 2020
7	301.1 – 400.1	Junio, julio 2020
8	101 – 200	Julio, agosto 2020
9	401.1 – 500.1	Agosto 2020
10	901 – 1000	Agosto, Setiembre 2020
11	301.3 – 400.3	Setiembre, octubre 2020
12	201 – 300	Octubre 2020
13	1.1 – 100.1	Octubre, noviembre 2020
14	801 – 900	Noviembre 2020
15	101.1 – 200.1	Noviembre, diciembre 2020
16	701 – 800	Diciembre 2020, enero 2021
17	201.4 – 300.4	Enero 2021
18	301 - 400	Enero, febrero 2021
19	401.2 – 500.2	Febrero, marzo 2021
20	501 – 600	Marzo 2021
21	601 – 700	Marzo, abril 2021
22	4701 – 4800	Abril, mayo 2021
23	4901 – 5000	Mayo, junio 2021
24	5001 – 5100	Junio, julio 2021
25	4501 – 4600	Julio, agosto 2021

Tabla de Boletas

Ante esto para que sea único estaremos agregando un numero al final para diferencial del resto que se repite, y una de las formas para poder agregar algo a un texto ya lleno es usar la función `=CONCATENAR([selección del texto];[entre comillas lo que queremos agregar])`.

A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P
	344	Cachito		18/08/2021		1	m3	Arena Grues	S/ 45.00	45	cancelado				
	344	Cachito		18/08/2021		1	m3	Piedra Chan	S/ 55.00	55	cancelado				
	346			18/08/2021		10		Varillas 8mr	S/ 16.80	168	cancelado				
	346			18/08/2021		3		Varillas 1/2	S/ 41.50	124.5	cancelado				
	347			18/08/2021		4		Eternit 1x3n	S/ 46.00	184	cancelado				
	201	Dori Gonzale	924340944	26/07/2021		1		Tanque Robplas		1020		cancelar		201.1	
	207	Vilam		4/08/2021		25	bolsa	Cemento SC	S/ 23.00	575	cancelado			207.1	
	210	Pedro - Sanc	901198931	10/08/2021		32		Varillas 5/8	S/ 62.00	1984		x cancelar		210.1	
	210	Pedro - Sanc	901198931	10/08/2021		40		Varillas 1/2	S/ 39.90	1598		x cancelar		210.1	
	210	Pedro - Sanc	901198931	10/08/2021		30		Varillas 8mr	S/ 16.80	504		x cancelar		210.1	
	210	Pedro - Sanc	901198931	10/08/2021		15	kilo	Alambre #16	S/ 6.50	97.5		x cancelar		210.1	
	210	Pedro - Sanc	901198931	10/08/2021		1		Bolsa de Agua		170		x cancelar		210.1	
	212	Madina Huarcayo Pepe		19/08/2021	Mz 71 Lt 27 -	6		Varillas 5/8	S/ 61.80	370.8	cancelado			212.1	
	212	Madina Huarcayo Pepe		19/08/2021	Mz 71 Lt 27 -	6		Varillas 3/8	S/ 21.80	130.8	cancelado			212.1	
	212	Madina Huarcayo Pepe		19/08/2021	Mz 71 Lt 27 -	12		Varillas 12n	S/ 36.80	441.6	cancelado			212.1	
	212	Madina Huarcayo Pepe		19/08/2021	Mz 71 Lt 27 -	5		Varillas 8mr	S/ 16.80	84	cancelado			212.1	
	212	Madina Huarcayo Pepe		19/08/2021	Mz 71 Lt 27 -	12	m3	Ormigón	S/ 45.00	500	cancelado			212.1	
	212	Madina Huarcayo Pepe		19/08/2021	Mz 71 Lt 27 -	20	bolsa	Cemento Ar	S/ 24.50	490	cancelado				
	212	Madina Huarcayo Pepe		19/08/2021	Mz 71 Lt 27 -	1	millar	18 huecos		850	cancelado				
	213	Ayala		19/08/2021		10	bolsa	Cemento APU		184.5					
	215	Ayala				7	bolsa	Cemento Ar	S/ 24.00	168	cancelado				
	221			20/08/2021		3		Varillas 12n	S/ 38.00	114	cancelado				

Figura: Generando Id única para cada boleta.

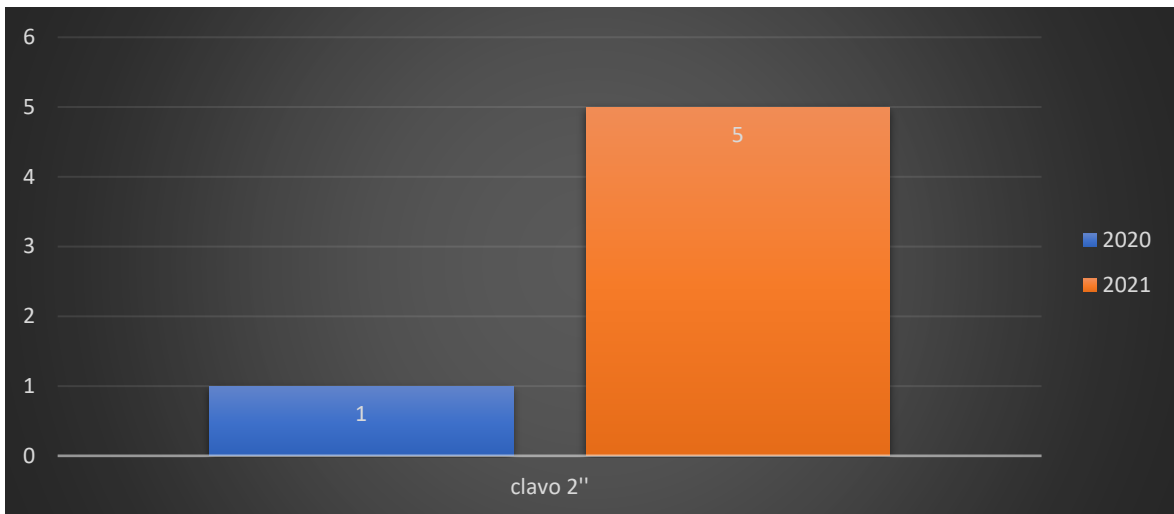
2.3. Explotación de los datos

Es esta fase, ya después de haber obtenido los datos de la base de datos, pero en nuestro caso, registrado los datos físicos a una hoja Excel, vamos examinar los datos encontrados para conocer mejor, y así ver qué tipo de valores y relaciones encontramos.

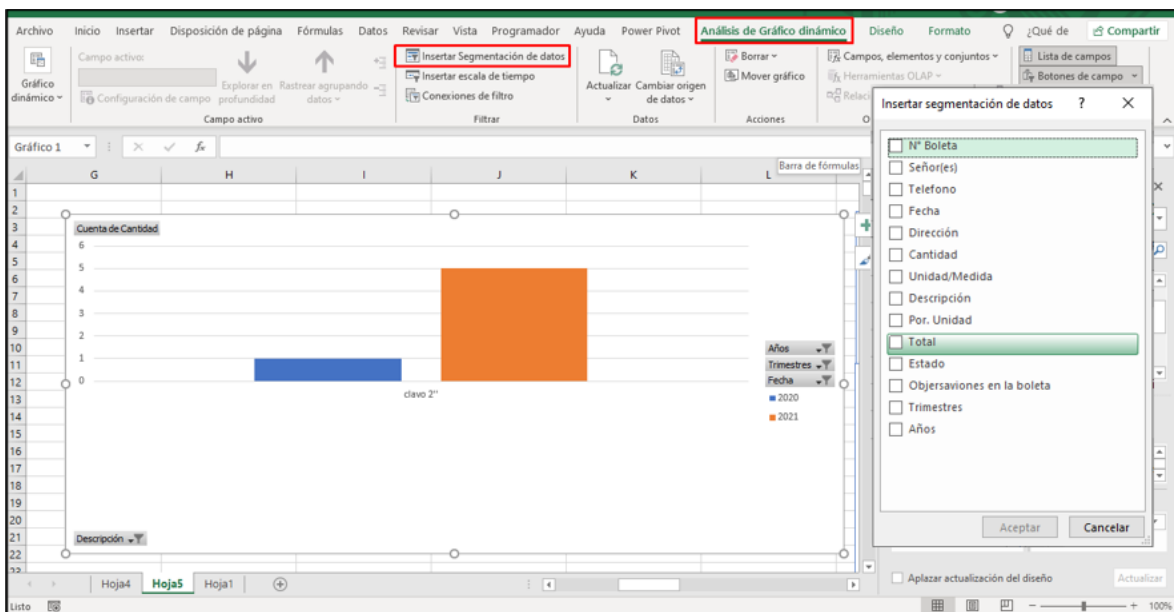
Primero veremos la cantada de productos, según cada producto, para ello estemos aplicando el *Grafico Ginámico*, y también se puede aplicar *PivotTable*

	N° Boleta	Señor(es)	Telefono	Fecha	Dirección	Cantidad	Unidad/Medida	Descripción	Por. Unid.	Total	Estado	Observavi
1	401			10/11/2019		6	metros cubicos (m3)	Hormigón	S/ 45.00	S/ 270.00		
2	401			10/11/2019		35	unidad	Bis. Cementi	S/ 19.00	S/ 665.00		
3	401			10/11/2019		10	unidad	Mangera Niv	S/ 1.50	S/ 15.00		
4	402		928318085	12/11/2019	ca. Sanchez c	5	unidad	cemento	S/ 20.00	S/ 100.00		
5	402		928318085	12/11/2019	ca. Sanchez c	1	metros cubicos (m3)	Arena fina	S/ 45.00	S/ 45.00		
6	402		928318085	12/11/2019	ca. Sanchez c	1	metros cubicos (m3)	Arena Grues	S/ 45.00	S/ 45.00		
7	403	Francisco Rexis		21/11/2019		1	millar	Millar de ad	S/ 250.00	S/ 250.00		
8	403	Francisco Rexis		21/11/2019		1	millar	millar kk rus	S/ 450.00	S/ 450.00		
9	403	Francisco Rexis		21/11/2019		1	millar	millar kk 18 l	S/ 530.00	S/ 530.00		
10	403	Francisco Rexis		21/11/2019				Bolsa de cen	S/ 23.00			
11	403	Francisco Rexis		21/11/2019		12	metros cubicos (m3)	arena gruesa	S/ 45.00	S/ 540.00		
12	403	Francisco Rexis		21/11/2019		12	metros cubicos (m3)	m3 ormigón	S/ 45.00	S/ 540.00		
13	403	Francisco Rexis		21/11/2019				Varillas de 1,	S/ 27.00			
14	403	Francisco Rexis		21/11/2019				Varillas de 5,	S/ 43.00			
15	405			22/11/2019		1		Tube 6" nara	S/ 78.00	S/ 78.00		
16	405			22/11/2019		1		tubo 1/2 p/s	S/ 12.00	S/ 12.00		
17	406	Victor Sigueñas		26/11/2019		12		Pernos con f	S/ 1.00	S/ 12.00		
18	407	Luis Tocas	981355678	27/11/2019		1	metros cubicos (m3)	arena gruesa	S/ 45.00	S/ 45.00		
19	408	Teodoro		27/11/2019		20	metros cubicos (m3)	Mangera Nivel		S/ 24.00		
20	409					1		llave N° 11	S/ 3.50	S/ 3.50		
21	410			28/11/2019		2		Galones de t	S/ 15.00	S/ 30.00		

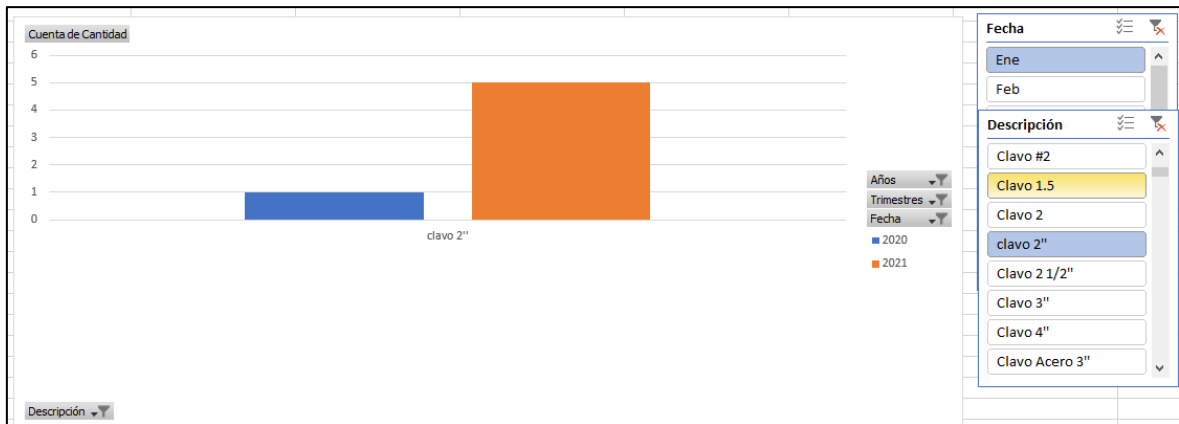
Ahora observaremos la cantidad de productos que se vendieron entre los años 2020 a setiembre de 2021.



Otra forma de explorar dentro de Excel es ingresando los filtros, seleccionamos el grafico y nos saldrá en herramientas > análisis de grafico dinámico > Insertar segmentación de datos



Después podemos combinar los diferentes filtros que queremos, como por ejemplo podemos ver en la imagen producto Clavo de 2'' en el mes de enero, existe un crecimiento en ventas de este producto en comparación del 2021 con el 2020.



2.4. Verificación de la calidad de datos

En esta parte es donde estaremos detallando como esta nuestra base de datos, que pudimos observar mientras explorábamos los datos.

Ya teniendo en cuenta que los datos han sido extraídos desde un formato físico a uno digital, podemos inferir que hay varios errores. En caso de ser digital podemos decir que no pudo ver este tipo de errores que se mencionaran a continuación. De los datos explorados, podemos mencionar que encontramos.

- Productos que son lo mismo pero escrito diferente que lo toma como diferente.
- Campos vacíos
- Algunas variables que no van a ser necesarios para nuestra investigación.

3. PREPARACION DE LOS DATOS

3.1. Selección de los datos.

En esta etapa tendremos que mencionar cuales son con los datos a trabajar o tablas de la base de datos, en nuestro caso como es un solo registro solo mencionaremos los campos de texto.

En la siguiente imagen podemos observar todos los campos de datos que tenemos, para este trabajo estaremos eliminando los datos de *N° Boleta*, *Señor(es)*, *Teléfono*, *Dirección*, *Estado*, *Observaciones*.

N° Boleta	Señor(es)	Teléfono	Fecha	Dirección	Cantidad	Unidad/M	Descripción	Por. Unid.	Total	Estado	Objersavi
443	Daniel		30/12/2019				Repuesto	S/	10.00	cancelado	
3	Javier		5/03/2021		1		Tubo 4"	S/ 11.00	S/ 11.00	cancelado	
658			13/04/2021		1		... INVIERNO	S/ 2.50	S/ 2.50	cancelado	Producto no
1733	Galindo		17/02/2020		2		1/2 PANALES	S/ 10.00	S/ 20.00	cancelado	
212	Madina Huarcayo Pepe		19/08/2021	Mz 71 Lt 27 -	1	millar	18 HUECOS	S/	850.00	cancelado	
723	Baltazar		28/12/2020		1	kilo	2	S/ 5.00	S/ 5.00	cancelado	
723	Baltazar		27/12/2020		2	kilo	2.5"	S/ 5.00	S/ 10.00	cancelado	
946	Katy				2		2X2X4 M	S/ 35.00	S/ 70.00	cancelado	Producto no
946	Katy				1		2X3X7	S/	70.00	cancelado	Producto no
915	Maritsa		31/08/2020		48		A/C TUBO DEVUELTO 4-2"			cancelado	
314	Jose		20/09/2020		2		ABRAZADERA	S/	1.00	cancelado	
345	Marino Machado		1/02/2021		8		ABRAZADERA	S/ 0.60	S/ 4.80	cancelado	
875	Katy		20/11/2020		200		ABRAZADERA	S/ 1.00	S/ 200.00	cancelado	
383	Jenni		8/02/2021		2		ABRAZADERA	S/	1.00	cancelado	
437							ACCESORIOS	S/	46.50	cancelado	
460			9/01/2020		10		ACCESORIOS	S/ 1.00	S/ 10.00	cancelado	
2	Estenio Goñas Calangos		5/03/2020	Mz 4 Lt 6 Apova			ACCESORIOS	S/	18.00	cancelado	
398.1					3		ACCESORIOS	S/ 1.00	S/ 3.00	cancelado	
155					4		ACCESORIOS	S/ 4.00	S/ 16.00	cancelado	

En esta etapa estaremos trabajando con los campos de *Descripción = productos*. *Precios producto, total, fecha, cantidad*. Opcional el de *unidad/medida*

Fecha	Cantidad	Unidad/M	PRODUCTO	Precio por	Total
30/12/2019			Repuesto		S/ 10.00
5/03/2021	1		Tubo 4"	S/ 11.00	S/ 11.00
13/04/2021	1		... INVIERNO	S/ 2.50	S/ 2.50
17/02/2020	2		1/2 PANALES	S/ 10.00	S/ 20.00
19/08/2021	1	millar	18 HUECOS	S/	850.00
28/12/2020	1	kilo	2	S/ 5.00	S/ 5.00
27/12/2020	2	kilo	2.5"	S/ 5.00	S/ 10.00
	2		2X2X4 M	S/ 35.00	S/ 70.00
	1		2X3X7	S/	70.00
31/08/2020	48		A/C TUBO DEVUELTO 4-2"		
20/09/2020	2		ABRAZADERA	S/	1.00
1/02/2021	8		ABRAZADERA	S/ 0.60	S/ 4.80
20/11/2020	200		ABRAZADERA	S/ 1.00	S/ 200.00
8/02/2021	2		ABRAZADERA	S/	1.00
			ACCESORIOS	S/	46.50
9/01/2020	10		ACCESORIOS	S/ 1.00	S/ 10.00
5/03/2020			ACCESORIOS	S/	18.00
	3		ACCESORIOS	S/ 1.00	S/ 3.00
	4		ACCESORIOS	S/ 4.00	S/ 16.00
	2		ACCESORIOS	S/	1.50

3.2. Limpieza de los datos.

Continuando, comenzamos primero a separar los registros que estén cancelados, de los que no, después se continuara a uniformizar los productos para que tengan un mismo nombre y estilo, al final a eliminar los campos vacíos, como las variables (columnas que no son necesarios para crear nuestro modelo).

Uniformizamos los datos a un mismo tipo de letra que en este caso vamos a poner todos en mayúscula “=MAYUSC(*campo texto*)”.

Daniel		30/12/2019		Repuesto		S/	10.00	cancelado	
Javier		5/03/2021	1	Tubo 4"		S/ 11.00	S/ 11.00	cancelado	
		13/04/2021	1		... INVIERNO	S/ 2.50	S/ 2.50	cancelado	Producto no cl
Galindo		17/02/2020	2		1/2 PANALES	S/ 10.00	S/ 20.00	cancelado	
Madina Huarayo Pepe		19/08/2021	1	millar	18 HUECOS		S/ 850.00	cancelado	
Baltazar		28/12/2020	1	kilo	2	S/ 5.00	S/ 5.00	cancelado	
Baltazar		27/12/2020	2	kilo	2.5"	S/ 5.00	S/ 10.00	cancelado	
Katy			2		2X2X4 M	S/ 35.00	S/ 70.00	cancelado	Producto no cl
Katy			1		2X3X7		S/ 70.00	cancelado	Producto no cl
Maritsa		31/08/2020	48		A/C TUBO DEVUELTO 4-2"			cancelado	
Jose		20/09/2020	2		ABRAZADERA	S/	1.00	cancelado	
Marino Machado		1/02/2021	8		ABRAZADERA	S/ 0.60	S/ 4.80	cancelado	
Katy		20/11/2020	200		ABRAZADERA	S/ 1.00	S/ 200.00	cancelado	
Jenni		8/02/2021	2		ABRAZADERA		S/ 1.00	cancelado	
					ACCESORIOS		S/ 46.50	cancelado	
		9/01/2020	10		ACCESORIOS	S/ 1.00	S/ 10.00	cancelado	
Estenio Goñas Calangos		5/03/2020		Mz 4 Lt 6 Apova	ACCESORIOS		S/ 18.00	cancelado	
			3		ACCESORIOS	S/ 1.00	S/ 3.00	cancelado	
			4		ACCESORIOS	S/ 4.00	S/ 16.00	cancelado	
			2		ACCESORIOS		S/ 1.50	cancelado	
Nely	995229425	7/10/2020	10	Calle Tayaca	ACCESORIOS	S/ 1.00	S/ 10.00	cancelado	

Ahora eliminamos algún espacio que pueda tener en campo *descripción = producto* algún campo del texto “=ESPACIOS (*campo texto*)”, y lo mismo con los demás. Con esto no debe eliminar campos vacíos de los textos y evitar lo que vemos en la imagen.

#	C	D	E	F	G	H	I	J
2	ADAPTADOR 3/4"					ADAPTADOR 3/4"		
3	ADAPTADOR 3/4 A 1"					ADAPTADOR 3/4 A 1"		
4	ADAPTADOR DE TOMACORRIENTE					ADAPTADOR DE TOMACORRIENTE		
5	AGUA					AGUA		
6	AGUJA					AGUJA		
7	ALAMBRE					ALAMBRE		
8	ALAMBRE					ALAMBRE #		
9	ALAMBRE #					ALAMBRE #16		
0	ALAMBRE #16					ALAMBRE #8		
1	ALAMBRE #8					ALAMBRE 2"		
2	ALAMBRE 2"					ALDABA		
3	ALDABA					ALICATE		
4	ALICATE					ARANDELAS A/ PRECION		
5	ARANDELAS A/ PRECION					ARCO SIERRA		
6	ARCO SIERRA					AREANA GRUESA		
7	AREANA GRUESA					AREMA GRUESA		
8	AREMA GRUESA					ARENA		
9	ARENA					ARENA CHANCADA		
0	ARENA CHANCADA					ARENA FINA		
1	ARENA FINA					ARENA GRUESA		
2	ARENA GRUESA					ARENA GRUEZA		

Ahora uniformizamos lo datos, de que los mismos productos, se escriban lo mismo

B	C	D	E	F	G	H	I	J	K	L	M
4912	Arebaló		10/05/2021	Vecino Jhoni	150		BLOQUETAS	S/ 1.35	S/ 202.50	cancelado	
4993	Maria Clorinda	927624584	1/06/2021	Miranda - An	1500		BLOQUETAS	S/ 1.30	S/ 1,950.00	cancelado	
5017	Maria Glorinda Lopez Rive		7/06/2021		1	millar	BLOQUETAS	S/ 1.30	S/ 1,300.00	cancelado	
4544	Maria Lopez Rivera		21/07/2021	Calle. Cesar,	300		BLOQUETAS	S/ 1.30	S/ 390.00	cancelado	
4567	Delia		28/07/2021		400		BLOQUETAS	S/ 1.35	S/ 540.00	cancelado	
460			9/01/2020		2		BOLAS DE CEMENTO	S/ 20.00	S/ 40.00	cancelado	
415	Yanet			Mz 128 Lt 17	100		BOLSA ANDINA ULTRA	S/ 22.30	S/ 2,230.00	cancelado	
229			21/08/2021	Ayacucho Mz	1		BOLSA DE AGUA	S/	S/ 270.00	cancelado	
721	Ludeña Abn	929315330	22/12/2020		1		BOLSA DE AGUA	S/	S/ 250.00	cancelado	
258	Santos		19/01/2021		1	bolsa	BOLSA DE AGUA	S/ 250.00	S/ 250.00	cancelado	
278	Segarra	920426665	20/01/2021	Vista Alegre	1		BOLSA DE AGUA	S/ 250.00	S/ 250.00	cancelado	
103	Yhoni	947950883	26/11/2020		1		BOLSA DE AGUA	S/ 250.00	S/ 250.00	cancelado	
448			4/01/2020		1	bolsa	CEMENTO APU	S/	S/ 20.00	cancelado	
721	Ludeña Abn	929315330	22/12/2020		4		BOLSA IMPRIMANTE	S/ 15.00	S/ 60.00	cancelado	
97			30/05/2020		1		BOLSA PINTURA	S/	S/ 4.00	cancelado	
721	Ludeña Abn	929315330	22/12/2020		8		BOLSA TEMPLE	S/ 15.00	S/ 120.00	cancelado	
447	Sr. Gabriel				50		BOLSAS DE CEMENTO	S/ 23.30	S/ 1,115.00	cancelado	
431	Gabriel		19/12/2019		50		BOLSAS DE CEMENTO ANDINO	S/ 22.30	S/ 1,115.00	cancelado	
433	Monica Sullana			Villamar Mz	9		BOLSAS DE CEMENTO APU	S/ 19.00	S/ 171.00	cancelado	
423	Grabiel Paucar				100		BOLSAS DE CEMENTO ULTRA	S/ 22.30	S/ 2,230.00	cancelado	
362	Ailin		29/09/2020		2		BOLSAS TEMPLE X 5K	S/ 3.50	S/ 7.00	cancelado	

Parecido al anterior en este caso observamos que algunos registros se menciona *Cemento Andino*, otro a *Cemento Andino Ultra*, y *cemento Ultra*, en este caso vamos uniformizar a *Cemento Andino Ultra*.

460.1					6	bolsa	CEMENTO ANDINO ULTRA				
482.1	Susi Faustino		24/08/2020		2	bolsa	CEMENTO ANDINO ULTRA				
482.1	Susi Faustino		24/08/2020		1	bolsa	CEMENTO ANDINO ULTRA				
931	Ventocilla				6	bolsa	CEMENTO ANDINO ULTRA				
941	Katy		4/09/2020		3	bolsa	CEMENTO ANDINO ULTRA				
334	Sahira		22/09/2020		10	bolsa	CEMENTO ANDINO ULTRA			S/ 23.80	
343	Jain Robles				1	bolsa	CEMENTO ANDINO ULTRA			S/ 24.50	
372	Edwin		2/10/2020		10	bolsa	CEMENTO ANDINO ULTRA			S/ 24.50	
375	Julisa		22/09/2020		30	bolsa	CEMENTO ANDINO ULTRA			S/ 24.00	
209	Sahira		7/10/2020		100	bolsa	CEMENTO ANDINO			S/ 23.80	
214	Ipollito deguz		7/10/2020		10	bolsa	CEMENTO ANDINO			S/ 24.00	
232	Enrique Huai	998001082	10/10/2020		10	bolsa	CEMENTO ANDINO			S/ 24.30	
7	Susi		21/10/2020		10	bolsa	CEMENTO ANDINO			S/ 24.50	
15	Delvis		24/10/2020		30	bolsa	CEMENTO ANDINO			S/ 24.20	
16	Deivis		19/10/2020		30	bolsa	CEMENTO ANDINO				
37	Shaira		27/10/2020		10	bolsa	CEMENTO ANDINO			S/ 23.80	
44	Deivis		28/10/2020		10	bolsa	CEMENTO ANDINO				
44	Deivis		28/10/2020		40	bolsa	CEMENTO ANDINO			S/ 24.20	
60	Deivis		31/10/2020		40	bolsa	CEMENTO ANDINO			S/ 24.20	
83	Divis		4/11/2020		4	bolsa	CEMENTO ANDINO			S/ 24.20	
823	Maly		9/11/2020		3	bolsa	CEMENTO ANDINO			S/ 24.30	

Otro ejemplo es el nombre en diminutivos que podemos encontrar

746	Baltazar		30/12/2020		1		CAJA RECTANGULAR		S/ 1.50	
182	Sosa		27/07/2020		4		CAJA RECTANGULAR S.			
970			13/09/2020		2		CAJA		S/ 3.00	
4977	Ferreteria Panorama		27/05/2021		2		CAJA DE DESAGUE X 6 - DE 4 PIESAS		S/ 85.00	
663	Lazarte		14/04/2021		5		CAJA DE PASO		S/ 17.00	
277	Campamento Dioses		20/01/2021		4		CAJA MODULAR		S/ 4.00	
1744			20/02/2020		2		CAJA OCTOGONAL		S/ 1.50	
526	Lazarte		11/03/2021		14		CAJA OCTOGONAL		S/ 1.60	
4716	Joaquin		24/04/2021		25		CAJA OCTOGONAL		S/ 1.50	
495	Flor Castillo		4/03/2021		2		CAJA OCTOGONAL		S/ 1.50	
4716	Joaquin		24/04/2021		30		CAJAS RECTANGULAR		S/ 1.50	
1744			20/02/2020		20		CAJAS RECTANGULAR		S/ 1.50	
48	Anibal Lara		13/03/2020		20		CAJAS RECTANGULAR		S/ 1.50	
244	Pepe Medina		26/08/2021		2		CAJITA LUZ PAVCO		S/ 2.50	
361			30/09/2020		1		CAJITA MODULAR		S/ 3.00	
148	Gloria Ravelo		6/12/2020		2		CAJITA PASO LUZ		S/ 5.00	
236	Katy		15/01/2021		21		CAJAS RECTANGULAR		S/ 1.20	
425			15/02/2021		2		CAJAS RECTANGULAR		S/ 1.50	
721	Ludeña Abn	929315330	22/12/2020		18		CAJAS RECTANGULAR		S/ 1.20	
5050	Walter		18/06/2021		6		CAJAS RECTANGULAR		S/ 1.50	
374.1	Claudio		26/06/2020		1	bolsa	CAL OBRA			

También otro de ejemplo que encontramos es *Galón Thinner*, que como observamos hace referencia al mismo producto, pero escrito de diferente forma.

441.1					1		GALON PINTURA 6 L		S/ 50.00	cancelado		
312.1			2/06/2020		1		GALON PRESERVANTE		S/ 18.00	cancelado		
494	Yodilia Jara				1		GALON SIKA		S/ 22.00	S/ 22.00	cancelado	Una palabra no
370.1	Javier Moran		25/06/2020		3		GALON THINNER		S/ 14.00	S/ 42.00	cancelado	
441.1					1		GALON THINNER		S/ 15.00		cancelado	
466.1			19/08/2020		2		GALON THINNER		S/ 30.00		cancelado	
834	Colegio Sagrada Familia		12/11/2020		1		GALON THINNER		S/ 15.00	S/ 15.00	cancelado	
4905	Mad Metal		10/05/2021		1		GALON THINNER		S/ 15.00	S/ 15.00	cancelado	
265	Katy		19/01/2021		1		GALON THINNER		S/ 15.00	S/ 15.00	cancelado	
289	Reategui		22/01/2021		2		GALONES DE PINTURA BASE GRISS		S/ 65.00	S/ 130.00	cancelado	
289	Reategui		22/01/2021		1		GALONES DE PINTURA GLLOS NEGRO		S/ 65.00	S/ 65.00	cancelado	
289	Reategui		22/01/2021		2		GALONES DE PINTURA GLLOSS		S/ 65.00	S/ 130.00	cancelado	
228	Campamento Dioses		14/01/2021		2		GALONES DE THINER		S/ 15.00	S/ 30.00	cancelado	
289	Reategui		22/01/2021		4		GALONES DE THINER		S/ 15.00	S/ 60.00	cancelado	
370.1	Javier Moran		25/06/2020		3		GALONES DE THINNER		S/ 14.00	S/ 42.00	cancelado	
178	Campamento Dioses		16/12/2020		3		GALONES DE THINNER		S/ 45.00		cancelado	
110	Sr. Moran		10/07/2020		5		GALONES GASOLINA		S/ 60.00		cancelado	
227	Servigen		14/01/2021		2		GALONES PINTURA NEGRO GLLOSS		S/ 40.00	S/ 80.00	cancelado	
227	Servigen		14/01/2021		3		GALONES THINER		S/ 15.00	S/ 45.00	cancelado	
686	Campamento Dioses		19/04/2021		2		GALONES THINER		S/ 15.00	S/ 30.00	cancelado	cancelado - por
420.1			10/08/2020		2		GALONES THINNER		S/ 30.00		cancelado	

Algunos productos no están claros o no son muy específicos a que producto hacen referencia.

4754	Marcos		30/04/2021		8	m	MANGUERA DE NIVEL		S/ 1.00	S/ 8.00	cancelado	
383	Jenni		8/02/2021		2	m	MANGUERA DE GAS		S/ 3.00	S/ 6.00	cancelado	
232	Medina		24/08/2021		15	m	MANGUERA		S/ 1.50	S/ 22.50	cancelado	
233	Medina		24/08/2021		15	m	MANGUERA		S/ 1.50	S/ 22.50	cancelado	
841					10	m	MANGUERA DE NIVEL		S/ 1.00	S/ 10.00	cancelado	
5044	Flavio Araujo		17/06/2021		20	m	MANGUERA DE NIVEL		S/ 1.00	S/ 20.00	cancelado	
143	Jenri Palomi	962844569	5/12/2020	Mz 59 Lt 16 -	6	m	MANGUERA DE NIVEL		S/ 1.00	S/ 6.00	cancelado	
450.1					1		MARTILLO		S/ 14.00		cancelado	
5044	Flavio Araujo		17/06/2021		1		MARTILLO		S/ 12.00	S/ 12.00	cancelado	
20			9/03/2020		1		MASILLA PARA META 2 ANIPSA		S/ 8.00		cancelado	
331.1							MASJE Y TEG		S/ 5.00		cancelado	
491	Carlos		3/03/2021				MATERIALES		S/ 70.00		cancelado	
453	Elita Muñoz		6/01/2020		1		MEDIDOR DIGITAL		S/ 46.00		cancelado	
105	Wilder		8/07/2020		1		MEDIDOR DIGITAL		S/ 40.00		cancelado	
658			13/04/2021		1		MEDIDOR DIGITAL		S/ 45.00	S/ 45.00	cancelado	
430	Elber		18/12/2019		3	m	PLASTICO		S/ 3.00	S/ 9.00	cancelado	
4771	Melva Espinoza		4/05/2021		2	mil	PANDERETA		S/ 690.00	S/ 1,380.00	cancelado	
5090	Eulalia Canc	910584700	6/07/2021		1		MURETE DE LUZ MONOFOSICO		S/ 90.00		cancelado	
660			14/04/2021		1		MULTICONECTOR		S/ 18.00	S/ 18.00	cancelado	
879	Campamento Dioses		21/11/2020		1		MULTICORERTOR		S/ 18.00	S/ 18.00	cancelado	
273	Orlando Sara	910278649	20/01/2021		1		MURETE DE LUZ MONOFOSICO		S/ 80.00	S/ 80.00	cancelado	

En esta parte estaremos reemplazando las “T” por “Tee” y todos sean iguales y mas entendibles para la organización.

424	Dioses	15/02/2021	2	SUMIDERO 2"	S/ 6.00	S/ 12.00	cancelado	
863		17/11/2020	1	SUMIDERO 4"	S/ 10.00	S/ 10.00	cancelado	
381.1	Frank	27/06/2020	1	SUMIDERO DE LAVADERO		S/ 22.00	cancelado	
668	Resead	15/04/2021	10	SUPER FLU	S/ 0.80	S/ 8.00	cancelado	
975	Elmer	14/09/2020	2	Tee	S/ 2.00	S/ 4.00	cancelado	
4978	Vita	27/05/2021	1	T	S/ 2.00	S/ 2.00	cancelado	
1745	Patrik	997359489	20/02/2020	5	T 1/2	S/ 1.50	S/ 7.50	cancelado
397.1			3/07/2020	3	T 1/2	S/ 1.50	S/ 4.50	cancelado
258	Nely	15/10/2020	2	T 1/2	S/ 1.00	S/ 2.00	cancelado	
842	Juan Cerna	12/11/2020	1	T 1/2	S/ 2.00	S/ 2.00	cancelado	
274			1	T 1/2"	S/	S/	cancelado	
329.1	Patrik	11/06/2020	5	T 1/2 A/PRECION	S/ 2.00	S/ 10.00	cancelado	
491.1	Moran	24/08/2020	3	T 1/2 A/PRECION	S/ 2.00	S/ 6.00	cancelado	
332.1		Asociacion H	1	T 2"	S/	S/ 3.00	cancelado	
975	Elmer	14/09/2020	1	T 2"	S/ 3.00	S/ 3.00	cancelado	
842	Juan Cerna	12/11/2020	1	T 2"	S/ 2.50	S/ 2.50	cancelado	
122		998983077	30/11/2020	5	T 2"	S/ 2.50	S/ 12.50	cancelado
796	Ocaña	9/01/2021	3	T 2"	S/ 2.00	S/ 6.00	cancelado	
4548	Primitiva Gusman Lopez	23/07/2021	Uchiza con Ti	6	T 2"	S/ 2.50	S/ 15.00	cancelado
4542	Vita Tapia	21/07/2021		1	T 2" A 2"	S/ 3.50	S/ 3.50	cancelado
491.1	Moran	24/08/2020	3	T 3/4 A/PRECION	S/ 4.00	S/ 12.00	cancelado	

- Ahora que ya hemos uniformizado los datos, vamos eliminar los productos que tengan campos vacíos, especialmente en las columnas de fecha, cantidad y producto no sea entendible o estén en blanco.

Recordar que normalmente las organizaciones ya tienen echo todo lo anterior que realizamos solo comenzaríamos de esta parte a delante.

Tener en cuenta que en la etapa de *Selección de Datos* debemos seleccionar los campos que usaremos o las tablas con campos a usar. En mi caso estera realizando esta operación en esta parte antes de continuar.

Fecha	Cantidad	Unidad/M	PRODUCTO	Precio por	Total
30/12/2019		Repuesto			S/ 10.00
5/03/2021	1	Tubo 4"		S/ 11.00	S/ 11.00
13/04/2021	1		... INVIERNO	S/ 2.50	S/ 2.50
17/02/2020	2		1/2 PANALES	S/ 10.00	S/ 20.00
19/08/2021	1	millar	18 HUECOS		S/ 850.00
28/12/2020	1	kilo	2	S/ 5.00	S/ 5.00
27/12/2020	2	kilo	2.5"	S/ 5.00	S/ 10.00
	2		2X2X4 M	S/ 35.00	S/ 70.00
	1		2X3X7		S/ 70.00
31/08/2020	48		A/C TUBO DEVUELTO 4-2"		
20/09/2020	2		ABRAZADERA		S/ 1.00
1/02/2021	8		ABRAZADERA	S/ 0.60	S/ 4.80
20/11/2020	200		ABRAZADERA	S/ 1.00	S/ 200.00
8/02/2021	2		ABRAZADERA		S/ 1.00
			ACCESORIOS		S/ 46.50
9/01/2020	10		ACCESORIOS	S/ 1.00	S/ 10.00
5/03/2020			ACCESORIOS		S/ 18.00
	3		ACCESORIOS	S/ 1.00	S/ 3.00
	4		ACCESORIOS	S/ 4.00	S/ 16.00
	2		ACCESORIOS		S/ 1.50
7/02/2020	10		ACCESORIOS	S/ 1.00	S/ 10.00

Esta parte dividiremos en varias partes para ir en orden, primero serán los campos vacíos de la columna Producto(descripción), después los de la fecha, los del total, pero en caso de los campos de cantidad y precio por unidad, trataremos de calcular la cantidad según el precio por unidad y el total, en caso de no poder se eliminará.

Eliminemos los campos vacíos de la columna PRODUCTO, en este caso, todo estaba lleno, pero ahora vemos, lo datos con correctos o no entendibles.

En la imagen siguientes podemos observar algunos productos que no son claros en este caso vamos a eliminarlos.

Fecha	Cantidad	Unidad/M	PRODUCTO	Precio por	Total
13/04/2021	1		... INVIERNO	S/ 2.50	S/ 2.50
17/02/2020	2		1/2 PANALES	S/ 10.00	S/ 20.00
19/08/2021	1	millar	18 HUECOS		S/ 850.00
28/12/2020	1	kilo	2	S/ 5.00	S/ 5.00
27/12/2020	2	kilo	2.5"	S/ 5.00	S/ 10.00
	2		2X2X4 M	S/ 35.00	S/ 70.00
	1		2X3X7		S/ 70.00
31/08/2020	48		A/C TUBO DEVUELTO 4-2"		
20/09/2020	2		ABRAZADERA		S/ 1.00
1/02/2021	8		ABRAZADERA	S/ 0.60	S/ 4.80
20/11/2020	200		ABRAZADERA	S/ 1.00	S/ 200.00
8/02/2021	2		ABRAZADERA		S/ 1.00
			ACCESORIOS		S/ 46.50
9/01/2020	10		ACCESORIOS	S/ 1.00	S/ 10.00
5/03/2020			ACCESORIOS		S/ 18.00
	3		ACCESORIOS	S/ 1.00	S/ 3.00
	4		ACCESORIOS	S/ 4.00	S/ 16.00
	2		ACCESORIOS		S/ 1.50
7/10/2020	10		ACCESORIOS	S/ 1.00	S/ 10.00
	3		ACCESORIOS	S/ 1.00	S/ 3.00

Ahora para agilizar el proceso vamos ir eliminando los productos no claros, campos vacíos, primero según fechas vacías., de lo cual se eliminaron un total de 429 registros.

1	Fecha	Cantidad	Unidad/M	PRODUCTO	Precio por	Total
				ACCESORIOS		S/ 46.50
				ACCESORIOS	S/ 1.00	S/ 3.00
				ACCESORIOS	S/ 4.00	S/ 16.00
				ACCESORIOS		S/ 1.50
				ACCESORIOS	S/ 1.00	S/ 3.00
				ACCESORIOS DE AGUA		S/ 46.50
				ACCESORIOS DE AGUA		S/ 19.00
				ACCESORIOS ELECTRICOS		S/ 10.00
				ADAPTADOR	S/ 1.00	S/ 5.00
				ADAPTADOR	S/ 1.50	S/ 3.00
				ADAPTADOR	S/ 3.00	S/ 3.00
		kilo		ALAMBRE	S/ 4.00	S/ 40.00
		kilo		ALAMBRE		S/ 5.00
		kilo		ALAMBRE	S/ 5.00	S/ 10.00
		kilo		ALAMBRE	S/ 5.00	S/ 100.00
		kilo		ALAMBRE	S/ 5.00	S/ 60.00
		kilo		ALAMBRE	S/ 5.00	S/ 25.00
		kilos		ALAMBRE	S/ 5.50	S/ 22.00
		kilos		ALAMBRE	S/ 6.00	S/ 18.00
		kilo		ALAMBRE #16	S/ 5.00	S/ 15.00

			ACCESORIOS		S/ 46.50
3			ACCESORIOS	S/ 1.00	S/ 3.00
4			ACCESORIOS	S/ 4.00	S/ 16.00
2			ACCESORIOS		S/ 1.50
3			ACCESORIOS	S/ 1.00	S/ 3.00
			ACCESORIOS DE AGUA		S/ 46.50
			ACCESORIOS DE AGUA		S/ 19.00
			ACCESORIOS ELECTRICOS		S/ 10.00
5			ADAPTADOR	S/ 1.00	S/ 5.00
2			ADAPTADOR	S/ 1.50	S/ 3.00
1			ADAPTADOR	S/ 3.00	S/ 3.00
10	kilo		ALAMBRE	S/ 4.00	S/ 40.00
1	kilo		ALAMBRE		S/ 5.00
2	kilo		ALAMBRE	S/ 5.00	S/ 10.00
20	kilo		ALAMBRE	S/ 5.00	S/ 100.00
12	kilo		ALAMBRE	S/ 5.00	S/ 60.00
5	kilo		ALAMBRE	S/ 5.00	S/ 25.00
4	kilos		ALAMBRE	S/ 5.50	S/ 22.00
3	kilos		ALAMBRE	S/ 6.00	S/ 18.00
3	kilo		ALAMBRE #16	S/ 5.00	S/ 15.00
10	kilos		ALAMBRE #16	S/ 5.00	S/ 50.00
3	kilo		ALAMBRE #8	S/ 6.00	S/ 18.00
3	kilo		ALAMBRE #8	S/ 5.00	S/ 15.00
1			ALICATE	S/ 8.00	S/ 8.00
1			ARCO SIERRA		S/ 8.00

4	m3		ARENA FINA	S/ 45.00	S/ 180.00
5	bolsa		ARENA FINA	S/ 3.00	S/ 15.00
2	m3		ARENA FINA	S/ 45.00	S/ 90.00
3	m3		ARENA GRUESA	S/ 45.00	S/ 135.00
1	m3		ARENA GRUESA	S/ 45.00	S/ 45.00
1/2	m3		ARENA GRUESA		S/ 23.00
9			ARENA GRUESA		S/ 405.00
			ARENA GRUESA		S/ 15.00
4	m		ARENA GRUESA	S/ 45.00	S/ 180.00
4	m3		ARENA GRUESA	S/ 45.00	S/ 180.00
2	m3		ARENA GRUEZA	S/ 45.00	S/ 90.00
10	m3		ARENA GUESA	S/ 45.00	S/ 450.00
1			ARO DE CERA	S/ 5.00	S/ 5.00
1			BALDE DE COLA SINTETICA	S/ 25.00	S/ 25.00
3			BALDE DE PINTURA	S/ 15.00	S/ 45.00
2			BALDE DE PINTURA	S/ 15.00	S/ 30.00
2			BALDE DE PINTURA LATEX		S/ 30.00
2			BALDE DE PINTURA SONAL	S/ 18.00	S/ 36.00
1			BALDE PINTURA ESMALTE BLANCO		S/ 35.00
1			BAÑO	S/ 18.00	S/ 18.00
4			BARETAS	S/ 18.00	S/ 72.00
			BARNIS		S/ 10.00
1			BASE FLOURECENTE	S/ 20.00	S/ 20.00
3			BISAGRA DE 3"		S/ 5.30
6			BISAGRA DE PORTON 3"	S/ 3.00	S/ 18.00

	250		BLOQUETA	S/ 1.30	S/ 325.00
	200		BLOQUETA	S/ 1.30	S/ 260.00
	20		BLOQUETA		S/ 28.60
	250		BLOQUETA	S/ 1.30	S/ 325.00
	3		BROCA 1/4		S/ 10.00
	2		BROCA FIERRO		S/ 5.00
	1		BROCHA		S/ 4.00
	1		BROCHA	S/ 5.00	S/ 5.00
	2		BROCHA		S/ 5.50
	13	m	CABLE	S/ 1.20	S/ 15.60
	22	m	CABLE #12	S/ 1.20	S/ 26.40
	30	m	CABLE #14	S/ 1.20	S/ 36.00
	20	m	CABLE MELLISO		S/ 20.00
	20	m	CABLE MELLISO	S/ 0.80	S/ 16.00
	10	m	CABLE MELLISO #14	S/ 1.00	S/ 10.00
	1		CAJA		S/ 3.00
	1		CAJA 20X20 LUZ	S/ 18.00	S/ 18.00
	1		CAJA MEDIDOR	S/ 20.00	S/ 20.00
	1	bolsa	CAL OBRA	S/ 6.00	S/ 6.00
	1	bolsa	CAL OBRA		S/ 5.00
	3		CALAMINA 3.6	S/ 20.00	S/ 60.00
	1		CANDADO		S/ 4.00
	1		CANT MAYOLICA		S/ 4.00
	2		CAÑO	S/ 12.00	S/ 24.00
	1		CAÑO	S/ 12.00	S/ 12.00
	1		CAÑO SCHUBET	S/ 12.00	S/ 12.00

	1	kilos	CARBON		S/ 16.00
			CARBON		S/ 160.00
	50	bolsa	CEMENTO	S/ 23.30	#####
	1	bolsa	CEMENTO	S/ 19.50	S/ 19.50
	3	bolsa	CEMENTO	S/ 20.00	S/ 60.00
	5	bolsa	CEMENTO	S/ 22.00	S/ 110.00
	1	bolsa	CEMENTO		S/ 22.00
	2	bolsa	CEMENTO		S/ 44.00
	2	kilos	CEMENTO		S/ 1.60
			CEMENTO		S/ 86.00
	10	kilos	CEMENTO		S/ 8.00
	2	kilos	CEMENTO		S/ 1.60
	2	bolsa	CEMENTO		S/ 38.00
	1	bolsa	CEMENTO	S/ 19.00	S/ 19.00
	6	bolsa	CEMENTO	S/ 19.00	S/ 114.00
	10	bolsa	CEMENTO		S/ 245.00
	6	bolsa	CEMENTO	S/ 22.00	S/ 132.00
	10	bolsa	CEMENTO	S/ 24.50	S/ 245.00
	2	bolsa	CEMENTO	S/ 19.50	S/ 39.00
	10		CEMENTO	S/ 24.30	S/ 243.00
	100	bolsa	CEMENTO ANDINA ULTRA	S/ 22.30	#####
	100	bolsa	CEMENTO ANDINO ULTRA	S/ 22.30	#####
	36	bolsa	CEMENTO ANDINO ULTRA	S/ 22.80	S/ 820.00
	1	bolsa	CEMENTO ANDINO ULTRA	S/ 23.00	S/ 23.00
	3	Bolsa	CEMENTO ANDINO ULTRA	S/ 23.00	S/ 69.00

	7	bolsa	CEMENTO ANDINO ULTRA	S/ 24.00	S/ 168.00
	2	bolsa	CEMENTO ANDINO ULTRA		S/ 49.00
	6	bolsa	CEMENTO ANDINO ULTRA		S/ 147.00
	6	bolsa	CEMENTO ANDINO ULTRA		S/ 147.00
	1	bolsa	CEMENTO ANDINO ULTRA	S/ 24.50	S/ 24.50
	10	bolsa	CEMENTO ANDINO ULTRA	S/ 24.30	S/ 243.00
	10	bolsa	CEMENTO ANDINO ULTRA	S/ 24.30	S/ 243.00
	15	bolsa	CEMENTO ANDINO ULTRA	S/ 24.00	S/ 360.00
	7	bolsa	CEMENTO ANDINO ULTRA		S/ 159.60
	6	bolsa	CEMENTO ANDINO ULTRA		S/ 147.00
	10	bolsa	CEMENTO ANDINO ULTRA	S/ 23.80	S/ 238.00
	3	bolsa	CEMENTO ANDINO ULTRA	S/ 24.50	S/ 73.50
	10	bolsa	CEMENTO ANDINO ULTRA	S/ 24.50	S/ 245.00
	12	bolsa	CEMENTO ANDINO ULTRA	S/ 24.00	S/ 288.00
	4	bolsa	CEMENTO ANDINO ULTRA		S/ 95.60
	1	bolsa	CEMENTO ANDINO ULTRA	S/ 23.00	S/ 23.00
	9	bolsa	CEMENTO APU	S/ 19.00	S/ 171.00
	1	bolsa	CEMENTO APU	S/ 22.00	S/ 22.00
	100	bolsa	CEMENTO APU	S/ 18.40	#####
	20	bolsa	CEMENTO APU	S/ 21.30	S/ 426.00
	2	bolsa	CEMENTO APU		S/ 44.00
	5	bolsa	CEMENTO APU	S/ 22.00	S/ 110.00
	6	bolsa	CEMENTO APU	S/ 22.00	S/ 132.00
	5	bolsa	CEMENTO APU	S/ 22.00	S/ 110.00
	45	bolsa	CEMENTO APU	S/ 22.00	S/ 990.00
	1	bolsa	CEMENTO APU		S/ 21.00

	2	bolsa	CEMENTO APU	S/ 21.00	S/ 42.00
	50	bolsa	CEMENTO APU	S/ 18.45	S/ 922.50
	1	bolsa	CEMENTO APU		S/ 21.00
	4	bolsa	CEMENTO APU	S/ 22.00	S/ 88.00
	2	bolsa	CEMENTO APU	S/ 22.00	S/ 44.00
	6	bolsa	CEMENTO APU	S/ 22.00	S/ 132.00
	20	bolsa	CEMENTO APU	S/ 21.50	S/ 430.00
	9	bolsa	CEMENTO APU	S/ 22.00	S/ 198.00
	5	bolsa	CEMENTO APU	S/ 22.00	S/ 110.00
	15	bolsa	CEMENTO NACIONAL	S/ 26.00	S/ 390.00
	5	bolsa	CEMENTO NACIONAL	S/ 22.50	S/ 112.50
	4	bolsa	CEMENTO NACIONAL	S/ 24.00	S/ 96.00
	3	bolsa	CEMENTO NACIONAL	S/ 19.00	S/ 57.00
	2	bolsa	CEMENTO NACIONAL	S/ 19.00	S/ 38.00
	4	bolsa	CEMENTO QUISQUEYA	S/ 22.00	S/ 88.00
	10	bolsa	CEMENTO QUISQUEYA		S/ 240.00
	15	bolsa	CEMENTO QUISQUEYA	S/ 19.50	S/ 292.50
	4	bolsa	CEMENTO QUISQUEYA	S/ 19.00	S/ 76.00
	2	bolsa	CEMENTO SOL	S/ 23.50	S/ 47.00
	3	bolsa	CEMENTO SOL	S/ 23.50	S/ 70.50
	3	bolsa	CEMENTO SOL	S/ 23.00	S/ 69.00
	3	bolsa	CEMENTO SOL	S/ 22.50	S/ 67.50
	1	bolsa	CEMENTO SOL	S/ 23.50	S/ 23.50
	2	bolsa	CEMENTO SOL		S/ 46.00
	6	bolsa	CEMENTO SOL	S/ 23.50	S/ 141.00

	4	bolsa	CEMENTO SOL	S/ 23.50	S/ 94.00
	1	bolsa	CEMENTO SOL	S/ 23.50	S/ 23.50
	6	bolsa	CEMENTO SOL	S/ 113.00	S/ 113.00
	30	bolsa	CEMENTO SOL	S/ 23.30	S/ 699.00
	10	bolsa	CEMENTO SOL	S/ 23.50	S/ 235.00
	2		CERA		S/ 10.00
	1		CERROJO DE PORTON		S/ 15.00
	1		CERROJO DE PUERTA	S/ 3.50	S/ 3.50
	1		CHAPA PHELIX	S/ 22.00	S/ 22.00
	1		CILINDRO		S/ 15.00
	3		CINTA		S/ 1.50
	1		CINTA AISLANTE		S/ 4.50
	1		CINTA AISLANTE		S/ 4.50
	1		CINTA AISLANTE	S/ 4.50	S/ 4.50
	1		CINTA AISLANTE		S/ 4.50
	1		CINTA AISLANTE	S/ 2.50	S/ 2.50
	2		CINTA AISLANTE	S/ 4.50	S/ 9.00
	12		CINTILLO	S/ 0.25	S/ 3.00
	20		CINTILLO		S/ 5.00
	1/2	kilo	CLAVO		S/ 2.50
	1	kilo	CLAVO	S/ 5.00	S/ 5.00
	1	kilo	CLAVO		S/ 10.00
			CLAVO		S/ 1.50
	5	kilos	CLAVO	S/ 5.00	S/ 25.00
	4	kilos	CLAVO	S/ 24.00	S/ 24.00
			CLAVO		S/ 25.00

	2	kilos	CLAVO 2 1/2"		S/ 10.00
	6	kilos	CLAVO 2 1/2", 3"	S/ 5.00	S/ 30.00
	1	kilo	CLAVO 3"	S/ 6.00	S/ 6.00
	2	kilo	CLAVO 3" Y 2"		S/ 10.00
	12		CLAVO ACERO 1"		S/ 1.20
	5		CLAVO ACERO 2"		S/ 1.00
			CLAVO CEMENTO 3"		S/ 3.00
	1/2	kilos	CLAVO ETERNIT		S/ 2.50
	3		CODO	S/ 1.50	S/ 4.50
	3		CODO	S/ 1.00	S/ 3.00
	2		CODO 1/2	S/ 1.50	S/ 3.00
	2		CODO 2		S/ 4.00
	1		CODO 2	S/ 2.00	S/ 2.00
	3		CODO 2	S/ 1.50	S/ 4.50
	7		CODO 2	S/ 1.50	S/ 10.50
	1		CODO 2 X 45		S/ 2.50
	1		CODO 4		S/ 5.00
	3		CODO 4	S/ 7.00	S/ 21.00
	1		CODO 4	S/ 7.00	S/ 7.00
	2		CODO 4	S/ 7.00	S/ 14.00
	1		CODO 4 A 2"		S/ 5.00
	1		CODO 4 X 45	S/ 9.00	S/ 9.00
	3		CODO BRONCE	S/ 3.00	S/ 12.00
	5		CODOS 1/2 MIXTO	S/ 1.50	S/ 7.50
	68	m	CORDEL	S/ 0.30	S/ 20.40

	1		CORDEL		S/ 8.00
	2		CRUSETA 2"	S/ 3.50	S/ 7.00
	1		CUCHILLO		S/ 8.00
	1		CUELLO DE CERA		S/ 5.00
	10		CURVA	S/ 0.30	S/ 3.00
	2		CURVA		S/ 1.00
	6		CURVA 3/4	S/ 0.60	S/ 3.60
	3		CURVA 3/4	S/ 0.50	S/ 1.50
	1		CUTER		S/ 1.00
	2		DISCO CORTE	S/ 5.00	S/ 10.00
	1		DISCO CORTE		S/ 10.00
	1		DISCO CORTE	S/ 7.00	S/ 7.00
	10		DISCO CORTE	S/ 7.00	S/ 70.00
	1		DISCO CORTE CONCRETO	S/ 10.00	S/ 10.00
	1		DISCO CORTE NORTON	S/ 5.00	S/ 5.00
	1		DOCENA DE TORNILLO	S/ 2.40	S/ 2.40
	1		DOCENA DE TORNILLO	S/ 2.40	S/ 2.40
	1		DUCHA	S/ 4.00	S/ 4.00
	1		DUCHA	S/ 5.00	S/ 5.00
			ELIMINACION DE DESMONTE		S/ 30.00
	2		ENCHUFE		S/ 2.00
	3		ENCHUFE	S/ 2.00	S/ 6.00
	1		ESCOBA	S/ 8.00	S/ 8.00
	1		ESCOBILLA DE ACERO	S/ 5.00	S/ 5.00
			ESPATULA		S/ 6.00
	3		EXTENCION 5M		S/ 8.00

	2		FIBRAFORTE	S/ 10.00	S/ 20.00
	1		FOCO	S/ 8.00	S/ 8.00
	2		FOCO		S/ 9.00
	2		FOCO	S/ 13.00	S/ 26.00
	3		FOCO		S/ 18.00
	2		FOCOS LED 9	S/ 6.00	S/ 12.00
	2		FRAGUA	S/ 5.00	S/ 10.00
	1		GALON DE THINNER		S/ 15.00
	1		GALON PINTURA 6 L		S/ 50.00
	1		GALON SIKA	S/ 22.00	S/ 22.00
	15		GRAPA		S/ 1.50
	4		HOJA LIJA	S/ 2.50	S/ 10.00
	2		HOJA SANDFLEX		S/ 12.00
	1		HOJA SIERRA	S/ 6.00	S/ 6.00
	1		HOJA SIERRA	S/ 7.00	S/ 7.00
	1		INODORO	S/ 2.00	S/ 20.00
	1		INODORO	S/ 20.00	S/ 20.00
	5		INTERRUPTOR	S/ 2.00	S/ 10.00
	3		INTERRUPTOR		S/ 6.00
	3		INTERRUPTOR	S/ 2.00	S/ 6.00
	2		KILOS DE CLAVO 2", 2.5"		S/ 10.00
	1		KINCHA	S/ 5.00	S/ 5.00
	250		KK 18 HUECOS	S/ 0.52	S/ 130.00
	1	millar	KK 18 HUECOS		S/ 800.00
	1/2	millar	KK 18 HUECOS		S/ 390.00
	250		KK 18 HUECOS	S/ 0.74	S/ 185.00

	1/2	millar	KK 18 HUECOS		S/ 390.00
	200		KK 18 HUECOS	S/ 0.76	S/ 152.00
	1	millar	KK 18 HUECOS	S/ 760.00	S/ 760.00
	1/2	Millar	KK 18 HUECOS		S/ 400.00
	1	millar	KK 18 HUECOS	S/ 840.00	S/ 840.00
	1		LAMPA		S/ 26.00
	1		LAMPA		S/ 24.50
	1		LAVADERO	S/ 40.00	S/ 40.00
	1		LINTERNA	S/ 8.00	S/ 8.00
			LISTON		S/ 30.00
	1		LLAVE DE DUCHA		S/ 18.00
	1		LLAVE PASO		S/ 4.00
	1		LLAVE PASO 1/2	S/ 5.00	S/ 5.00
			LLAVE PASO 3/4		S/ 12.00
	1		LT. THINNER		S/ 5.00
	2		LUNAS DE SOLDAR TRANSPARENTE	S/ 1.50	S/ 3.00
	12	m3	MALLA	S/ 8.00	S/ 96.00
			MANGUERA		S/ 4.00
	10	m	MANGUERA DE NIVEL	S/ 1.00	S/ 10.00
	1		MARTILLO		S/ 14.00
			MASJE Y TEG		S/ 5.00
	1		NIVEL	S/ 14.00	S/ 14.00
	15		OREJAS	S/ 0.25	S/ 3.75
	10	m3	ORMIGON		
	2	m3	ORMIGON	S/ 45.00	S/ 90.00

	4	m3	ORMIGON	S/ 45.00	S/ 180.00
	6	m3	ORMIGON	S/ 45.00	S/ 270.00
	12	m3	ORMIGON		S/ 520.00
	3	m3	ORMIGON	S/ 45.00	S/ 135.00
	3		PANEL	S/ 44.00	S/ 132.00
	6		PANEL CERRADO	S/ 38.00	S/ 228.00
	1		PANEL CONPUERTA		S/ 48.00
	3		PAR DE BISAGRA		S/ 24.00
	1		PAR DE PERNOS DE INODORO	S/ 3.50	S/ 3.50
	3		PAVCO	S/ 1.00	S/ 3.00
	1		PEGAMENTO	S/ 10.00	S/ 10.00
	1		PEGAMENTO		S/ 9.00
	1		PEGAMENTO		S/ 5.00
	1		PEGAMENTO	S/ 5.00	S/ 5.00
	1		PEGAMENTO	S/ 5.00	S/ 5.00
	1		PEGAMENTO	S/ 3.00	S/ 3.00
	1		PEGAMENTO	S/ 9.00	S/ 9.00
	1		PEGAMENTO	S/ 9.00	S/ 9.00
	2		PEGAMENTO		S/ 6.00
	1		PEGAMENTO	S/ 10.00	S/ 10.00
	1		PEGAMENTO MAYOLICA		S/ 15.00
	10	bolsa	PEGAMENTO MAYOLICA	S/ 18.00	S/ 180.00
	15	bolsa	PEGAMENTO MAYOLICA	S/ 18.00	S/ 270.00
	1		PEGAMENTO MAYOLICA	S/ 18.00	S/ 18.00
	1	saco	PEGAMENTO MAYOLICA CELIMA	S/ 18.00	S/ 18.00
	2		PERNO	S/ 0.70	S/ 1.40

	8		PERNO CAMA		S/ 8.00
	1		PICO		S/ 28.00
	10	metros cub	PIEDRA BASE	S/ 45.00	S/ 450.00
	3	m3	PIEDRA BASE	S/ 45.00	S/ 135.00
	10	metros cub	PIEDRA CHANCADA	S/ 55.00	S/ 550.00
	1/2	m3	PIEDRA CHANCADA		S/ 27.50
	1	m3	PIEDRA CHANCADA	S/ 55.00	S/ 55.00
	5	bolsa	PIEDRA CHANCADA	S/ 4.00	S/ 20.00
	2	m3	PIEDRA CHANCADA	S/ 55.00	S/ 110.00
	1	m3	PIEDRA CHANCADA	S/ 55.00	S/ 55.00
	1		PINTURA		S/ 30.00
	1		PINTURA	S/ 18.00	S/ 18.00
	1/4		PINTURA BLANCA		S/ 9.00
	1		PINTURA BLANCA	S/ 15.00	S/ 15.00
	1/4		PINTURA GLOSS		S/ 10.00
	1		PLANCHA DE 2X3X4.30		S/ 37.00
	6		PLANCHA DE ETERNIT	S/ 35.00	S/ 210.00
	1		PLANCHA TECNOPOR	S/ 18.00	S/ 18.00
	6		PLANCHA TECNOPOR	S/ 18.00	S/ 108.00
	4	m	PLASTICO	S/ 4.00	S/ 16.00
	2	m	PLASTICO		S/ 3.00
	5	m	PLASTICO		S/ 15.00
	10	m	PLASTICO	S/ 3.00	S/ 30.00
	10	m	PLASTICO	S/ 2.00	S/ 20.00
	7 1/2	m	PLASTICO	S/ 2.00	S/ 15.00
	5	m	PLASTICO	S/ 3.00	S/ 15.00
	1		PUERTA		S/ 48.00

	2		RASTRILLOS	S/ 11.00	S/ 22.00
	1		RECOGEDOR	S/ 3.00	S/ 3.00
	2		REDUCCION 2 A 1"	S/ 1.50	S/ 3.00
	1		REDUCCION 4 A 2"		S/ 1.50
	1		REGISTRO 2"	S/ 5.00	S/ 5.00
	1		REJILLA		S/ 5.00
	1		RODILLO		S/ 6.00
	1		RODILLO 10"		S/ 5.00
	2		RODOMETAL		S/ 5.00
	1	kilo	SOLDADURA PUNTO AZUL	S/ 15.00	S/ 15.00
	1		SOQUET		S/ 1.50
	3		SOQUET		S/ 9.00
	2		SOQUET		S/ 7.00
	4		SOQUET	S/ 3.00	S/ 12.00
	3		SOQUET	S/ 1.50	S/ 4.50
	2		SOQUET DE PARED	S/ 3.00	S/ 6.00
	3		SPRINT COLOR BLANCO		S/ 24.00
	1		SUMIDERO 2"	S/ 5.00	S/ 5.00
			TANGO PARA BAÑO		S/ 3.50
	1		TAPON 1"		S/ 4.00
	1		TAPON 1/2		S/ 1.00
	1		TEE 1/2		S/ 2.00
	1		TEE 2"		S/ 3.00
	3		TEE 3/4	S/ 1.50	S/ 4.50
	1		TEE 4 A 2		S/ 2.00

1		TEFLON		S/ 1.50
1		TEFLON	S/ 1.00	S/ 1.00
1		TEFLON	S/ 1.50	S/ 1.50
1		TEFLON	S/ 1.50	S/ 1.50
1		TEKNOPOR	S/ 18.00	S/ 18.00
	Litro	THINNER		S/ 3.00
1		THINNER 1/2 LITRO		S/ 3.00
1		TINA BATEA		S/ 15.00
1	caja	TIZA		S/ 5.00
1		TOMACORRIENTE DE 5		S/ 10.00
3		TOMACORRIENTE DOBLE	S/ 3.00	S/ 9.00
3		TOMACORRIENTE TRIPLE	S/ 3.50	S/ 10.50
4		TOMACORRIENTE TRIPLE		S/ 22.00
1		TOMACORRIENTE TRIPLE	S/ 8.00	S/ 8.00
1		TRAMPA	S/ 12.00	S/ 12.00
1		TRAMPA 2"		S/ 12.00
1		TRAMPA 2"		S/ 5.00
3	kilos	TRAPO INDUSTRIAL		S/ 10.50
7		TUBO	S/ 1.50	S/ 10.50
5		TUBO 1"	S/ 3.00	S/ 15.00
4		TUBO 1"	S/ 5.00	S/ 20.00
1		TUBO 1/2		S/ 12.00
1		TUBO 2	S/ 6.00	S/ 6.00
1		TUBO 2"		S/ 12.00
1		TUBO 2"		S/ 6.00
1		TUBO 2"		S/ 10.00
8		TUBO 2"	S/ 10.00	S/ 80.00

1		TUBO 2" P/SADO		S/ 12.00
14		TUBO 3/4	S/ 12.00	S/ 168.00
5		TUBO 3/4	S/ 1.80	S/ 9.00
2		TUBO 3/4	S/ 22.00	S/ 44.00
4		TUBO 3/4	S/ 12.00	S/ 48.00
5		TUBO 3/4 LUZ	S/ 1.50	S/ 7.50
20		TUBO 3/4 LUZ	S/ 3.50	S/ 70.00
3		TUBO 3/4 LUZ	S/ 3.50	S/ 10.50
1		TUBO 3/4 LUZ	S/ 4.00	S/ 4.00
6		TUBO 3/4 LUZ NICOLL	S/ 5.00	S/ 30.00
2		TUBO 3/4 LUZ NICOLL	S/ 3.50	S/ 7.00
1		TUBO 4"	S/ 15.00	S/ 15.00
3		TUBO 4"	S/ 16.00	S/ 48.00
1		TUBO 4 P/SADO		S/ 22.00
8		TUBO 6"		S/ 48.00
5		TUBO LUZ	S/ 1.50	S/ 7.50
5		TUBO LUZ 3/4	S/ 1.50	S/ 4.50
1		TUBO NICOLL 1"	S/ 28.00	S/ 28.00
5		TUBOS LUZ	S/ 3.50	S/ 17.50
1		UNION	S/ 1.50	S/ 1.50
1		UNION	S/ 5.00	S/ 5.00
15		VARILLA 1/2	S/ 27.00	S/ 459.00
2		VARILLA 1/2	S/ 28.50	S/ 57.00
1		VARILLA 1/2		S/ 27.50
25		VARILLA 1/2 13MM	S/ 38.80	S/ 970.00
3		VARILLA 1/2 13MM	S/ 28.50	S/ 85.50
1		VARILLA 1/4		S/ 7.00

	6		VARILLA 1/4 6MM	S/ 7.50	S/ 45.00
	2		VARILLA 3/8	S/ 22.00	S/ 44.00
	5		VARILLA 3/8	S/ 17.00	S/ 85.00
	4		VARILLA 3/8	S/ 17.50	S/ 70.00
	50		VARILLA 3/8	S/ 17.00	S/ 850.00
	7		VARILLA 3/8	S/ 17.00	S/ 119.00
	1		VARILLA 8MM	S/ 13.00	S/ 13.00
	32		VARILLA 8MM	S/ 13.70	S/ 438.40
	2		VARILLA 8MM	S/ 16.50	S/ 33.00
	2		VIDRIO		S/ 3.00
	1		WINCHA CABLEADO	S/ 7.00	S/ 7.00
	1		WINCHA X 30		S/ 25.00
	1		YEE 2"		S/ 2.80
	1	kilo	YESO	S/ 1.00	S/ 1.00
	3	kilo	YESO	S/ 1.00	S/ 3.00

Ahora vamos a verificar los campos vacíos en el campo *cantidad*. En donde se encontraron 50 registros y fueron eliminados.

	A	B	C	D	E	F	G
1	Fecha	Cantidad	Unidad/l	PRODUCTO	Precio pd	Total	
				ACCESORIOS		S/ 18.00	
				ACCESORIOS DE AGUA		S/ 10.00	
				ACCESORIOS DE AGUA		S/ 53.20	
				ACCESORIOS DE AGUA 3/4		S/ 24.50	
				ACCESORIOS ELECTRICOS		S/ 120.00	
				ACEITE AZUL		S/ 1.50	
				ALAMBRE		S/ 2.50	
				BISAGRA		S/ 7.00	
				BROCAS DE METAL		S/ 9.00	
				CERROJO		S/ 3.00	
				CLAVO		S/ 6.00	
				CLAVO 1"		S/ 1.00	
				CLAVO 1/4", 1 1/2"		S/ 2.00	
				CLAVO 2"		S/ 1.00	
				CLAVO 2"		S/ 0.50	
				CLAVO 2 1/4"		S/ 1.50	
				CLAVO 4"		S/ 1.50	
				CLAVO CONCRETO		S/ 15.00	
				CONECTOR AGUA		S/ 15.00	
				CORDEL		S/ 6.00	
				CORDEL		S/ 5.00	
				CORDEL		S/ 0.90	
				ELIMINACION DE DESMONTE		S/ 220.00	
				GASOLINA		S/ 32.00	
				HOJA LIJS + KEYTE		S/ 18.00	
				HUACHA		S/ 2.00	

Cantidad	Unidad/m	PRODUCTO	Precio po	Total
		ACCESORIOS		S/ 18.00
		ACCESORIOS DE AGUA		S/ 10.00
		ACCESORIOS DE AGUA		S/ 53.20
		ACCESORIOS DE AGUA 3/4		S/ 24.50
		ACCESORIOS ELECTRICOS		S/ 120.00
		ACEITE AZUL		S/ 1.50
	kilo	ALAMBRE		S/ 2.50
		BISAGRA		S/ 7.00
		BROCAS DE METAL		S/ 9.00
		CERROJO		S/ 3.00
		CLAVO		S/ 6.00
		CLAVO 1"		S/ 1.00
		CLAVO 1/4", 1 1/2"		S/ 2.00
		CLAVO 2"		S/ 1.00
		CLAVO 2"		S/ 0.50
	kilos	CLAVO 2 1/4"		S/ 1.50
	kilos	CLAVO 4"		S/ 1.50
		CLAVO CONCRETO		S/ 15.00
		CONECTOR AGUA		S/ 15.00
		CORDEL		S/ 6.00
		CORDEL		S/ 5.00
		CORDEL		S/ 0.90
		ELIMINACION DE DESMONTE		S/ 220.00
		GASOLINA		S/ 32.00
		HOJA LIJS + KEYTE		S/ 18.00

7/06/2021		HUACHA		S/ 2.00
7/07/2021		LADRILLO		S/ 440.00
3/03/2021		MATERIALES		S/ 70.00
29/12/2019		PAGO DE 200 SOLES AL SR. POR CONCEPTO DE UNA BATERIA DE 2DA		S/ 100.00
2/06/2020		PARIHUELA		S/ 64.00
2/06/2020		PERNO		S/ 1.00
14/09/2020		PERNO		S/ 4.00
20/08/2020		PERNO (PERNITOS)		S/ 4.80
17/04/2021		PERNO AUTOROSCANTE CON ARANDELA		S/ 3.00
3/07/2020		PERNO BAÑO		S/ 3.00
30/05/2020		PERNO BISAGRA		S/ 6.50
28/12/2020		PERNOS SPAK		S/ 5.00
17/06/2020		REPOSICION DE CAJA DE DESAGUE		S/ 15.00
30/12/2019		Repuesto		S/ 10.00
25/02/2021		SALDO		S/ 8.00
6/04/2021		SALDO		S/ 69.00
4/02/2021		SOLDADO DE LA POUURETA		S/ 30.00
19/10/2020		TARUGO NARANJA		S/ 4.00
21/12/2019		TIZA		S/ 0.50
26/08/2021		TIZA COLOR		S/ 1.00
23/01/2021		TORNILLO		S/ 1.00
17/11/2020		TRABAJO		S/ 20.00
4/03/2020		TRAMPA PARA CABLE		S/ 2.00
5/08/2021		TURA GO		S/ 9.00
18/01/2021	-	VARILLA 5/8		S/ 56.00

Lo siguiente que vemos serán los productos no claros, y los estaremos eliminando, tener en cuenta que el registro recogido fue de físico a digital y puede haber algún error.

8/07/2020	1		BROCA DE 3/8	S/ 6.00
8/07/2020	1		BROCA DE 5/16	S/ 4.00
1/03/2021	1		BROCA MADERA	S/ 6.00
21/10/2020	1		BROCA MADERA	S/ 4.00
28/02/2021	1		BROCA MADERA	S/ 6.00
1/03/2021	1		BROCA MADERA	S/ 5.00
26/10/2020	10		BROCAS 1/4 METAL	S/ 3.00
15/07/2020	1		BROCHA	S/ 2.00
25/02/2020	2		BROCHA	S/ 9.00
27/02/2021	1		BROCHA 3"	S/ 4.00
6/06/2020	1		BROCHA 3"	S/ 15.00
14/02/2020	1		BROCHA 4"	S/ 18.00
29/01/2021	1		BROCHA SCHUBERT 6"	S/ 26.00
20/01/2020	1		BUGUI P. CHUDA	S/ 6.50
14/09/2020	5	m	CABLE	S/ 1.20
20/10/2020	10	m	CABLE #12	S/ 1.20
20/01/2021	20	m	CABLE #12	S/ 1.20
15/12/2020	20	m	CABLE #14	S/ 1.00
20/12/2020	8	m	CABLE #14	S/ 1.00
20/12/2020	25	m	CABLE MELLISO #12	S/ 1.20
1/02/2021	30	m	CABLE MELLISO #12	S/ 1.20
6/10/2020	5	m	CABLE MELLISO #14	S/ 0.80
15/01/2020	6	m	CADENA	S/ 12.00

En la siguiente imagen podemos ver alguna

15/01/2020	6	m	CADENA	S/ 12.00
12/02/2021	1	m	CADENA	S/ 12.00
23/07/2021	2	m	CADENA	S/ 5.00
13/09/2020	2		CAJA	S/ 3.00
27/08/2020	1		CAJA 20X20 LUZ	S/ 18.00
28/09/2020	1		CAJA DE DESAGUE	S/ 78.00
27/05/2021	2		CAJA DE DESAGUE X 6 - DE 4 PIESAS	S/ 85.00
8/07/2020	1		CAJA DE LLAVE TERMICA	S/ 5.00
13/04/2021	1		CAJA DE MEDIDOR	S/ 22.00
14/04/2021	5		CAJA DE PASO	S/ 17.00
6/12/2020	2		CAJA DE PASO PEQUEÑO	S/ 5.00
15/07/2021	2		CAJA DE DESAGUE	S/ 85.00
20/01/2021	4		CAJA MODULAR	S/ 4.00
30/09/2020	1		CAJA MODULAR	S/ 3.00
20/11/2020	40		CAJA OCTOGONAL	S/ 1.50
26/06/2020	20		CAJA OCTOGONAL	S/ 1.50
20/02/2020	2		CAJA OCTOGONAL	S/ 1.50
11/03/2021	14		CAJA OCTOGONAL	S/ 1.60
24/04/2021	25		CAJA OCTOGONAL	S/ 1.50
4/03/2021	2		CAJA OCTOGONAL	S/ 1.50
15/07/2020	2		CAJA PARA LLAVE TERMICA	S/ 18.00
25/06/2020	4		CAJA PARA LLAVE TERMICA DE 1	S/ 12.00
3/08/2020	15		CAJA RECTANGULAR	S/ 1.50
24/07/2020	15		CAJA RECTANGULAR	S/ 1.50
20/11/2020	30		CAJA RECTANGULAR	S/ 1.50
30/12/2020	1		CAJA RECTANGULAR	S/ 1.50

En la siguiente imagen observamos un agrupamiento de pedidos. Lo cual vamos a separar.

26/01/2021	1	kilo	CLAVO 2"
4/02/2021	1/2	kilo	CLAVO 2"
11/03/2021	1	kilo	CLAVO 2"
24/03/2021	6	kilo	CLAVO 2"
4/06/2021	2	kilo	CLAVO 2"
4/06/2021	1	kilo	CLAVO 2"
10/09/2020	2	kilo	CLAVO 2"
4/03/2021	1	kilo	CLAVO 2 1/2
18/01/2021	3	kilo	CLAVO 2 1/2"
4/06/2021	2	kilo	CLAVO 2 1/2"
8/12/2020	2	kilo	CLAVO 2 1/2"
10/03/2020	2	kilo	CLAVO 2 Y 2.5"
23/11/2020	3	kilo	CLAVO 2 Y 2.5"
10/02/2021	6	kilo	CLAVO 2, 2 1/2, 3
5/03/2020	4	kilo	CLAVO 3"
5/10/2020	1	kilo	CLAVO 3"
7/10/2020	1	kilo	CLAVO 3"
20/10/2020	16		CLAVO 3"
8/12/2020	1	kilo	CLAVO 3"
18/01/2021	2	kilo	CLAVO 3"
22/01/2021	1/2	kilo	CLAVO 3"
4/03/2021	1	kilo	CLAVO 3"
24/03/2021	4	kilo	CLAVO 3"
6/05/2021	1	kilo	CLAVO 3"

Los datos de amarillo son las columnas que estaban y ahora los emos separados agregando nuevas columnas, los datos emos dividido en partes iguales.

24/03/2021	6	kilo	CLAVO 2"	S/ 6.00
4/06/2021	2	kilo	CLAVO 2"	S/ 6.00
4/06/2021	1	kilo	CLAVO 2"	S/ 6.00
10/09/2020	2	kilo	CLAVO 2"	S/ 5.00
4/03/2021	1	kilo	CLAVO 2 1/2"	S/ 6.00
18/01/2021	3	kilo	CLAVO 2 1/2"	S/ 6.00
4/06/2021	2	kilo	CLAVO 2 1/2"	S/ 6.00
8/12/2020	2	kilo	CLAVO 2 1/2"	S/ 5.00
10/03/2020	1	kilo	CLAVO 2"	S/ 5.00
23/11/2020	1 1/2	kilo	CLAVO 2"	S/ 5.00
10/02/2021	2	kilo	CLAVO 2"	S/ 6.00
10/03/2020	1	kilo	CLAVO 2 1/2"	S/ 5.00
23/11/2020	1 1/2	kilo	CLAVO 2 1/2"	S/ 5.00
10/02/2021	2	kilo	CLAVO 2 1/2"	S/ 6.00
10/02/2021	2	kilo	CLAVO 3"	S/ 6.00
5/03/2020	4	kilo	CLAVO 3"	S/ 5.00
5/10/2020	1	kilo	CLAVO 3"	S/ 5.00
7/10/2020	1	kilo	CLAVO 3"	S/ 5.00
20/10/2020	16		CLAVO 3"	S/ 0.30
8/12/2020	1	kilo	CLAVO 3"	S/ 5.00
18/01/2021	2	kilo	CLAVO 3"	S/ 6.00
22/01/2021	1/2	kilo	CLAVO 3"	S/ 3.00
4/03/2021	1	kilo	CLAVO 3"	S/ 6.00
24/03/2021	4	kilo	CLAVO 3"	S/ 6.00

En la siguiente imagen observamos como los separamos.

22/01/2021	1/2	kilo	CLAVO 3"	S/ 3.00	S/	3.00
4/03/2021	1	kilo	CLAVO 3"	S/ 6.00	S/	6.00
24/03/2021	4	kilo	CLAVO 3"	S/ 6.00	S/	24.00
6/05/2021	1	kilo	CLAVO 3"	S/ 6.00	S/	6.00
4/06/2021	1	kilo	CLAVO 3"	S/ 6.00	S/	6.00
7/06/2021	1	kilo	CLAVO 3"	S/ 6.00	S/	6.00
24/05/2021	1	kilo	CLAVO 3"	S/ 6.00	S/	6.00
24/05/2021	1	kilo	CLAVO 2"	S/ 6.00	S/	6.00
7/08/2020	6	kilo	CLAVO 4"	S/ 5.00	S/	30.00
7/10/2020	1	kilo	CLAVO 4"	S/ 5.00	S/	5.00
9/11/2020	1		CLAVO 4"	S/ 1.00	S/	1.00
28/12/2020	2	kilo	CLAVO 4"	S/ 5.00	S/	10.00
18/01/2021	2	kilo	CLAVO 4"	S/ 6.00	S/	12.00
23/07/2021	4	kilo	CLAVO 4"	S/ 5.50	S/	22.00
15/03/2020	1/2	kilo	CLAVO 4"	S/ 5.00	S/	2.50
15/03/2020	1/2	kilo	CLAVO 3"	S/ 5.00	S/	2.50
29/04/2021	30		CLAVO ACERO 2"	S/ 0.30	S/	9.00
29/08/2020	10		CLAVO ACERO 2 1/2"	S/ 0.30	S/	3.00
1/12/2020	25		CLAVO ACERO 3"	S/ 0.40	S/	10.00
22/04/2021	30		CLAVO ACERO 3"	S/ 0.30	S/	9.00
19/06/2020	10		CLAVO ACERO 3"	S/ 0.30	S/	3.00
10/05/2021	10		CLAVO ACERO 3"	S/ 0.40	S/	4.00
8/01/2021	1/4		CLAVO CALAMINA	S/ 10.00	S/	2.50
27/02/2021	6		CLAVO CEMENTO 2 1/2"	S/ 0.33	S/	2.00

Otros nombres no claros encontrado es las siguientes, los cuales vamos eliminar.

17/06/2021	2		COMBA 4 LIBRAS	S/ 14.00	S/	28.00
1/10/2020	4		CONECTOR	S/ 0.50	S/	2.00
26/08/2021	3		CONECTOR 3/4	S/ 0.50	S/	1.50
23/09/2020	36		CONQUITO DE ATENUANTE	S/ 0.30	S/	10.80
1/06/2020	1		CORDEL DE NAYLON	S/ 8.00	S/	8.00
5/04/2021	2		CRUCETA 3"	S/ 4.00	S/	8.00
5/04/2021	4		CRUSETA 2"	S/ 3.50	S/	14.00

14/01/2021	2		GALONE DE PINTURA NEGRO GLOSS	S/ 40.00	S/	80.00
10/07/2020	5		GALONES DE GASOLINA	S/ 12.00	S/	60.00
29/12/2020	1/2	kilo	GAMEZAN	S/ 16.00	S/	8.00
3/06/2020	17		GANCHO DE TECHO	S/ 1.20	S/	20.40
22/02/2021	6		GUAYAQUILO	S/ 13.33	S/	80.00
11/01/2021	25	m	HILO	S/ 0.20	S/	5.00
14/01/2021	1		HILO DE PESCAR	S/ 5.00	S/	5.00

27/03/2021	2		HOJA SIERRA SANDFLEX	S/ 5.00	S/	10.00
18/07/2020	3		HORAS TRABAJO MAQUINA	S/ 200.00	S/	600.00
19/06/2020	1		HORMIGUI SIDA	S/ 3.00	S/	3.00
12/11/2020	18		HUACHA PLANA	S/ 0.20	S/	3.20
12/11/2020	12		HUACHA PRECION	S/ 0.15	S/	1.80

20/03/2021	7		NIPLES BRONCE 1/2	S/ 4.00	S/	28.00
5/12/2020	1		NIVEL	S/ 13.00	S/	13.00
5/05/2021	1		NYLON	S/ 6.00	S/	6.00
9/01/2021	39		O 3/8	S/ 20.30	S/	791.70
13/01/2020	1	bolsa	OCRE	S/ 10.00	S/	10.00
7/10/2020	1 1/2	kilo	OCRE	S/ 10.00	S/	15.00

6/06/2020	5		PAR ANGULO DE 1/2	S/ 2.00	S/	10.00
23/03/2020	1		PAR DE 2	S/ 2.00	S/	2.00
23/03/2020	1		PAR DE 2.5	S/ 2.50	S/	2.50
23/03/2020	2		PAR DE BISAGRA	S/ 4.00	S/	8.00
17/04/2021	2		PAR DE BISAGRA	S/ 4.00	S/	8.00

19/06/2020	4		PAR VISAGRA DE 4"	S/ 5.00	S/	20.00
11/08/2020	2		PARCHE DE LLANTA DE CARRETILLA	S/ 5.00	S/	10.00
17/02/2020	1		PARIHUELA	S/ 10.00	S/	10.00
2/07/2020	1		PARIHUELA	S/ 10.00	S/	10.00
10/10/2020	2		PAVCO	S/ 1.00	S/	2.00
9/01/2020	1		PEGAMENTO	S/ 50.00	S/	50.00

21/07/2021	1		PEGAMENTO OATEY	S/ 20.00	S/	20.00
20/11/2020	1		PEGAMENTO OATEY	S/ 9.00	S/	9.00
16/01/2021	1		PEGAMENTO OD	S/ 5.00	S/	5.00
1/12/2020	3		PEGAMENTO PARA MOSCAS	S/ 2.00	S/	6.00
7/06/2021	2		PEGAMENTO PORCELANATO	S/ 18.00	S/	36.00

18/07/2020	12		PERNO COMPLETO	S/ 1.20	S/	12.00
20/08/2020	6		PERNOS TUERCA Y HUACHA	S/ 1.50	S/	9.00
16/12/2020	1	m	PI...	S/ 55.00	S/	55.00
5/03/2020	2		PIE DERECHO	S/ 14.00	S/	28.00
27/02/2020	20	m3	PIEDRA BASE	S/ 43.00	S/	860.00

17/11/2020	1		SUMIDERO 4"	S/ 10.00	S/	10.00
27/06/2020	1		SUMIDERO DE LAVADERO	S/ 22.00	S/	22.00
15/04/2021	10		SUPER FLU	S/ 0.80	S/	8.00
16/01/2021	2		T... CAÑO	S/ 8.50	S/	17.00
5/08/2021	1		TANQUE 1100 LT	S/ 410.00	S/	410.00
24/06/2020	1		TANQUE 1100 LT	S/ 350.00	S/	350.00

En la siguiente imagen vamos a remplazar *fierro por varillas*. En la primera como estaban y la segunda como queda con el cambio.

Fecha	Día	Mes	Año	Cantidad	PRODUCTO	IT	Precio po	Total
19/02/2020	19		2020	2	DISCOS FIERRO		S/ 5.00	S/ 10.00
7/10/2020	7	10	2020	1	ESCOBILLA DE FIERRO		S/ 5.00	S/ 5.00
9/03/2020	9	3	2020	6	FIERRO DE 1/4		S/ 7.00	S/ 42.00
9/03/2020	9	3	2020	4	FIERRO DE 12MM		S/ 27.00	S/ 108.00
6/06/2020	6	6	2020	3	FIERRO DE 3/4		S/ 17.00	S/ 51.00
13/08/2021	13	8	2021	3	FIERRO DE 3/8		S/ 22.50	S/ 67.50
15/07/2020	15	7	2020	8	FIERRO DE 3/8		S/ 17.00	S/ 136.00
18/01/2020	18	1	2020	6	FIERRO DE 8MM		S/ 12.00	S/ 72.00

Fecha	Día	Mes	Año	Cantidad	PRODUCTO	IT	Precio po	Total
19/02/2020	19		2020	2	DISCOS FIERRO		S/ 5.00	S/ 10.00
7/10/2020	7	10	2020	1	ESCOBILLA DE FIERRO		S/ 5.00	S/ 5.00
9/03/2020	9	3	2020	6	VARILLA 1/4		S/ 7.00	S/ 42.00
9/03/2020	9	3	2020	4	VARILLA 12MM		S/ 27.00	S/ 108.00
6/06/2020	6	6	2020	3	VARILLA 3/4		S/ 17.00	S/ 51.00
13/08/2021	13	8	2021	3	VARILLA 3/8		S/ 22.50	S/ 67.50
15/07/2020	15	7	2020	8	VARILLA 3/8		S/ 17.00	S/ 136.00
18/01/2020	18	1	2020	6	VARILLA 8MM		S/ 12.00	S/ 72.00

Algunos servicios que se han brindado también se eliminarán.

1/06/2020	6		TORNILLO DE 60X1"	S/ 0.30	S/	1.80
28/12/2020	20		TORNILLO STOVEBOLT	S/ 0.90	S/	18.00
10/05/2021	1		TORTOL	S/ 3.00	S/	3.00
11/07/2020	4	horas	TRABAJO MAQUINA	S/ 200.00	S/	800.00
12/09/2020	1		TRAMPA	S/ 5.00	S/	5.00
10/05/2021	1		TRAMPA	S/ 10.00	S/	10.00

25/05/2021	6		VARILLA 8MM	S/ 16.50	S/ 99.00
4/06/2021	2		VARILLA 8MM	S/ 18.50	S/ 37.00
11/07/2020	1		VIAJE DE DESMONTE	S/ 200.00	S/ 200.00
25/07/2021	5		VIAJES DE DESMONTE REALIZADO CON DOS MAQUINAS	S/ 250.00	S/ 1,250.00
2/11/2020	2		VIAJES DE ELIMINACION DE DESMONTES	S/ 200.00	S/ 400.00
30/05/2020	1		VOLQUETADA DE RIPEO 15M3	S/ 300.00	S/ 300.00
1/06/2020	1		VOLQUETADA DE RIPIO	S/ 300.00	S/ 300.00
21/02/2020	1		WINCHA 5M	S/ 5.00	S/ 5.00
1/06/2020	1		WINCHA 5M	S/ 5.00	S/ 5.00
13/01/2021	1		WINCHA 5M	S/ 5.00	S/ 5.00
2/11/2020	1		Y SANITARIO	S/ 15.00	S/ 15.00
26/06/2020	4		YEE 2"	S/ 3.00	S/ 12.00
18/04/2020	1		YEE 2"	S/ 5.00	S/ 5.00

Fecha	Día	Mes	Año	Cantidad	PRODUCTO	Precio po	Total
18/07/2020	18	7	2020	3	HORAS TRABAJO MAQUINA	S/ 200.00	S/ 600.00

- Después de realizar la limpieza nos queda los siguientes productos, total de 494, de un total de 2546 registros.

PRODUCTO
ABRAZADERA
ACCESORIOS
ACEITE 3 EN 1
ACEITE AZUL
ACIDO MURIATICO
ADAPTADOR
ADAPTADOR 1"
ADAPTADOR 1/2"
ADAPTADOR 3/4"
ADAPTADOR 3/4 A 1"
ADAPTADOR DE TOMACORRIENTE
ALAMBRE
ALAMBRE #16
ALAMBRE #2
ALAMBRE #8
ALDABA
ARANDELAS A/ PRECION
ARCO SIERRA
ARENA FINA
ARENA GRUESA
ARMELLA
AGREGADOS
BADILEJO MANGO MADERA
BALDE DE PINTURA
BALDE DE PINTURA BLANCA
BALDE DE PINTURA NEGRA X 4 GALONES

BALDE SIKA IMPERMEABLE
BALDE
BALDE VACIOS
BANDE DE PINTURA
BASE FLOURECENTE
BATEAS
BISAGRA 4"
BLOQUETA
BLOQUETAS
BOLSA DE AGUA
BOLSA IMPRIMANTE
BOLSA PINTURA
BOLSA TEMPLE
BOLSAS VACIAS DE CEMENTO
BOQUETAS
BOTE DE 100 UNIDADES DE SOU FLEXCIERRA
BOTE VACIO PARA PINTURA
BROCA
BROCA CEMENTO 1/2
BROCA DE 3/8
BROCA DE 5/16
BROCA MADERA
BROCAS 1/4 METAL
BROCHA
BROCHA 3"
BROCHA 4"
BROCHA SCHUBERT 6"
CABLE
CABLE #12
CABLE #14
CABLE MELLISO #12
CABLE MELLISO #14
CADENA
CAJA
CAJA 20X20 LUZ
CAJA DE DESAGUE
CAJA DE DESAGUE X 6 - DE 4 PIESAS
CAJA DE LLAVE TERMICA
CAJA DE MEDIDOR
CAJA DE PASO
CAJA DE PASO PEQUEÑO
CAJA MODULAR
CAJA OCTOGONAL
CAJA DE LLAVE TERMICA DE 1

CAJA RECTANGULAR
CAJITA LUZ PAVCO
CAL NIEVE
CAL OBRA
CALAMINA 3.6
CANALETA
CANDADO
CANDADOS
CAÑO
CAÑO 1/2
CAÑO 1/2 AGUA
CAÑO BRONCE
CAÑO S. CHASER
CARBON
CARRETILLA (BUGUI)
CEMENTO
CEMENTO ANDINO ULTRA
CEMENTO APU
PEGAMENTO MAYOLICA
CEMENTO NACIONAL
CEMENTO NACIONAL - ANTI SALITRE AZUL
CEMENTO QUISQUEYA
CEMENTO QUISQUEYA NARANJA
CEMENTO QUISQUEYA TIPO I
CEMENTO QUISQUEYA VERDE
CEMENTO SOL
CERROJO DE PUERTA
CHAPA
CHAPA SCOLTA
CHAPAFORTE
CILINDRO
CINTA AISLANTE
CINTA MASKING
CINTA TEFLON
CLAVO
CLAVO 1"
CLAVO 1 1/2"
CLAVO 2"
CLAVO 2 1/2"
CLAVO 3"
CLAVO 4"
CLAVO ACERO 2"
CLAVO ACERO 2 1/2"
CLAVO ACERO 3"

CLAVO CALAMINA
CLAVO CEMENTO 2 1/2"
CLAVO DE ACERO
CLAVO ETERNIT
CLAVO PUA.
CODO
CODO 1
CODO 1/2
CODO 1/2 A/PRECION
CODO 1/2 BRONCE
CODO 1/2 MIXTO AGUA
CODO 1/2 ROSCA
CODO 2
CODO 2 X 45
CODO 3/4
CODO 4
CODO 4 A 2"
CODO 4 X 2"
CODO 4 X 4"
CODO 4 X 45
CODO 4 X 90
CODO 6
CODO MIXTO 1/2
COMBA 4 LIBRAS
CONECTOR
CONECTOR 3/4
CORDEL DE NAYLON
CRUCETA 3"
CRUSETA 2"
CUADRO
CUELLO DE CERA
CURVA 3/4
CURVA 1"
CURVA 1/2 SAP
CURVA 3"
CUTER
DESARMADOR
DESINFECTANTE DE KRESSO
DISCO
DISCO COPA
DISCO COPA TROUSA
DISCO CORTE
DISCO CORTE 14"
DISCO CORTE 4"

DISCO CORTE 4.5"
DISCO CORTE 7"
DISCO CORTE 7.5"
DISCO CORTE CONCRETO
DISCO CORTE MADERA
DISCO CORTE MAYOLICA
DISCO CORTE METAL
DISCO CORTE METAL 14"
DISCO CORTE NORTON
DISCO DESGASTE
DISCO PULIR
DISCOS FIERRO
DUCHA
ELECTRODO PUNTO AZUL
ELIMINACION DE DESMONTE
ENCHUFE
ESCOBA
ESCOBA DE ACERO
ESCOBILLA DE ACERO
ESCOBILLA DE FIERRO
ESPATULA
ESPONJA
ESPRAY Y MATA MOSCAS
ETERNIT
ETERNIT 1 X 3M
VARILLA
VARILLA 1/2
FLUORESCENTE COMPLEX
FOCO
FOCO AHORRADOR
FOCO LED
FRAGUA
FRAGUA PORCELANA EXTRA FUERTE
GALON DE LEJIA
GALON DE PINTURA GLOSS
GALON DE THINNER
GALON PINTURA
GALON PRESERVANTE
GALONE DE PINTURA BASE GRISS
GALONE DE PINTURA GLLOSS
GALONE DE PINTURA NEGRO GLOSS
GALONES DE GASOLINA
GAMEZAN
GANCHO DE TECHO

HILO
HILO DE PESCAR
HILO DE PESCAR 0.8
HISOPO PARA BAÑO
HOJA DE SANDFLEX
HOJA LIJA
HOJA LIJA # 100
HOJA LIJA #40
HOJA LIJA #80
HOJA SANDFLEX
HOJA SIERRA
HOJA SIERRA SANDFLEX
HUACHA PLANA
HUACHA PRECION
IMPRIMANTE
IMPRIMANTE BLANCO
INODORO
INODORO WINDO
INTERRUPTOR
INTERRUPTOR BTICINO
INTERRUPTOR DOBLE
INTERRUPTOR TRIPLE
INTURRUPTOR
ISOPO PARA BAÑO
JUEGO BROCA - SACABOCADO
JUEGO BROVA MAFRTS
KILO SIKA
KK 18 HUECOS
KK 18 HUECOS LARK
KK RUSTICO
KOBRE ROJO
LADRILLO BLOQUETAS
LADRILLO PANDERETA
LADRILLO TECHO 30X30X15
LAMPA
LAMPA TRAMONTINA
LAMPARA PRETUL
LAPIZ DE CARPINTERIA
LAVADERO
LAVADERO - CON FREGADERA
LAVADERO DE 1M X 50 CUT
LENTE
LIMA
LIMA PLANA

LISTON
LISTON 2X3X4M
LISTON 2X3X5M
LISTON 3X2X3.7
LISTON 3X2X5M
LISTON 4X2X3.5
LISTON DE 5MM DE 3X2
LLANTA CARRETILLA
LLANTA CON CAMARA PARA CARRETILLA
LLAVE
LLAVE 1/2 AUTOMATICO
LLAVE ASA
LLAVE DE DUCHA
LLAVE DE GAS
LLAVE DE PASO METAL
LLAVE DIFERENCIAL
LLAVE DIFERENCIAL 30MA
LLAVE ESTILSON
LLAVE PASO
LLAVE PASO 1"
LLAVE PASO 1/2
LLAVE PASO 3/4
LLAVE TERMICA
LLAVE TERMICA C32 - TICINO
LLAVE TRIFASICA
LLAVE UNIVERSAL
LLAVE UNIVERSAL 1"
LLAVE UNIVERSAL 1/2
LLAVE UNIVERSAL PAVCO
LLAVES JARDINERO
MALLA RACHEL
MANGUERA
MANGUERA DE GAS
MANGUERA DE NIVEL
MARTILLO
MASILLA PARA META 2 ANIPSA
MEDIDOR DIGITAL
MULTICONECTOR
MURETE DE LUZ MONOFOSICO
NIPLE 1/2
NIPLE 1/2 X 1.5"
NIPLES 3/4
NIPLES BRONCE 1/2
NIVEL

NYLON
OCRE
OCRE AMARILLO DE 2 kg
OCRE ROJO
ORMIGON
PABILO
PALETA DE BOTIN
PALO DE ESCOBA
PALOS
PANALES 1/2
PANEL
PANEL ABIERTO
PANEL CERRADO
PANEL PUERTA Y VENTANA
PANEL VERTANA
PAQUETE DE ESPONJA
PAR ANGULO DE 1/2
PAR DE BISAGRA
PAR DE BISAGRA 1/2
PAR DE BISAGRA 3"
PAR DE GUANTE
PAR GUANTE BADANA
PAR VISAGRA DE 4"
PARIHUELA
PEGAMENTO
PEGAMENTO 118ML
PEGAMENTO AZUL
PEGAMENTO AZUL 1/4
PEGAMENTO GRIS
PEGAMENTO OATEY
PEGAMENTO PARA MOSCAS
PEGAMENTO PORCELANATO
PEGAMENTO PVC
PERNO
PERNO 1 1/2
PERNO 6"
PERNO CAMA
PERNO COMPLETO
PERNOS TUERCA Y HUACHA
PIE DERECHO
PIEDRA BASE
PIEDRA CHANCADA
PIEDRA GRUESA
PIEDRA ZANJA

PINTURA BLANCA
PINTURA BOLSA
PINTURA ESPRAY
PINTURA GLOSS
PINTURA LATEX COLOR MAIZ
PINTURA N.A
PINTURA VERDE
PLANCHA
PLANCHA DE ETERNIT
PLANCHA DE TECNOPOR
PLANCHA FIBRA FORTE TRAPLUCIDO
PLANCHA PULIR
PLANCHA TECNOPOR
PLANCHA TECNOPOR 3/4
PLANCHAS CALAMINA
PLANCHAS FIBRA TRASFUCIDO
PLANCHAS FIBRAFORTE
PLASTICO
PLASTICO 2 M - DOBLE ANCHO
PLOMADA
PLOMADO
PUERTA
REDUCCION 1 A 3/4
REDUCCION 2 A 1.25"
REDUCCION 3/4 A 1/2
REDUCCION 4 A 2"
REDUCCION 6 A 4"
REDUCCION DE 3/9 A 1/2 CON ROSCA
REGADERA
REGLA ALUMINIO
REJILLA 2"
RIPIO
RODILLO
RODOPLAST
ROLLO CABLE #12
ROLLO CABLE #14
ROLLO CABLE #16
TANQUE ROTOPLAS 1100 Lt
SACA GRASA
SACA OXIDO
SIKA
SIKA EN POLVO
SILICONA
SINCEL

SODA CAUSTICA
SOLDADURA PUNTO AZUL
SOQUET
SUMIDERO
SUMIDERO 2"
SUMIDERO 4"
SUMIDERO DE LAVADERO
TANQUE ETERNIT 1100 LT
TANQUE ROTOPLAS 2500 Lt
TANQUE AZUL
TANQUE DE AGUA
TANQUE DE INODORO
TANQUE FARPLAS 1100 Lt
TANQUE FARPLAS 2500 Lt
TAPAS CIEGA
TAPON 1/2
TAPON 2"
TARUGO
TARUGO 1/2
TEE
TEE 1/2
TEE 2"
TEE 2" A 2"
TEE 3/4
TEE 4
TEE 4 A 2
TEE 4X4
TEE 6 A 4
TEE 6 A 4"
TEFLON
TEMPLE X 25 K.
TEMPLE X 5K
TIRAFON
TIRAFON 5/8 X 1.5
TIRAFON DE 1/2
TIZA
TOMACORRIENTE
TOMACORRIENTE 3
TOMACORRIENTE DE 4
TOMACORRIENTE DOBLE
TOMACORRIENTE TRIPLE
TORNILLO
TORNILLO #30
TORNILLO DE 50X3"

TORNILLO DE 60X1"
TORNILLO STOVEBOLT
TORTOL
TRAMPA
TRAMPA 2"
TRAPO INDUSTRIAL
TRAPO INDUSTRIAL
TRIPLAY
TUBO
TUBO 1"
TUBO 1/2
TUBO 2"
TUBO 2" MATUSITA
TUBO 2" NICOLL
TUBO 2" PESADO
TUBO 3/4
TUBO 3/4 AGUA
TUBO 3/4 AGUA P/D
TUBO 3/4 LUZ
TUBO 3/4 LUZ SAP
TUBO 3/4 NICOLL
TUBO 4"
TUBO 4" DESAGUE
TUBO 4" LIVIANA
TUBO 4" NICOLL
TUBO 4" PESADO
TUBO 5/4
TUBO 6"
TUBO 6" PESADO
TUBO ABASTO
TUBO ABASTO 1/2
TUBO ABASTO 3/4
TUBO CURVOS LUZ
TUBO DE 2
TUBO DE 2"
TUBO DE 3/4
TUBO DE 4"
TUBO DE AGUA DE 1/2
TUBO DEVUELTO 4-2" A/C
TUBO LUZ
TUBO LUZ 1"
TUBO MARCA AMERICAN
TUBO VISOR
TUBOS AGUA A/P

TUBOS LUZ
UNION
UNION 1"
UNION 1/2
UNION 1/5
UNION 2"
UNION 3/4
UNION 4"
UNION 6"
VADILEJO
VALVULA DE COCINA A GAS DOMESTICO
VARILLA 1/2 12MM
VARILLA 1/2 13MM
VARILLA 1/4
VARILLA 1/4 6MM
VARILLA 12MM
VARILLA 3/4
VARILLA 3/8
VARILLA 5/8
VARILLA 8MM
WINCHA 5M
YEE 2"
YEE 4 A 2"
YEE 6"
YEE DE 4X2
YEE DE 4X4
YESO
ZAPOLIO

Si realizamos la correcta relación con la cantidad según la fecha que corresponde nos saldría los gimiente.

- En la siguiente vamos agrupar el producto en un solo, de varios nombres.

Siguiendo ahora vamos a nombrar genérico, en unos más específicos *soldadura punto azul = Electrodo punto azul*.

Cantidad	PRODUCTO	Precio po
1	ACEITE AZUL	S/ 3.00
100	CEMENTO NACIONAL - ANTI SALITRE AZUL	S/ 22.00
120	CEMENTO NACIONAL - ANTI SALITRE AZUL	S/ 22.00
12	CEMENTO NACIONAL - ANTI SALITRE AZUL	S/ 22.00
1	ELECTRODO PUNTO AZUL	S/ 17.00
1	PEGAMENTO AZUL	S/ 12.00
1	PEGAMENTO AZUL	S/ 34.00
1	PEGAMENTO AZUL	S/ 5.00
1	PEGAMENTO AZUL	S/ 5.00
1	PEGAMENTO AZUL	S/ 42.00
1	PEGAMENTO AZUL	S/ 5.00
1	PEGAMENTO AZUL 1/4	S/ 36.00
1	PEGAMENTO AZUL 1/4	S/ 38.00
2	SOLDADURA PUNTO AZUL	S/ 15.00
1	SOLDADURA PUNTO AZUL	S/ 16.00
1	SOLDADURA PUNTO AZUL	S/ 16.00
2	SOLDADURA PUNTO AZUL	S/ 15.00
1	SOLDADURA PUNTO AZUL	S/ 15.00
1	SOLDADURA PUNTO AZUL	S/ 15.00
1	SOLDADURA PUNTO AZUL	S/ 15.00
5	SOLDADURA PUNTO AZUL	S/ 16.00

Adaptador va ser iguala a adaptador 1", adaptador 1/2", adaptador 3/4, etc.

Cantidad	PRODUCTO	Precio po	Total
3	ADAPTADOR 1"	S/ 2.00	S/ 6.00
1	ADAPTADOR 1/2"	S/ 1.00	S/ 1.00
5	ADAPTADOR 1/2"	S/ 1.00	S/ 5.00
2	ADAPTADOR 1/2"	S/ 0.50	S/ 1.00
8	ADAPTADOR 1/2"	S/ 1.40	S/ 11.20
2	ADAPTADOR 1/2"	S/ 1.00	S/ 2.00
4	ADAPTADOR 1/2"	S/ 1.00	S/ 4.00
1	ADAPTADOR 1/2"	S/ 1.00	S/ 1.00
1	ADAPTADOR 1/2"	S/ 1.50	S/ 1.50
1	ADAPTADOR 1/2"	S/ 1.00	S/ 1.00
10	ADAPTADOR 1/2"	S/ 1.50	S/ 15.00
4	ADAPTADOR 1/2"	S/ 1.00	S/ 4.00
2	ADAPTADOR 1/2"	S/ 2.00	S/ 4.00
1	ADAPTADOR 3/4"	S/ 3.00	S/ 3.00
1	ADAPTADOR 3/4"	S/ 1.50	S/ 1.50
1	ADAPTADOR 3/4"	S/ 1.00	S/ 1.00
2	ADAPTADOR 3/4"	S/ 10.00	S/ 20.00
2	ADAPTADOR 3/4 A 1"	S/ 4.00	S/ 8.00
1	ADAPTADOR 3/4 A 1"	S/ 4.00	S/ 4.00

Cantidad	PRODUCTO
3	ADAPTADOR
1	ADAPTADOR
5	ADAPTADOR
2	ADAPTADOR
8	ADAPTADOR
2	ADAPTADOR
4	ADAPTADOR
1	ADAPTADOR
1	ADAPTADOR
1	ADAPTADOR
10	ADAPTADOR
4	ADAPTADOR
2	ADAPTADOR
1	ADAPTADOR
1	ADAPTADOR
1	ADAPTADOR
2	ADAPTADOR
2	ADAPTADOR
1	ADAPTADOR

El otro producto es *alambre*, que fue agrupado de alambre, alambre 8, alambre 16.

Cantidad	PRODUCTO
6	ALAMBRE
5	ALAMBRE
11	ALAMBRE
1	ALAMBRE
7	ALAMBRE
10	ALAMBRE
5	ALAMBRE
1	ALAMBRE
2	ALAMBRE
1	ALAMBRE
2	ALAMBRE
1	ALAMBRE
1	ALAMBRE
3	ALAMBRE
20	ALAMBRE
2	ALAMBRE
10	ALAMBRE
30	ALAMBRE
20	ALAMBRE
10	ALAMBRE
30	ALAMBRE
5	ALAMBRE
2	ALAMBRE

Ahora, aremos los mismo con tubo, ya que contamos con tubo 1", tubo 1/2, tubo 2", tubo 2" matusita, tubo 2" nicoll, tubo 2" pesado, tubo 3/4, tubo 4", tubo 5/4, etc.

3	TUBO 6"
2	TUBO 6"
3	TUBO 6"
1	TUBO 6" PESADO
1	TUBO 6" PESADO
1	TUBO ABASTO
1	TUBO ABASTO 1/2
1	TUBO ABASTO 3/4
16	TUBO CURVOS LUZ
5	TUBO DE 2
1	TUBO DE 2"
2	TUBO DE 2"
1 4/9	TUBO DE 3/4
1	TUBO DE 4"
1	TUBO DE 4"
2	TUBO DE AGUA DE 1/2
48	TUBO DEVUELTO 4-2" A/C
1	TUBO LUZ
3	TUBO LUZ 1"
5	TUBO LUZ 1"
1	TUBO MARCA AMERICAN
1	TUBO VISOR
4	TUBOS AGUA A/P

2	TUBO
3	TUBO
1	TUBO
1	TUBO
1	TUBO
1	TUBO
1	TUBO
16	TUBO
5	TUBO
1	TUBO
2	TUBO
1 4/9	TUBO
1	TUBO
1	TUBO
2	TUBO
48	TUBO
1	TUBO
3	TUBO
5	TUBO
1	TUBO
1	TUBO
4	TUBO
2	TUBO

Con el producto codo es casi parecido, teniendo varias variantes de la misma, vamos agruparlo en los mismo para evitar campos vacíos.

Cantidad	PRODUCTO
2	CODO
2	CODO
6	CODO
1	CODO 1
4	CODO 1
4	CODO 1
3	CODO 1
6	CODO 1
2	CODO 1
2	CODO 1/2
1	CODO 1/2
1	CODO 1/2
3	CODO 1/2
10	CODO 1/2
4	CODO 1/2
2	CODO 1/2
2	CODO 1/2
3	CODO 1/2
2	CODO 1/2
6	CODO 1/2
10	CODO 1/2
2	CODO 1/2

Cantidad	PRODUCTO
2	CODO
2	CODO
6	CODO
1	CODO
4	CODO
4	CODO
3	CODO
6	CODO
2	CODO
2	CODO
1	CODO
1	CODO
3	CODO
10	CODO
4	CODO
2	CODO
2	CODO
3	CODO
2	CODO
6	CODO
10	CODO
2	CODO

El que sigue es el producto reducción 1 a $\frac{3}{4}$, reducción 2 a $1 \frac{1}{4}$, etc.

Cantidad	PRODUCTO
2	REDUCCION 1 A $\frac{3}{4}$
1	REDUCCION 2 A 1.25"
1	REDUCCION $\frac{3}{4}$ A $\frac{1}{2}$
1	REDUCCION $\frac{3}{4}$ A $\frac{1}{2}$
1	REDUCCION $\frac{3}{4}$ A $\frac{1}{2}$
1	REDUCCION $\frac{3}{4}$ A $\frac{1}{2}$
1	REDUCCION $\frac{3}{4}$ A $\frac{1}{2}$
1	REDUCCION $\frac{3}{4}$ A $\frac{1}{2}$
1	REDUCCION $\frac{3}{4}$ A $\frac{1}{2}$
1	REDUCCION 4 A 2"
4	REDUCCION 6 A 4"
1	REDUCCION DE $\frac{3}{9}$ A $\frac{1}{2}$ CON ROSCA

Cantidad	PRODUCTO
2	REDUCCION
1	REDUCCION
1	REDUCCION
1	REDUCCION
1	REDUCCION
1	REDUCCION
1	REDUCCION
1	REDUCCION
1	REDUCCION
4	REDUCCION
1	REDUCCION

Con siguiente es el tee, de los que tenemos tee ½, tee 2, tee 2 a 2, te ¾, tee 4, etc.

Cantidad	PRODUCTO
2	TEE
1	TEE
5	TEE 1/2
3	TEE 1/2
2	TEE 1/2
1	TEE 1/2
5	TEE 1/2
3	TEE 1/2
1	TEE 1/2
2	TEE 1/2
1	TEE 1/2
1	TEE 2"
1	TEE 2"
5	TEE 2"
3	TEE 2"
6	TEE 2"
1	TEE 2" A 2"
3	TEE 3/4
1	TEE 4
2	TEE 4
1	TEE 4
10	TEE 4

Cantidad	PRODUCTO
2	TEE
1	TEE
5	TEE
3	TEE
2	TEE
1	TEE
5	TEE
3	TEE
1	TEE
2	TEE
1	TEE
1	TEE
1	TEE
5	TEE
3	TEE
6	TEE
1	TEE
3	TEE
1	TEE
2	TEE
1	TEE
10	TEE

El producto Unión que se deriva en unión 1", unión 1/2, unión 1/5, unión 2", unión de 3/4, unión de 4" y unión de 6". Todo estos variables vamos agruparlas.

8	UNION
1	UNION
2	UNION 1"
8	UNION 1"
1	UNION 1"
1	UNION 1/2
2	UNION 1/2
20	UNION 1/2
2	UNION 1/2
10	UNION 1/2
2	UNION 1/2
1	UNION 1/2
2	UNION 1/2
3	UNION 1/5
1	UNION 2"
3	UNION 2"
1	UNION 3/4
3	UNION 3/4
5	UNION 3/4
1	UNION 3/4
1	UNION 3/4
3	UNION 4"
4	UNION 4"
1	UNION 4"

8	UNION
1	UNION
2	UNION
8	UNION
1	UNION
1	UNION
2	UNION
20	UNION
2	UNION
10	UNION
2	UNION
1	UNION
2	UNION
3	UNION
1	UNION
3	UNION
1	UNION
3	UNION
5	UNION
1	UNION
1	UNION
3	UNION
4	UNION
1	UNION

El tubo yee es otro producto que encontramos en yee de 2", yee 4 a 2, yee de 6", y yee de 3x2 y 4x4, estos los vamos unir en un solo producto.

Cantidad	PRODUCTO
4	YEE 2"
1	YEE 2"
1	YEE 4 A 2"
3	YEE 4 A 2"
1	YEE 6"
3	YEE DE 4X2
2	YEE DE 4X4
2	YEE DE 4X4

Cantidad	PRODUCTO
4	YEE
1	YEE
1	YEE
3	YEE
1	YEE
3	YEE
2	YEE
2	YEE

Otro producto encontrado es lavadero que puede indicarse con fregadera o medida como 1m x 50 cut

Cantidad	PRODUCTO
1	LAVADERO
1	LAVADERO
1	LAVADERO
1	LAVADERO
1	LAVADERO
1	LAVADERO
1	LAVADERO
1	LAVADERO
1	LAVADERO
1	LAVADERO
1	LAVADERO
1	LAVADERO
1	LAVADERO
1	LAVADERO - CON FREGADERA
1	LAVADERO DE 1M X 50 CUT

Cantidad	PRODUCTO
1	LAVADERO
1	LAVADERO
1	LAVADERO
1	LAVADERO
1	LAVADERO
1	LAVADERO
1	LAVADERO
1	LAVADERO
1	LAVADERO
1	LAVADERO
1	LAVADERO
1	LAVADERO
1	LAVADERO
1	LAVADERO
1	LAVADERO
1	LAVADERO

En los clavos tenemos varios registros como clavo 1", clavo 1 ½", clavo, 2", clavo 2 ½", clavo 3", clavo 4", clavo calamina, clavo pua, clavo eternit. Este tipo de clavo se vender por kilos.

2	CLAVO
1/2	CLAVO
4	CLAVO
8	CLAVO
1	CLAVO 1"
1/4	CLAVO 1"
100	CLAVO 1"
1	CLAVO 1 1/2"
1/4	CLAVO 1 1/2"
1/2	CLAVO 1 1/2"
1	CLAVO 1 1/2"
1	CLAVO 2"
4	CLAVO 2"
1/2	CLAVO 2"
1	CLAVO 2"
3	CLAVO 2"
1	CLAVO 2"
2	CLAVO 2"
10	CLAVO 2"
2	CLAVO 2"
2	CLAVO 2"
4	CLAVO 2"
1	CLAVO 2"

6	CLAVO
1	CLAVO
5	CLAVO
4	CLAVO
6	CLAVO
2	CLAVO
2	CLAVO
5	CLAVO
2	CLAVO
2	CLAVO
2	CLAVO
1/2	CLAVO
4	CLAVO
8	CLAVO
1	CLAVO
1/4	CLAVO

Ahora los que son clavo que se vender por unidad, como clavo de acero 2", clavo acero 2 ½, clavo acero 3", clavo cemento 2 ½".

30	CLAVO ACERO 2"
10	CLAVO ACERO 2 1/2"
25	CLAVO ACERO 3"
30	CLAVO ACERO 3"
10	CLAVO ACERO 3"
10	CLAVO ACERO 3"
1/4	CLAVO
6	CLAVO CEMENTO 2 1/2"
12	CLAVO DE ACERO
32	CLAVO DE ACERO
1	CLAVO ETERNIT
1	CLAVO ETERNIT
1	CLAVO PUA.

6	CLAVO
1	CLAVO
1	CLAVO
2	CLAVO
2	CLAVO
4	CLAVO
1/2	CLAVO
30	CLAVO UNIDAD
10	CLAVO UNIDAD
25	CLAVO UNIDAD
30	CLAVO UNIDAD
10	CLAVO UNIDAD
10	CLAVO UNIDAD
1/4	CLAVO
6	CLAVO UNIDAD
12	CLAVO UNIDAD
32	CLAVO UNIDAD
1	CLAVO
1	CLAVO
1	CLAVO

Llave paso es otro producto donde tenemos llave paso 1", llave paso 1/2, llave paso 3/4.

1	LLAVE PASO
1	LLAVE PASO
1	LLAVE PASO 1"
1	LLAVE PASO 1/2
1	LLAVE PASO 1/2
3	LLAVE PASO 1/2
1	LLAVE PASO 1/2
1	LLAVE PASO 1/2
1	LLAVE PASO 3/4
1	LLAVE PASO 3/4
1	LLAVE PASO 3/4
1	LLAVE PASO 3/4

1	LLAVE ESTILSON
1	LLAVE PASO
1	LLAVE PASO
1	LLAVE PASO
1	LLAVE PASO
1	LLAVE PASO
3	LLAVE PASO
1	LLAVE PASO
1	LLAVE PASO
1	LLAVE PASO
1	LLAVE PASO
1	LLAVE PASO
1	LLAVE PASO

Llave universal, tiene las variantes de valle 1", llave 1/2, llave Pavco.

Cantidad	PRODUCTO
6	LLAVE UNIVERSAL
8	LLAVE UNIVERSAL
1	LLAVE UNIVERSAL 1"
1	LLAVE UNIVERSAL 1/2
4	LLAVE UNIVERSAL 1/2
1	LLAVE UNIVERSAL PAVCO

Cantidad	PRODUCTO
6	LLAVE UNIVERSAL
8	LLAVE UNIVERSAL
1	LLAVE UNIVERSAL
1	LLAVE UNIVERSAL
4	LLAVE UNIVERSAL
1	LLAVE UNIVERSAL

Niple, tenemos niple 1/2, niple x 1.5, niple 3/4, niple bronce 1/2.

Cantidad	PRODUCTO
1	NIPLE 1/2
6	NIPLE 1/2
2	NIPLE 1/2
4	NIPLE 1/2 X 1.5"
2	NIPLES 3/4
7	NIPLES BRONCE 1/2

Cantidad	PRODUCTO
1	NIPLE
6	NIPLE
2	NIPLE
4	NIPLE
2	NIPLE
7	NIPLE

Bisagras, encontramos las bisagras de 1/2, de 3" y 4

Año	Cantidad	PRODUCTO
2020	2	PAR DE BISAGRA
2021	2	PAR DE BISAGRA
2020	1	PAR DE BISAGRA
2021	1	PAR DE BISAGRA 1/2
2020	1	PAR DE BISAGRA 3"
2020	4	PAR DE BISAGRA DE 4"

Año	Cantidad	PRODUCTO
2020	2	PAR DE BISAGRA
2021	2	PAR DE BISAGRA
2020	1	PAR DE BISAGRA
2021	1	PAR DE BISAGRA
2020	1	PAR DE BISAGRA
2020	4	PAR DE BISAGRA

Los pernos de los cuales encontramos los que solo indica perno, pero otros pernos 1.5, perno 6, perno cama, pernos completos, perno con tuerca y huacha.

Año	Cantidad	PRODUCTO
2020	12	PERNO
2020	1	PERNO
2020	16	PERNO
2020	4	PERNO
2020	3	PERNO
2021	10	PERNO
2020	8	PERNO 1 1/2
2020	60	PERNO 6"
2020	20	PERNO 6"
2020	8	PERNO CAMA
2020	12	PERNO COMPLETO
2020	6	PERNOS TUERCA Y HUACHA

Año	Cantidad	PRODUCTO
2020	12	PERNO
2020	1	PERNO
2020	16	PERNO
2020	4	PERNO
2020	3	PERNO
2021	10	PERNO
2020	8	PERNO
2020	60	PERNO
2020	20	PERNO
2020	8	PERNO
2020	12	PERNO
2020	6	PERNO

Tornillos son los productos que terminan en punta y son los que se roscan en algún material a diferencia del perno que necesita una tuerca o en casos huacha.

Tornillos encontramos esta el tonillo #30, 50x3, 60x1 y stovebolt.

Año	Cantidad	PRODUCTO
2020	2	TORNILLO
2020	24	TORNILLO #30
2020	12	TORNILLO DE 50X3"
2020	6	TORNILLO DE 60X1"
2020	20	TORNILLO STOVEBOLT

Año	Cantidad	PRODUCTO
2020	2	TORNILLO
2020	24	TORNILLO
2020	12	TORNILLO
2020	6	TORNILLO
2020	20	TORNILLO

Baldes de pintura, solo agruparemos los baldes de pintura, pintura blanca, negra, bolsas de pintura y Galones de pintura, de los cuales encontramos pintura gloss, base gris, negra gloss.

Cantidad	PRODUCTO
2	PINTURA
2	BALDE DE PINTURA
1	BALDE DE PINTURA
1	BALDE DE PINTURA
1	BALDE DE PINTURA
2	BANDE DE PINTURA
1	BOLSA PINTURA
1/2	GALON DE PINTURA
1	GALON DE PINTURA
2	GALON DE PINTURA
2	GALON DE PINTURA
1	GALON DE PINTURA
2	GALON DE PINTURA
1	PINTURA BLANCA
4	PINTURA BLANCA
1	PINTURA BOLSA
1	PINTURA GLOSS
2	PINTURA LATEX COLOR MAIZ

Cantidad	PRODUCTO
2	PINTURA
2	PINTURA
1	PINTURA
1	PINTURA
1	PINTURA
2	PINTURA
1	PINTURA
1/2	PINTURA
1	PINTURA
2	PINTURA
2	PINTURA
1	PINTURA
2	PINTURA
1	PINTURA
4	PINTURA
1	PINTURA
1	PINTURA
2	PINTURA

Pegamento, al la cual tenemos pegamento 118ml, azul, azul ¼, gris, mayólica, oatey, porcelanato, pvc.

1	PEGAMENTO
1	PEGAMENTO
1	PEGAMENTO
1	PEGAMENTO
1	PEGAMENTO
1	PEGAMENTO
1	PEGAMENTO
1	PEGAMENTO 118ML
1	PEGAMENTO 118ML
1	PEGAMENTO 118ML
1	PEGAMENTO AZUL
1	PEGAMENTO AZUL
1	PEGAMENTO AZUL
1	PEGAMENTO AZUL 1/4
1	PEGAMENTO AZUL 1/4
1	PEGAMENTO AZUL 1/4
1	PEGAMENTO AZUL 1/4
4	PEGAMENTO GRIS
8	PEGAMENTO MAYOLICA
5	PEGAMENTO MAYOLICA
2	PEGAMENTO MAYOLICA
2	PEGAMENTO MAYOLICA
3	PEGAMENTO MAYOLICA
3	PEGAMENTO MAYOLICA

1	PEGAMENTO
1	PEGAMENTO
1	PEGAMENTO
1	PEGAMENTO
1	PEGAMENTO
2	PEGAMENTO
1	PEGAMENTO
1	PEGAMENTO
1	PEGAMENTO
1	PEGAMENTO
2	PEGAMENTO
1	PEGAMENTO
1	PEGAMENTO
1	PEGAMENTO
1	PEGAMENTO
12	PEGAMENTO
1	PEGAMENTO
1	PEGAMENTO
1	PEGAMENTO
1	PEGAMENTO
1	PEGAMENTO

Brocha, encontramos de 3, 4 y de 6”.

Cantidad	PRODUCTO
1	BROCHA
2	BROCHA
1	BROCHA 3”
1	BROCHA 3”
1	BROCHA 4”
1	BROCHA SCHUBERT 6”

Cantidad	PRODUCTO
1	BROCHA
2	BROCHA
1	BROCHA
1	BROCHA
1	BROCHA
1	BROCHA

De todo este proceso nos quedó un total de 298 producto únicos.

PRODUCTO
ABRAZADERA
ACCESORIOS
ACEITE AZUL
ACIDO MURIATICO
ADAPTADOR

AGREGADOS
ALAMBRE
ALDABA
ARANDELAS A/ PRECION
ARCO SIERRA
ARENA FINA
ARENA GRUESA
ARMELLA
BADILEJO MANGO MADERA
BALDE SIKA IMPERMEABLE
BALDE VACIOS
BASE FLOURECENTE
BATEAS
BLOQUETA
BOLSA DE AGUA
BOLSA TEMPLE
BOTE DE 100 UNIDADES DE SOU FLEXCIERRA
BROCA 1/4 METAL
BROCA CEMENTO 1/2
BROCA DE 3/8
BROCA DE 5/16
BROCA MADERA
BROCHA
CABLE
CABLE #12
CABLE #14
CABLE MELLISO #12
CABLE MELLISO #14
CADENA
CAJA
CAJA 20X20 LUZ
CAJA DE DESAGUE
CAJA DE DESAGUE X 6 - DE 4 PIESAS
CAJA DE LLAVE TERMICA
CAJA DE LLAVE TERMICA DE 1
CAJA DE MEDIDOR
CAJA DE PASO
CAJA DE PASO PEQUEÑO
CAJA MODULAR
CAJA OCTOGONAL
CAJA RECTANGULAR
CAJITA LUZ PAVCO
CAL NIEVE
CAL OBRA

CALAMINA 3.6
CANALETA
CANDADO
CAÑO
CAÑO 1/2
CAÑO 1/2 AGUA
CAÑO BRONCE
CAÑO S. CHASER
CARBON
CARRETILLA (BUGUI)
CEMENTO ANDINO ULTRA
CEMENTO APU
CEMENTO NACIONAL
CEMENTO QUISQUEYA
CEMENTO SOL
CERROJO DE PUERTA
CHAPA
CHAPA SCOLTA
CHAPAFORTE
CILINDRO
CINTA AISLANTE
CINTA MASKING
CINTA TEFLON
CLAVO
CLAVO UNIDAD
CODO
COMBA 4 LIBRAS
CONECTOR 3/4
CORDEL DE NAYLON
CRUCETA 3"
CRUSETA 2"
CUADRO
CUELLO DE CERA
CURVA 1"
CURVA 1/2 SAP
CURVA 3"
CURVA 3/4
CUTER
DESARMADOR
DESINFECTANTE DE KRESSO
DISCO COPA
DISCO CORTE
DISCO DESGASTE
DISCO PULIR

DUCHA
ELIMINACION DE DESMONTE
ENCHUFE
ESCOBA
ESCOBA DE ACERO
ESCOBILLA DE ACERO
ESCOBILLA DE FIERRO
ESPATULA
ESPONJA
ESPRAY Y MATA MOSCAS
ETERNIT
ETERNIT 1 X 3M
FLUORESCENTE COMPLEX
FOCO
FOCO AHORRADOR
FOCO LED
FRAGUA
FRAGUA PORCELANA EXTRA FUERTE
GALON DE GASOLINA
GALON DE LEJIA
GALON DE THINNER
GALON PRESERVANTE
GAMEZAN
GANCHO DE TECHO
HILO
HILO DE PESCAR
HILO DE PESCAR 0.8
HISOPO PARA BAÑO
HOJA LIJA
HOJA SIERRA SANDFLEX
HUACHA PLANA
HUACHA PRECION
IMPRIMANTE
INODORO
INODORO WINDO
INTERRUPTOR
INTERRUPTOR BTICINO
INTERRUPTOR DOBLE
INTERRUPTOR TRIPLE
INTURRUPTOR
JUEGO BROCA - SACABOCADO
JUEGO BROVA MAFRTS
KILO SIKA
KK 18 HUECOS

KK 18 HUECOS LARK
KK RUSTICO
KOBRE ROJO
LADRILLO PANDERETA
LADRILLO TECHO 30X30X15
LAMPA
LAMPA TRAMONTINA
LAMPARA PRETUL
LAPIZ DE CARPINTERIA
LAVADERO
LENTE
LIMA
LIMA PLANA
LISTON
LISTON 2X3X4M
LISTON 2X3X5M
LISTON 3X2X3.7
LISTON 3X2X5M
LISTON 4X2X3.5
LISTON DE 5MM DE 3X2
LLANTA CARRETILLA
LLANTA CON CAMARA PARA CARRETILLA
LLAVE
LLAVE 1/2 AUTOMATICO
LLAVE ASA
LLAVE DE DUCHA
LLAVE DE GAS
LLAVE DE PASO METAL
LLAVE DIFERENCIAL
LLAVE DIFERENCIAL 30MA
LLAVE ESTILSON
LLAVE PASO
LLAVE TERMICA
LLAVE TERMICA C32 - TICINO
LLAVE TRIFASICA
LLAVE UNIVERSAL
LLAVES JARDINERO
MALLA RACHEL
MANGUERA
MANGUERA DE GAS
MANGUERA DE NIVEL
MARTILLO
MASILLA PARA META 2 ANIPSA
MEDIDOR DIGITAL

MULTICONECTOR
MURETE DE LUZ MONOFOSICO
NIPLE
NIVEL
NYLON
OCRE
OCRE AMARILLO DE 2 kg
OCRE ROJO
ORMIGON
PABILO
PALETA DE BOTIN
PALO DE ESCOBA
PALOS
PANALES 1/2
PANEL
PANEL ABIERTO
PANEL CERRADO
PANEL PUERTA Y VENTANA
PANEL VERTANA
PAQUETE DE ESPONJA
PAR ANGULO DE 1/2
PAR DE BISAGRA
PAR DE GUANTE
PAR GUANTE BADANA
PARIHUELA
PEGAMENTO
PEGAMENTO PARA MOSCAS
PERNO
PIE DERECHO
PIEDRA BASE
PIEDRA CHANCADA
PIEDRA GRUESA
PIEDRA ZANJA
PINTURA
PLANCHA
PLANCHA DE ETERNIT
PLANCHA DE TECNOPOR
PLANCHA FIBRA FORTE TRAPLUCIDO
PLANCHA PULIR
PLANCHA TECNOPOR
PLANCHA TECNOPOR 3/4
PLANCHAS CALAMINA
PLANCHAS FIBRA TRASFUCIDO
PLANCHAS FIBRAFORTE

PLASTICO
PLASTICO 2 M - DOBLE ANCHO
PLOMADA
PLOMADO
PUERTA
REDUCCION
REGADERA
REGLA ALUMINIO
REJILLA 2"
RIPIO
RODILLO
RODOPLAST
ROLLO CABLE #12
ROLLO CABLE #14
ROLLO CABLE #16
SACA GRASA
SACA OXIDO
SIKA
SIKA EN POLVO
SILICONA
SINCEL
SODA CAUSTICA
SOLDADURA PUNTO AZUL
SOQUET
SUMIDERO
SUMIDERO 2"
SUMIDERO 4"
SUMIDERO DE LAVADERO
TANQUE DE AGUA
TANQUE DE INODORO
TANQUE ETERNIT 1100 LT
TANQUE FARPLAS 1100 Lt
TANQUE FARPLAS 2500 Lt
TANQUE ROTOPLAS 1100 Lt
TANQUE ROTOPLAS 2500 Lt
TAPAS CIEGA
TAPON 1/2
TAPON 2"
TARUGO
TARUGO 1/2
TEE
TEFLON
TEMPLE X 25 K.
TEMPLE X 5K

TIRAFON
TIRAFON 5/8 X 1.5
TIRAFON DE 1/2
TIZA
TOMACORRIENTE
TOMACORRIENTE 3
TOMACORRIENTE DE 4
TOMACORRIENTE DOBLE
TOMACORRIENTE TRIPLE
TORNILLO
TORTOL
TRAMPA
TRAMPA 2"
TRAPO INDUSTRIAL
TRIPLAY
TUBO
UNION
VADILEJO
VALVULA DE COCINA A GAS DOMESTICO
VARILLA 1/2
VARILLA 1/4
VARILLA 3/4
VARILLA 3/8
VARILLA 5/8
VARILLA 8MM
WINCHA 5M
YEE
YESO
ZAPOLIO

Como observamos tenemos muchos productos por ellos vamos a que darlos solo con 168 productos de un total de 298 y los cuales fueron seleccionados a los 168 primeros que generan mayor ganancia. Tener en cuenta que esta selección se puede definir antes, pero como en ocasiones suceden en este caso recién vamos delimitar ahora.

En la siguiente imagen ya observamos el cálculo de la muestra.

CALCULO DE TAMAÑO DE MUESTRA FINA	
PARAMETRO	INSERTAR VALOR
N	298
Z	1.96
P	0.5
Q	0.5
e	0.05

Tamaño de muestra

168

Numerador	286.1992
Denominador	1.7029

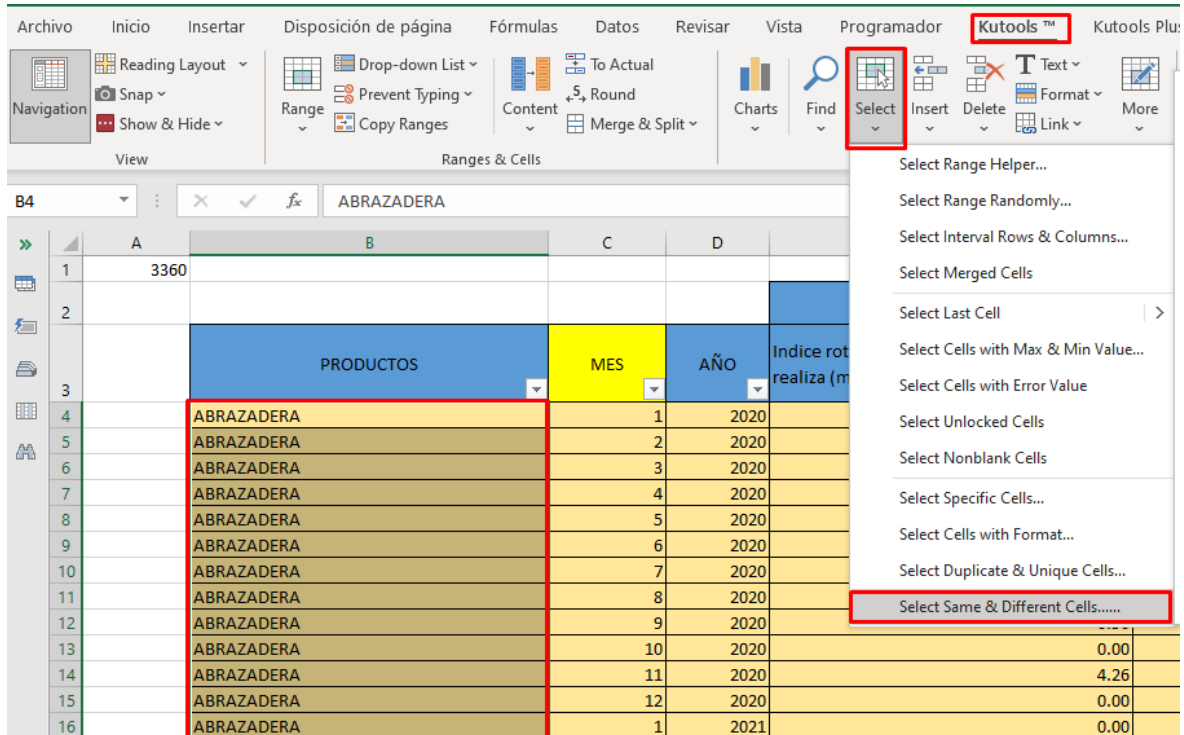
n = Tamaño de muestra Buscado
 N = Tamaño de población o Universo
 Z = Parametro Estadístico que depende el N
 e = Erro de estimacion Maximo Aceptado
 p = Probabilidad de que ocurra el evento
 q = (1 - p) = Probabilidad de que ocurra

Antes de llegar al resultado previo deber haberse hecho la seleccionar los productos a mantener y los que se eliminan. En este caso como se usó Excel hemos usado un complemento llamando *Kutools para Excel* que se tiene varias funciones en nuestro caso usamos para eliminar los productos que no vamos usar y lo que vamos usar.

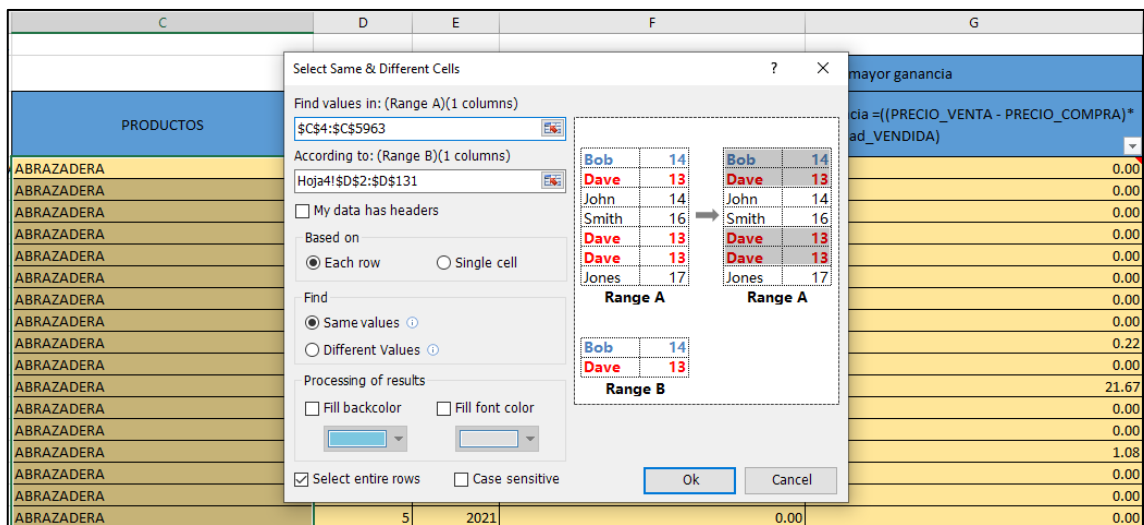
- Primero se creó las 2 listas de loque se mantendrá y los que se eliminaran, pero el más importante es el de eliminar ya que este se usara para eliminar de nuestro registro principal.

PRODUCTOS	PRODUCTO ELIMINAR
ABRAZADERA	ACCESORIOS
ADAPTADOR	ACEITE AZUL
ALAMBRE	ACIDO MURIATICO
ARENA FINA	AGREGADOS
ARENA GRUESA	ALDABA
BALDE SIKA IMPERMEABLE	ARANDELAS A/ PRECION
BALDE VACIOS	ARCO SIERRA
BATEAS	ARMELLA
BLOQUETA	BADILEJO MANGO MADERA
BOLSA DE AGUA	BASE FLOURECENTE
BOLSA TEMPLE	BROCA DE 3/8
BOTE DE 100 UNIDADES DE SOU FLEXCIERRA	BROCA DE 5/16
BROCA 1/4 METAL	BROCA MADERA
BROCA CEMENTO 1/2	CABLE
BROCHA	CABLE MELLISO #14
CABLE #12	CAJA
CABLE #14	CAJA 20X20 LUZ
CABLE MELLISO #12	CAJA DE PASO PEQUEÑO
CADENA	CAJA MODULAR
CAJA DE DESAGUE	CAJITA LUZ PAVCO
CAJA DE DESAGUE X 6 - DE 4 PIESAS	CAL NIEVE

- Seleccionar la columna de la cual queremos que busque y eliminar, después ir a herramientas>kutools>select>selec same & Different Cell...




- Donde indica (Rango B) debemos seleccionar la lista de los productos a eliminar. Los demás dejarlos como esta. Y el ultimo donde indica *Select entre rows* marcarlo. Y OK



- Después nos mostrara un mensaje, cuantos encontró, ok. Y después podremos eliminarlos.

ABRAZADERA	10	20
ABRAZADERA	11	20
ABRAZADERA	12	20
ABRAZADERA	1	20
ABRAZADERA	2	20
ABRAZADERA	3	20
ABR	4	20
ABR	5	20
ABR	6	20
ABR	7	20
ABR	8	20
ACC	1	20
ACC	2	20
ACC	3	20
ACCESORIOS	4	20

Select Same & Different Cells

 2600 rows have been selected.

OK

3.3. Estructurar los datos

Atributos derivados

No se derivado ningún registro nuevo para este caso.

Registros nuevos

Del registro obtenidos de las ventas, y del registro de compra que en este caso se ha generado de una forma aleatoria ya que no se logró obtener los datos reales, de estos registros se van obtener nuevos atributos con los que estaremos trabajando, como será el *índice de rotación*, *Ganancia de producto*, *frecuencia de compra* y el *stock*.

3.4. Integración de datos

Se crearon los campos de *índice de rotación*, *ganancia*. El registro de compras también se ha tenido de generar aleatoria mente, ya que como anterior mente se mencionó no se logró conseguir la data.

3.5. Formateo de datos

Para tener un mejor resultado en WEKA, trataremos de codificar algunos valores.

Los productos se les asignara codificara en un número.

<i>ABRAZADERA</i>	1
<i>ADAPTADOR</i>	2
<i>ALAMBRE</i>	3
<i>ARENA FINA</i>	4
<i>ARENA GRUESA</i>	5
<i>BALDE SIKA IMPERMEABLE</i>	6
<i>BALDE VACIOS</i>	7
<i>BATEAS</i>	8
<i>BLOQUETA</i>	9
<i>BOLSA DE AGUA</i>	10
<i>BOLSA TEMPLE</i>	11
<i>BOTE DE 100 UNIDADES DE SOU FLEXCIERRA</i>	12
<i>BROCA 1/4 METAL</i>	13
<i>BROCA CEMENTO 1/2</i>	14
<i>BROCHA</i>	15
<i>CABLE #12</i>	16
<i>CABLE #14</i>	17
<i>CABLE MELLISO #12</i>	18
<i>CADENA</i>	19
<i>CAJA DE DESAGUE</i>	20
<i>CAJA DE DESAGUE X 6 - DE 4 PIESAS</i>	21
<i>CAJA DE LLAVE TERMICA</i>	22
<i>CAJA DE LLAVE TERMICA DE 1</i>	23
<i>CAJA DE MEDIDOR</i>	24
<i>CAJA DE PASO</i>	25
<i>CAJA OCTOGONAL</i>	26
<i>CAJA RECTANGULAR</i>	27
<i>CAL OBRA</i>	28
<i>CANDADO</i>	29
<i>CAÑO</i>	30
<i>CAÑO 1/2</i>	31
<i>CARBON</i>	32
<i>CARRETILLA (BUGUI)</i>	33

<i>CEMENTO ANDINO ULTRA</i>	<i>34</i>
<i>CEMENTO APU</i>	<i>35</i>
<i>CEMENTO NACIONAL</i>	<i>36</i>
<i>CEMENTO QUISQUEYA</i>	<i>37</i>
<i>CEMENTO SOL</i>	<i>38</i>
<i>CHAPA</i>	<i>39</i>
<i>CHAPA SCOLTA</i>	<i>40</i>
<i>CHAPAFORTE</i>	<i>41</i>
<i>CILINDRO</i>	<i>42</i>
<i>CINTA AISLANTE</i>	<i>43</i>
<i>CLAVO</i>	<i>44</i>
<i>CLAVO UNIDAD</i>	<i>45</i>
<i>CODO</i>	<i>46</i>
<i>COMBA 4 LIBRAS</i>	<i>47</i>
<i>CURVA 3"</i>	<i>48</i>
<i>CURVA 3/4</i>	<i>49</i>
<i>DESARMADOR</i>	<i>50</i>
<i>DISCO COPA</i>	<i>51</i>
<i>DISCO CORTE</i>	<i>52</i>
<i>ELIMINACION DE DESMONTE</i>	<i>53</i>
<i>ESCOBA</i>	<i>54</i>
<i>ETERNIT</i>	<i>55</i>
<i>ETERNIT 1 X 3M</i>	<i>56</i>
<i>FLUORESCENTE COMPLEX</i>	<i>57</i>
<i>FOCO</i>	<i>58</i>
<i>GALON DE GASOLINA</i>	<i>59</i>
<i>GALON DE THINNER</i>	<i>60</i>
<i>HILO DE PESCAR 0.8</i>	<i>61</i>
<i>HOJA LIJA</i>	<i>62</i>
<i>HOJA SIERRA SANDFLEX</i>	<i>63</i>
<i>IMPRIMANTE</i>	<i>64</i>
<i>INODORO</i>	<i>65</i>
<i>JUEGO BROCA - SACABOCADO</i>	<i>66</i>
<i>KILO SIKA</i>	<i>67</i>
<i>KK 18 HUECOS</i>	<i>68</i>
<i>KK 18 HUECOS LARK</i>	<i>69</i>
<i>KK RUSTICO</i>	<i>70</i>
<i>LADRILLO PANDERETA</i>	<i>71</i>

LADRILLO TECHO 30X30X15	72
LAMPA	73
LAMPA TRAMONTINA	74
LAMPARA PRETUL	75
LAVADERO	76
LIMA PLANA	77
LISTON 2X3X4M	78
LISTON 2X3X5M	79
LISTON 3X2X3.7	80
LISTON 3X2X5M	81
LISTON 4X2X3.5	82
LISTON DE 5MM DE 3X2	83
LLANTA CARRETILLA	84
LLAVE 1/2 AUTOMATICO	85
LLAVE DE DUCHA	86
LLAVE DIFERENCIAL	87
LLAVE DIFERENCIAL 30MA	88
LLAVE PASO	89
LLAVE TERMICA	90
LLAVE TRIFASICA	91
LLAVE UNIVERSAL	92
LLAVES JARDINERO	93
MALLA RACHEL	94
MANGUERA	95
MANGUERA DE NIVEL	96
MEDIDOR DIGITAL	97
MULTICONECTOR	98
MURETE DE LUZ MONOFOSICO	99
NIPLE	100
OCRE	101
OCRE AMARILLO DE 2 kg	102
ORMIGON	103
PANEL	104
PANEL ABIERTO	105
PANEL CERRADO	106
PANEL PUERTA Y VENTANA	107
PANEL VERTANA	108
PAQUETE DE ESPONJA	109

<i>PAR DE BISAGRA</i>	<i>110</i>
<i>PAR GUANTE BADANA</i>	<i>111</i>
<i>PEGAMENTO</i>	<i>112</i>
<i>PERNO</i>	<i>113</i>
<i>PIE DERECHO</i>	<i>114</i>
<i>PIEDRA BASE</i>	<i>115</i>
<i>PIEDRA CHANCADA</i>	<i>116</i>
<i>PIEDRA GRUESA</i>	<i>117</i>
<i>PIEDRA ZANJA</i>	<i>118</i>
<i>PINTURA</i>	<i>119</i>
<i>PLANCHA DE ETERNIT</i>	<i>120</i>
<i>PLANCHA FIBRA FORTE TRAPLUCIDO</i>	<i>121</i>
<i>PLANCHA TECNOPOR</i>	<i>122</i>
<i>PLANCHAS CALAMINA</i>	<i>123</i>
<i>PLANCHAS FIBRA TRASFUCIDO</i>	<i>124</i>
<i>PLANCHAS FIBRAFORTE</i>	<i>125</i>
<i>PLASTICO</i>	<i>126</i>
<i>PLASTICO 2 M - DOBLE ANCHO</i>	<i>127</i>
<i>PLOMADA</i>	<i>128</i>
<i>PUERTA</i>	<i>129</i>
<i>REDUCCION</i>	<i>130</i>
<i>REGLA ALUMINIO</i>	<i>131</i>
<i>RIPIO</i>	<i>132</i>
<i>RODILLO</i>	<i>133</i>
<i>RODOPLAST</i>	<i>134</i>
<i>ROLLO CABLE #12</i>	<i>135</i>
<i>ROLLO CABLE #14</i>	<i>136</i>
<i>ROLLO CABLE #16</i>	<i>137</i>
<i>SIKA</i>	<i>138</i>
<i>SIKA EN POLVO</i>	<i>139</i>
<i>SOLDADURA PUNTO AZUL</i>	<i>140</i>
<i>SOQUET</i>	<i>141</i>
<i>TANQUE DE AGUA</i>	<i>142</i>
<i>TANQUE DE INODORO</i>	<i>143</i>
<i>TANQUE ETERNIT 1100 LT</i>	<i>144</i>
<i>TANQUE FARPLAS 1100 Lt</i>	<i>145</i>
<i>TANQUE FARPLAS 2500 Lt</i>	<i>146</i>
<i>TANQUE ROTOPLAS 1100 Lt</i>	<i>147</i>

TANQUE ROTOPLAS 2500 Lt	148
TEE	149
TEFLON	150
TEMPLE X 25 K.	151
TIRAFON	152
TOMACORRIENTE	153
TOMACORRIENTE DOBLE	154
TOMACORRIENTE TRIPLE	155
TORNILLO	156
TRAMPA	157
TRAMPA 2"	158
TRIPLAY	159
TUBO	160
UNION	161
VARILLA 1/2	162
VARILLA 1/4	163
VARILLA 3/4	164
VARILLA 3/8	165
VARILLA 5/8	166
VARILLA 8MM	167
YEE	168

Tabla de codificación de productos.

Trataremos de realizar lo mismo con los datos de índice de rotación y ganancia los cuales los agruparemos en *NULA, BAJA, MEDIA, ALTA Y MUUY ALTA*.

Nota: Para realizar este tipo de agrupamientos o segmentaciones debe ser en base a los que la organización tenga como parámetros o indique los rangos. Otro factor a tener en cuenta es si la organización tiene prestamos en un banco, esto debe ser basado en lo el banco trabaja.

Calculamos los datos mínimos, máximo, mediana, promedio(media) y moda del índice de rotación. Ya que estaremos cogiendo el promedio como la *MEDIA*.

INDICE ROTACION			
Mínimo	0	Mediana	0.00
máximo	218.75	Promedio	0.79

		Moda	0.00
--	--	------	------

Tabla de cálculo de media, moda, mediana de Índice de rotación.

0	Nula	$0 \geq x < 0.1$
0.40	Baja	$0.1 \leq x \leq 0.78$
0.79	Media	$0.79 \geq x \leq 1.18$
1.19	Alta	$1.19 \geq x \leq 218.74$
218.75	Muy Alta	$x \geq 218.75$

Tabla de Agrupamiento de los datos de Índice de rotación.

Calculamos los datos mínimos, máximo, mediana, promedio(media) y moda de Ganancia. Y estaremos poniendo el promedio como la MEDIA.

GANANCIA			
Mínimo	0	Mediana	0.00
máximo	218.75	Promedio	20.31
		Moda	0.00

Tabla de cálculo de media, moda, mediana de Ganancia.

0	Nula	$0 \geq x < 0.1$
10.16	Baja	$0.1 \geq x \leq 20.30$
20.31	Media	$20.31 \geq x \leq 30.46$
30.47	Alta	$30.47 \geq x \leq 2026.92$
2026.93	Muy Alta	$X \geq 2026.93$

Tabla de Agrupamiento de los datos de Ganancia.

Elaboramos la matriz del cruce de índice de rotación con ganancia y ver en que intersecciones no comprar, evaluar y comprar el producto.

		GANANCIA				
		NULA	BAJA	MEDIA	ALTA	MUY ALTA
ÍNDICE DE ROTACIÓN	NULA	NO COMPRAR	NO COMPRAR	EVALUAR	EVALUAR	EVALUAR
	BAJA	NO COMPRAR	EVALUAR	EVALUAR	COMPRAR	COMPRAR
	MEDIA	NO COMPRAR	EVALUAR	COMPRAR	COMPRAR	COMPRAR
	ALTA	EVALUAR	EVALUAR	COMPRAR	COMPRAR	COMPRAR
	MUY ALTA	EVALUAR	EVALUAR	COMPRAR	COMPRAR	COMPRAR

Matriz para determinar si compramos el producto o no.

- **DATOS FINALES A USAR.**

Logramos obtener un total de 3360 registros.

La primera imagen es en formato Excel y la siguiente es ya pasado en el archivo .arff

PRODUCTOS	MES	CODIFICACION INDICE DE ROTACION	CODIFICACION DE GANANCIA	STOCK
1	1	NULA	NULA	NO_COMPRAR
1	2	NULA	NULA	NO_COMPRAR
1	3	NULA	NULA	NO_COMPRAR
1	4	NULA	NULA	NO_COMPRAR
1	5	NULA	NULA	NO_COMPRAR
1	6	NULA	NULA	NO_COMPRAR
1	7	NULA	NULA	NO_COMPRAR
1	8	NULA	NULA	NO_COMPRAR
1	9	BAJA	BAJA	EVALUAR
1	10	NULA	NULA	NO_COMPRAR
1	11	ALTA	MEDIA	COMPRAR
1	12	NULA	NULA	NO_COMPRAR
1	1	NULA	NULA	NO_COMPRAR
1	2	ALTA	BAJA	EVALUAR
1	3	NULA	NULA	NO_COMPRAR
1	4	NULA	NULA	NO_COMPRAR
1	5	NULA	NULA	NO_COMPRAR
1	6	NULA	NULA	NO_COMPRAR
1	7	NULA	NULA	NO_COMPRAR
1	8	NULA	NULA	NO_COMPRAR
2	1	NULA	NULA	NO_COMPRAR

Imagen del registro final en excel

```
@data
1,1,NULA,NULA,NO_COMPRAR
1,2,NULA,NULA,NO_COMPRAR
1,3,NULA,NULA,NO_COMPRAR
1,4,NULA,NULA,NO_COMPRAR
1,5,NULA,NULA,NO_COMPRAR
1,6,NULA,NULA,NO_COMPRAR
1,7,NULA,NULA,NO_COMPRAR
1,8,NULA,NULA,NO_COMPRAR
1,9,BAJA,BAJA,EVALUAR
1,10,NULA,NULA,NO_COMPRAR
1,11,ALTA,MEDIA,COMPRAR
1,12,NULA,NULA,NO_COMPRAR
1,1,NULA,NULA,NO_COMPRAR
1,2,ALTA,BAJA,EVALUAR
1,3,NULA,NULA,NO_COMPRAR
1,4,NULA,NULA,NO_COMPRAR
1,5,NULA,NULA,NO_COMPRAR
1,6,NULA,NULA,NO_COMPRAR
1,7,NULA,NULA,NO_COMPRAR
1,8,NULA,NULA,NO_COMPRAR
2,1,NULA,NULA,NO_COMPRAR
2,2,BAJA,BAJA,EVALUAR
2,3,BAJA,BAJA,EVALUAR
2,4,NULA,NULA,NO_COMPRAR
2,5,NULA,NULA,NO_COMPRAR
2,6,NULA,NULA,NO_COMPRAR
2,7,ALTA,BAJA,EVALUAR
2,8,NULA,NULA,NO_COMPRAR
2,9,BAJA,BAJA,EVALUAR
2,10,BAJA,BAJA,EVALUAR
2,11,BAJA,BAJA,EVALUAR
2,12,NULA,NULA,NO_COMPRAR
2,1,BAJA,BAJA,EVALUAR
2,2,NULA,NULA,NO_COMPRAR
2,3,BAJA,BAJA,EVALUAR
2,4,BAJA,BAJA,EVALUAR
2,5,BAJA,BAJA,EVALUAR
2,6,BAJA,BAJA,EVALUAR
2,7,NULA,NULA,NO_COMPRAR
2,8,NULA,NULA,NO_COMPRAR
3,1,ALTA,ALTA,COMPRAR
3,2,ALTA,ALTA,COMPRAR
3,3,ALTA,ALTA,COMPRAR
3,4,NULA,NULA,NO_COMPRAR
3,5,BAJA,BAJA,EVALUAR
3,6,ALTA,BAJA,EVALUAR
3,7,ALTA,BAJA,EVALUAR
3,8,ALTA,ALTA,COMPRAR
3,9,ALTA,ALTA,COMPRAR
3,10,ALTA,ALTA,COMPRAR
3,11,MEDIA,BAJA,EVALUAR
3,12,ALTA,ALTA,COMPRAR
3,1,ALTA,ALTA,COMPRAR
3,2,ALTA,ALTA,COMPRAR
3,3,ALTA,ALTA,COMPRAR
3,4,ALTA,ALTA,COMPRAR
```

Imagen de los datos concatenados en un achito .arff

4. MODELADO

4.1. Seleccionar Técnica de modelado

El aprendizaje que se aplicara en la investigación es supervisado, ya que se ingresaran los datos de entrada y los resultados. Esto permite que la maquina aprenda y este sepa cuales son las posibles salidas.

Este tipo de aprendizaje va con los regresión lineal y logístico, árbol siendo esto el más usado para este tipo de aprendizaje, entre otros.

El modelo que aplicaremos es el del árbol, pero antes de esto tratemos de probar otros modelos con WEKA para ver cuál es que se adecua mejor a mis datos de entrada.

Las variables a ingresar van ser los siguientes.

- Índice de rotación
- Ganancia
- Stock (comprar o no)

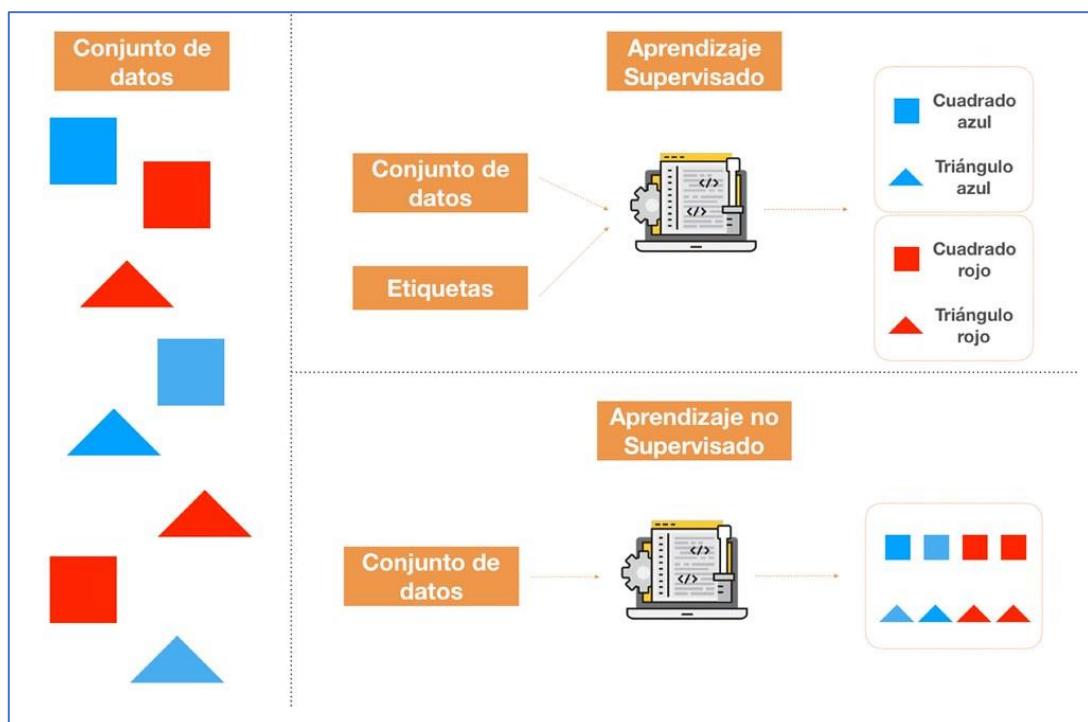


Figura: Aprendizaje supervisado y no supervisado

Fuente: https://live.staticflickr.com/65535/48054954426_a75cf0ea67_b.jpg

4.2. Generar plan de prueba

Esteremos usando el programa WEKA para comparar los diferentes algoritmos que nos ofrece y ver cuál es mejor para nuestros datos.

Como ya sabemos cuáles van ser nuestros datos ahora creamos nuestro archivo .arff el cual reconoce WEKA. Donde primer se declara la cabecera o nombre de la relación, después creamos nuestros atributos y al final nuestra data en el orden que creamos nuestros atributos separados con una coma, cada fina es un registro.

```
File Edit Format View Help
@relation PRODUCTO

@attribute producto INTEGER
@attribute MES {1,2,3,4,5,6,7,8,9,10,11,12}
@attribute IndiceRotacion {NULA,BAJA,MEDIA,ALTA,MUY_ALTA}
@attribute Ganancia {NULA,BAJA,MEDIA,ALTA,MUY_ALTA}
@attribute PREDICCION_COMPRA {NO_COMPRAR,EVALUAR,COMPRAR}

@data
1,1,NULA,NULA,NO COMPRAR
```

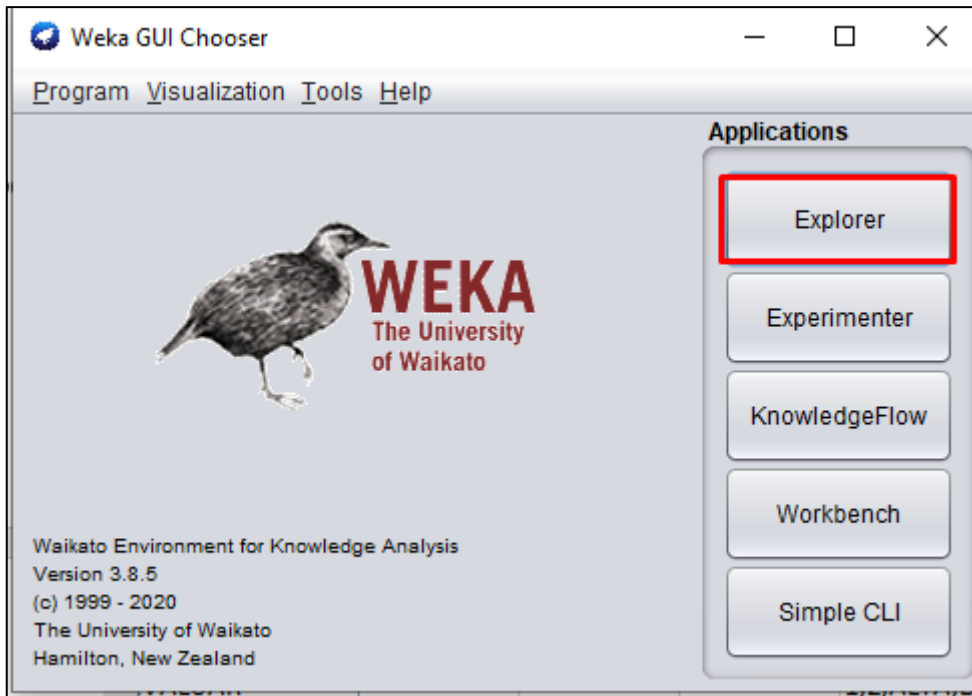
```
1, 2, NULA, NULA, NO_COMPRAR
1, 3, NULA, NULA, NO_COMPRAR
1, 4, NULA, NULA, NO_COMPRAR
1, 5, NULA, NULA, NO_COMPRAR
1, 6, NULA, NULA, NO_COMPRAR
1, 7, NULA, NULA, NO_COMPRAR
1, 8, NULA, NULA, NO_COMPRAR
1, 9, BAJA, BAJA, EVALUAR
1, 10, NULA, NULA, NO_COMPRAR
1, 11, ALTA, MEDIA, COMPRAR
1, 12, NULA, NULA, NO_COMPRAR
1, 1, NULA, NULA, NO_COMPRAR
1, 2, ALTA, BAJA, EVALUAR
1, 3, NULA, NULA, NO_COMPRAR
1, 4, NULA, NULA, NO_COMPRAR
1, 5, NULA, NULA, NO_COMPRAR
1, 6, NULA, NULA, NO_COMPRAR
1, 7, NULA, NULA, NO_COMPRAR
1, 8, NULA, NULA, NO_COMPRAR
2, 1, NULA, NULA, NO_COMPRAR
2, 2, BAJA, BAJA, EVALUAR
2, 3, BAJA, BAJA, EVALUAR
2, 4, NULA, NULA, NO_COMPRAR
2, 5, NULA, NULA, NO_COMPRAR
2, 6, NULA, NULA, NO_COMPRAR
2, 7, ALTA, BAJA, EVALUAR
2, 8, NULA, NULA, NO_COMPRAR
2, 9, BAJA, BAJA, EVALUAR
2, 10, BAJA, BAJA, EVALUAR
2, 11, BAJA, BAJA, EVALUAR
2, 12, NULA, NULA, NO_COMPRAR
```

4.3. Construir el modelo

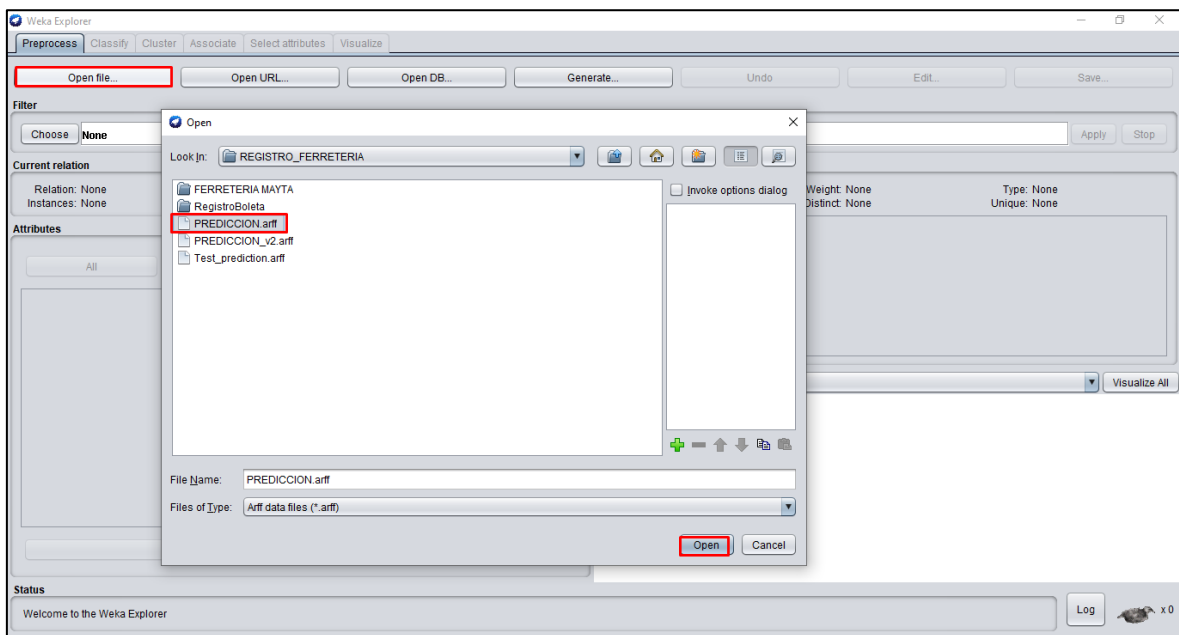
Ahora es donde vamos a probar los diferentes algoritmos, en especial los de árbol para ver cuál es mejor.

Nota: Realizare directo las funciones, si quiere saber los funcionamientos de WEKA o que características tiene, recomiendo revisar el manual o videos.

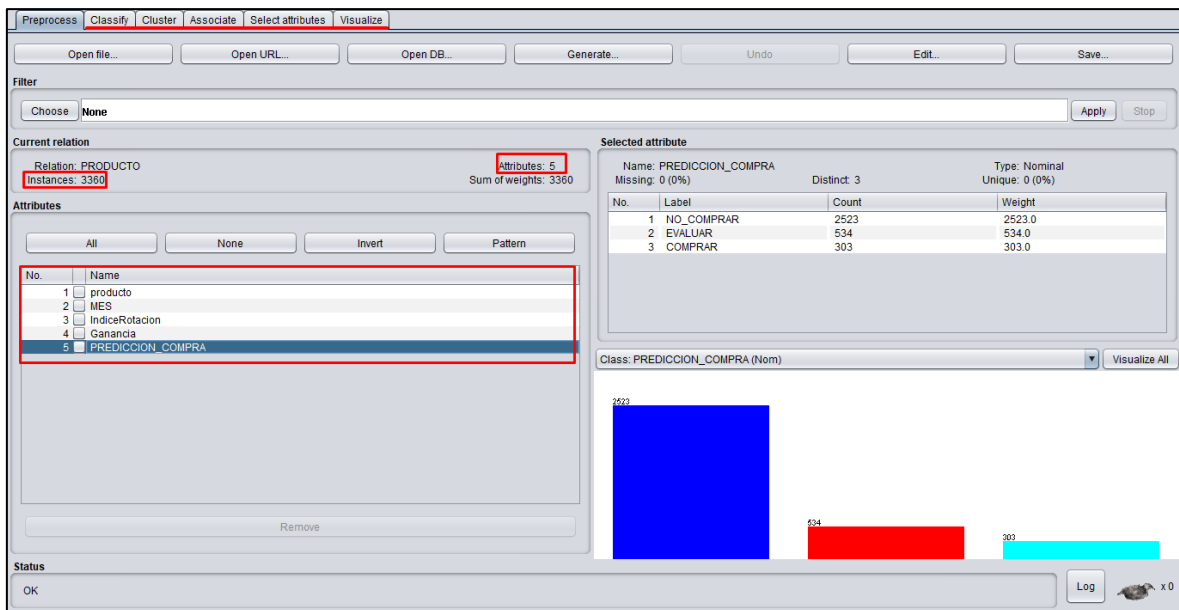
Ingresamos a WEKA, previo haber [descargado](#) en la página oficial. Después nos dirigimos a la opción *Explorer*.



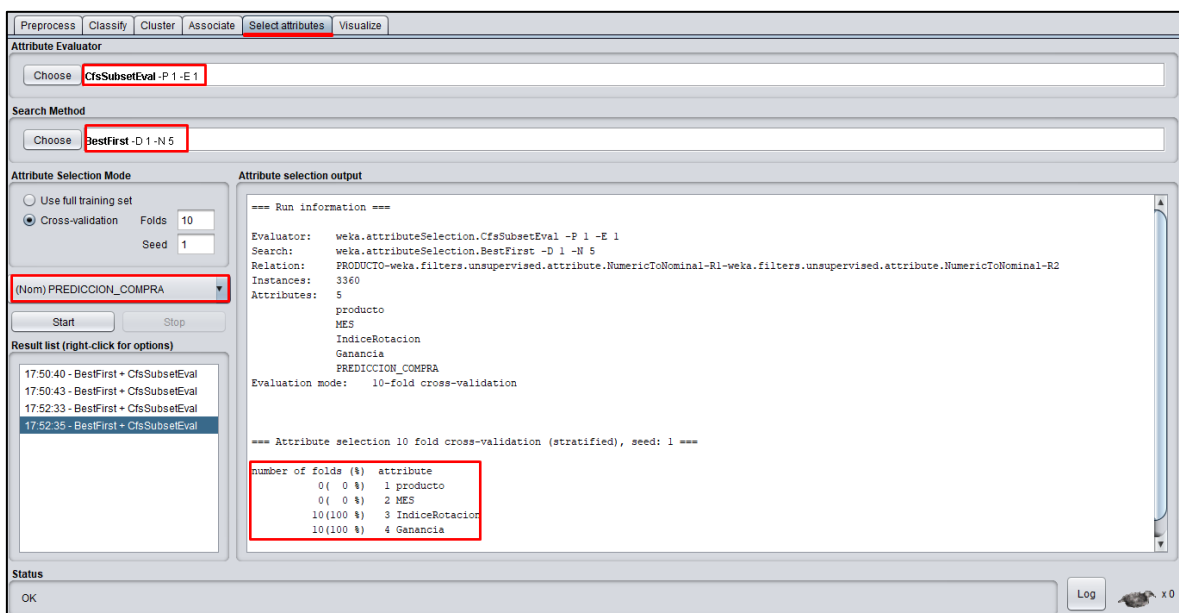
Después debemos cargar nuestra base de datos.



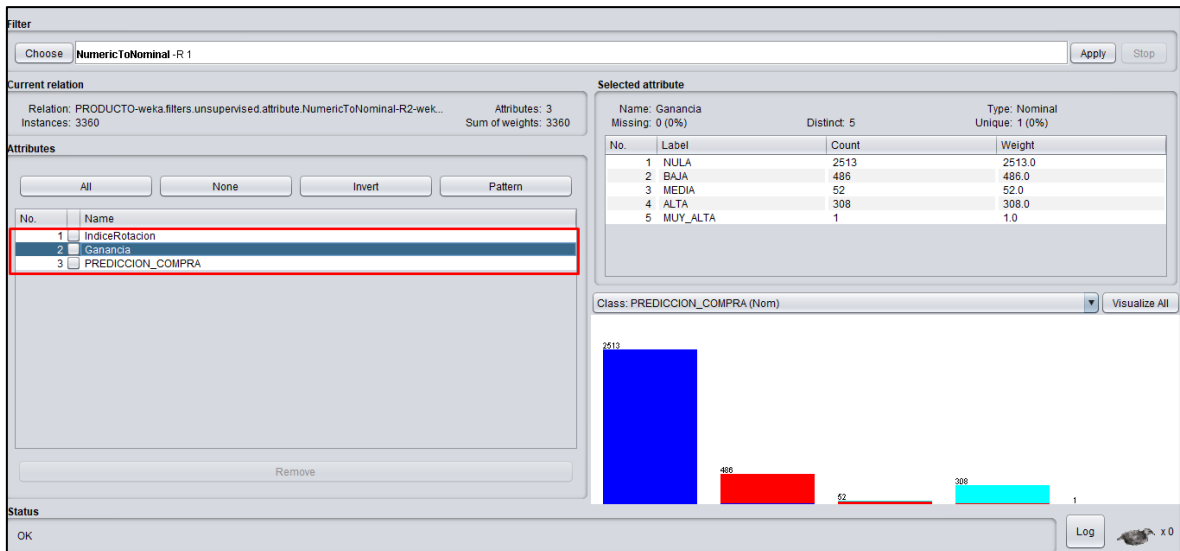
Cuando nos carga nos mostrara los siguiente, donde podremos observar los atributos e instancias, y también en la parte superior se nos activa diversas funciones.



Para ver que atributos son los que más relevantes o hayamos agregado algún atributo, pero no es muy importante, nos dirigimos a *Select Attribute* donde podremos ver que atributo los eliminamos y los que conservamos. Solo se mantendrán todos lo que tenga 10(100%) todo lo demás vamos a eliminarlos.



De todo lo demás observamos que nos quedamos con los siguientes atributos y ahora comenzamos a aplicar lo ya planificado.



- Primero estaremos probando con el J48 y también conocido como C4.5 que es el más usado para este tipo de predicciones. Y en Test Opción seleccionaremos la opción Porcentaje Split, donde asignaremos un 80% para entrenar el modelo y el restante para el test.

Preprocess **Classify** Cluster Associate Select attributes

Classifier

weka

- classifiers
 - bayes
 - functions
 - lazy
 - meta
 - misc
 - rules
 - trees
 - DecisionStump
 - HoeffdingTree
 - J48**
 - LMT
 - M5P
 - RandomForest
 - RandomTree
 - REPTree

Classifier: Choose **J48 -C 0.25 -M 2**

Test options

Use training set

Supplied test set [Set...]

Cross-validation Folds 10

Percentage split % 80

[More options...]

(Nom) PREDICCION_COMPRA

[Start] [Stop]

Después de realizado nos mostrar el resultado, en la parte superior nos muestra, nombre del algoritmo aplicado, base de datos, instancias, atributos, y el modo de tes aplicado.

```
Classifier output

=== Run information ===

Scheme:      weka.classifiers.trees.J48 -C 0.25 -M 2
Relation:    PRODUCTO-weka.filters.unsupervised.attribute.NumericToN
Instances:   3360
Attributes:  3
             IndiceRotacion
             Ganancia
             PREDICCION COMPRA
Test mode:   split 80.0% train, remainder test
```

Continuando veremos las interacciones o lógica que siguió el modelo también el tamaño y nivel del árbol.

```
J48 pruned tree
-----

Ganancia = NULA: NO_COMPRAR (2513.0)
Ganancia = BAJA
|  IndiceRotacion = NULA: NO_COMPRAR (10.0)
|  IndiceRotacion = BAJA: EVALUAR (411.0)
|  IndiceRotacion = MEDIA: EVALUAR (25.0)
|  IndiceRotacion = ALTA: EVALUAR (40.0)
|  IndiceRotacion = MUY_ALTA: EVALUAR (0.0)
Ganancia = MEDIA
|  IndiceRotacion = NULA: EVALUAR (0.0)
|  IndiceRotacion = BAJA: EVALUAR (43.0)
|  IndiceRotacion = MEDIA: COMPRAR (5.0)
|  IndiceRotacion = ALTA: COMPRAR (4.0)
|  IndiceRotacion = MUY_ALTA: EVALUAR (0.0)
Ganancia = ALTA
|  IndiceRotacion = NULA: EVALUAR (15.0)
|  IndiceRotacion = BAJA: COMPRAR (167.0)
|  IndiceRotacion = MEDIA: COMPRAR (18.0)
|  IndiceRotacion = ALTA: COMPRAR (107.0)
|  IndiceRotacion = MUY_ALTA: COMPRAR (1.0)
Ganancia = MUY_ALTA: COMPRAR (1.0)

Number of Leaves :    17

Size of the tree :    21
```

Lo más importante de ver la correcta clasificación de las instancias se recomienda que mientras más cerca de 100% mejor va ser el modelo. En este caso observamos que obtuvimos un 99.85% de asertividad. Obtenemos también Error absoluto

medio, error cuadrático medio, error absoluto relativo y error cuadrático medio de la raíz.

```

=== Summary ===

Correctly Classified Instances      671          99.8512 %
Incorrectly Classified Instances     1           0.1488 %
Kappa statistic                    0.9962
Mean absolute error                 0.0009
Root mean squared error             0.0265
Relative absolute error              0.357 %
Root relative squared error         7.3557 %
Total Number of Instances          672

=== Detailed Accuracy By Class ===

          TP Rate  FP Rate  Precision  Recall  F-Measure  MCC      ROC Area  PRC Area  Class
          1.000   0.006   0.998     1.000   0.999     0.996   1.000    1.000    NO_COMPRAR
          1.000   0.000   1.000     1.000   1.000     1.000   1.000    1.000    EVALUAR
          0.981   0.000   1.000     0.981   0.990     0.990   1.000    1.000    COMPRAR
Weighted Avg.   0.999   0.005   0.999     0.999   0.999     0.996   1.000    1.000
    
```

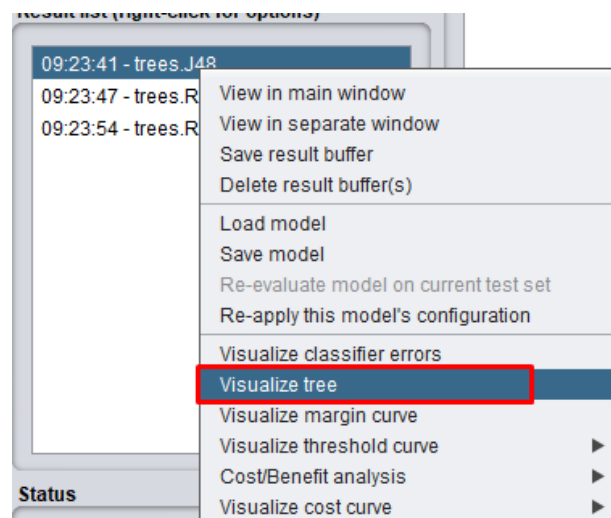
Otro muy importante también es nuestra matriz de confusión, donde la diagonal comenzando de la a, la cual son las acertadas y los demás no.

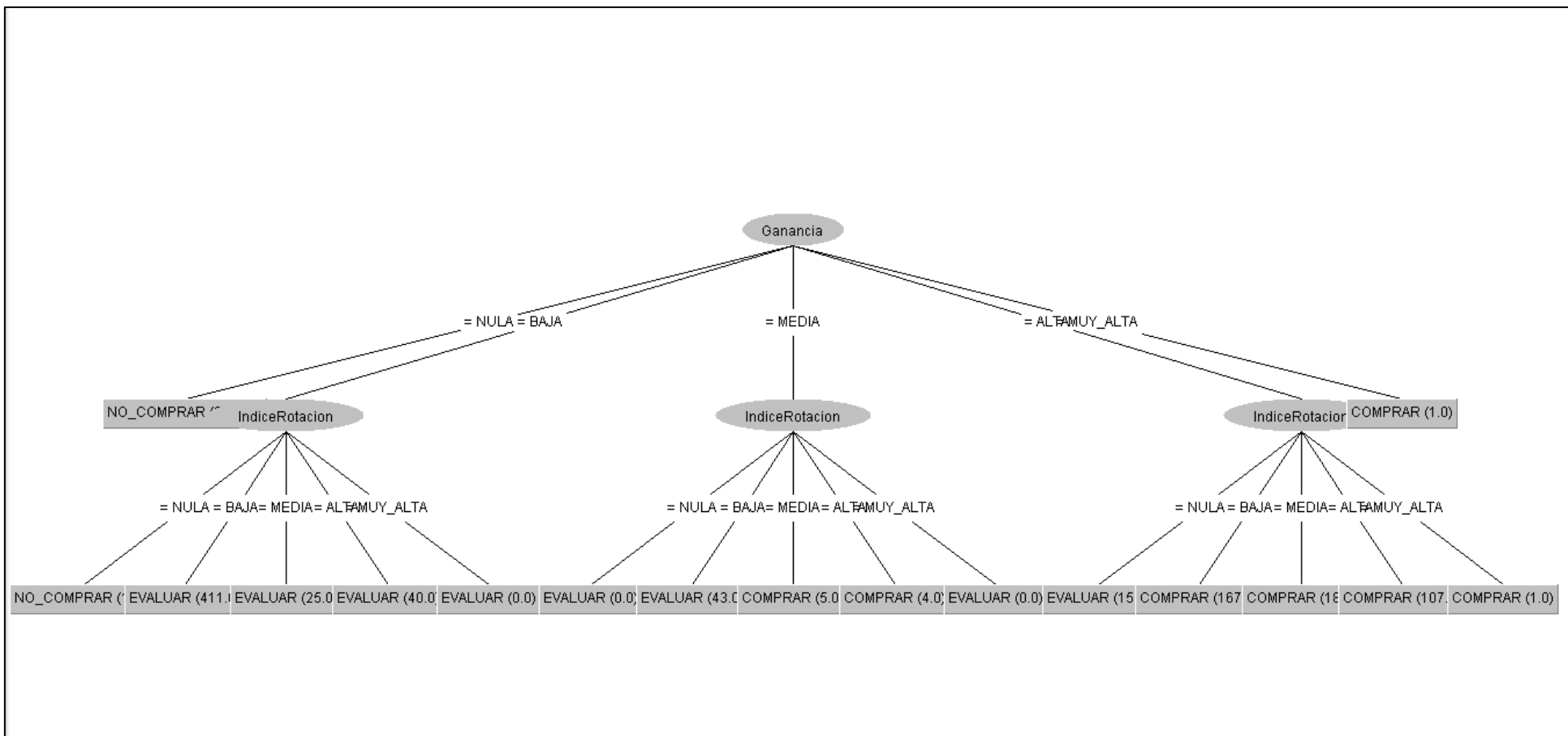
```

=== Confusion Matrix ===

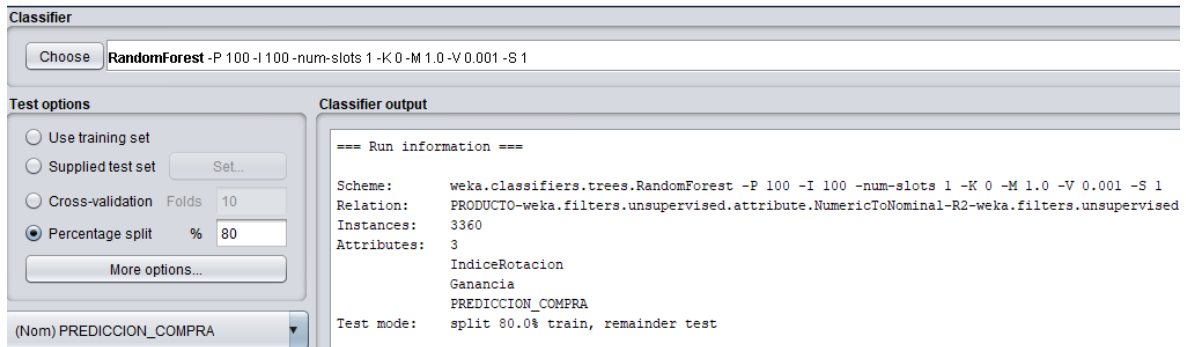
   a  b  c  <-- classified as
512  0  0 | a = NO_COMPRAR
  0 107  0 | b = EVALUAR
  1  0 52 | c = COMPRAR
    
```

Una opción más que podemos encontrar en WEKA es y en árbol, es poder visualizar en forma gráfica. En donde en una forma más grafica podemos observar que comienza con ganancia y según el valor no indica que resultado tomar.





- Ahora veamos el algoritmo RandomForest, aplicaremos el mismo Test Options que el anterior.



En la correcta clasificación obtenemos un 99.85%. Obtenemos también Error absoluto medio, error cuadrático medio, error absoluto relativo y error cuadrático medio de la raíz.

```

=== Summary ===
Correctly Classified Instances      671          99.8512 %
Incorrectly Classified Instances      1           0.1488 %
Kappa statistic                     0.9962
Mean absolute error                  0.001
Root mean squared error              0.0266
Relative absolute error              0.3933 %
Root relative squared error          7.3804 %
Total Number of Instances           672

=== Detailed Accuracy By Class ===

```

	TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area	Class
	1.000	0.006	0.998	1.000	0.999	0.996	1.000	1.000	NO_COMPRAR
	1.000	0.000	1.000	1.000	1.000	1.000	1.000	1.000	EVALUAR
	0.981	0.000	1.000	0.981	0.990	0.990	1.000	1.000	COMPRAR
Weighted Avg.	0.999	0.005	0.999	0.999	0.999	0.996	1.000	1.000	

La matriz donde muestra de 512 NO_COMPRAR se clasificó correctamente, pero 1 no, 107 Evaluar se clasificaron correctamente, 52 Comprar están bien,

```

=== Confusion Matrix ===

```

	a	b	c	<-- classified as
512	0	0		a = NO_COMPRAR
0	107	0		b = EVALUAR
1	0	52		c = COMPRAR

- Como tercero veamos del árbol veamos el RandomTree.

```

=== Run information ===

Scheme:      weka.classifiers.trees.RandomTree -K 0 -M 1.0 -V 0.001 -S 1
Relation:    PRODUCTO-weka.filters.unsupervised.attribute.Remove-R1-2
Instances:   3360
Attributes:  3
              IndiceRotacion
              Ganancia
              PREDICCION_COMPRA
Test mode:   split 80.0% train, remainder test

```

También podemos observar las diferentes relaciones e interacciones que a realizado el algoritmo y el tamaño del árbol es de 21.

```

=== Classifier model (full training set) ===

RandomTree
=====

Ganancia = NULA : NO_COMPRAR (2513/0)
Ganancia = BAJA
|  IndiceRotacion = NULA : NO_COMPRAR (10/0)
|  IndiceRotacion = BAJA : EVALUAR (411/0)
|  IndiceRotacion = MEDIA : EVALUAR (25/0)
|  IndiceRotacion = ALTA : EVALUAR (40/0)
|  IndiceRotacion = MUY_ALTA : NO_COMPRAR (0/0)
Ganancia = MEDIA
|  IndiceRotacion = NULA : NO_COMPRAR (0/0)
|  IndiceRotacion = BAJA : EVALUAR (43/0)
|  IndiceRotacion = MEDIA : COMPRAR (5/0)
|  IndiceRotacion = ALTA : COMPRAR (4/0)
|  IndiceRotacion = MUY_ALTA : NO_COMPRAR (0/0)
Ganancia = ALTA
|  IndiceRotacion = NULA : EVALUAR (15/0)
|  IndiceRotacion = BAJA : COMPRAR (167/0)
|  IndiceRotacion = MEDIA : COMPRAR (18/0)
|  IndiceRotacion = ALTA : COMPRAR (107/0)
|  IndiceRotacion = MUY_ALTA : COMPRAR (1/0)
Ganancia = MUY_ALTA : COMPRAR (1/0)

Size of the tree : 21

```

Imagen del árbol RandomTree en forma texto.

Continuando igual a los anteriores vemos que la correcta clasificación es de un 99.85%. Obtenemos también Error absoluto medio, error cuadrático medio, error absoluto relativo y error cuadrático medio de la raíz.

```

=== Summary ===
Correctly Classified Instances      671          99.8512 %
Incorrectly Classified Instances    1            0.1488 %
Kappa statistic                    0.9962
Mean absolute error                0.0009
Root mean squared error            0.0265
Relative absolute error            0.357 %
Root relative squared error        7.3557 %
Total Number of Instances          672

=== Detailed Accuracy By Class ===

```

	TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area	Class
	1.000	0.006	0.998	1.000	0.999	0.996	1.000	1.000	NO_COMPRAR
	1.000	0.000	1.000	1.000	1.000	1.000	1.000	1.000	EVALUAR
	0.981	0.000	1.000	0.981	0.990	0.990	1.000	1.000	COMPRAR
Weighted Avg.	0.999	0.005	0.999	0.999	0.999	0.996	1.000	1.000	

La matriz de confusión.

```

=== Confusion Matrix ===

```

	a	b	c	<-- classified as
512	0	0		a = NO_COMPRAR
0	107	0		b = EVALUAR
1	0	52		c = COMPRAR

Este algoritmo nos permite también visualizar el modelo en una forma gráfica

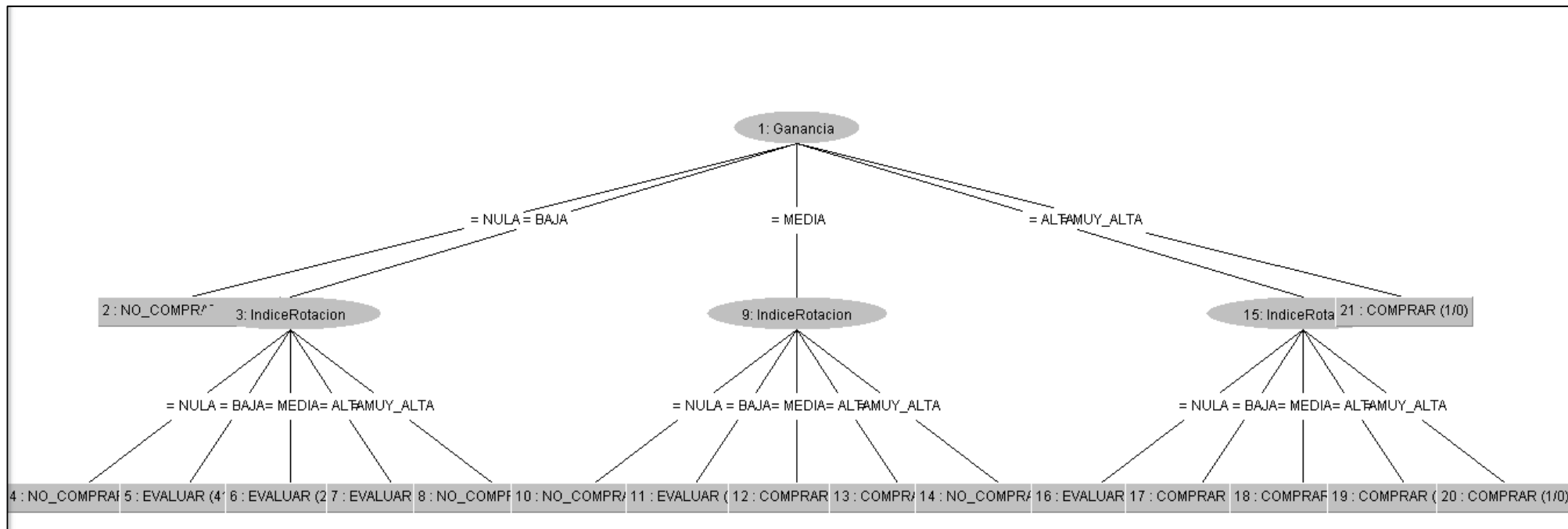


Imagen del grafico del Modelo de arbol, RandomTree

- Un último que probaremos será un clasificador *rules* y el algoritmo JRip pero normalmente conocido como RIPPER. Dejamos las opciones los mismo que los anteriores, teniendo 3360 instancias y 3 atributos. 80% para entrenar y el restante para entrenar.

```

=== Run information ===

Scheme:      weka.classifiers.rules.JRip -F 3 -N 2.0 -O 2 -S 1
Relation:    PRODUCTO-weka.filters.unsupervised.attribute.Remove-R1-2
Instances:   3360
Attributes:  3
              IndiceRotacion
              Ganancia
              PREDICCION_COMPRA
Test mode:   split 80.0% train, remainder test

```

En JRIP rules nos muestra las relaciones que existen, como ejemplo podemos ver que si la ganancia es ALTA y el índice de rotación es BAJA la predicción es que se compraron todos (167) y así los demás. También nos indica que son 11 reglas

```

JRIP rules:
=====

(Ganancia = ALTA) and (IndiceRotacion = BAJA) => PREDICCION_COMPRA=COMPRAR (167.0/0.0)
(Ganancia = ALTA) and (IndiceRotacion = ALTA) => PREDICCION_COMPRA=COMPRAR (107.0/0.0)
(IndiceRotacion = MEDIA) and (Ganancia = ALTA) => PREDICCION_COMPRA=COMPRAR (18.0/0.0)
(Ganancia = MEDIA) and (IndiceRotacion = MEDIA) => PREDICCION_COMPRA=COMPRAR (5.0/0.0)
(IndiceRotacion = ALTA) and (Ganancia = MEDIA) => PREDICCION_COMPRA=COMPRAR (4.0/0.0)
(Ganancia = BAJA) and (IndiceRotacion = BAJA) => PREDICCION_COMPRA=EVALUAR (411.0/0.0)
(Ganancia = BAJA) and (IndiceRotacion = ALTA) => PREDICCION_COMPRA=EVALUAR (40.0/0.0)
(IndiceRotacion = BAJA) => PREDICCION_COMPRA=EVALUAR (43.0/0.0)
(IndiceRotacion = MEDIA) => PREDICCION_COMPRA=EVALUAR (25.0/0.0)
(Ganancia = ALTA) => PREDICCION_COMPRA=EVALUAR (16.0/1.0)
=> PREDICCION_COMPRA=NO_COMPRAR (2524.0/1.0)

Number of Rules : 11

Time taken to build model: 0.2 seconds

=== Evaluation on test split ===

Time taken to test model on test split: 0 seconds

```

Ahora observaremos que tenemos un 99.55 de asertividad con el algoritmo JRip. Que sería 669 de aciertos y 3 incorrectos. Obtenemos también Error absoluto medio, error cuadrático medio, error absoluto relativo y error cuadrático medio de la raíz.

```

=== Summary ===

Correctly Classified Instances      669          99.5536 %
Incorrectly Classified Instances    3            0.4464 %
Kappa statistic                    0.9885
Mean absolute error                 0.0052
Root mean squared error             0.0546
Relative absolute error             1.9559 %
Root relative squared error         15.179 %
Total Number of Instances          672

=== Detailed Accuracy By Class ===

          TP Rate  FP Rate  Precision  Recall  F-Measure  MCC      ROC Area  PRC Area  Class
          0.998   0.006   0.998     0.998   0.998     0.992   0.996   0.998   NO_COMPRAR
          1.000   0.004   0.982     1.000   0.991     0.989   0.998   0.975   EVALUAR
          0.962   0.000   1.000     0.962   0.981     0.979   0.981   0.965   COMPRAR
Weighted Avg.   0.996   0.005   0.996     0.996   0.996     0.990   0.995   0.992

```

Y por último la matriz, observamos que en NO_COMPRAR 551 se clasificaron correctamente, pero 1 no, en EVALUAR 107 correctas y 2 no y en COMPRAR 51 correctos.

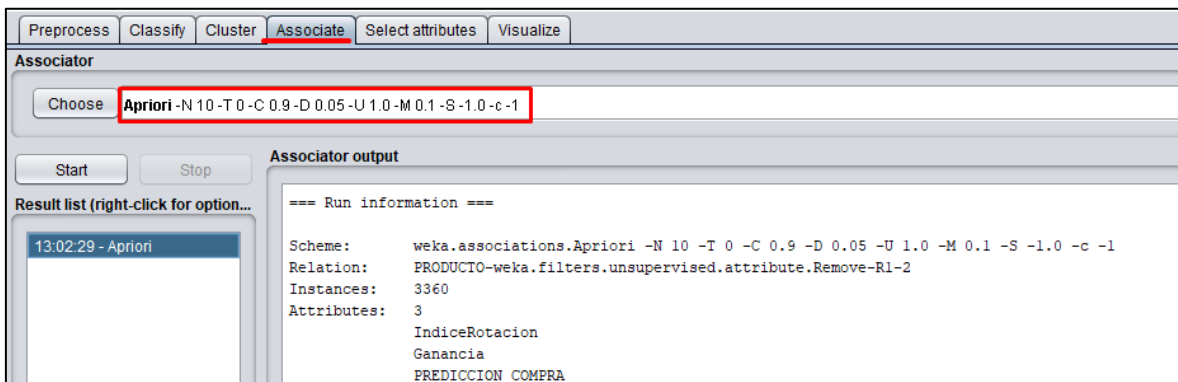
```

=== Confusion Matrix ===

  a  b  c  <-- classified as
511  1  0 |  a = NO_COMPRAR
  0 107  0 |  b = EVALUAR
  1  1 51 |  c = COMPRAR

```

- Otro punto también donde podemos ver algunos reglas o relaciones que tal vez que están ocultas podemos verla en la pestaña asociación.



Encontramos 10 reglas. El conf() del casi el final de cada regla indica el porcentaje de asertividad. El 1 es un 100%.


```

Best rules found:
1. PREDICCION_COMPRA=NO_COMPRAR 2523 ==> IndiceRotacion=NULA 2523 <conf:(1)> lift:(1.32) lev:(0.18) [617] conv:(617.23)
2. Ganancia=NULA 2513 ==> IndiceRotacion=NULA 2513 <conf:(1)> lift:(1.32) lev:(0.18) [614] conv:(614.79)
3. Ganancia=NULA 2513 ==> PREDICCION_COMPRA=NO_COMPRAR 2513 <conf:(1)> lift:(1.33) lev:(0.19) [626] conv:(626.01)
4. Ganancia=NULA PREDICCION_COMPRA=NO_COMPRAR 2513 ==> IndiceRotacion=NULA 2513 <conf:(1)> lift:(1.32) lev:(0.18) [614] conv:(614.79)
5. IndiceRotacion=NULA Ganancia=NULA 2513 ==> PREDICCION_COMPRA=NO_COMPRAR 2513 <conf:(1)> lift:(1.33) lev:(0.19) [626] conv:(626.01)
6. Ganancia=NULA 2513 ==> IndiceRotacion=NULA PREDICCION_COMPRA=NO_COMPRAR 2513 <conf:(1)> lift:(1.33) lev:(0.19) [626] conv:(626.01)
7. PREDICCION_COMPRA=NO_COMPRAR 2523 ==> Ganancia=NULA 2513 <conf:(1)> lift:(1.33) lev:(0.19) [626] conv:(57.82)
8. IndiceRotacion=NULA PREDICCION_COMPRA=NO_COMPRAR 2523 ==> Ganancia=NULA 2513 <conf:(1)> lift:(1.33) lev:(0.19) [626] conv:(57.82)
9. PREDICCION_COMPRA=NO_COMPRAR 2523 ==> IndiceRotacion=NULA Ganancia=NULA 2513 <conf:(1)> lift:(1.33) lev:(0.19) [626] conv:(57.82)
10. IndiceRotacion=NULA 2538 ==> PREDICCION_COMPRA=NO_COMPRAR 2523 <conf:(0.99)> lift:(1.32) lev:(0.18) [617] conv:(39.51)

```

1. La primera indica que los productos a *no comprar* tienen un índice de rotación NULA,
2. El segundo es que toda ganancia sea NULA tiene un índice de rotación NULA.
3. Los que tienen Ganancia NULA, el producto no se compra.
4. Cuando la ganancia es NULA y No se compra es porque el índice de rotación es NULA
5. Cuando el índice de rotación es NULA y Ganancia NULA, No comprar el producto.
6. Cuando la ganancia es NULA, el índice de rotación es NULA y No se compra el producto.
7. Cuando se indique que no se compra es porque la ganancia es NULA.
8. Cuando el índice de rotación es NULA y Predicción de Compra indique NO_COMPRAR es porque la Gancia nula.
9. Cuando predicción compra Indique NO_COMPRAR es que el índice de rotación es NULA y Ganancia NULA.
10. El índice de rotación NULA, no comprar el producto.

4.4. EVALUAR EL MODELO

Ahora veremos el resumen de los diferentes modelos o clasificadores aplicados. En donde presentamos el porcentaje de asertividad y error absoluto obtenido por cada uno. Pero gracias al programa WEKA también podemos obtener error cuadrático medio, error absoluto relativo y error cuadrático medio de la raíz.

Nº	CLASIFICADOR	ALGORITMO	PORCENTAJE DE ASERTIVIDAD	ERROR ABSOLUTO MEDIO
1	Árbol	J48	99.85%	0.0009 %
2	Árbol	RANDOMFOREST	99.85%	0.001 %
3	Árbol	RANDOMTREE	99.85%	0.0009 %

4	Rules	JRIP	99.85%	0.0052 %
---	-------	------	--------	----------

Podemos observar que los Modelos de Árbol son los mejores que se adecuan a nuestros datos en especial el J48 y el RandomTree. Los dos tiene un 99.855% y tienen un error absoluto medio 0.0009 % que es la distancia promedio de los puntos y la recta encontrada.

- Vamos a realizar un pequeño testeo con para ello también tenemos que crear un archivo arff. el atributo que queremos que nos prediga lo aginamos con un '?',

```

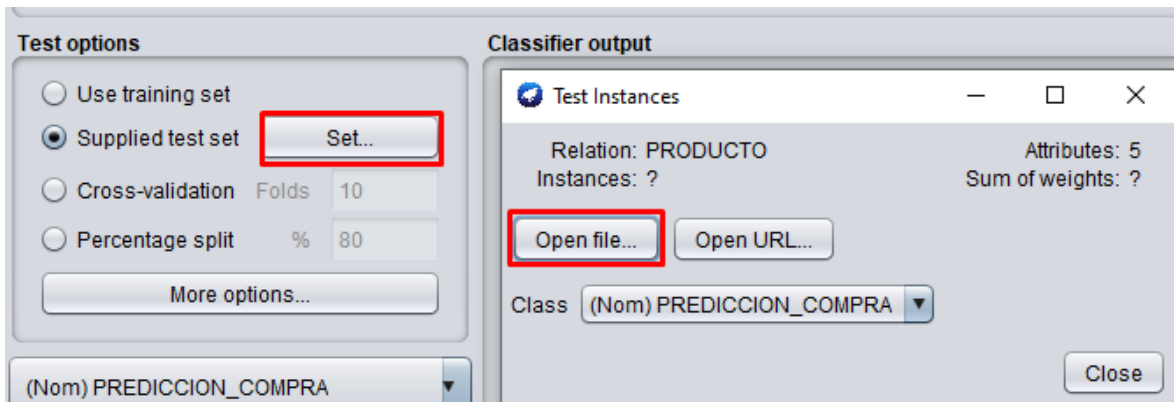
Test_prediction.arff - Notepad
File Edit Format View Help
@relation PRODUCTO

@attribute producto INTEGER
@attribute MES {1,2,3,4,5,6,7,8,9,10,11,12}
@attribute IndiceRotacion {NULA,BAJA,MEDIA,ALTA,MUY_ALTA}
@attribute Ganancia {NULA,BAJA,MEDIA,ALTA,MUY_ALTA}
@attribute PREDICCION_COMPRA {NO_COMPRAR,EVALUAR,COMPRAR}

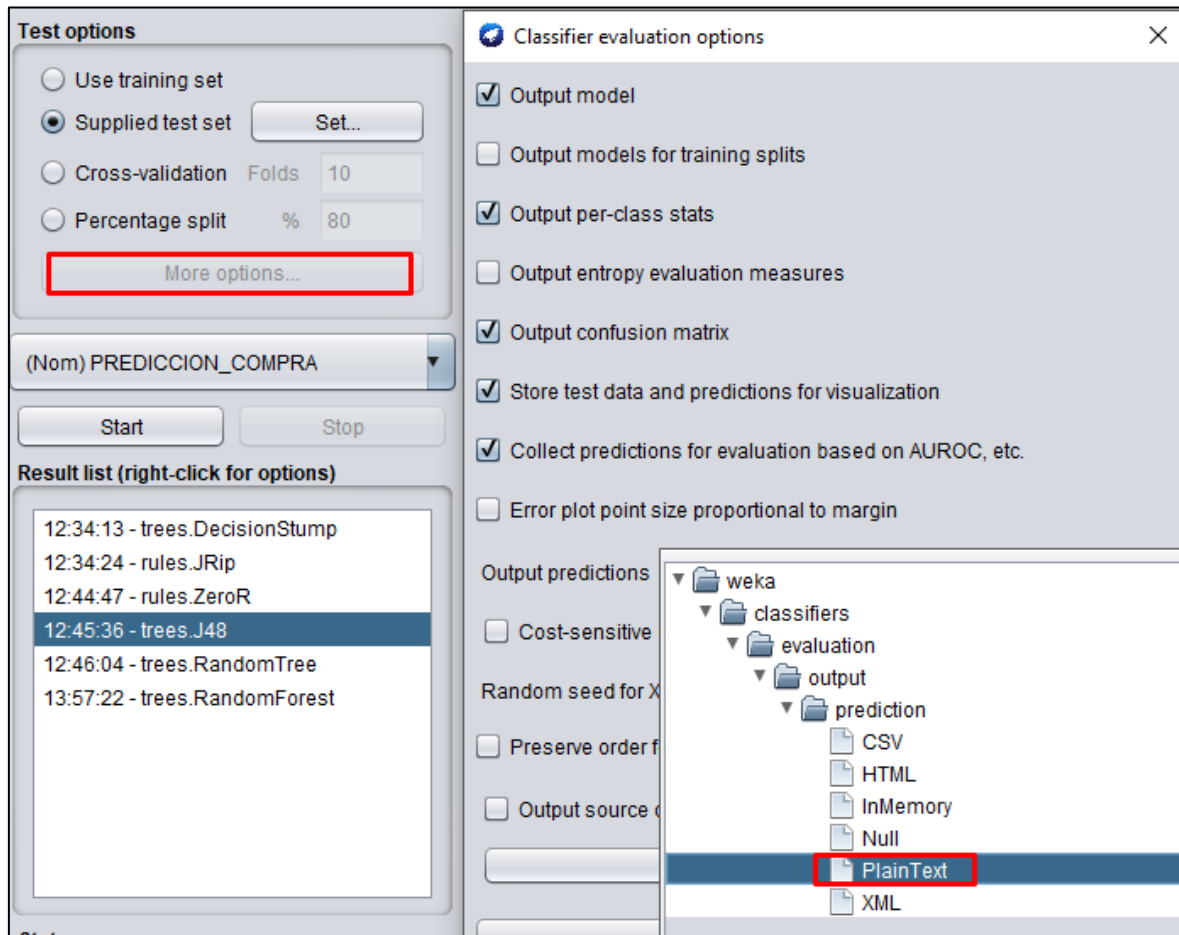
@data
3,6,ALTA,BAJA,?
3,7,ALTA,BAJA,?
3,8,ALTA,ALTA,?
3,9,ALTA,ALTA,?
3,10,ALTA,ALTA,?
6,10,NULA,NULA,?
6,11,NULA,NULA,?
6,12,NULA,NULA,?
6,1,NULA,NULA,?
6,2,BAJA,BAJA,?
10,4,NULA,NULA,?
10,5,NULA,NULA,?
10,6,NULA,NULA,?
10,7,NULA,NULA,?
10,8,NULA,NULA,?
15,12,NULA,NULA,?
15,1,BAJA,BAJA,?
15,2,BAJA,BAJA,?
15,3,NULA,NULA,?
15,4,NULA,NULA,?

```

En WEKA cambiamos el *test options* y ponemos *Supplied test set* y cargamos nuestro archivo tes creado y también en class podemos seleccionar lo que vamos a predecir que en este caso lo dejamos como esta.



No dirigimos en más opciones, escogemos la salida en texto plano.



El resultado, observamos que todos tiene una predicción del 100% con lo que el modelo si tiene una buena precisión.

Test options

Use training set
 Supplied test set Set...
 Cross-validation Folds
 Percentage split %
More options...

(Nom) PREDICCION_COMPRA ▼

Start Stop

Result list (right-click for options)

- 12:34:13 - trees.DecisionStump
- 12:34:24 - rules.JRip
- 12:44:47 - rules.ZeroR
- 12:45:36 - trees.J48
- 12:46:04 - trees.RandomTree
- 13:57:22 - trees.RandomForest
- 16:28:15 - misc.InputMappedClassifier

Classifier output

Time taken to build model: 0 seconds

=== Predictions on test set ===

inst#	actual	predicted	error	prediction
1	1:?	2:EVALUAR		1
2	1:?	2:EVALUAR		1
3	1:?	3:COMPRAR		1
4	1:?	3:COMPRAR		1
5	1:?	3:COMPRAR		1
6	1:?	1:NO_COMPRAR		1
7	1:?	1:NO_COMPRAR		1
8	1:?	1:NO_COMPRAR		1
9	1:?	1:NO_COMPRAR		1
10	1:?	2:EVALUAR		1
11	1:?	1:NO_COMPRAR		1
12	1:?	1:NO_COMPRAR		1
13	1:?	1:NO_COMPRAR		1
14	1:?	1:NO_COMPRAR		1
15	1:?	1:NO_COMPRAR		1
16	1:?	1:NO_COMPRAR		1
17	1:?	2:EVALUAR		1
18	1:?	2:EVALUAR		1
19	1:?	1:NO_COMPRAR		1
20	1:?	1:NO_COMPRAR		1

5. EVALUACIÓN

Esta evaluación a diferencia el anterior se iba con los objetivos de minería de datos, este se hace desde el punto de vista de los objetivos de negocio.

5.1. Evaluar los resultados

Se consideró que él va ser factible y aceptable cuando este supere el 70% de asertividad en la clasificación y cómo podemos ver en la fase previa los diferentes algoritmos de clasificaciones superan el 99%

Modelos aprobados.

Los 4 están aprobados, pero en este caso seleccionaremos 2, el primero y principal el modelo J48 siendo este el más conocido y usado, y el segundo RandomTree, teniendo igual error absoluto medio.

5.2. Revisar el proceso

El proceso se realizó según los pasos establecidos, pero si hay algo en tener en cuenta es que el tiempo otro no contar con datos reales de la organización en su totalidad, solo parcial puede que no proporcione cálculos reales pero el proceso se ha seguido según lo planificado.

5.3. Determinar Próximos pasos

Los siguientes pasos de la investigación sería la fase de implantación.

6. IMPLANTACIÓN

Se procederá a explicar al aliado u organización como poner en funcionamiento el proyecto que se ha construido en las fases anteriores y también exponer los resultados obtenidos de una forma que lo pueda entender. Otro objetivo de la fase es crear una estrategia para el mantenimiento del proyecto y generar un informe en que se incluyan posibles mejoras para el futuro y una lista de dificultades encontradas a la hora de realizarlo.

6.1. Plan de implementación

En primera parte para la implementación será tener acceso a los datos completos y reales del negocio. A partir de ahí sería aplicar lo mismo que se realizó en esta investigación. Cabe mencionar que contando con los datos reales pueda tener nuevos requerimientos o factores que puedan intervenir y generar que el proyecto sea más grande.

En segundo lugar (opcional) sería en el caso tenga los registros en físico, pasarlos a una base de datos como Oracle, MySQL, Microsoft SQL Server y Excel.

Y por último revisar el manual de usuario de WEKA y realizar lo que se realizó en el proyecto desde la comprensión del negocio hasta la implementación.

6.2. Plan de monitoreo y mantenimiento

Esta fase es una de las más importantes ya que como sabemos los datos pueden ser modificados o cambiados por diferentes razones, como codificación incorrecta, haber asignado incorrectamente valores o rango de valores. Teniendo en cuenta esto, siempre hay que realizar una verificación por segunda vez los datos extraídos. La realización del modelo para verificar que funcione bien se recomienda realizar cada un año, pero esto variar según el plan de estudio que queramos realizar.

A diferencia de realizar la momera de datos, el modelo de predicción solo se aplica una vez y esto puede ser usado por un tiempo largo, pero se recomienda para mayor precisión del modelo que cuando el historial de datos crezca, realizar de nuevo el proceso de crear el modelo.

Los pasos para mantener el modelo:

- Cuando se extraiga los datos siempre realizar la limpieza.
- Realizar de nuevo los pasos para crear el modelo, desde comprensión del negocio hasta la implementación.
- Realizar de nuevo el testeado del modelo creado.

6.3. Informe final

Con el uso de la metodología CRISP-DM en este proyecto pudimos predecir la compra del producto a partir del índice de rotación del producto y ganancia que genera este producto.

De los objetivos planteados se lograron realizar los 3 donde se indica identificar los productos que tienen índice de rotación y ganancia más alta, la segunda seleccionar el mejor modelo de clasificación, y el ultimo usar el 80% de los datos para entrenar y el resto para testear el modelo.

Repodando las etapas para llegar a encontrar el modelo:

La primera etapa donde se tenía que conseguir los datos fue las más difícil ya que el negocio solo tenía los datos en físico o no los tenía. Esto género que usemos los datos que tengamos disponibles y los que faltaban genera forma aleatoria.

Cuando ya se disponía de los datos se realizó la exploración y la limpieza de esto, y como se obtuvieron datos de físico en la etapa de limpieza fue la más laboriosa de todas.

Cuando se tuvieron los datos ser realizo la preparación para cargarlas a WEKA y seleccionar el mejor modelo de clasificación y se determinó que el mejor modelo es el J48 y en RandomTree.

Realizados estos pasos se logró alcanzar el objetivo que era crear el modelo de predicción para la compra.

6.4. Revisión del proyecto

Uno de los principales inconvenientes he tenido y ya mencioné en anteriores apartados es que los datos no son 100% reales y esto es una condicional de los resultados. A pesar de completar los datos faltantes con datos generados por mí

esto puede que no se vieran si usara los datos reales, o variables nuevas que se puedan generar.

Un factor positivo es que probamos 4 modelos para ver cuál es mejor para nuestros datos y estos salieron con asertividad alta lo que quiere decir que si usáramos los datos reales estos resultados serían mejores.