



# UNIVERSIDAD ANDINA DEL CUSCO

FACULTAD DE INGENIERÍA Y ARQUITECTURA  
ESCUELA PROFESIONAL DE INGENIERÍA DE  
SISTEMAS



## UAC

**“BÚSQUEDA DE PATRONES DE COMPORTAMIENTO USANDO MACHINE LEARNING, PARA LA TOMA DECISIONES GERENCIALES EN LA EMPRESA CHUCHUHUASI”.**

**PRESENTADO POR:**

Christian Mitchell Nina Asto

Jean Adriano Vilca Malpartida



Para optar al Título Profesional de Ingeniero de  
Sistemas.

**ASESOR:**

Ing. Adriel Ramírez Vargas

**CUSCO - PERÚ**

**2018**



## PRESENTACIÓN

**SEÑOR DECANO DE LA FACULTAD DE INGENIERÍA Y ARQUITECTURA.**

**SEÑORES MIEMBROS DEL JURADO.**

Cumpliendo con los preceptos del Reglamento de Grados y Títulos de la Universidad Andina Del Cusco, presentamos a su consideración nuestro trabajo de investigación intitulado:

**“BÚSQUEDA DE PATRONES DE COMPORTAMIENTO USANDO  
MACHINE LEARNING, PARA LA TOMA DECISIONES GERENCIALES EN  
LA EMPRESA CHUCHUHUASI”.**

El presente trabajo utilizó un sub campo de las ciencias computacionales que es parte de la rama de la inteligencia artificial llamado Machine learning. El cual nos facilitará crear un programa que generalizará patrones de comportamiento de acuerdo a la cantidad de datos ingresados, estos datos serán obtenidos en periodos de tiempo prolongados (anualmente) y de forma (incremental). Dichos datos serán consumidos por la aplicación y nos mostrarán información implícita de las distintas actividades de la empresa ChuchuHuasi. Dicha información contribuirá al mejoramiento de la toma de decisiones gerenciales en los procesos internos la empresa.

Atentamente.

**Los tesistas.**



## DEDICATORIA

Esta tesis se la dedico a Dios, que me guio por el camino correcto, me dio fuerzas para seguir adelante y no rendirme ante los tropiezos que se presentaron, enseñándome a encarar la adversidad sin desfallecer en el intento.

A mi familia quienes por ellos soy lo que soy, pero sobre todo a mi madre por su amor, apoyo, comprensión y consejo además por ser mi ejemplo y mi motivación principal de mi constante superación.

A mis compañeros y docentes porque sin duda fueron parte crucial para la culminación de este trabajo.

Jean Adriano Vilca Malpartida

Ante todo a Dios, por la fortaleza brindada en toda mi labor, siempre guiando mis pasos y permitir llegar a este momento tan esperado.

A quienes siempre han estado a mi lado, son mis pilares y ejemplo a seguir mis padres y familia amada. Son el motivo de todo lo que hago, para ellos este homenaje.

A aquellas personas involucradas en este proyecto, amigos, docentes, por el apoyo recibido y saber que puedo contar con ellos en cualquier momento.

Gracias

Christian Mitchell Nina Asto



## AGRADECIMIENTOS

A nuestra Universidad Andina del Cusco por habernos dado la oportunidad de educarnos y tenerla como nuestra casa de estudios alma mater.

A nuestra querida escuela profesional de INGENIERÍA DE SISTEMAS que a lo largo de más de 25 años de actividad académica sigue formando profesionales.

A nuestros docentes por su abnegada dedicación y por transmitirnos sus conocimientos en el transcurso de nuestra vida universitaria con el único fin de hacernos útiles a la sociedad.



## RESUMEN

Actualmente la utilización de Machine learning tiene mucha vigencia en el tratamiento de datos los cuales pueden ser de distintas características. El aprendizaje automático tiene una amplia gama de aplicaciones, incluyendo motores de búsqueda, diagnósticos médicos, detección de fraude en el uso de tarjetas de crédito, análisis del mercado de valores, clasificación de secuencias de ADN, reconocimiento del habla y del lenguaje escrito, juegos y robótica.

Se ha desarrollado una aplicación que utiliza modelos probabilísticos, lógicos, los cuales son parte del aprendizaje automático (Machine learning). Esta aplicación generalizara patrones de comportamiento. En este contexto eso quiere decir (identificar patrones complejos en miles o millones de datos). Nuestro aplicativo utiliza un algoritmo que revisa los datos y es capaz de predecir comportamientos futuros.

Los datos obtenidos para el trabajo de investigación en la Empresa Chuchuasi S.A. llegaron a una cantidad de más de 5000 obtenidos con las fichas de recolección de datos, estos han sido procesados anualmente en los periodos (2016-2017). A través del algoritmo predictivo por patrones de comportamiento (KVecino más cercano) de Machine learning. La información obtenida ha mostrado una diversidad de preferencias en los gustos de los clientes. Estas diversidades de gustos inducen al comportamiento de la empresa en los procesos del negocio (Gestión, funcional, económicos, y operativos). Como en toda empresa existen diversas áreas enmarcadas en los procesos antes mencionados. La información obtenida después del proceso de (KVecino más cercano) de Machine learning es de gran ayuda para la toma de decisiones para la gerencia de la empresa en el (Proceso de gestión), ya que, a través del aplicativo en la Web desarrollado en la investigación, muestra características implícitas de los gustos de los clientes que están estructuradas en los horarios de atención, nacionalidad de los clientes, tipo de tragos, acompañamiento, edad etc.

1. Por último, la información obtenida recomienda de una mejor manera en las de decisiones de los diferentes procesos de la empresa Chuchuhuasi



tales como (de gestión, funcionales, económicos y operativos). Una muestra clara de lo mencionado es una descripción detallada de variables de compra de insumos, y contrato de personal (Proceso Económico).



## ABSTRACT

Currently the use of Machine learning is very valid in the processing of data which can be of different characteristics. Machine learning has a wide range of applications, including search engines, medical diagnostics, fraud detection in the use of credit cards, stock market analysis, classification of ADN sequences, recognition of speech and written language, games and robotics.

An application has been developed that uses probabilistic, logical models, which are part of machine learning. This application will generalize behavior patterns. In this context, that means (identifying complex patterns in thousands or millions of data). Our application uses an algorithm that reviews the data and is able to predict future behavior.

The data has been processed in the periods (2016-2017). Through the predictive algorithm by behavior patterns (Learning closer) of Machine Learning. The information obtained has shown a diversity of preferences in the tastes of the clients. This diversity of tastes induces the behavior of the company in business processes (Management, functional, economic, and operational). As in any company there are several areas framed in the aforementioned processes. The information was made after the process of (closest KVecino) of Machine Learning is of great help for decision making for the management of the company in the (Management process). investigation, shows implicit characteristics of the tastes of the clients that are structured in the hours of attention, nationality of the clients, type of drinks, accompaniment, age, etc. (Statistical Figures). Finally, the information for a better way in the decisions in the different processes (functional, economic, operative). A clear sample of the aforementioned is a detailed description of inputs purchase variables, and personnel contract (Economic Process).



## INTRODUCCIÓN

La utilización de Machine Learning como una herramienta computacional en la búsqueda de patrones de comportamientos. Es un tipo de extracción de información está encaminado al desarrollo de técnicas para que las computadoras nos faciliten información adecuada por ellas sin intervención nuestra. Este aprendizaje se puede realizar mediante la detección de patrones de comportamiento dentro de un grupo de datos de manera que es el propio programa es el que predice qué situaciones podrían aparecer o no. (Gonzalez, 2014) Estos datos (cálculos) nos facilitaran conocer para, finalmente, iniciar toma de decisiones y conclusiones fiables para la gerencia. Su impacto está siendo enorme. A corto plazo lo más probable es que la herramienta Machine Learning se siga realizando como una solución adquirida. Sin embargo, a largo plazo será muy probable encontrar métodos de Machine Learning diseñados a de acuerdo o según las necesidades de cada empresa.

En la actualidad la empresa ChuchuHuasi tiene: datos de los clientes como consumos diarios, servicio de atención los cuales los usa solo para facturar y para hacer estadísticas, con estos datos se pueden hacer muchas más cosas, con Machine Learning se puede recopilar datos históricos, para organizarlos, formatearlos de una manera transaccional, lo que generara una base de datos que se pueda manejar para predecir futuros comportamientos que favorecerán los objetivos de negocio y evitar el mal uso de recursos.

La cantidad de datos que se maneja en una Organización para sacar conclusiones o toma de decisiones es muy difícil, pero en cambio para Machine Learning detectar estos patrones es más fácil, analizar datos de gran volumen y, además, los sistemas proporcionan resultados rápidos que optimizan recursos y procesos necesarios para su funcionamiento.

Y como resultado, lo que obtendremos para la empresa serán beneficios y mejoras de la utilización de su información las cuales brindaran predicciones de alto valor que nos dará solución al problema que se plantea, tomando como base la información necesaria que requiere la empresa ChuchuHuasi.





**INDICE DE CONTENIDO**

**PRESENTACIÓN** ..... 1

**DEDICATORIA** ..... 2

**AGRADECIMIENTOS** ..... 3

**RESUMEN** ..... 4

**ABSTRACT** ..... 6

**INTRODUCCIÓN** ..... 7

**INDICE DE ILUSTRACIONES** ..... 11

**INDICE DE TABLAS** ..... 12

**CAPITULO I. ASPECTOS GENERALES** ..... 14

**1.1. DESCRIPCIÓN DE LA SITUACIÓN ACTUAL** ..... 14

1.2. FORMULACION DEL PROBLEMA ..... 16

**1.3. OBJETIVOS** ..... 16

1.3.1. OBJETIVO GENERAL ..... 16

1.3.2. OBJETIVOS ESPECÍFICOS ..... 16

**1.4. HIPÓTESIS** ..... 16

**1.5. VARIABLES** ..... 16

**1.6. JUSTIFICACION** ..... 17

**1.7. METODOLOGIA DE LA TESIS:** ..... 17

**1.8. Matriz de Consistencia de la Investigación:** ..... 19

**1.9. CUADRO DE OPERACIONALIZACIÓN DE VARIABLES** ..... 20

1.9.1. Cuadro de Operacionalización de Variable Independiente..... 20

1.9.2. Cuadro de Operacionalización de Variable Dependiente ..... 21

**2. CAPITULO II: MARCO TEÓRICO** ..... 23

2.1. Aspectos Teóricos Pertinentes..... 23

2.1.1. Aprendizaje ..... 23

2.1.2. Aprendizaje Automático ..... 24

2.1.3. Aprendizaje Supervisado..... 25

2.1.4. Aprendizaje No Supervisado ..... 25

2.1.5. Aprendizaje por Refuerzo ..... 25

2.1.6. Algoritmos Utilizados ..... 25

A. Regresión Lineal ..... 26

B. KNN o KVecino más Cercanos ..... 26

B.1. FUNCIONAMIENTO ESPECÍFICO DEL ALGORITMO KNN ALGORITHM:(DATOS CHUCHUHUASI S.A.) ..... 27



- C.           **ARBOLES DE DECISIÓN** ..... 28
- D.           **K-MEANS**..... 28
- 2.2.11. **GESTOR DE BASE DE DATOS Y DISEÑADOR DE MODELOS:** ..... 35
- BASE DE DATOS MYSQL..... 35
- 2.2.12..... LENGUAJE DE PROGRAMACIÓN:  
          37
- **LENGUAJE DE PROGRAMACIÓN PHP:** ..... 37
- **LO QUE SE PUEDE HACER CON PHP** ..... 37
- **SEGURIDAD PHP** ..... 37
- **ADAPTABILIDAD**..... 38
- **SOPORTE PARA BASE DE DATOS** ..... 38
- **JAVASCRIPT**..... 38
- **HTML5** ..... 39
- **CSS3**..... 40
- **AJAX**..... 41
- **ARQUITECTURA DEL SISTEMA**..... 42
- 2.3.       **ANTECEDENTES DE LA INVESTIGACIÓN.**..... 43
- 2.3.1.     **ANTECEDENTES A NIVEL NACIONAL** ..... 43
- 2.3.2.     **Antecedentes a Nivel Internacional**..... 44
- 3.         **CAPITULO III: METODOLOGIA**..... 48
- 3.1.       **TIPO DE LA INVESTIGACIÓN:**..... 48
- 3.2.       **NIVEL DE LA INVESTIGACIÓN:**..... 48
- 3.3.       **DISEÑO DE LA INVESTIGACIÓN:** ..... 49
- 3.3.1.     **APLICACIÓN DEL ALGORITMO PARA EL CONSUMO DE DATOS: “VECINO MÁS CERCANO”** ..... 49
- 3.3.2.     **ENTENDIMIENTO DEL ALGORITMO** ..... 49
- A.         **FUNCIONAMIENTO GENERICO DEL ALGORITMO KNN ALGORITHM:**..... 49
- 3.4.       **POBLACIÓN Y MUESTRA** ..... 53
- 3.5.       **INSTRUMENTOS**..... 54
- 3.5.1.     **FUNCIONAMIENTO DEL ALGORITMO KNN ALGORITHM CODIFICACIÓN PYTHON:**.. 54
- 3.6.       **RECOLECIÓN Y ANALISIS DE DATOS**..... 59
- 3.6.1.     **FUNCIONAMIENTO DEL ALGORITMO KNN ALGORITHM CODIFICACIÓN WEKA EN EMPRESA CHUCHUASI S.A:**..... 59
- CAPITULO IV: RESULTADOS**..... 76
- DESARROLLO DEL APLICATIVO FUNCIONAL PARA MOSTRAR EL RESULTADO DE LA INVESTIGACION.** ..... 76



- 4.1. ANÁLISIS DEL PROCESO DEL NEGOCIO DESPUÉS DE APLICAR MACHINE LEARNING MEDIANTE PATRONES DE COMPORTAMIENTO EN LA TOMA DE DECISIONES: ..... 76
- 4.2. PROPUESTA DE MEJORA DEL PROCESO DE TOMA DE DECISIONES ..... 76
- 4.3. MOSTRAR LOS DATOS OBTENIDOS EN EL CAPÍTULO III, MEDIANTE UN APLICATIVO FUNCIONAL..... 77
  - 4.3.1. HERRAMIENTA UTILIZADA. BIZAGI:..... 77
  - 4.3.2. PROCESO DE TOMA DE DECISIONES. BIZAGI:..... 78
  - 4.3.3. PROCESO DE VALIDACION DE USUARIOS BIZAGI: ..... 79
  - 4.3.4. TECNOLOGÍAS UTILIZADAS:..... 80
    - a. SCRUM. .... 80
  - 4.3.4. LA ARQUITECTURA DEL SISTEMA SE PUEDE DIFERENCIAR EN 3 CAPAS:..... 82
    - a. Nivel de Persistencia ..... 82
    - b. Nivel Lógico ..... 85
    - c. Nivel de Presentación ..... 86
- CONCLUSIONES..... 95
- RECOMENDACIONES ..... 96
- GLOSARIO ..... 97
- REFERENCIAS ..... 101
- ANEXOS ..... 105



## INDICE DE ILUSTRACIONES

Ilustración 1: Organigrama de la Empresa Chuchuhuasi S.A ..... 14

Ilustración 2: Referencia BPM ..... 30

Ilustración 3: Referencia BPM. .... 31

Ilustración 4: Referencia Bizagi ..... 32

Ilustración 5: Referencia Pagina Materialize ..... 33

Ilustración 6: Esquema Cliente/Servidor ..... 41

Ilustración 7: Datos en texto Plano listos para ser trabajados y pasarlo por el algoritmo que utiliza Weka. Autoría Propia..... 62

Ilustración 8: Muestra la carga de datos en texto plano listo para que Weka haga el trabajo. .... 63

Ilustración 9: Muestra de datos corrida de ejecución 70.025%, autoría propia..... 64

Ilustración 10 En el siguiente cuadro de confusión, Autoría Propia. .... 65

Ilustración 11: En el siguiente cuadro podemos observar cómo se clasifico nuestra data. Autoría Propia ..... 66

Ilustración 12 En el siguiente cuadro podemos observar más detalladamente las coincidencias de los datos. Autoría Propia..... 67

Ilustración 13: Autoría Propia..... 68

Ilustración 14: Se observa de acuerdo al día y la edad del cliente. Autoría Propia ..... 69

Ilustración 15: Se observa el cruce entre el género y la bebida Autoría propia. .... 70

Ilustración 16. Se observa que la bebida y la edad. Autoría propia ..... 71

Ilustración 17: Se observa la música y bebida. Autoría propia ..... 72

Ilustración 18: Se observa si los clientes vienen acompañados o no y cuanto consumen. Autoría propia..... 73

Ilustración 19: Se observa que los clientes comienzan tomando bebidas tradicionales y luego cambian a bebidas exóticas, esto nos ayuda que nuestro clientes les ofertemos los tragos exóticos con ofertas o incentivarlos para que ellos les guste más este tipo de bebida. Autoría propia..... 74

Ilustración 20: Tabla de la base de datos dbBitácora (autoría propia). .... 84

Ilustración 21: Página de Presentación de Matriz de confusión (Autoría Propia) ..... 86

Ilustración 22: Login Usuario (autoría propia)..... 87

Ilustración 23: Autoría propia..... 87

Ilustración 24: Autoría Propia..... 88

**Formato de clave:** Ilustración 25: Autoría Propia ..... 88

Ilustración 26: Autoría Propia..... 89

Ilustración 27: Autoría propia..... 91

Ilustración 28: Autoría propia..... 93

Ilustración 29: Capas de la arquitectura del sistema..... 94



## INDICE DE TABLAS

Tabla 1: Operacionalización Autoría Propia .....	20
Tabla 2: Operacionalización Autoría Propia .....	21
Tabla 3: Datos Excel en el cual se muestra toda la data recolectada mediante la encuesta realizada a los clientes que concurren a la empresa. Autoría Propia	60
Tabla 4: Estos son los datos que pasaran a ser normalizados. Autoría Propia	61
Tabla 5:Product Backlog .....	82
Tabla 6: (Autoría Propia) .....	89
Tabla 7: Autoría Propia .....	90
Tabla 8: Autoría propia.....	92



# CÁPITULO I:



## CAPITULO I. ASPECTOS GENERALES

### 1.1. DESCRIPCIÓN DE LA SITUACIÓN ACTUAL

La empresa ChuchuHuasi S.A. hoy en día no cuenta con una información gerencial que direcciona sus lineamientos funcionales. La toma de decisiones en sus procesos no es proyectada correctamente ni ejecutada a tiempo. Lo cual se manifiesta en pérdidas económicas por no saber identificar, utilizar, o mejorar sus procesos. Esto también es percibido por los clientes de la empresa al no tener un servicio funcional solicitado por falta de insumos.

La empresa ChuchuHuasi en la actualidad brindar el servicio de venta de tragos exóticos, tradicionales y tropicales de la selva peruana en la ciudad del cusco, la empresa cuenta con datos recolectados de los clientes que concurren a su local. El manejo no adecuado de la información con la que cuenta la empresa es un problema, ya que esta no es precisa he instantánea para la toma de decisiones y que mejoren los procesos del negocio.

#### Organigrama de la Empresa Chuchuasi S.A.:

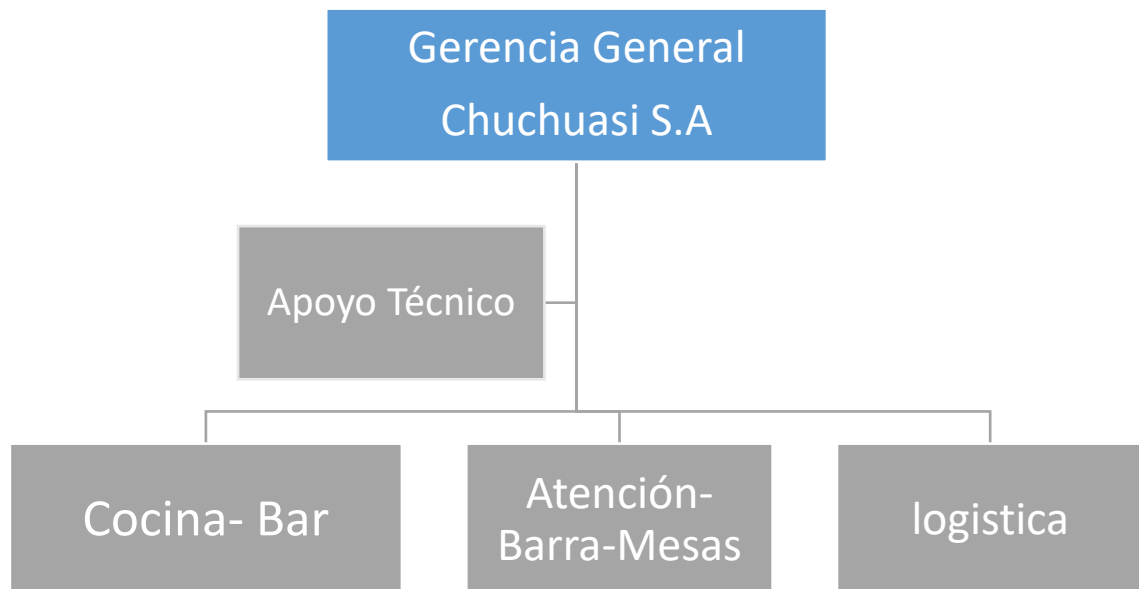


Ilustración 1: Organigrama de la Empresa Chuchuhuasi S.A

Luego de haber realizado el análisis situacional en el ámbito de los procesos del negocio como: las **actividades gerenciales, actividades funcionales, transaccionales y operativas** del negocio se ha



identificado las principales debilidades y deficiencias, las cuales podemos detallarlas como variables que contribuirán a la formulación del problema:

**Procesos del negocio:**

Los procesos más importantes de la Empresa Chuchuhuasi S.A. se han podido definir en el siguiente resumen:

- Proceso de Gestión
- Procesos Económico
- Proceso Operativo
- Proceso Funcional

**Toma de decisiones gerenciales:**

La potencialidad en el crecimiento de la cartera de clientes, está en función del mejoramiento de los procesos del negocio los cuales incluyen a los recursos como información, tiempo, y dinero.

Se realizó el análisis situacional de los procesos funciones de la Empresa Chuchuhuasi, sobre el conocimiento del negocio en referencia a:

- Gustos del cliente
- Frecuencias de visitas.
- Tiempo de permanencia de los clientes en el local
- Cantidad de recursos para la atención a los clientes tanto en aspecto material como de servicio de atención
- Potencialidad en el crecimiento de la cartera de clientes.

Esta información o datos no están adecuadamente canalizados en los procesos del negocio. Lo que no permite contar con herramientas estratégicas para la toma de sus decisiones que alineen correctamente la visión del negocio para el crecimiento de su cartera de clientes.





## 1.2. FORMULACION DEL PROBLEMA

¿Cómo influye en la toma de decisiones gerenciales, **la aplicación de Machine Learning mediante los patrones de comportamiento**, en los **proceso del negocio** de la empresa ChuchuHuasi S.A.?

## 1.3. OBJETIVOS

### 1.3.1. OBJETIVO GENERAL

Determinar la influencia de **Machine Learning mediante patrones de comportamiento**, en los **procesos del negocio**, para recomendar en la toma de decisiones gerenciales de la empresa ChuchuHuasi S.A.

### 1.3.2. OBJETIVOS ESPECÍFICOS

- Recolectar una cantidad de información anual a través de una muestra de datos, para la utilización o consumo del algoritmo predictivo por patrones de comportamiento (KVecino más cercano).
- Implementar y validar el algoritmo predictivo por patrones de comportamiento (KVecino más cercano). Ha generado una herramienta de recomendaciones sobre la variable dependiente.
- Desarrollar un aplicativo que muestre el resultado del algoritmo predictivo por patrones de comportamiento (KVecino más cercano).

## 1.4. HIPÓTESIS

La aplicación de **Machine Learning mediante patrones de comportamiento**, recomienda y aporta información relevante en los **procesos del negocio** para la toma de decisiones gerenciales en la empresa ChuchuHuasi S.A.

## 1.5. VARIABLES

- **Machine Learning mediante patrones de comportamiento.**
- **Procesos del Negocio**

### 1.5.1. VARIABLE INDEPENDIENTE

- **Machine Learning mediante patrones de comportamiento.**

### 1.5.2. VARIABLE DEPENDIENTE

- **Procesos del Negocio**

### 1.6. JUSTIFICACION

El propósito de la utilización de Machine Learning mediante patrones de comportamiento en la empresa ChuchuHuasi influencia, recomienda, aporta información relevante en la toma de decisiones gerenciales. Lo cual tiene una gran repercusión en los procesos del negocio.

El manejo adecuado y singular de los datos nos brindara una información adecuada en los procesos internos del funcionamiento en el aspecto de tiempos de proceso, contrato de personal y utilización de materiales de producción de bebidas que ofrece la empresa.

La información que se brinda tiene hasta 70% de certeza gracias al análisis predictivo proporcionado por la herramienta machine learning, la cual ayudará a la gerencia en la toma de decisiones y esto permitirá tener una ventaja competitiva sobre las demás empresas del rubro.

- **Relevancia Social**
- **Relevancia Aplicativa**
- **Relevancia Teórica**

### 1.7. METODOLOGIA DE LA TESIS:

La metodología de la investigación que se desarrolló en nuestra tesis fue de tipo cuantitativa. Según (Sampieri, 2014), con el objeto de analizar, conocer, y aplicar Machine Learning, utilizando datos y trabajarlos mediante algoritmos computaciones tales como patrones de comportamiento los cuales generalizan patrones de comportamiento. Las variables generadas en la investigación han sido evaluadas de acuerdo a su funcionalidad operativa en los procesos ubicados dentro de la empresa, y cuál es la relación que tienen entre ellas. Lo cual generará indicadores después del uso del Aprendizaje Automatizado (Machine Learning) en el manejo de datos. Los cuales se mostrarán a través de una interface de aplicación Web que será utilizado por la gerencia de la Empresa.

De la pregunta se establecen hipótesis y determinan variables; se traza un plan para probarlas a través de un diseño de tipo cuantitativo, se



miden las variables en un determinado contexto mostrado por las dos variables definidas

- Machine Learning mediante patrones de comportamiento.
- Procesos del Negocio

Se analizan las mediciones obtenidas utilizando métodos estadísticos, y se extrae una serie de conclusiones respecto a la hipótesis.

La investigación fue descriptiva por la situación de eventos que se encontraron a lo largo de la investigación los cuales identificaron de manera clara al problema en estudio. Los conocimientos obtenidos han sido recopilados de fuentes de información tales como libros, trabajos de investigación de pregrado y post grado, el Internet. Lo cual genero conocimiento cognoscitivo del tema de investigación: Aprendizaje Automatizado (Machine learning), en sus distintos niveles de investigación. El alcance más significativo de la investigación es el monitoreo de procesos del negocio a través de datos recolectados de los clientes los cuales fueron analizados por algoritmos computacionales proporcionados por Machine learning.



1.8. Matriz de Consistencia de la Investigación:

MATRIZ DE CONSISTENCIA DE LA INVESTIGACIÓN				
Título: BÚSQUEDA DE PATRONES DE COMPORTAMIENTO USANDO MACHINE LEARNING, PARA LA TOMA DECISIONES GERENCIALES EN LA EMPRESA CHUCHUHUASI				
Problemas	Objetivos	Hipótesis	Variables/Dimensiones	Metodología
General	General	General	Variable Independiente	
¿Cómo influye en la toma de decisiones gerenciales, la aplicación de Machine learning mediante los patrones de comportamiento, en el proceso del negocio de la empresa ChuchuHuasi S.A.?	Determinar la influencia de Machine Learning mediante patrones de comportamiento, en los procesos del negocio, para recomendar en la toma de decisiones gerenciales de la empresa ChuchuHuasi S.A.	La aplicación de Machine Learning mediante patrones de comportamiento, recomendará y aportará información relevante para mejorar los procesos del negocio y la toma de decisiones gerenciales en la empresa ChuchuHuasi S.A.	Machine Learning en patrones de comportamiento por patrones de comportamiento. <b>Dimensiones</b> <ul style="list-style-type: none"> <li>• Accesibilidad a los datos recolectados para el proceso de análisis predictivo por patrones.</li> <li>• Accesibilidad de estructura tecnológica.</li> <li>• Accesibilidad a los algoritmos de machine learning.</li> </ul>	<b>*Tipo</b> - <u>Recopilación</u> de Datos <b>*Población</b> - * <b>*Técnicas e Instrumento</b> <b>*Técnica</b> - Encuesta <b>*Instrumento:</b> - Cuestionario o test <b>* Técnicas para Análisis de Datos</b> - Estadística Descriptiva - Estadística Inferencial
Específicos	Específicos	Específicos	Variable Dependiente	
	Recolectar una cantidad de información anual a través de una muestra de datos, para la utilización o consumo del algoritmo predictivo por patrones de comportamiento (KVecino más cercano).		Plan Operativo Anual <b>Dimensiones:</b> <ul style="list-style-type: none"> <li>• Indicadores del Negocio</li> <li>• Indicadores funcionales</li> </ul>	
	Implementar y validar el algoritmo predictivo por patrones de comportamiento (KVecino más cercano). Ha generado una herramienta de recomendaciones sobre la variable dependiente.		<ul style="list-style-type: none"> <li>• Indicadores económicos</li> <li>• Indicadores operativos.</li> </ul>	
	Desarrollar un aplicativo que muestre el resultado del algoritmo predictivo por patrones de comportamiento (KVecino más cercano).			



### 1.9. CUADRO DE OPERACIONALIZACIÓN DE VARIABLES

#### 1.9.1. Cuadro de Operacionalización de Variable Independiente

Variable	Definición conceptual	Dimensiones	Indicadores
Machine Learning en patrones de comportamiento.	Es una herramienta basada en algoritmos computacionales, que se aplican sobre un conjunto de datos abundantes con la finalidad analizar predicciones generalizando patrones de comportamiento. (Tinaquero, 2018)	<ul style="list-style-type: none"> <li>• Accesibilidad a los datos recolectados para el proceso de análisis predictivo por patrones.</li> <li>• Accesibilidad de estructura tecnológica.</li> <li>• Accesibilidad a los algoritmos de machine learning.</li> </ul>	<ul style="list-style-type: none"> <li>• El periodo de captación de información obtenida diariamente de los clientes que visitaron el local en los años. Años (2015, 2016, 2017)</li> <li>• Para poder dar seguimiento y consumo de los datos se contó con la disponibilidad de estructuras hardware y software computacional</li> <li>• El análisis y aplicación de los algoritmos de solución (KVecino más cercano) con los datos recolectados. Estableció parámetros de referencia en el cumplimiento del objetivo principal de la investigación lo cual es mostrado en un aplicativo Web.</li> </ul>

Tabla 1: Operacionalización Autoría Propia



1.9.2. Cuadro de Operacionalización de Variable Dependiente

Variable	Definición conceptual	Dimensiones	Indicadores	Criterios de valoración
Procesos del Negocio	Un proceso es un conjunto de tareas lógicamente interrelacionadas para encontrar resultados favorables y bien definidos dentro de un negocio; por lo tanto, toman una entrada y le agregan valor para producir una salida (Barrios, 2010).	<ul style="list-style-type: none"> <li>• Procesos de gestión</li> <li>• Procesos funcionales</li> <li>• Procesos económicos</li> <li>• Procesos operativos</li> </ul>	<ul style="list-style-type: none"> <li>• Decisiones de gestión de acuerdo a los resultados anuales mostrados en el sistema Web.</li> <li>• Indicadores o cartas de productos (tragos o bebidas), ofrecidas al cliente en función de los gustos de los clientes, distribuidas en los días de la semana.</li> <li>• Indicadores para la compra de insumos. Según el gusto del cliente, se realiza mensualmente.</li> <li>• Indicadores operativos por cantidad de personal, e insumos utilizados, distribuidos por día de la semana.</li> </ul>	Nominal Inadecuado, poco adecuado, adecuado

Tabla 2: Operacionalización Autoría Propia



# CÁPITULO II:

---





## 2. CAPITULO II: MARCO TEÓRICO

### 2.1. Aspectos Teóricos Pertinentes

#### 2.1.1. Aprendizaje

Según Agustín Díaz define aprendizaje como: <sup>1</sup> Que el aprendizaje es la ejecución de cambios en el sistema que se acondicionan, en las cuales estos nos permitan realizar las tareas de un modo más eficiente y eficaz. En la práctica, el aprendizaje se usa para resolver problemas y puede representar la diferencia entre la resolución rápida y la imposibilidad de resolverlo. La idea de poder aprender de la propia experiencia en la resolución de problemas conlleva a esperar obtener mejores soluciones en un futuro. Además, está relacionado con el conocimiento. Puede definirse como el cambio permanente de la adquisición de conocimientos. Estos conocimientos son brindados por un emisor que en este caso sería el docente o profesor y por otra parte también puede adquirirse de manera propia. El aprendizaje es también el estudio del comportamiento del ser humano en el cual se va ganando experiencias, habilidades para que este tenga un mejor conocimiento y se desempeñe de la mejor manera posible.

Otra característica muy importante en el aprendizaje es la experiencia, ya que se tiene una estrecha relación en la práctica, esto quiere decir que cuando uno aprende necesita de determinadas reglas y la disponibilidad de material el cual ayudara a mejorar ese aprendizaje, como en nuestro caso la data que va ser muy importante para generar información necesaria. El aprendizaje refleja la forma de como una va aprendiendo y como este se va moldeando para construir nuevas formas de información.

---

<sup>1</sup> Agustín Alejandro Ortiz Díaz, (Tesis Doctoral) ALGORITMO MULTICLASIFICADOR CON APRENDIZAJE INCREMENTAL QUE MANIPULA CAMBIOS DE CONCEPTOS, 2014.





### 2.1.2. Aprendizaje Automático

El Aprendizaje Automático, Aprendizaje Computacional o Aprendizaje de Máquina (del inglés, Machine Learning) tiene como propósito el desarrollo de sistemas (programas de computador) de forma que las propias maquinas aprendan, o sea, que estas mejoren su propio funcionamiento para realizar sus actividades de forma automática, basándose en la adquisición de experiencia.

Un concepto importante de aprendizaje automático es que es una rama de la inteligencia artificial que se encarga de generar algoritmos que se encargan de aprender, esto nos sirve para que no tengamos que programar los algoritmos de manera explícita es decir no necesitamos que un desarrollador se sienta horas y horas a pensar los escenarios posibles y excepciones para hacer un programa, simplemente alimentamos a un algoritmo con un volumen gigantesco de data y que el algoritmo aprenda y sepa que hacer en cada una de estas excepciones.

Entonces podemos decir que, el Aprendizaje Automático nos ayuda a encontrar un patrón de acuerdo a una función obtenida de un algoritmo de computacional. El Aprendizaje Automático se habla desde los años 50 pero es hoy en día más usado por el aumento de la capacidad de computación y así como la gran cantidad de volúmenes de datos que las empresas tienen que manejar

Al analizar y diseñar un sistema de aprendizaje automático se debe tener en cuenta muchos elementos: el tipo de experiencia de entrenamiento, la medida de desempeño, la función objetivo y su representación, así como el algoritmo para aproximarla.

Machine Learning o Aprendizaje Automático se clasifican en tres grupos (Russell, Norving, & Gutierrez, 1996):

- Supervisado
- No supervisado
- Por refuerzo



### 2.1.3. Aprendizaje Supervisado

El aprendizaje supervisado consiste en tener un algoritmo en el cual brinda una información que nos ayudara a entender los datos cuando estos sean utilizados, quiere decir que nosotros entrenamos al algoritmo con preguntas que son las características y las respuestas que son las etiquetas que para que en un futuro estos algoritmos nos ayuden a tomar decisiones o hacer predicciones. Un ejemplo de esto es cuando nuestros docentes nos enseñan casos de problemas y como estos pueden ser resueltos, para que luego nosotros aplicando lo aprendido y sus métodos podamos resolver situaciones iguales (Harrington, Machine Learning in Action, 2012).

### 2.1.4. Aprendizaje No Supervisado

En el aprendizaje no supervisado, el conjunto de datos que se le es ingresado no contienen ninguna etiqueta o característica de lo que representa lo que se quiere es que el algoritmo aprenda por sí mismo y busque sus propios patrones para poder entender el conjunto de datos. Un ejemplo claro de esto son los bebés, que en un inicio cuando se les habla no entienden nada, pero mientras van escuchando más conversaciones cada vez su cerebro empezara a crear relaciones y estos empezaran a formar modelos de la funcionalidad del lenguaje y él bebe reconocerá esos patrones ya y así podrá responder o tener un estímulo. (Harrington, Machine Learning in Action, 2012)

### 2.1.5. Aprendizaje por Refuerzo

En el aprendizaje por refuerzo el algoritmo aprende observando toda la información que maneja, y utiliza la retroalimentación que obtiene respuesta a las acciones que se realizó para así aprender de sus propios errores y así poder encontrar nuevas estrategias que ayuden a mejorar la solución. (Harrington, Machine Learning in Action, 2012)

### 2.1.6. Algoritmos Utilizados

Entre los algoritmos más utilizados son los siguientes:



### A. Regresión Lineal

Se utiliza para calcular valores reales basadas en variables continuas, es un método causal en que una variable dependiente está relacionada con una o más variables independiente por medio de una ecuación lineal. Según (Shai Shalev Shwartz, 2014) La regresión lineal es una herramienta estadística común para modelar la relación entre algunas variables explícitas y algunos resultados reales valorados, como un problema de aprendizaje.

### B. KNN o KVecino más Cercanos

Es un algoritmo de clasificación no paramétrico, la regresión KNN este se realiza mediante el cálculo del promedio del punto  $k$  más cercano al punto que se está probando. Según (Shai Shalev Shwartz, 2014) la idea es memorizar el conjunto de entrenamiento y luego predecir la etiqueta de cualquier nueva instancia sobre la base de las etiquetas de sus vecinos más cercanos en el conjunto de entrenamiento, la razón detrás de este método se basa en la suposición de que las características que se utilizan para describir los puntos de domino son relevantes para su etiquetado de una manera que hace que los puntos cercanos probablemente tengan la misma etiqueta. Además, en algunas situaciones, incluso cuando el conjunto de entrenamiento es inmenso, encontrar un vecino más cercano puede hacerse extremadamente rápido.

Se puede mostrar un panorama en función de la variable  $x$  que se quiere aproximar a una similitud.

En este caso la distancia entre muestras y clases se define como sigue:

$$d_{VMP}(x, c_i) = \min_{\forall p \in c_i} d(x, p); \quad x \in E, c_i \in C \quad (1.5)$$

La muestra  $x$  se clasificará en la clase que contiene el prototipo más cercano.

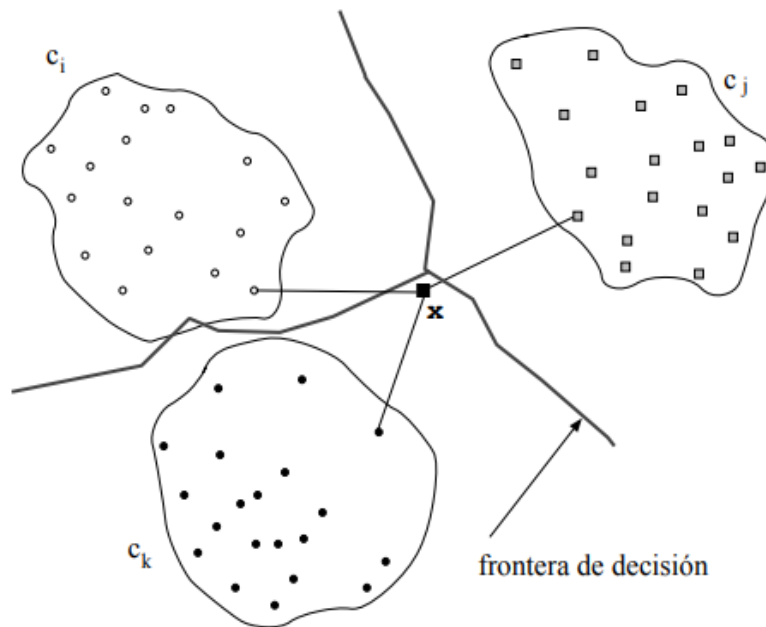


Ilustración 1: Referencia Algoritmos de búsqueda de vecinos más próximos en espacios métricos

## B.1. FUNCIONAMIENTO ESPECÍFICO DEL ALGORITMO KNN ALGORITHM:(DATOS CHUCHUHUASI S.A.)

Según (Diagramas, 2010) Se puede decir que un algoritmo es una serie de pasos que interpreta un modelos de solución para determinados problemas. O también como un grupo de instrucciones hechas de manera ordena que nos ayudan a solucionar un problema.

- ✓ **Recolectar:** Cualquier método.
- ✓ **Prepare:** los valores numéricos son necesarios para un cálculo de distancia. Una información estructurada el formato es mejor.
- ✓ **Analizar:** Cualquier método.
- ✓ **Entrenamiento:** no se aplica al algoritmo KNN.
- ✓ **Prueba:** calcula la tasa de error.
- ✓ **Uso:** esta aplicación necesita obtener algunos datos de entrada y valores numéricos estructurados de salida. A continuación, la aplicación ejecuta el algoritmo KNN en esta entrada de datos y

determina a qué clase pertenecen los datos de entrada. La aplicación entonces toma alguna acción en la clase calculada.

### C. ARBOLES DE DECISIÓN

Es una técnica que se utiliza para la toma de decisiones, toma en cuenta la complejidad de las situaciones, quiere decir que se trate de la resolución de problemas, la búsqueda minuciosa y de su aplicación para poder elegir la mejor alternativa posible que se pueda dar, y nos plantea problemas de diferentes perspectivas, y analiza completamente las posibles soluciones y cuantifica el costo y las probabilidades de uso que se pueda conocer.

Esta técnica nos muestra una secuencia de reglas las cuales nos ayudan a una clase de datos, con los que nosotros podamos examinarlo para poder sacar un resultado y poder predecir futuras toma de decisiones. Los árboles de decisión poseen una estructura que está conformada por: (Lior Rokach)

- Nodos: Nombre o identificadores de los atributos que caracterizan al conjunto de datos.
- Ramas: condiciones o variables de decisión que cumplen los objetos para poder separarse unos de otros.
- Hojas: conjunto o grupos de datos resultantes de la división que realiza el algoritmo.

### D. K-MEANS

Es un algoritmo de aprendizaje no supervisado más utilizado, utiliza algoritmos de agrupación más amplia. Según (Harrington, Machine Learning in action, 2012) K-Means es un algoritmo que encontrará k clúster para un conjunto de datos dado. El número de clústeres k está definido por el usuario. Cada grupo se describe por un solo punto conocido como el centroide (Que está en el centro de todos los puntos del clúster). El algoritmo k-means funciona de este modo. En primer lugar, los k centroides se asignan aleatoriamente en un punto. A continuación, cada punto del conjunto de datos se asigna a un clúster.

La asignación se realiza encontrando el centroide más cercano y asignado el punto a ese grupo. Después de este paso, los centroides se actualizan tomando el valor medio de todos los puntos de ese grupo.

### 2.2.7. METODOLOGÍA BPM:RAD

(Gianni, 2011)

Rapid Analysis & Designes es una metodología determinada y práctica, para el análisis y diseño de procesos orientados a la automatización con tecnologías BPM. Es un proceso en el cual un grupo de tareas que están relacionadas entre sí, en el que los usuarios y los analistas del proceso de desarrollo puedan relacionar para una serie de actividades.

Esta nueva forma de proceso nos ayuda a ver qué tan exitoso o no será nuestro proyecto de negocio, y es por eso que nos brinda la interacción entre los actores y recursos del negocio.

Para la Automatización de un proyecto utilizando BPM o BPM Suite es obligatorio seguir con sus fases en la cual la parte más problemática será la fase de análisis y diseño, el proceso de esta metodología BPM es de agilizar y simplificar cada proceso que se encuentre en la organización para poder optimizarlo.

Las ventajas de aplicar BPM: RAD son las siguientes:

- La fase de inicio se agiliza de BPM entre un 50% y un 70%.
- Entender y simplificarlos procesos del negocio.
- Modelizar y diseñar los procesos en su totalidad, holísticamente, con recursos, servicios, datos, reglas de negocio e indicadores.
- Diseñar procesos orientados a tecnologías BPM y de forma independiente del software que se implemente.
- Lograr una gestión del cambio más rápida y efectiva, para el desarrollo de capacidades y conocimiento en gestión por procesos y tecnologías BPM en la organización.

Facilitar, organizar y relacionar los recursos humanos como:

- ✓ Realizar técnicas en las cuales se aproveche al máximo el

conocimiento y talento del personal que está participando en el proceso.

- ✓ La construcción de una Arquitectura Empresarial, de abajo hacia Arriba.
- ✓ Garantizar la calidad de los modelos y diseños.

**Alcance**

Para conocer de mejor manera el alcancé que nos brinda la metodología BPM: RAD, se ilustrara las fases de un proyecto de análisis, desarrollo y puesta en marcha de un sistema BPM.

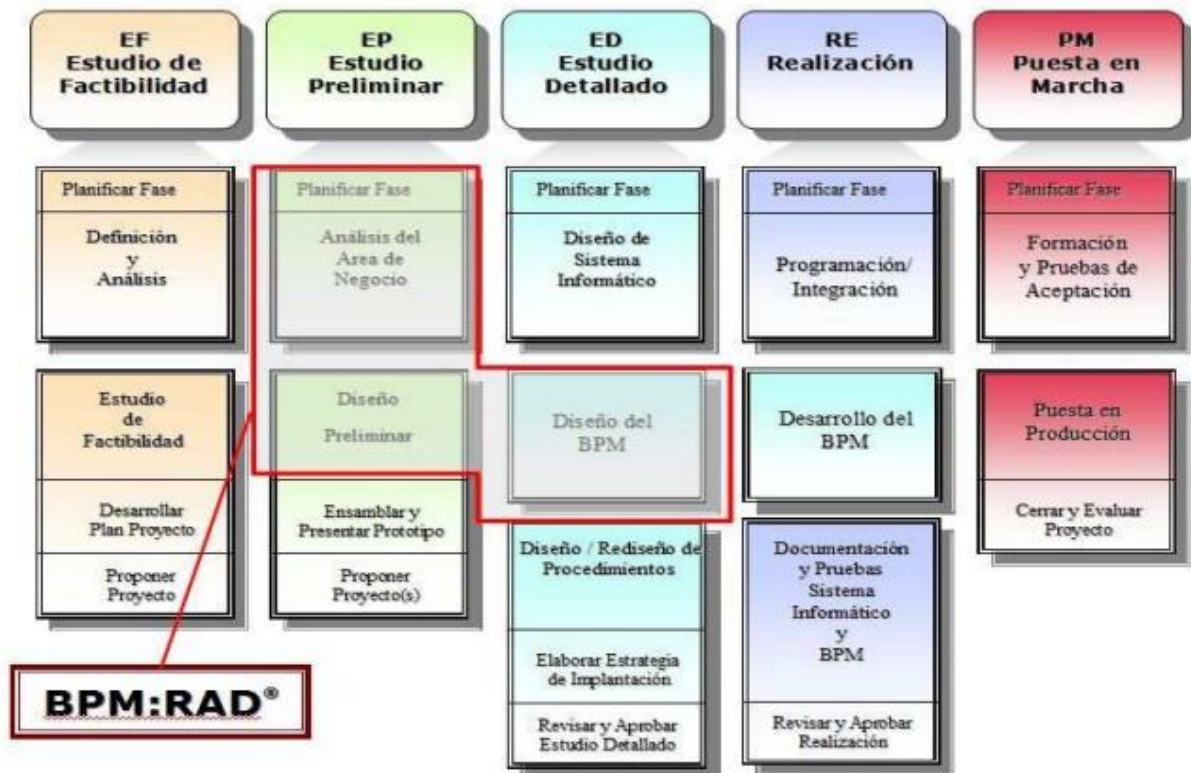


Ilustración 2: Referencia BPM

Esta metodología tiene tres etapas que son las siguientes:

- ✓ La modelización lógica
- ✓ Diseño preliminar
- ✓ Diseño BPM.

Al realizar los proyectos BPM no implementaremos de manera inmediata el empleo de las TIC, sino que consideraremos los procesos del negocio y como



este va funcionando. En BPM sin dejar de lado el uso de tecnología ordenaremos anticipadamente las soluciones de tecnología con los aspectos de negocio, no solo para la mejora y rentabilización de nuestras inversiones en TIC, sino para que estas realicen su tarea adecuadamente y den soporte y se pueda mejorar los modelos y características del negocio. Entonces BPM se transforma en una disciplina como primer punto ejecutar correctamente los procesos del negocio y luego utilizar de manera adecuada la tecnología para poder automatizar y mejorar el control de los procesos de diseño.

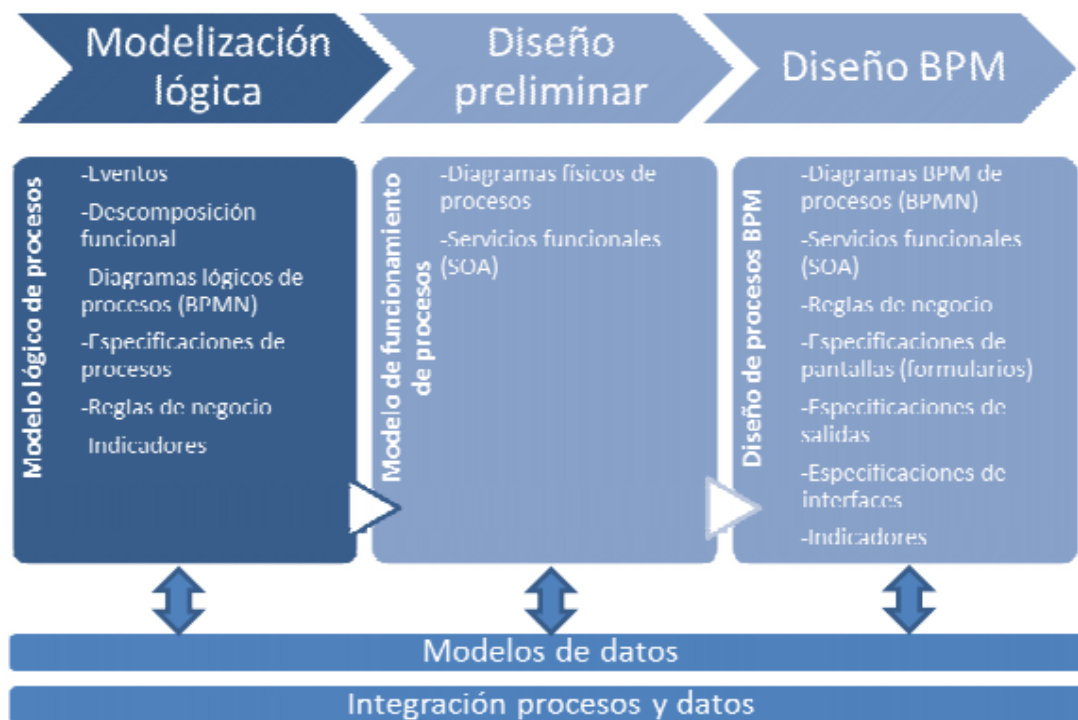


Ilustración 3: Referencia BPM.

## 2.2.8. HERRAMIENTA UTILIZADA. BIZAGI MODELER

La herramienta que se utilizara para el modelado de los procesos del negocio propuesto se seleccionó Bizagi Modeler BPMS ya que esta herramienta nos ofrece múltiples ventajas para el desarrollo de nuestro proyecto. Esta herramienta nos brinda las siguientes funcionalidades:

- ✓ Modelación de procesos (con notación BPMS de BOC y BPMN)
- ✓ Documentación de procesos
- ✓ Análisis de procesos
- ✓ Simulación de procesos



En este proyecto lo primero que realizaremos será la modelación y la documentación de los procesos básicos con la herramienta Bizagi. Este software nos brinda diferentes opciones de modelos y características de acuerdo a los escenarios posibles de gestión en el que se trabaje, los más utilizados y con bastante frecuencia son los que se menciona a continuación:

- ✓ El mapa de procesos
- ✓ El modelo de proceso de negocios
- ✓ El modelo de documento
- ✓ El modelo del entorno de trabajo

Estos cuatro modelos nos permitirán desarrollar de mejor manera el diagramado de procesos de la empresa y dirigirlos de mejor manera como tal. Y la ventaja de estos es posible determinar el entorno de trabajo con los modelos creados y así estos modelos poderlos interactuarlos con la mayoría de los modelos creados.

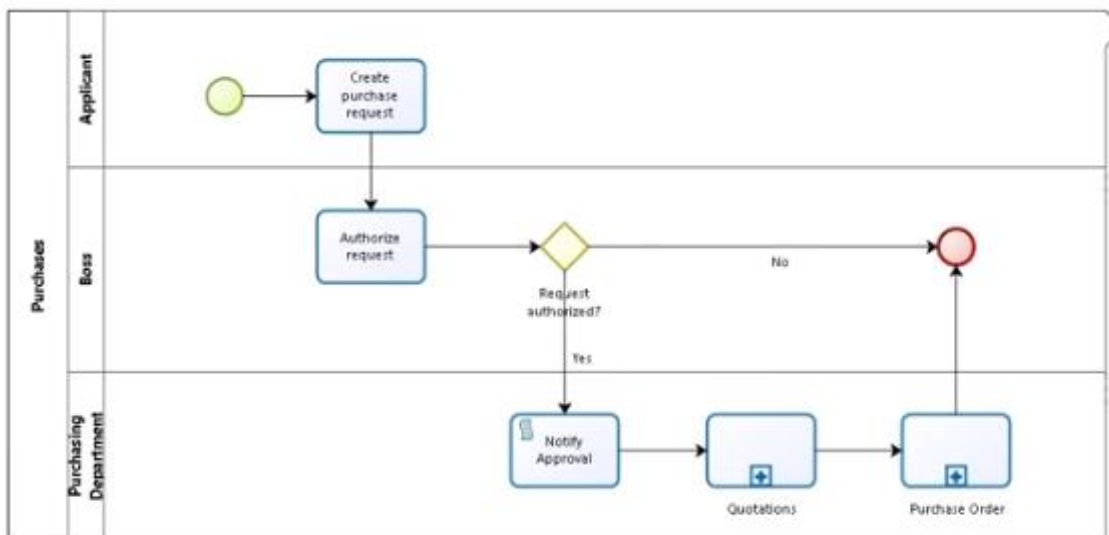


Ilustración 4: Referencia Bizagi

### 2.2.9. MATERIAL DESIGN:

(Design, 2018) Creado y diseñado por Google, Material Design es una renovación de la interfaz en la que se denominan animaciones y transiciones de respuesta, el relleno y los efectos de profundidad, tales como la iluminación y la sombra. El objetivo de Google es desarrollar un sistema con un diseño que permita al usuario la mayor experiencia con relación de todos sus productos de la plataforma de google.

La metáfora del material define la relación entre el espacio y el movimiento. La imagen de la tecnología ilumina en hoja y tinta, y que facilite al usuario la mejor creatividad y su innovación. La interfaz y diseño nos proporcionan marcas visuales que se nos familiarizan y esto hace que el usuario conozca la tecnología de una manera rápida para comprender el mundo físico mas allá.

Ubicación del framework en la web es: <https://materializecss.com/about.html>.

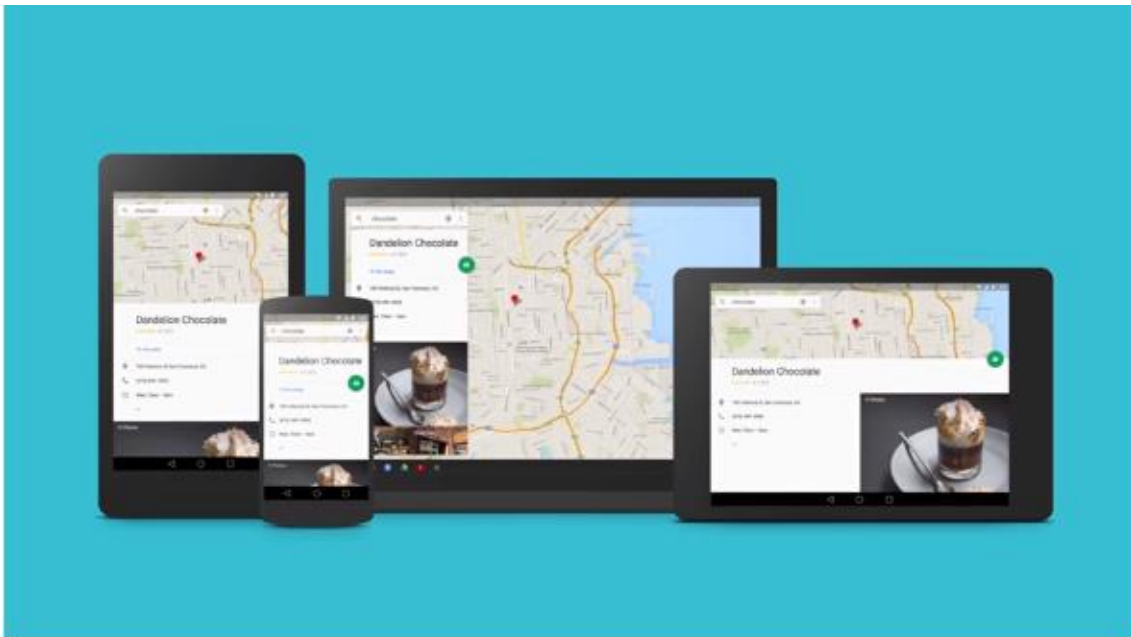


Ilustración 5: Referencia Pagina Materialize

### 2.2.10. WEKA 3:

(Pal, 2016) Weka es una herramienta con un conjunto de algoritmos de aprendizaje automático el cual nos ayudara con el trabajo de minería de datos. Los algoritmos que se presentan pueden ser utilizados directamente en un grupo de datos. Weka tiene los mecanismos para el pre procesamiento de datos, clasificación, regresión, clustering, reglas de asociación y visualización. Y nos brinda un correcto manejo para su desarrollo con nuevos esquemas de aprendizaje automático.

La referencia del nombre del software está vinculada a que, en las islas de Nueva Zelanda, el Weka es un ave no voladora con una naturaleza inquisitiva. El nombre se pronuncia así, y el pájaro suena así. Weka es un software de código libre el cual está bajo la Licencia Publica. Hemos estudiado el aprendizaje y bajo

la Licencia Pública General de GNU. Hemos reunido varios cursos los cuales gratuitos en línea que enseñan el aprendizaje automático y la minería de datos con Weka.



## MINERÍA DE DATOS: HERRAMIENTAS Y TÉCNICAS PRÁCTICAS DE APRENDIZAJE AUTOMÁTICO

(Pal, 2016) El aprendizaje automático proporciona un conjunto emocionante de tecnologías que incluye herramientas prácticas para analizar datos y hacer predicciones, pero también potencia los últimos avances en inteligencia artificial. Hemos escrito un libro que proporciona una introducción muy accesible al área, pero también está dirigido a los lectores que quieran profundizar en las técnicas más matemáticas disponibles en el modelado probabilístico moderno y los enfoques de aprendizaje profundo. Chris Pal se unió a Ian Witten, Eibe Frank y Mark Hall en la cuarta edición, y su experiencia en modelos probabilísticos y aprendizaje profundo ha ampliado enormemente la cobertura del libro. Para hacer espacio para el nuevo material, ahora ofrecemos un apéndice en línea sobre el software Weka. Es una versión extendida de una breve descripción de Weka incluida como un apéndice en el libro. El libro continúa proporcionando referencias a implementaciones Weka de algoritmos que describe. Los Weka MOOC proporcionan actividades similares a los ejercicios tutoriales en la 3ª



edición. Ahora también proporcionamos información sobre otro software: el ecosistema computacional para el aprendizaje automático ha crecido enormemente desde que escribimos la tercera edición en 2011.

## 2.2.11. GESTOR DE BASE DE DATOS Y DISEÑADOR DE MODELOS:

### BASE DE DATOS MYSQL

#### ***DEFINICIÓN DE MYSQL***

MySQL es un conjunto de elementos que ayudan a gestionar base de datos de código abierto, la cual tiene gran acogida y nos brindan la generación asequible y fiable de las aplicaciones de base de datos y que utiliza un lenguaje de programación llamado SQL basado en la web de un nivel muy alto y fácil de aplicarlo.

Este SGBD a diferencia de otros carece de características necesarias pero eso no implica ser una mala opción, al contrario es atractiva por los distintos sistemas comerciales, ayuda bastante para la práctica y desarrollo por su fácil uso y poco tiempo que necesita para poder ser puesto en marcha. Un beneficio de este es su distribución de licencia GPL y que nos brinda un grado de seguridad, escalabilidad, conectividad y un desarrollo rápido. (Casillas Santillan, Ginesta, & Perez Mora, 2009, pág. 5)

Para el desarrollo de las actividades, es necesario contar con el acceso del usuario administrador de gestor de base de datos, pero no siempre serán necesarios estos privilegios por ser menores.

El dominio de C/C++ y UNIX se puede contemplar de la misma forma en su sintaxis. Un ejemplo de esto sería, el uso de declaraciones regulares, la diferenciación de funciones por los paréntesis, los valores lógicos como 0 y 1, la utilización del tabulador para completar sentencias, por mencionar algunos. (Casillas Santillan, Ginesta, & Perez Mora, 2009, pág. 7)

#### ***MYSQL WORKBENCH***



Es un programa de entorno gráfico (IDE) de base de datos que nos ayuda a poder visualizar el desarrollo del sistema, administrar la base de datos, modelo de diagrama de base de datos, la creación y la conservación del sistema de base de datos MySQL. (Java - Oracle, 2016)

Lo sencillo de MySQL Workbench nos permite abarcar cinco temas principales:

- **DESARROLLO DE SQL**

Concede la facilidad de administrar y crear las relaciones entre los servidores de base de datos. Y esto permite la configuración de los parámetros de conexión, nos permite también el volumen de consultas SQL en la base de datos con sus conexiones empleando el built-in editor de SQL. (Java - Oracle, 2016)

- **MODELADO DE DATOS (DISEÑO)**

Aprueba la creación de prototipos de base de datos en forma visual en la que podemos ver el esquema de trabajo, y este a su vez no permite editar en tiempo real los aspectos de las tablas de la base de datos de manera completa. El editor de tablas nos brinda el fácil uso para la edición de las tablas, columnas, activadores, índices, inserciones, eliminación, creación de privilegios, rutinas y vistas. (Java - Oracle, 2016)

- **ADMINISTRACIÓN DEL SERVIDOR**

Permite administrar los usuarios e instancias del servidor MySQL, con este administrador accede a la visualización de la salud de la base de datos, la ejecución de las copias de seguridad, rescatar la información de la base de datos, la fiscalización de copias de seguridad, y el seguimiento de la productividad del servidor MySQL Backup y Auditoría de MySQL. (Affiliates, 2016)

- **LA MIGRACIÓN DE DATOS**

Facilita el cambio de Microsoft SQL Server, Microsoft Access, Sybase ASE, SQLite, SQL Anywhere, PostgreSQL, y otras tablas RDBMS, objetos y datos a MySQL. El cambio también es compatible con la



migración desde versiones anteriores de MySQL a las últimas versiones. Backup y Auditoría de MySQL. (Affiliates, 2016)

- **MYSQL ENTERPRISE SUPPORT**

Apoya en los recursos empresariales tales como MySQL Enterprise Backup y Auditoría de MySQL. (Affiliates, 2016)

## 2.2.12. LENGUAJE DE PROGRAMACIÓN:

### ➤ LENGUAJE DE PROGRAMACIÓN PHP:

PHP, acrónimo de "Hypertext Preprocessor", es un lenguaje de 'scripting' este es código libre en el cual esta adecuado para el uso de aplicaciones web el cual esta conectado en las páginas web HTML. Su sintaxis recurre a C, Java y Perl, y es fácil de aprender. La principal acción de este lenguaje es de proporcionar a los desarrolladores web, que puedan manejar de manera dinámica y de fácil acceso a las paginas generadas, aunque se pueda hacer muchas más cosas con PHP. (Group, The PHP, 2018)

- **LO QUE SE PUEDE HACER CON PHP**

Con PHP se puede hacer todo lo necesario ya que se enfoca primordialmente en scripts de parte del servidor, y esto ayuda a q pueda hacer muchas cosas como otros programas de CGI, entre estas características principales tenemos la recopilación de información de formularios, la generación de páginas dinámicas, y el envío y recepción de cookies. (Group, The PHP, 2018)

- **SEGURIDAD PHP**

PHP es un lenguaje e interprete muy potente, con los módulos web y binarios CGI, con este se puede acceder a los diferentes ficheros, realizar comandos o abrir conexiones del servidor mediante la red. Bueno este último hace que sea inseguro por parte del servidor web. Y es por ese motivo que PHP es lenguaje más seguro para escribir aplicaciones CGI, pero este tiene que tener un ajuste en el tiempo de ejecución como en el tiempo de compilación, y la correcta forma de



programación estos dos pueden mejorar la combinación de seguridad y libertad que se necesita. (Group, The PHP, 2018)

- **ADAPTABILIDAD**

Una de las principales características de PHP es que se puede aplicarse a los diferentes sistemas operativos que existen ya que al ser OpenSource se puede ejecutar en, Linux, Unix, Microsoft Windows, Mac OSX, RISC OS y en los otros más. Otra de sus adaptabilidades es que PHP se conecta con la mayoría de servidores web conocidos, entre estos los más conocidos como IIS, Apache. Esto incluye cualquier servidor web que pueda utilizar el binario de PHP FastCGI, como lighttpd y nginx. PHP funciona tanto como módulo como procesador de CGI. (Group, The PHP, 2018)

Con PHP, se tiene la facilidad de elegir el sistema operativo que desee y el servidor web que maneje. Y la ventaja de esta es que se puede aplicar POO o procedimientos almacenados o la mezcla entre estos. (Group, The PHP, 2018)

- **SOPORTE PARA BASE DE DATOS**

Una de las características más importantes, robustas y conocidas de PHP es de su extenso soporte de base de datos. Con PHP es de fácil acceso a una base de datos y de manera muy simple se puede utilizar extensiones específicas de bases de datos (p.ej., para mysql), o manejar una forma de abstracción de PDO, o de conectarse de maneras extensas o conexiones abiertas por la extensión de ODBC. Otras bases de datos podrían utilizar cURL o sockets, como lo hace CouchDB. (Achour, 2013)

➤ **JAVASCRIPT**

Javascript conocido también como JS es un lenguaje de programación interpretado es decir que no requiere compilación si no que es analizado por otro programa específicamente los navegadores web dicho de otra manera javascript está diseñado para que se ejecute en los navegadores web. Cabe resaltar que para insertar código javascript en un documento





HTML se debe utilizar la etiqueta `<script>` de apertura como de cierre de esta manera el navegador sabrá donde se encuentra el código javascript. Este lenguaje es generalmente para ejecutar acciones por lado del cliente, aunque también existen formas de como ejecutarlos por lados del servidor. Su principal característica es de brindar interactividad en las interfaces.

Este lenguaje también ayuda a los programadores dar efectos especiales dinámicos a las páginas web para proporcionarlos mayor viscosidad. Javascript está definido como un lenguaje de Programación Orientada a Objetos POO que se base en prototipo ya que en javascript no existen las clases, y también es débilmente tipado es decir que no hace falta ser especificado el tipo de dato que se almacene a diferencia de otros lenguajes de programación.

Estas nuevas APIs (como Web Storage, Canvas, y otras) son presentaciones que están incluidas en las librerías de los navegadores. Lo importantes es hacer que las funcionalidades de programación sean sencillas y tengan un estándar atravez de técnicas poderosas, expandiendo el alcance del lenguaje y facilitando la creación de programas útiles para la web. (Gauchat, 2012, pág. 84)

#### ➤ **HTML5**

HTML (HyperText Markup Language) es un formato de documentos para hipertexto quiere decir que es un documento en el cual se puede enlazar con otros documentos, HTML es el fundamento de World Wide Web, que es el lenguaje en el que la mayoría de las páginas web se editan o generan. (Pérez, Introducción a XHTML, 2010, pág. 5). Los documentos que se generan en HTML es de texto ASCII, quiere decir que pueden ser escritos en diferentes editores de textos básicos como complejos.

HTML5 nos brinda características importantes como el estilo, estructura y funcionalidad. Esto no fue oficialmente dicho o declarado, pero algunas APIs (Interface de Programación de Aplicaciones) y la especificación de CSS3 por completo no son parte de este, HTML5 es el producto combinado de HTML, CSS y Javascript. Cada una de estas tecnologías





son independientes una de otras pero actúan como una sola unidad en la cual se relacionan bajo una especificación de HTML5.

Entonces podemos decir que estas tres herramientas toman un papel importante ya que cada uno se encarga de algo en específico como HTML se encarga exclusivamente en la estructura del cuerpo del documento, CSS nos muestra la estructura diseñada y moldeada y Javascript hace que esta funcionalidad puesta sea dinámica para el usuario. Fuera de la integración de estos elementos en nuestra estructura HTML sigue siendo muy importante para la realización del documento. Ya que este nos muestra elementos muy importantes para ubicar y definir la integración de la estructura del documento. Y esta brinda elementos con los cuales podamos ubicarlos y darle a estos elementos funcionalidad estática o dinámica. Es por eso que HTML es muy importante en la estructura de una página web. (Gauchat, 2012, pág. 18)

➤ **CSS3**

CCS es una herramienta muy importante a la hora de trabajar HTML ya que nos provee de estilos visuales con los cuales se puede trabajar en el documento, tamaño, fondo, estilos, bordes, etc...

Esta herramienta es un complemento con el cual podemos superar las limitaciones y todas las dificultades que nos presenta HTML. Al comienzo este lenguaje proporcionaba atributos dentro de las etiquetas HTML lo cuales nos brindaban estilos a cada elemento, pero con el transcurso del tiempo esto evolucionó y cada vez se volvió más complejo y es por eso que HTML no pudo más con estos códigos complejos que nos mostraba CSS y es por esta razón que CSS se estuvo adaptando con una forma aparte estructurada de como presentarse.

La nueva versión de CSS3 nos muestra el mismo camino pero con un mayor compromiso, que este se encargue exclusivamente del diseño de nuestro HTML5 y ahora es una parte muy importante a la hora de desarrollar un sitio web y se podría decir que es parte vital de HTML5 y a pesar que son dos cosas muy diferentes. (Gauchat, 2012, pág. 42)

### ➤ AJAX

El término AJAX es un acrónimo de Asynchronous JavaScript y XML; se presentó por primera vez en el artículo "Ajax: A New Approach to Web Applications" publicado por Jesse James Garrett el 18 de Febrero de 2005. (Pérez, Introducción a AJAX, 2010, pág. 5)

En realidad, AJAX puede estar definido de la siguiente forma:

“Ajax es una herramienta que trabaja por sí sola, si no se trata de la combinación de diferentes tecnologías que luego de un trabajo bien organizado y estructurado se combinan en formas nuevas.”

Una de sus principales características de Ajax, es que nos muestra una interacción diferente con el usuario haciendo que las aplicaciones sean más dinámicas, y estas evitan las sobrecargas de las páginas.

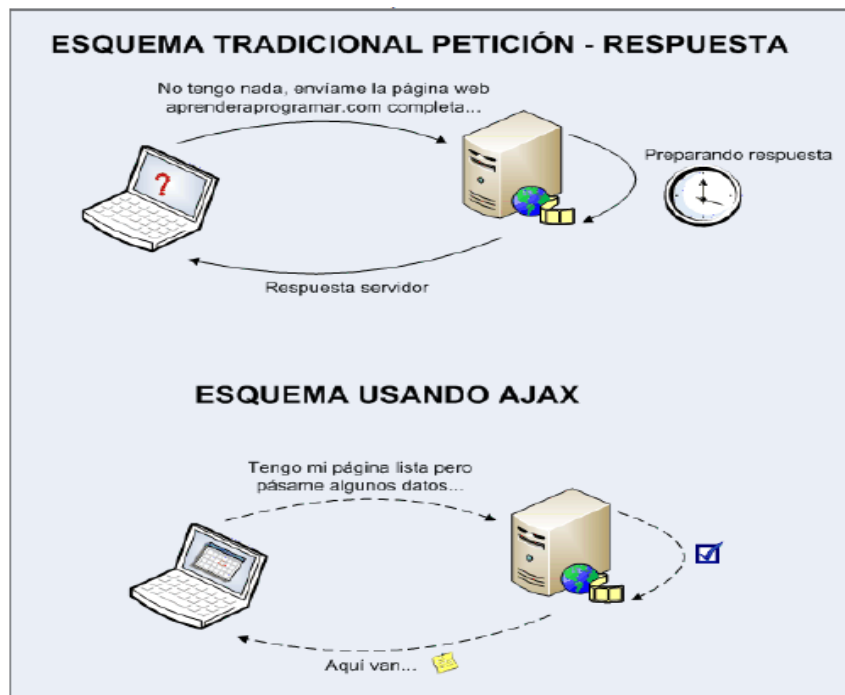


Ilustración 6: Esquema Cliente/Servidor

Fuente: (Rancel, 2013)

➤ **ARQUITECTURA DEL SISTEMA.**

El estilo arquitectónico utilizado para desarrollar el aplicativo del proceso de Toma de Decisiones esta realizado mediante el modelo cliente/servidor. Es un prototipo para el desarrollo del sistema de información, que realiza la tarea de brindar información, facilitando la mejora del rendimiento de nuestro sistema global de información. A ahora mostraremos la ilustración de este sistema:





**2.3. ANTECEDENTES DE LA INVESTIGACIÓN.**

**2.3.1. ANTECEDENTES A NIVEL NACIONAL**

<b>TÍTULO</b>	Modelo algorítmico para la clasificación de una hoja de planta en base a sus características de forma y textura.
<b>AUTOR</b>	Susana Milagros Malca Bulnes
<b>AÑO</b>	2015
<b>LUGAR</b>	Lima, Perú
<b>DESCRIPCIÓN</b>	Dado que el proceso de Minería de datos incluye el uso de algoritmos para hallar patrones en los datos y finalmente generar conocimiento, teniendo en cuenta que dentro del proceso de minería de datos existen una diversidad de métodos de clasificación, agrupación y regresión, el análisis desarrollado en el presente proyecto se limita a trabajar con 4 métodos de clasificación: Árboles de decisión, Redes Neuronales, Redes Bayesianas y KVecino más cercano. Estos cuatro métodos se seleccionaron teniendo como base de investigaciones realizadas en temas similares al presente proyecto y de las cuales se obtuvieron resultados positivos. Por otro lado, con el conjunto de datos debidamente estructurado, se adaptan y evalúan los modelos de clasificación obteniéndose de estos una serie de criterios de precisión que son utilizados para evaluar su rendimiento.



**APORTE PERSONAL** La utilización de estos 4 métodos ayudan que el software sea flexible; que este nos ayude a modificar el modelo que se tenga cada cierto tiempo, pero esto también dependerá mucho de la modificación del conjunto de datos que se le ingrese, el cual se irá incrementando, la utilización de la minería de datos ayuda bastante por sus técnicas empleadas.

### 2.3.2. Antecedentes a Nivel Internacional

<b>TÍTULO</b>	Machine Learning in Scientometrics
<b>AUTOR</b>	Ibáñez Martín, Alfonso (Martin, 2015)
<b>AÑO</b>	2015
<b>LUGAR</b>	Madrid, España
<b>DESCRIPCIÓN</b>	El aprendizaje automático trata sobre la construcción y el estudio de algoritmos que puedan aprender a partir de datos, mientras que la cienciometría se ocupa principalmente del análisis de la ciencia desde una perspectiva cuantitativa. Hoy en día, los avances en el aprendizaje automático proporcionan las herramientas matemáticas y estadísticas para trabajar correctamente con la gran cantidad de datos cienciométricos almacenados en bases de datos bibliográficas. En este contexto, el uso de nuevos métodos de aprendizaje automático en aplicaciones de cienciometría es el foco de atención. Esta tesis propone nuevas contribuciones en el aprendizaje automático que podrían arrojar luz sobre el área de la cienciometría.



**APORTE PERSONAL** A diferencia de otros estudios en este se presenta un nuevo método para descubrir un subconjunto importante para la predicción de un conjunto de índices las cuales luego se convertirán en indicadores muy importantes, pero al ser indicadores va depender mucho de su tamaño que a veces se comportan de una manera contra intuitiva debido a las inconsistencias de sus indicadores.

<b>TÍTULO</b>	Machine Learning Algorithms with Applications in Finance
<b>AUTOR</b>	Eyal Gofer
<b>AÑO</b>	2014
<b>LUGAR</b>	United States of America
<b>DESCRIPCIÓN</b>	Predicting future outcomes is a chief objective of statistics and machine learning. It is therefore reasonable to appeal to those disciplines for methods of coping with uncertainty. For the example above, it would seem natural to suggest a statistical model for the euro-dollar exchange rate that is based on past values. One might also employ sophisticated machine learning algorithms for predicting the future rate using any number of relevant financial indicators as input. Given such tools, one could hope to quantify the risk using a prediction of the exchange rate along



with an estimate of the accuracy of the prediction.

**APORTE PERSONAL**

Muestra que los límites son muy importantes a la hora de utilizar los algoritmos de aprendizaje. Los algoritmos de minimización son muy necesarios para cada escenario específico ya que su presentación utiliza límites los cuales nos ayudan a aplicar ampliamente los algoritmos existentes y además nuestros resultados robustos al comprender con límites estos nos ayudaran a que los resultados sean más claros.



# CÁPITULO III:

---







### 3. CAPITULO III: METODOLOGIA

#### 3.1. TIPO DE LA INVESTIGACIÓN:

La metodología de investigación es no experimental longitudinal o evolutivo ya que se está utilizando la recopilación de datos abundantes. (Sampieri, 2014) Y el uso del Aprendizaje Automatizado (Machine learning) mediante patrones de comportamiento generalizando tipos específicos de comportamiento. Describiendo las variables y analizando la incidencia de los flujos de datos analizados en los procesos del negocio. Esta información obtenida después de la aplicación de Machine Learning muestra datos implícitos, explícitos. No observados, referentes a los gustos del cliente, frecuencias de visitas, tiempo de permanencia de los clientes en el local, cantidad de recursos para la atención a los clientes tanto en aspecto material como de servicio de atención, potencialidad en el crecimiento de la cartera de clientes.

#### 3.2. NIVEL DE LA INVESTIGACIÓN:

El nivel de la investigación fue descriptiva ya que se realizó en función del objetivo general de la tesis, citándolo textualmente: “Determinar la influencia de **Machine Learning mediante patrones de comportamiento**, en los **procesos del negocio**, para recomendar en la toma de decisiones gerenciales de la empresa ChuchuHuasi S.A.”.

Donde se detalla o describe los componentes principales tales como “Procesos del Negocio” y “Machine Learning mediante patrones de comportamiento” para el logro del objetivo general planteado.

Según: (Hernandez, 2016) La Investigación descriptiva: se efectúa cuando se desea describir, en todos sus componentes principales una realidad.



### 3.3. DISEÑO DE LA INVESTIGACIÓN:

En el diseño de la investigación se ha concretado después del análisis y estudio de los algoritmos computacionales un esquema de funcionamiento secuencial que se muestra a partir del punto 3.3.1:

#### 3.3.1. APLICACIÓN DEL ALGORITMO PARA EL CONSUMO DE DATOS: “VECINO MÁS CERCANO”

#### 3.3.2. ENTENDIMIENTO DEL ALGORITMO

##### A. FUNCIONAMIENTO GENÉRICO DEL ALGORITMO KNN ALGORITHM:

Según (Harrington, Machine Learning in action, 2012) El algoritmo los K vecinos más cercanos (KNN) es un algoritmo de clasificación supervisada. Este algoritmo clasifica una gran cantidad de datos en un grupo en lo que este algoritmo ya conoce sus etiquetas o clases, el primer paso de este algoritmo es de conocer la distancia que tiene los diferentes datos para poder clasificarlos y seleccionada los K con la menor distancias. La instancia que se quiere agrupar será la que más se repita en los k que asigne.

La descripción genérica del algoritmo está basada en el manejo de un conjunto existente de datos adecuándolo al siguiente ejemplo:

- a. Conjunto de datos: Gusto del Cliente: Tenemos etiquetas para todos estos datos; sabemos en qué clase debe caer cada parte de los datos.
- b. Recibimos una nueva información sin una etiqueta: Comparamos esa nueva pieza de datos a los datos existentes, cada pieza de datos existentes.
- c. Tomamos el más similar: a las piezas de datos (los vecinos más cercanos) y mira sus etiquetas.
- d. Fijamos en la parte superior k: la mayoría de los datos similares de nuestro conjunto de datos conocido; aquí es de donde provienen los k menores. (k es un número entero y generalmente es menos de 20.)



e. Por último, tomamos una mayoría de votos de  $k$ , la mayoría de los datos similares, y la mayoría es la nueva clase que asignamos a los datos que se les pidió que clasificaran.

¿Clasificar las películas en románticas o películas de acción? Alguien vio muchas películas y contó la cantidad de patadas y besos en cada película. Se ha trazado seis películas para saber la cantidad de besos y patadas en cada película, Y encuentras una película que aún no has visto y quieres saber si es una película romántica o una película de acción. Para determinar esto, usaremos el algoritmo Vecino más cercano (KNN algorithm). El cual obtendrá datos cuantitativos, que demuestren los patrones de predicción.

En el enfoque de solución utilizando Machine learning, encontramos diversos algoritmos que nos permiten dar solución a nuestro consumo de datos que nos permite obtener características de la información que contribuirá al área de la gerencia de la Empresa Chuchuasi S.A.

Es así que en nuestra investigación de pregrado determinamos el uso del algoritmo llamado KNN algorithm (Vecino más cercano). Teniendo como conjunto de datos los RECOLECTADOS EN LA EMPRESA CHUCHUHUASI en función de gustos del cliente que consume las diferentes bebidas que ofrece el establecimiento. El cual tiene un enfoque de desarrollo enmarcado en los siguientes puntos secuenciales:

- ✓ **Recolección de datos:** Cualquier método.
- ✓ **Prepare:** los valores numéricos son necesarios para un cálculo de distancia. Una información estructurada el formato es mejor.
- ✓ **Analizar:** Cualquier método.
- ✓ **Entrenamiento:** no se aplica al algoritmo KNN.
- ✓ **Prueba:** calcula la tasa de error.
- ✓ **Uso:** esta aplicación necesita obtener algunos datos de entrada y valores numéricos estructurados de salida. A continuación, la aplicación ejecuta el algoritmo KNN en esta entrada de datos y determina a qué clase pertenecen los datos de entrada. La aplicación entonces toma alguna acción en la clase calculada.



✓ **Recolección de datos:**

El recojo o la acumulación de datos se obtuvo de las fichas o encuestas las cuales contaron con una serie de preguntas cerradas a los clientes que visitaban al establecimiento.

Toda esta data recolectada fue en un periodo de 1 a 2 años donde se utilizaron herramientas de captura de datos como encuestas, y un software para la parte estadística como Excel, ya que nos facilitaron los registros de la data buscada, desde abril de 2017 a noviembre de 2017. Los datos relevantes para la investigación se hizo un registro exhaustivo y en muchos casos se completó los datos según criterio del que toma la muestra (por ejemplo, la edad del consumidor). Otro dato que no tenemos referencia es el tiempo de permanencia en el local ya que muchos grupos se van reacomodando en el transcurso del tiempo. Adicionalmente se incluyó datos del cuaderno de bitácora, estos datos suministraron información de incidentes y nos es una referencia de control del que pudimos incluir muestras.

✓ **Prepare:**

Del proceso de recolección de datos se han establecido 8 variables que han sido definidos de la siguiente forma:

- Edad
- Procedencia
- Acompañado
- Genero
- Tipo de bebida
- Fecha (día de semana)
- Hora
- Cantidad (bebidas consumidas)

✓ **Analizar (Análisis de la data):**

Existen diferentes softwares en ámbito comercial y el software libre que permiten el análisis de datos en grandes cantidades.



Para realizar el análisis de data empleamos el software WEKA en su versión 3.8, con el cual se obtuvo las reglas de Clustering (KNN o KVecino más cercano) para agrupamiento de datos. La visualización de la data también nos permitió determinar la correlación entre variables.

Describir el funcionamiento de WEKA. Exige la utilización de archivos con formatos cvs (archivo plano). La exportación de la data se adoptó a través de la funcionalidad de Excel en la creación de formatos Cvs.

En la utilización de la herramienta Excel con sus diferentes funciones convierte las variables fecha, día de semana, así como para la variable hora establecerla como un tipo de dato real. Dado que la variable hora tenía una frecuencia mayoritaria a las 10 pm se hizo un ajuste para centrar esta frecuencia y mejorar la distribución esto es tomamos el valor máximo y mínimo e hicimos una nueva distribución. En cuanto a la variable procedencia para establecerla como variable de entrada asignamos la distancia en Km con referencia al Cusco, hemos de mencionar que esta estrategia no distinguía a visitantes nacionales del norte del país con por ejemplo visitantes extranjeros del sur del país, de modo que tuvimos que hacer una ponderación en la variable distancia de los visitantes extranjeros. Con esta data modificada se obtuvo

✓ **Entrenamiento:**

No se aplica al algoritmo KNN, por tanto al estudio tampoco.

✓ **Prueba:**

Para validar nuestro modelo tuvimos que probar en la mitad de los datos, es decir con la mitad de los datos se trató de hacer un modelo y en con la otra mitad se validó el **clasificador**. Dado que se trata de modelos supervisados tuvimos que escoger el modelo con los siguientes parámetros menor tasa de **aciertos en falsos**

**positivos**, esto es tenemos puntos que no corresponden a su clase, pero el modelo los establece como pertenecientes, otra métrica que utilizamos es **la tasa de aciertos** esto es cuando el modelo nos dice que cierto punto pertenece a una clase de la que estamos seguros que corresponde.

Para hacer las pruebas se escogió en orden intercalado la mitad de los datos que no formaron parte de los datos que fueron útiles para establecer el **clasificador**, esto con el fin de hacer pruebas con distribuciones homogéneas.

✓ **Uso:**

Con el modelo validado se puede hacer patrones de comportamiento, esta misma data puede pasar a formar data histórica con el que se puede reajustar el modelo, afinándose en cada iteración este proceso. Este proceso no es automático. Pues como dijimos se trata de aprendizaje supervisado y requiere la validación del modelo en cada iteración.

### 3.4. POBLACIÓN Y MUESTRA

✓ Población de estudio:

La población objetivo está conformada por los consumos de los clientes que visitan diariamente al negocio, es decir toda la información recolectada (Chuchuasi S.A.) en un periodo de dos años (2016 -.2017).

✓ Muestra de estudio:

La muestra de estudio corresponde a muestras establecidas en la matriz de confusión, tanto para la población de clientes del negocio (Chuchuasi S.A.), los cuales manifiestan sus gustos a través de sus pedidos o consumos dentro local. Se estableció el tamaño de la muestra y el tipo de muestreo que se utilizó tras el resultado de la matriz de confusión la cual hizo que los datos se confundieran y este



se clasifique gracias al algoritmo KNN con la colaboración de la herramienta Weka 3.8.

### 3.5. INSTRUMENTOS

#### 3.5.1. FUNCIONAMIENTO DEL ALGORITMO KNN ALGORITHM CODIFICACIÓN PYTHON:

El funcionamiento del algoritmo KNN o vecino más cercano da como resultado en la codificación presentada y la sintaxis del PROC (Procedimientos), DISCRIM (Discriminante).

Los datos son procesados en PROC lo cual no es muy compleja, necesitamos conjunto de datos de test y validación y generamos un Dataset de salida que denominamos puntuación en Method especificamos que la clasificación la llevamos a cabo con un método no paramétrico y en ese momento necesitamos especificar la  $k$  que es el número de vecinos que vamos a evaluar para clasificar cada observación y evaluamos como funciona nuestra clasificación mediante validación cruzada.

El resultado de esta clasificación no paramétrica lo podemos ver en el gráfico del código mostrado en el código Python, el resultado muy satisfactorio.

Es necesario tratar con más determinación el número de vecinos ( $k$ ) pero esta tarea se deja pendiente para otro momento sea automatizada.

El caso en los datos de la empresa Chuchuhuasi S.A. se asumió utilizado datos preferidos el uso de  $k$  medias pero knn solicito seleccionar las semillas iniciales para nuestra clasificación por  $k$  medias.

Distintos clientes clasificados por los gustos que son determinados por las semillas iniciales de nuestra clasificación a los cuales llamaremos patrones, o alineamientos de formación de tendencias.



Código basado en el libro (Harrington, Machine Learning in action,

2012)//**Definición de conjunto de datos o Clasificador**

```
from numpy import *
```

```
import operator
```

```
from os import listdir
```

```
def classify0(inX, dataSet, labels, k):
```

```
    dataSetSize = dataSet.shape[0]
```

```
    diffMat = tile(inX, (dataSetSize,1)) - dataSet
```

```
    sqDiffMat = diffMat**2
```

```
    sqDistances = sqDiffMat.sum(axis=1) #diferencia euclidiana
```

```
    distances = sqDistances**0.5
```

```
    sortedDistIndicies = distances.argsort()
```

```
    classCount={}
```

```
    for i in range(k):
```

```
        votellabel = labels[sortedDistIndicies[i]]
```

```
        classCount[votellabel] = classCount.get(votellabel,0) + 1
```

```
        sortedClassCount=sorted(classCount.iteritems()),
```

```
key=operator.itemgetter(1), reverse=True)
```

```
return sortedClassCount[0][0]
```





**//Datos definidos, en arreglos y etiquetas para probar y entender el algoritmo**

```
def createDataSet():  
  
    group = array([[1.0,1.1],[1.0,1.0],[0,0],[0,0.1]])  
  
    labels = ['A','A','B','B']  
  
    return group, labels
```

**//Datos de un archivo Excel se automatizo para ingresarlos a Python como matrices:**

```
def file2matrix(filename):  
  
    fr = open(filename)  
  
    numberOfLines = len(fr.readlines()) //atrapa el número de líneas del archivo  
  
    returnMat = zeros((numberOfLines,3)) //prepara la matriz a devolver  
  
    classLabelVector = [] //Prepara las etiquetas a devolver  
  
    fr = open(filename)  
  
    index = 0  
  
    for line in fr.readlines():  
  
        line = line.strip()  
        listFromLine = line.split('\t')  
        returnMat[index,:] = listFromLine[0:3]  
        classLabelVector.append(int(listFromLine[-1]))  
        index += 1
```



```
return returnMat,classLabelVector
```

**//Función para equalizar los datos**

```
def autoNorm(dataSet):  
    minVals = dataSet.min(0)  
    maxVals = dataSet.max(0)  
    ranges = maxVals - minVals  
    normDataSet = zeros(shape(dataSet))  
    m = dataSet.shape[0]  
    normDataSet = dataSet - tile(minVals, (m,1))  
  
    normDataSet = normDataSet/tile(ranges, (m,1)) # elemento sabio divide  
  
    return normDataSet, ranges, minVals
```

**//Función para validar el clasificador**

```
def datingClassTest():  
    hoRatio = 0.50  
    datingDataMat,datingLabels = file2matrix('datingTestSet2.txt') //Cargamos los  
datos de prueba  
    normMat, ranges, minVals = autoNorm(datingDataMat)  
    m = normMat.shape[0]  
    numTestVecs = int(m*hoRatio)  
    errorCount = 0.0  
    for i in range(numTestVecs):  
        classifierResult =  
classify0(normMat[i,:],normMat[numTestVecs:m,:],datingLabels[numTestVecs:m],3)  
        print "El clasif  
icador arroja: %d, El valor exacto es: %d" % (classifierResult, datingLabels[i])  
        if (classifierResult != datingLabels[i]): errorCount += 1.0  
    print "La tasa de error es: %f" % (errorCount/float(numTestVecs))
```



```
print errorCount
```

**//Función que permite probar y mostrar el resultado del algoritmo con un único dato:**

```
def classifyPerson():
    resultList = ['Etiqueta 1', 'Etiqueta 2', 'Etiqueta 3']
    percentTats = float(raw_input("Variable 1?"))
    ffMiles = float(raw_input("Variable 2?"))
    iceCream = float(raw_input("Variable 3?"))
    datingDataMat, datingLabels = file2matrix("datingTestSet2.txt") //archivo que
contiene la data
    normMat, ranges, minVals = autoNorm(datingDataMat)
    inArr = array([ffMiles, percentTats, iceCream])
    classifierResult = classify0((inArr - minVals) / ranges, normMat, datingLabels, 3)
    print "La probabilidad que esta persona consume ? es: ",
    resultList[classifierResult - 1]
```



### 3.6. RECOLECIÓN Y ANÁLISIS DE DATOS

#### 3.6.1. FUNCIONAMIENTO DEL ALGORITMO KNN ALGORITHM CODIFICACIÓN WEKA EN EMPRESA CHUCHUASI S.A:

La utilización de Weka 3.8. Como herramienta de simulación en el consumo de datos y poder generar patrones de comportamiento. A través de los datos obtenidos en la empresa Chuchuhuasi SA.

1. Ingresar a Weka 3.8 muestra un menú de opciones en la cual también podemos visualizar sus elementos de Weka. En el que los que necesitaremos para el desarrollo y es muy importante será el Explorer (para explorar los datos) y Experimenter (para realizar experimentos que comparen estadísticamente distintos algoritmos en distintos conjuntos de datos, de manera automatizada). Se utilizará el Explorer.
2. En la siguiente ilustración se muestra los datos recolectados a través del tiempo en un periodo de uno a dos años consecutivos los cuales se obtuvieron de las visitas de los clientes al local de la empresa. Estos datos fueron normalizados para su utilización en formato de texto plano. Y luego la normalización de los obtenidos en el formato Weka pueda utilizar o la preparación de la Data. Convertir el archivo Chuchuhuasi.xls a Chuchuhuasi.csv

Apellidos y Nombres	DNI o CE	País	Ciudad	Genero	Edad	Dirección	Celular	Correo Electrónico	Bebida	Fecha	Hora	Musica	Día Especial	Local	Amiente	Acompañado	Cantidad de Bebidas	Tipo de Bebida	Precio Aproximado de Gasto
Contreras Perez Estrella		Peru	Lima	Femenino	21	s/n			Mojito	19/01/2017	12:00:00 a.m.	Merengue	Sabado	Bueno	Bueno	no		1 Tradicional	
Tomes Bellido Elias		Peru		Masculino	43	s/n			Cerveza	21/01/2017	9:43:00 p.m.	Rock	Domingo	Bueno	Bueno	no		1 Normal	
SANCHEZ PEÑA HUGO		Peru	Lima	Masculino	38				3.59	11/08/2017	9:40:00 p.m.	Rock	Lunes	Regular	Bueno	no		1 Exótica	
RAMIREZ CHOTA Lucy Esther		Peru	Lima	Femenino	40				Leche de monja	27/10/2017	10:30:00 p.m.	Merengue	Lunes	Bueno	Bueno	no		1 Exótica	
SOTO BARBARAN GILBERTO		Peru	Ucayali	Masculino	33				Virgen colada	11/06/2017	11:35:00 p.m.	Alternativo	Miercoles	Bueno	Bueno	si		2 Exótica	
ENCISO MAYNAS OSCAR ELEDODORO		Peru	San Martín	Masculino	41				Coctel de café	15/10/2017	9:34:00 p.m.	Cumbia	Miercoles	Regular	Bueno	si		1 Exótica	
SANGAMA MONTES ELVIS		Peru	Lima	Masculino	27				Pisco sour	14/04/2017	7:25:00 p.m.	Salsa	Jueves	Regular	Bueno	si		2 Tradicional	
LECCA TORRES DANIEL		Peru	Ucayali	Masculino	30				Pisco sour	02/06/2017	8:30:00 p.m.	Salsa	Sabado	Bueno	Bueno	si		3 Tradicional	
ROJAS ELEGIANO JAIME AUGUSTO		Peru	Ucayali	Masculino	32				Maricucha	26/02/2017	9:50:00 p.m.	Alternativo	Lunes	Bueno	Bueno	si		2 Exótica	
GABAYAR LOAYZA YONNY EDUARDO		Peru	Tarapoto	Masculino	22				Esperame en el Suelo	10/08/2017	12:08:00 p.m.	Merengue	Martes	Bueno	Bueno	si		2 Exótica	
TANG TORRES SANDRA		Peru	Ucayali	Femenino	37				Mojito	23/09/2017	1:53:00 a.m.	Cumbia	Sabado	Bueno	Bueno	si		3 Tradicional	
PORTOCARRERO FUCHS WINDSOR		Peru	Ucayali	Masculino	40				Mojito	12/04/2017	10:00:00 p.m.	Salsa	Martes	Regular	Regular	si		3 Tradicional	
MORI AREVALO JESSICA		Peru	Ucayali	Femenino	42				Piña Colada	12/10/2017	2:30:00 a.m.	Rock	Miercoles	Bueno	Bueno	si		4 Tradicional	
BUENAFICO CUMAPA FRANCISCA		Peru	Lima	Femenino	23				Maricucha	08/09/2017	10:41:00 p.m.	Bachata	Domingo	Bueno	Regular	si		3 Exótica	
AYLLON PINCHI MARIANELA		Chile	Chile	Femenino	22				Charapita	14/09/2017	11:35:00 a.m.	Electro	Domingo	Regular	Regular	si		2 Exótica	
SORIA CORDOVA CARMEN KARINA		Chile	Chile	Femenino	31				Daquiri	14/01/2017	12:20:00 a.m.	Cumbia	Jueves	Bueno	Bueno	si		2 Tradicional	
RUIZ OLDRTEGUI CARLO MAGNO		Argentina	Argentina	Masculino	27				Tequila	09/02/2017	7:10:00 p.m.	Rock	Sabado	Bueno	Regular	si		2 Tradicional	
PANDURO SANDOVAL GUIDO		Argentina	Argentina	Masculino	30				Coctel de café	16/04/2017	8:14:00 p.m.	Electro	Sabado	Bueno	Bueno	si		3 Exótica	
CHAVEZ RENGIFO JANNET		Peru	Lima	Femenino	32				Charapita	08/03/2017	11:56:00 p.m.	Cumbia	Lunes	Regular	Bueno	si		3 Exótica	
DAZ AREVALO CESAR		Peru	Lima	Masculino	20				Mojito	05/04/2017	12:30:00 a.m.	Electro	Domingo	Bueno	Bueno	si		1 Tradicional	
GAMARRA GUTIERREZ GILDA		Colombia		Femenino	25				Chuchu Wasi	19/09/2017	11:15:00 p.m.	Rock	Miercoles	Bueno	Regular	si		3 Exótica	
QUISPE ARAGON NISEY DAIALA		Chile		Femenino	23				Pisco Sour	18/08/2017	9:40:00 p.m.	Electro	Jueves	Regula	Bueno	si		2 Tradicional	
BERRIOS APAZA, MARIA FLOR		México		Femenino	32				Virgen colada	16/06/2017	10:16:00 p.m.	Cumbia	Viernes	Bueno	Bueno	si		3 Exótica	
CONCHOY OCHOA, MARIA DELA		Argentina		Femenino	29				Virgen colada	26/07/2017	7:00:00 p.m.	Cumbia	Viernes	Bueno	Regular	si		2 Exótica	
HERREIRA GALLEGOS CARLOS		Peru	Cajamarca	Masculino	30				Levanta Muertos	07/10/2017	8:35:00 p.m.	Cumbia	Sabado	Regular	Bueno	si		3 Exótica	
ROJAS CACERES ANDRE		Peru	Lima	Masculino	25				Pisco Sour	11/02/2017	8:16:00 p.m.	Salsa	Sabado	Bueno	Bueno	si		1 Tradicional	
MORALES PEREZ EDITH		Peru	Lima	Femenino	28				Virgen colada	04/07/2017	6:16:00 p.m.	Salsa	Sabado	Bueno	Bueno	si		3 Exótica	
HUACALLO FARFAN JANET MARISELA		Peru	Lima	Femenino	24				Virgen colada	19/07/2017	6:16:00 p.m.	Cumbia	Sabado	Regular	Bueno	si		2 Exótica	
AÇHATA PALOMINO HDON		Peru	Cusco	Masculino	32				Esperame en el Suelo	06/01/2017	7:20:00 p.m.	Alternativo	Sabado	Bueno	Bueno	si		3 Exótica	
LEVA CASCAMAYTA AUCIA		Peru	Cusco	Femenino	31				Leche de monja	01/12/2017	7:20:00 p.m.	Alternativo	Sabado	Bueno	Bueno	si		4 Exótica	
FLÓREZ DELGADO RUTH		Venezuela		Femenino	27				Virgen colada	29/07/2017	8:50:00 p.m.	Bachata	Domingo	Regular	Bueno	si		3 Exótica	
ARMAS MUÑEZ DEL PRADO FERNANDO		Venezuela		Masculino	33				Levanta Muertos	19/08/2017	8:50:00 p.m.	Bachata	Domingo	Regular	Bueno	si		3 Exótica	
DEZA CRUZ KORINA		Venezuela		Femenino	29				Leche de monja	01/06/2017	8:50:00 p.m.	Salsa	Domingo	Bueno	Bueno	si		2 Exótica	
GUTIERREZ GOMEZ NEILER		México		Masculino	36				Siete veces sin sacarlo	29/11/2017	9:08:00 p.m.	Cumbia	Domingo	Bueno	Bueno	si		3 Exótica	
MANCHEGO MERCADO INGRID SABEL		México		Femenino	32				Piña Colada	02/12/2017	9:08:00 p.m.	Cumbia	Domingo	Regular	Bueno	si		2 Tradicional	
FLORES SOTO JACKELINE		México		Femenino	30				Virgen colada	14/09/2017	9:08:00 p.m.	Cumbia	Domingo	Bueno	Bueno	si		3 Exótica	
JAUSTO TUPAYACHI SANDRA		Peru	Tacna	Femenino	37				Pisco Sour	06/06/2017	6:10:00 p.m.	Merengue	Lunes	Bueno	Bueno	si		2 Tradicional	
FLORES GUTIERREZ ALEJANDRO		Peru	Tacna	Masculino	38				Siete veces sin sacarlo	12/05/2017	6:10:00 p.m.	Merengue	Lunes	Regular	Bueno	si		3 Exótica	
CÁRDENAS ZEGARRA OSCAR		EEUU	Cusco	Masculino	30				Levanta Muertos	11/05/2017	7:27:00 p.m.	Cumbia	Lunes	Bueno	Bueno	no		2 Exótica	
CAVO LEDWI		EEUU	Cusco	Masculino	20				Siete Raices	16/06/2017	9:27:00 p.m.	Cumbia	Lunes	Bueno	Bueno	no		2 Exótica	
MENDOZA ZARAVIA MICHAEL		EEUU	Cusco	Masculino	24				Esperame en el Suelo	08/11/2017	5:27:00 p.m.	Cumbia	Lunes	Bueno	Bueno	no		1 Exótica	
FUENTES FLORES LUZ		Argentina		Femenino	36				Pisco Sour	01/03/2017	9:30:00 p.m.	Balada	Lunes	Regular	Bueno	si		1 Tradicional	
MEZA GALARZA POOL		Argentina		Masculino	35				Pisco Sour	17/09/2017	9:30:00 p.m.	Balada	Lunes	Regular	Bueno	si		2 Tradicional	

Tabla 3: Datos Excel en el cual se muestra toda la data recolectada mediante las fichas o encuestas que son más de 5000 datos realizada a los clientes que concurren a la empresa. Autoría Propia

### 3. Discriminar variables que tenga mayor incidencia de relación mostrados en la Tabla N°5

Pais	Genero	Edad	Bebida	Fecha	Hora	Musica	Dia Especial	Local	Acompañado	Cantidad de Bebidas	Tipo de Bebida
Peru	Masculino	28	Amaserado de Suri	4/06/16	6:00:00 p. m.	Rock	sábado	Bueno	si		4 Exotica
Peru	Masculino	27	Amaserado de Uvachado	4/06/16	6:00:00 p. m.	Baladas	sábado	Bueno	si		3 Exotica
Peru	Masculino	28	Rom Puro	13/06/16	8:35:00 p. m.	Regueton	lunes	Bueno	no		2 Normal
Peru	Masculino	28	Coctel Pisco	11/06/16	7:21:00 p. m.	Rock	sábado	Bueno	no		1 Exotica
Peru	Femenino	22	Machupicchu	22/10/16	9:23:00 p. m.	Regueton	sábado	Bueno	si		3 Tradicional
Peru	Femenino	31	Coctel de Aguaje	17/08/16	11:12:00 p. m.	Electro	miércoles	Bueno	si		2 Exotica
Peru	Masculino	46	Esperame en el Suelo	20/11/16	11:30:00 p. m.	Rock	domingo	Bueno	si		3 Exotica
Colombia	Femenino	34	Pisco Sour	12/12/16	10:00:00 p. m.	Bachata	lunes	Bueno	si		2 Tradicional
Peru	Femenino	22	Machupicchu	2/07/16	8:00:00 p. m.	Bachata	sábado	Regular	si		3 Tradicional
Peru	Masculino	29	Rom con Gaseosa	17/09/16	9:45:00 p. m.	Rock	sábado	Bueno	si		4 Normal
Ecuatoriano	Masculino	26	Levanta Muertos	4/07/16	8:15:00 p. m.	Metal	lunes	Regular	si		2 Exotica
Peru	Femenino	20	Coctel de leche	12/09/16	9:15:00 p. m.	Balada	lunes	Bueno	si		3 Exotica
Italia	Femenino	26	Pisco Sour	23/11/16	10:32:00 p. m.	Bachata	miércoles	Bueno	si		2 Tradicional
Brasil	Masculino	33	Amaserado de Uvachado	17/08/16	11:02:00 p. m.	Rock	miércoles	Bueno	si		3 Exotica
Italia	Masculino	29	Cerveza	26/09/16	8:12:00 p. m.	Rock	lunes	Bueno	no		1 Normal
Italia	Masculino	33	Esperame en el Suelo	18/10/16	10:02:00 p. m.	Alternativo	martes	Regular	no		2 Exotica
Peru	Masculino	22	Chilcano de Pisco	13/12/16	21:00:00	Rock	martes	Bueno	no		1 Tradicional
Peru	Femenino	29	Mojito	23/12/16	10:32:00 p. m.	Bachata	viernes	Bueno	no		2 Tradicional
Peru	Femenino	32	Esperame en el Suelo	23/12/16	10:32:00 p. m.	Bachata	viernes	Bueno	si		1 Exotica
Chile	Femenino	26	Machupicchu	14/10/16	8:12:00 p. m.	Cumbia	viernes	Bueno	si		3 Tradicional
Mexico	Femenino	37	Amaserado de Suri	25/11/16	9:23:00 p. m.	Rock	viernes	Bueno	si		4 Exotica
Peru	Femenino	19	Mojito	13/11/16	10:45:00 p. m.	Alternativo	domingo	Regular	no		2 Tradicional
Colombia	Femenino	45	Esperame en el Suelo	26/12/16	11:12:00 p. m.	Rock	lunes	Bueno	no		4 Exotica
Italia	Masculino	34	Levanta Muertos	13/11/16	10:12:00 p. m.	Alternativo	domingo	Bueno	no		4 Exotica
Peru	Femenino	38	Pisco Sour	28/09/16	9:56:00 p. m.	Cumbia	miércoles	Bueno	no		2 Tradicional
Peru	Femenino	21	Mojito	19/01/17	0:00:00	Merengue	jueves	Bueno	no		1 Tradicional
Colombia	Masculino	30	Esperame en el Suelo	11/11/16	1:23:00 a. m.	Rock	viernes	Regular	no		1 Exotica
Italia	Masculino	43	Cerveza	19/11/16	12:45:00 a. m.	Rock	sábado	Bueno	no		2 Normal
Italia	Masculino	51	Cerveza	20/12/16	11:23:00 p. m.	Alternativo	martes	Bueno	no		3 Normal
Mexico	Femenino	35	Coctel de Aguaje	23/10/16	10:34:00 p. m.	Bachata	domingo	Bueno	no		3 Exotica
Peru	Femenino	32	Machupicchu	17/10/16	8:23:00 p. m.	Alternativo	lunes	Bueno	no		4 Tradicional
Cuba	Masculino	41	Cerveza	12/09/16	2:23:00 a. m.	Alternativo	lunes	Regular	si		2 Normal
Peru	Masculino	30	Cerveza	11/07/16	8:12:00 p. m.	Cumbia	lunes	Bueno	si		2 Normal
EEUU	Femenino	27	Machupicchu	21/08/16	9:54:00 p. m.	Electro	domingo	Bueno	no		2 Tradicional
Peru	Femenino	23	Coctel de Aguaje	30/12/16	11:32:00 p. m.	Bachata	viernes	Bueno	no		2 Exotica
Peru	Femenino	20	Mojito	15/11/16	7:32:00 p. m.	Merengue	martes	Bueno	si		3 Tradicional
Peru	Masculino	43	Cerveza	21/01/17	21:43:00	Rock	sábado	Bueno	no		1 Normal
Italia	Femenino	38	Rom con Gaseosa	28/12/16	1:23:00 a. m.	Alternativo	miércoles	Bueno	si		3 Normal
Ecuador	Masculino	46	Esperame en el Suelo	17/11/16	12:01:00 a. m.	Metal	jueves	Regular	no		4 Exotica
Bolivia	Masculino	34	Para Para	4/06/16	11:21:00 p. m.	Rock	sábado	Bueno	si		3 Exotica
Peru	Masculino	26	Tequila	10/07/16	10:12:00 p. m.	Rock	domingo	Bueno	si		2 Tradicional
Peru	Masculino	31	Cerveza	27/08/16	11:54:00 p. m.	Rock	sábado	Bueno	no		3 Normal

Tabla 4: Estos son los datos que pasaran a ser normalizados. Autoría Propia

Después de haber recolectado toda la data, solo se seleccionarán los datos con las etiquetas ya establecidas para encontrar los patrones que necesitamos.

- Los datos ya están listos y guardados en formato .csv como se visualiza en la Ilustración 8 para utilizarlos en Weka



Peru,Masculino,27,Daiquiri,1/04/17,18:16:00,Cumbia,s-bado,Bueno,si,5,Tradicional  
Peru,Femenino,29,Chuchuhuasi,6/07/17,18:16:00,Cumbia,jueves,Bueno,si,6,Exotica  
Peru,Masculino,19,7 Raices,7/01/17,19:20:00,Cumbia,s-bado,Bueno,si,6,Exotica  
Peru,Femenino,34,macerado de anis,31/08/17,19:20:00,Cumbia,jueves,Bueno,si,6,Exotica  
Peru,Femenino,29,Macerado de canela,18/08/17,21:23:00,Rock,viernes,Regular,si,3,Exotica  
Peru,Masculino,20,Pispachito,29/09/17,23:12:00,Rock,viernes,Bueno,si,2,Exotica  
Peru,Masculino,38,3.59,17/02/17,23:30:00,Cumbia,viernes,Bueno,si,4,Exotica  
Peru,Masculino,32,Daiquiri,12/08/17,22:00:00,Cumbia,s-bado,Bueno,si,2,Tradicional  
Peru,Masculino,41,Charapita,1/01/17,20:00:00,Salsa,domingo,Bueno,si,4,Exotica  
Peru,Masculino,24,Chuchu Exotic,1/12/17,21:45:00,Rock,viernes,Bueno,si,2,Exotica  
Peru,Femenino,29,Levantate Afasi,16/12/17,20:15:00,Rock,s-bado,Bueno,si,3,Exotica  
Peru,Femenino,38,Pisco Sour,7/01/17,21:15:00,Regueton,s-bado,Bueno,si,2,Tradicional  
Peru,Femenino,35,Machupicchu,27/07/17,22:32:00,Regueton,jueves,Bueno,si,4,Tradicional  
Peru,Masculino,19,Daiquiri,17/08/17,23:02:00,Cumbia,jueves,Bueno,si,2,Tradicional  
Peru,Masculino,20,mojito,15/10/17,20:12:00,Rock,domingo,Bueno,si,3,Tradicional  
Peru,Masculino,38,Charapita,23/06/17,22:02:00,Salsa,viernes,Bueno,si,2,Exotica  
Peru,Masculino,56,tramposa,26/08/17,21:34:00,Salsa,s-bado,Bueno,si,4,Exotica  
Peru,Femenino,51,Chuchu Exotic,23/07/17,22:32:00,Salsa,domingo,Bueno,si,3,Exotica  
Peru,Femenino,37,Tramposa,28/09/17,22:32:00,Salsa,jueves,Regular,si,3,Exotica  
Peru,Femenino,56,Levantate Afasi,3/06/17,20:12:00,Cumbia,s-bado,Regular,si,2,Exotica  
Peru,Masculino,30,Macerado de Suri,6/01/17,21:23:00,Cumbia,viernes,Bueno,si,4,Exotica  
Peru,Femenino,40,coctel de uva,24/06/17,22:45:00,Cumbia,s-bado,Bueno,si,5,Exotica  
Peru,Femenino,31,Esperame en el Suelo,12/05/17,23:12:00,Salsa,viernes,Bueno,si,2,Exotica  
Peru,Femenino,38,Rompe Calzon,24/02/17,22:12:00,Electro,viernes,Bueno,si,3,Exotica  
Peru,Masculino,26,7 Raices,6/04/17,21:56:00,Electro,jueves,Bueno,si,5,Exotica  
Peru,Femenino,34,Chuchu Exotic,27/05/17,0:00:00,Salsa,s-bado,Bueno,si,5,Exotica  
Peru,Femenino,30,Cerveza,18/11/17,1:23:00,Salsa,s-bado,Bueno,si,3,Normal  
Peru,Femenino,29,sangria,29/06/17,0:45:00,Cumbia,jueves,Bueno,si,2,Tradicional  
Peru,Femenino,36,sanfria,1/11/17,23:23:00,Salsa,miÉrcoles,Bueno,si,2,Tradicional  
Peru,Masculino,30,Pisco Sour,18/11/17,22:34:00,Cumbia,s-bado,Bueno,si,4,Tradicional  
Peru,Masculino,46,coctel de uva,4/11/17,20:23:00,Cumbia,s-bado,Bueno,no,2,Exotica  
Peru,Femenino,41,Cocona Sour,25/03/17,2:23:00,Cumbia,s-bado,Bueno,no,3,Tradicional  
Peru,Femenino,19,Machupicchu,12/07/17,20:12:00,Rock,miÉrcoles,Bueno,no,2,Tradicional  
Peru,Femenino,37,Mojito,31/12/17,21:54:00,Baladas,domingo,Regular,si,1,Tradicional  
Peru,Femenino,36,Mojito,17/11/17,23:32:00,Regueton,viernes,Bueno,si,5,Tradicional  
Peru,Femenino,30,Pisco Sour,2/07/17,19:32:00,Rock,domingo,Bueno,no,2,Tradicional  
Peru,Masculino,19,Pispachito,10/12/17,21:43:00,Regueton,domingo,Bueno,no,3,Exotica  
Peru,Femenino,26,3.59,11/02/17,1:23:00,Electro,s-bado,Bueno,si,1,Exotica  
Peru,Masculino,27,Pisco Sour,5/08/17,0:01:00,Rock,s-bado,Bueno,si,2,Tradicional  
Peru,Femenino,30,Cuba Libre,27/06/17,23:21:00,Bachata,martes,Bueno,no,5,Normal  
Peru,Femenino,44,Tequila,15/10/17,22:12:00,Bachata,domingo,Bueno,no,2,Tradicional  
Peru,Masculino,35,coctel de uva,28/01/17,23:54:00,Rock,s-bado,Bueno,no,3,Exotica  
Peru,Masculino,26,macerado de anis,17/12/17,21:22:00,Metal,domingo,Regular,no,3,Exotica  
Peru,Masculino,31,coctel de uva,9/04/17,0:34:00,Balada ,domingo,Bueno,no,3,Exotica  
Peru,Femenino,35,Coctel de cafÉ,23/03/17,23:12:00,Bachata,jueves,Bueno,si,2,Exotica  
Peru,Masculino,33,Machupicchu,17/06/17,21:34:00,Rock,s-bado,Bueno,si,4,Tradicional  
Peru,Femenino,31,Pisco Sour,10/03/17,22:01:00,Rock,viernes,Bueno,si,1,Tradicional  
Peru,Femenino,37,Cocona Sour,28/06/17,21:35:00,Alternativo,miÉrcoles,Bueno,si,3,Tradicional  
Peru,Masculino,20,Cocona Sour,19/07/17,21:40:00,Rock,miÉrcoles,Regular,si,8,Tradicional  
Peru,Masculino,22,Machupicchu,27/10/17,22:30:00,Bachata,viernes,Bueno,no,3,Tradicional  
Peru,Masculino,37,Pescheco Azul,12/07/17,23:35:00,Bachata,miÉrcoles,Bueno,no,2,Exotica  
Peru,Masculino,42,3.59,9/09/17,21:34:00,Cumbia,s-bado,Bueno,no,1,Exotica  
Peru,Femenino,19,Coctel de cafÉ,21/12/17,19:25:00,Rock,jueves,Bueno,no,1,Exotica  
Peru,Femenino,36,Charapita,16/12/17,20:30:00,Alternativo,s-bado,Regular,no,1,Exotica  
Peru,Masculino,50,Mojito,10/08/17,21:50:00,Rock,jueves,Bueno,no,4,Tradicional  
Peru,Masculino,47,Chuchu Wasi,25/08/17,12:00:00,Alternativo,viernes,Regular,no,4,Exotica  
Peru,Femenino,62,Pisco Sour,19/09/17,1:53:00,Cumbia,martes,Bueno,no,3,Tradicional  
Peru,Masculino,37,Virgen colada,26/05/17,22:00:00,Merengue,viernes,Regular,no,5,Exotica  
Peru,Femenino,19,Virgen colada,8/07/17,2:30:00,Rock,s-bado,Regular,si,2,Exotica  
Peru,Masculino,35,Levanta Muertos,6/08/17,22:41:00,Rock,domingo,Bueno,si,7,Exotica  
Peru,Femenino,26,Pisco Sour,27/12/17,11:35:00,Alternativo,miÉrcoles,Bueno,si,3,Tradicional  
Peru,Masculino,38,Virgen colada,12/04/17,0:20:00,Bachata,miÉrcoles,Bueno,si,2,Exotica  
Peru,Masculino,49,Virgen colada,18/05/17,19:10:00,Alternativo,jueves,Bueno,si,3,Exotica

*Ilustración 7: Datos en texto Plano listos para ser trabajados y pasarlo por el algoritmo que utiliza Weka.  
Autoría Propia*

Una vez tenido toda la data trabajada se tiene que pasar a texto plano en el cual se tiene que tener mucho cuidado, en que en esta parte no exista comas de más o espacio entre comas por que perjudicaría en el análisis de los datos.

5. Realizamos la carga de datos ya normalizados en formato .csv a Weka como se muestra en la ilustración 9

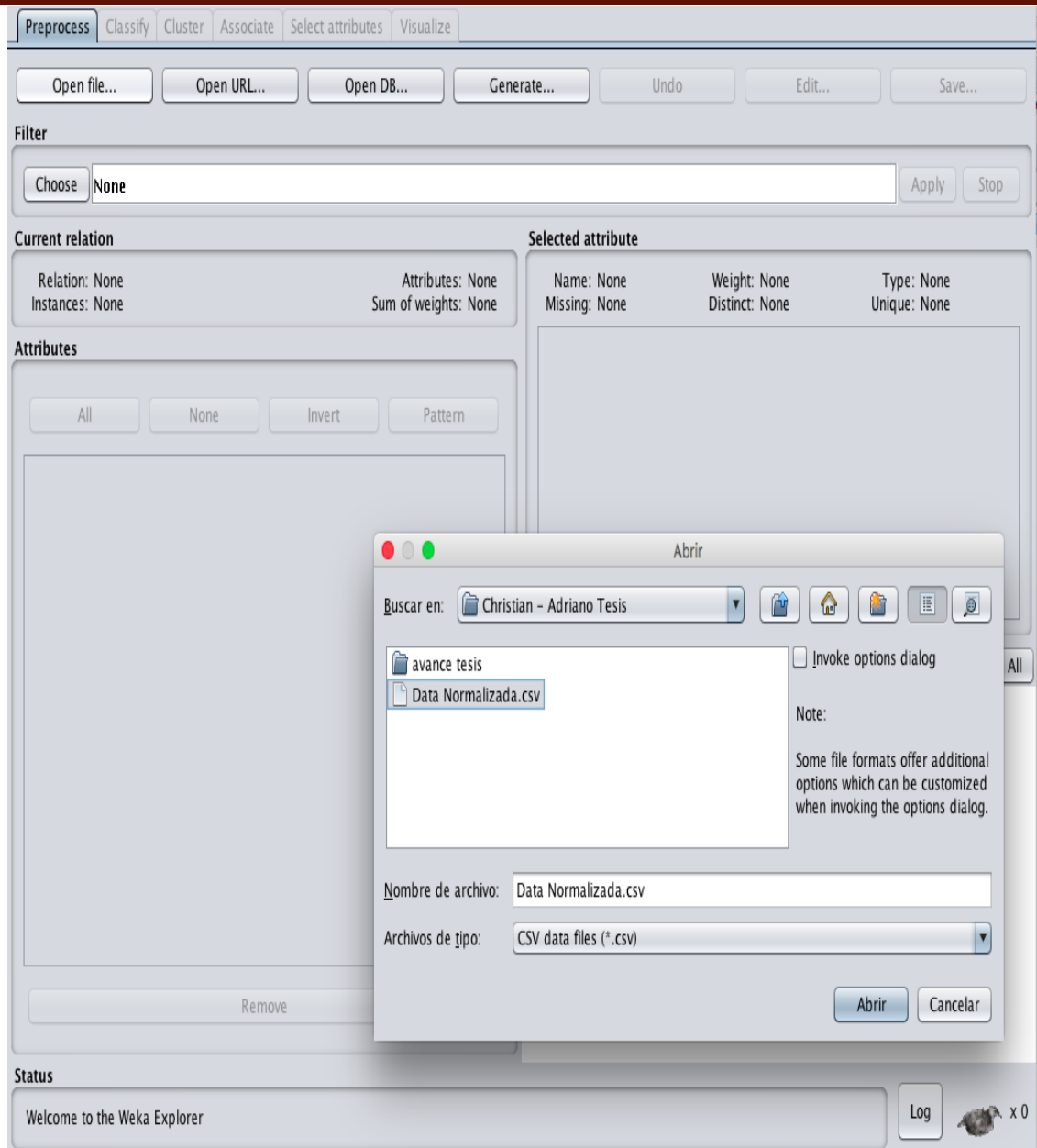


Ilustración 8: Muestra la carga de datos en texto plano listo para que Weka haga el trabajo.

En esta parte se subió la data ya normalizada y preparada con las etiquetas que nosotros le asignamos, para que el algoritmo pueda clasificar ya que utilizamos la clasificación supervisada.



6. Instancia de clasificación del modelo de datos y lo ejecutamos la muestra de datos que será clasificada por el tipo de bebida como se muestra en el Ilustración N°10

The screenshot shows the Weka Classifier window. The classifier selected is 'Ibk -K 1 -W 0 -A "weka.core.neighboursearch.LinearNNSearch -A "weka.core.EuclideanDistance -R first-last"'. The test options are set to 'Cross-validation' with 'Folds' set to 20. The classifier output shows the following details:

```

Relation: Encuestas-adri 2
Instances: 4000
Attributes: 12
Pais
Genero
Edad
Bebida
Fecha
Hora
Musica
Dia Especial
Local
Acompañado
Cantidad de Bebidas
Tipo de Bebida
Test mode: 20-fold cross-validation

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

Ib1 instance-based classifier
using 1 nearest neighbour(s) for classification

Time taken to build model: 0 seconds

=== Stratified cross-validation ===
=== Summary ===

Correctly Classified Instances      2801      70.025 %
Incorrectly Classified Instances    1199      29.975 %
Kappa statistic                    0.458
Mean absolute error                 0.1212
Root mean squared error             0.3438
Relative absolute error              54.9956 %
Root relative squared error         103.6005 %
Total Number of Instances          4000

=== Detailed Accuracy By Class ===

      TP Rate  FP Rate  Precision  Recall  F-Measure  MCC   ROC Area  PRC Area  Class
0.742   0.273   0.768     0.742   0.755     0.468  0.741     0.719     Exotica
0.454   0.040   0.470     0.454   0.462     0.420  0.708     0.262     Normal
0.688   0.224   0.651     0.688   0.669     0.460  0.738     0.576     Tradicional
    
```

Ilustración 9: Muestra de datos corrida de ejecución 70.025%, autoría propia

En esta ilustración podemos ver el resultado de la matrix de confusión tras la clasificación supervisada del algoritmo KNN con la cantidad de 4000 datos.



7. Validación de modelo con Matriz de Confusión Ilustración N°11.9

=== Confusion Matrix ===

a	b	c		<-- classified as
1611	87	498		a = Exotica
91	137	63		b = Normal
396	66	1051		c = Tradicional

Ilustración 10 En el siguiente cuadro de confusión, Autoría Propia.

Luego de haber hecho las corridas o iteración correspondientes este es el resultado de la matriz de confusión, podemos observar que de 2196(es la sumatoria de a, b, c) tragos exóticos 87 son tragos normales y 498 son tradicionales, de los 291(es la sumatoria de a, b, c) tragos normales predijo que 91 era exóticos y 63 son tradicionales, y por último de los 1513 (es la sumatoria de a, b, c) tragos tradicionales 396 son exóticos y 66 son normales. La matriz puede distinguir razonablemente bien entre las diferentes categorías de tragos.

8. Grafica de resultados de cada etiqueta de acuerdo al tipo de bebida como se muestra en la Tabla N° 8

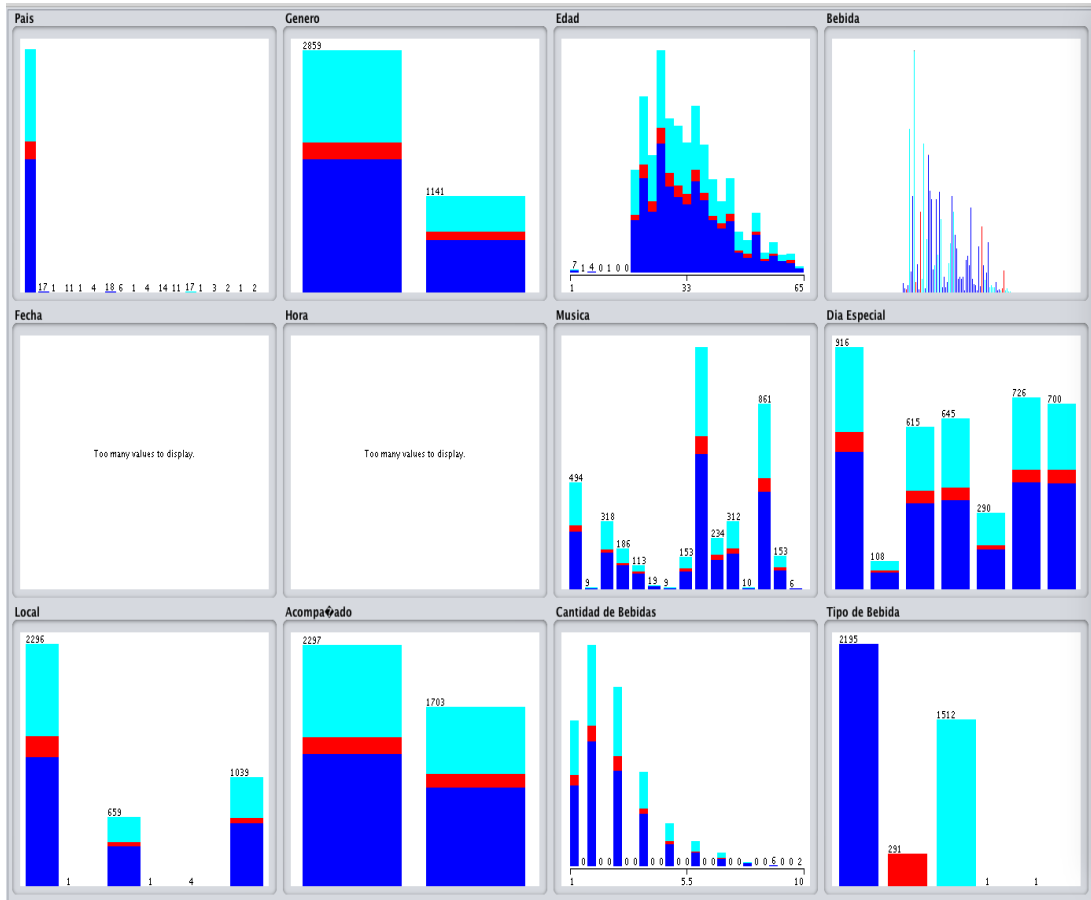


Ilustración 11: En el siguiente cuadro podemos observar cómo se clasificó nuestra data. Autoría Propia

Luego de haber creado los patrones de comportamiento el algoritmo KNN con la herramienta Weka, esta nos muestra los histogramas con los cuales lo ha clasificado por el tipo de bebida de acuerdo a cada etiqueta para hacer los cruces de necesarios para encontrar el patrón adecuado.

## 9. Los resultados de nuestra corrida de datos en la Ilustración N°13

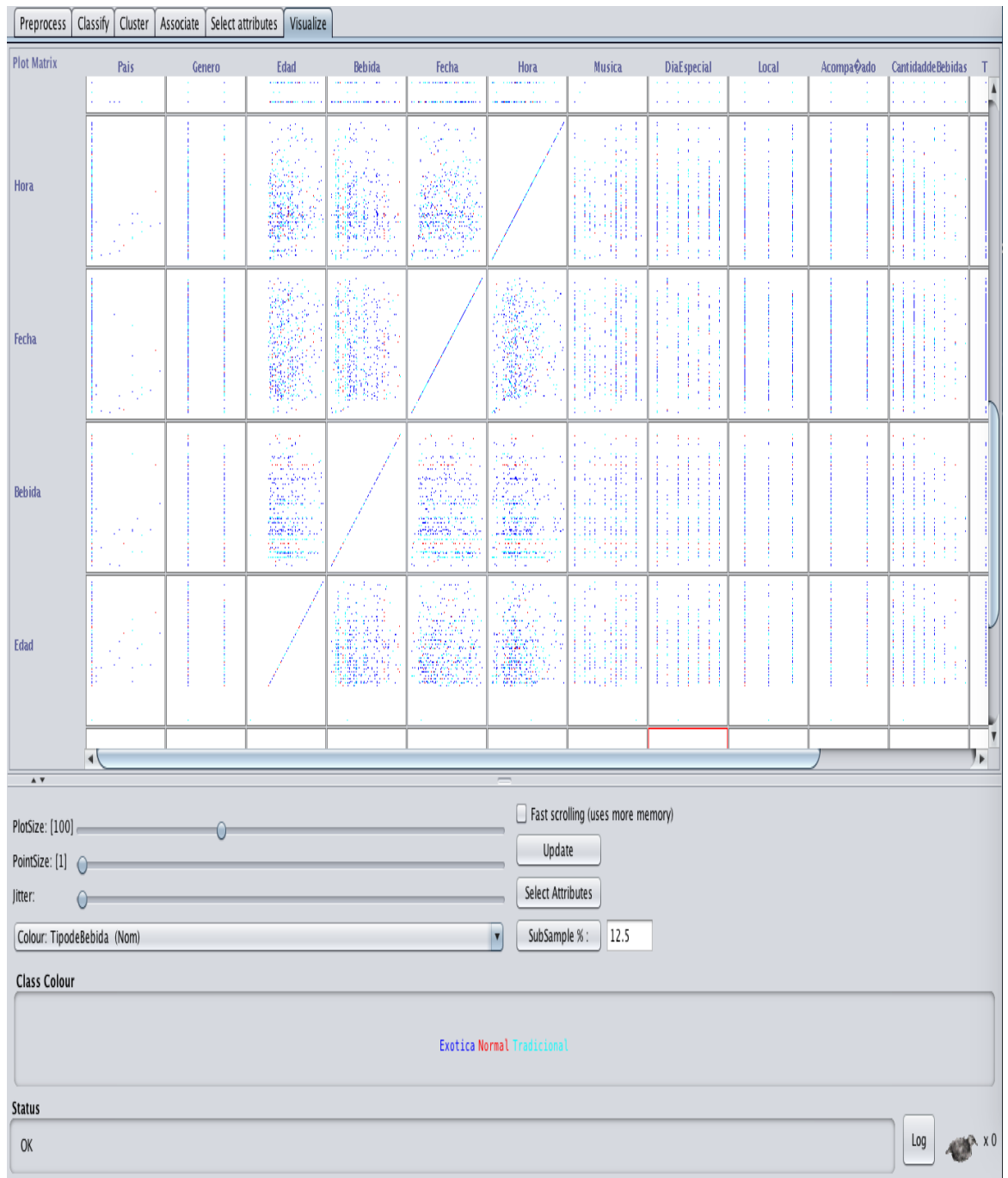
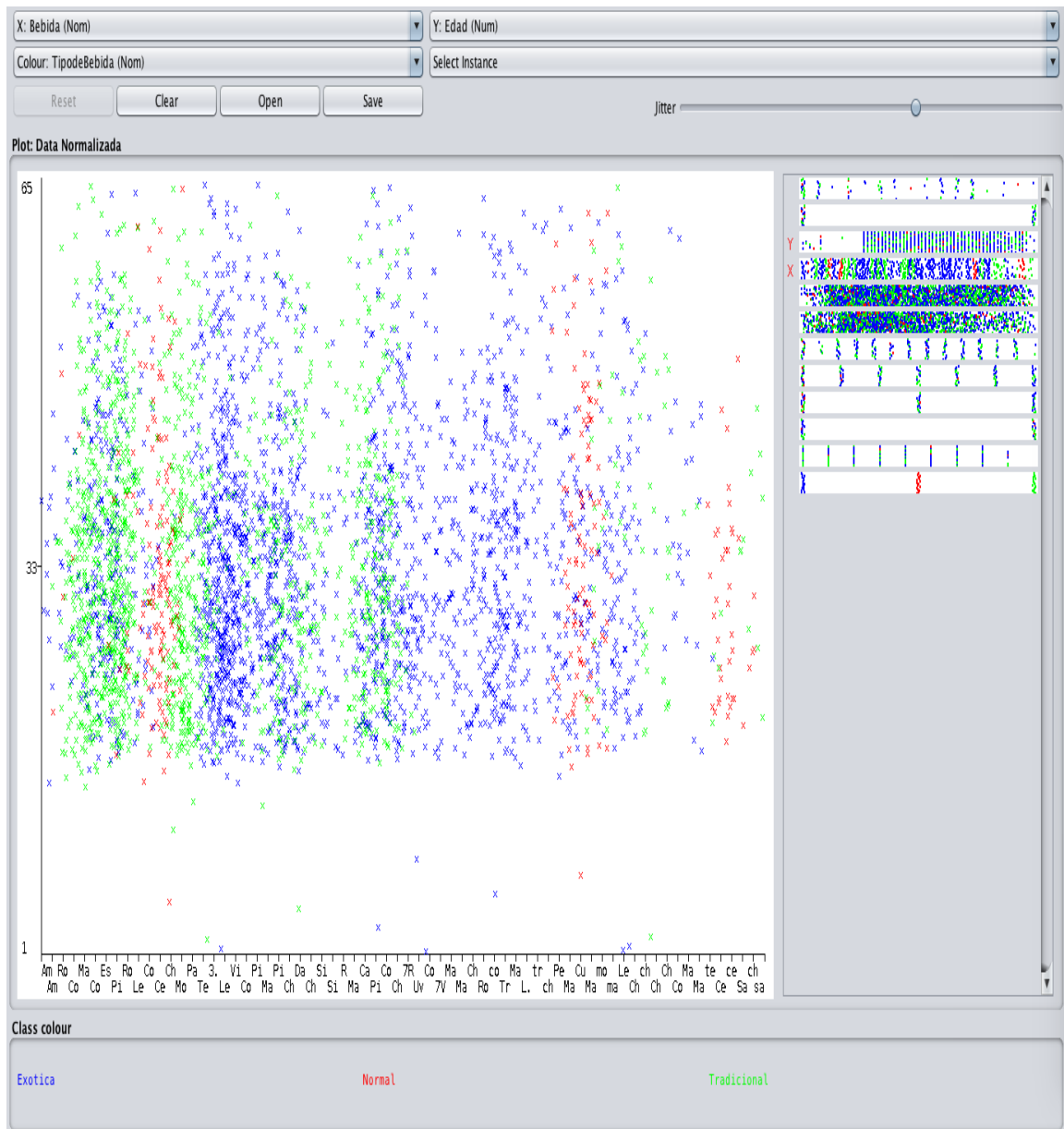


Ilustración 12 En el siguiente cuadro podemos observar más detalladamente las coincidencias de los datos. Autoría Propia.

En la siguiente ilustración podemos ver como es el cruce de nuestras etiquetas o clases de acuerdo al tipo de bebida y lo clasifica el tipo de bebida con las distintas etiquetas planteadas y como cada una de estas tiene un comportamiento distinto de acuerdo a la etiqueta dada.

## 10. Cuadro cruzado Bebida con la edad como se muestra en la ilustración

Nº 14

*Ilustración 13: Autoría Propia*

En el siguiente podemos observar como entre las edades de 25 a 35 años de edad son los que consumen más tragos exóticos y tradicionales a diferencia de los que las personas de mayor de 50 que su consumo es menor y esto puede ayudarnos a que la gerencia tome en consideración hacer promociones de acuerdo a las edades de sus clientes, y así brindar un mejor servicio por parte de su personal.

## 11. Cuadro cruzado Día con la edad como se muestra en la ilustración N°

15

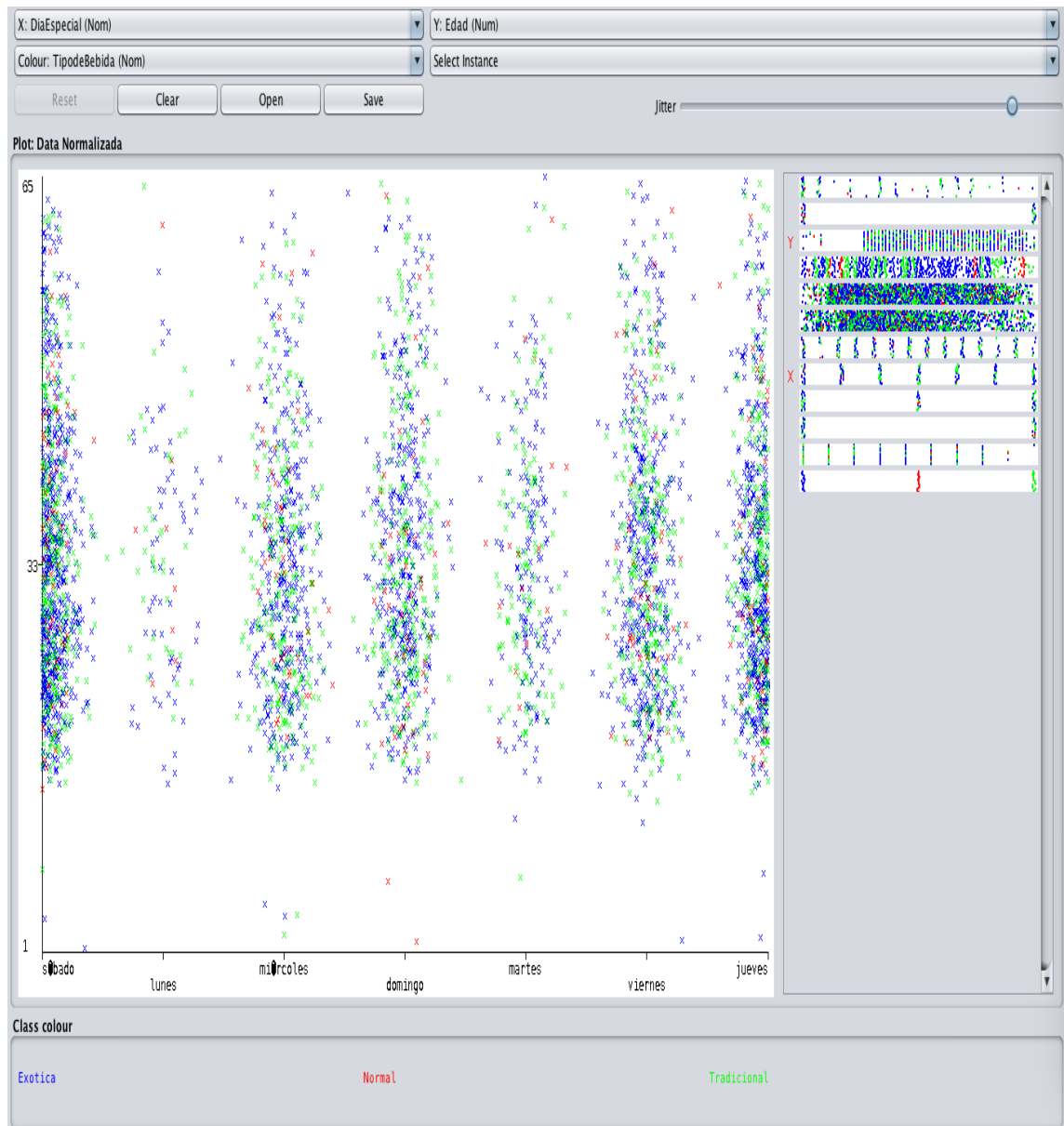


Ilustración 14: Se observa de acuerdo al día y la edad del cliente. Autoría Propia

Cruce del día y la edad de los clientes, los cuales podemos observar que hay días que se tiene más afluencia o concurrencia de clientes al local indicando que lunes y martes son los días con menos concurrencia y con esta información la gerencia tome la decisión de contratar personal de acuerdo a los días en los cuales tiene menos afluencia de personas y este pueda contar con la cantidad de personal necesario

12. Cuadro cruzado Bebida con el género como se muestra en la ilustración N° 16.

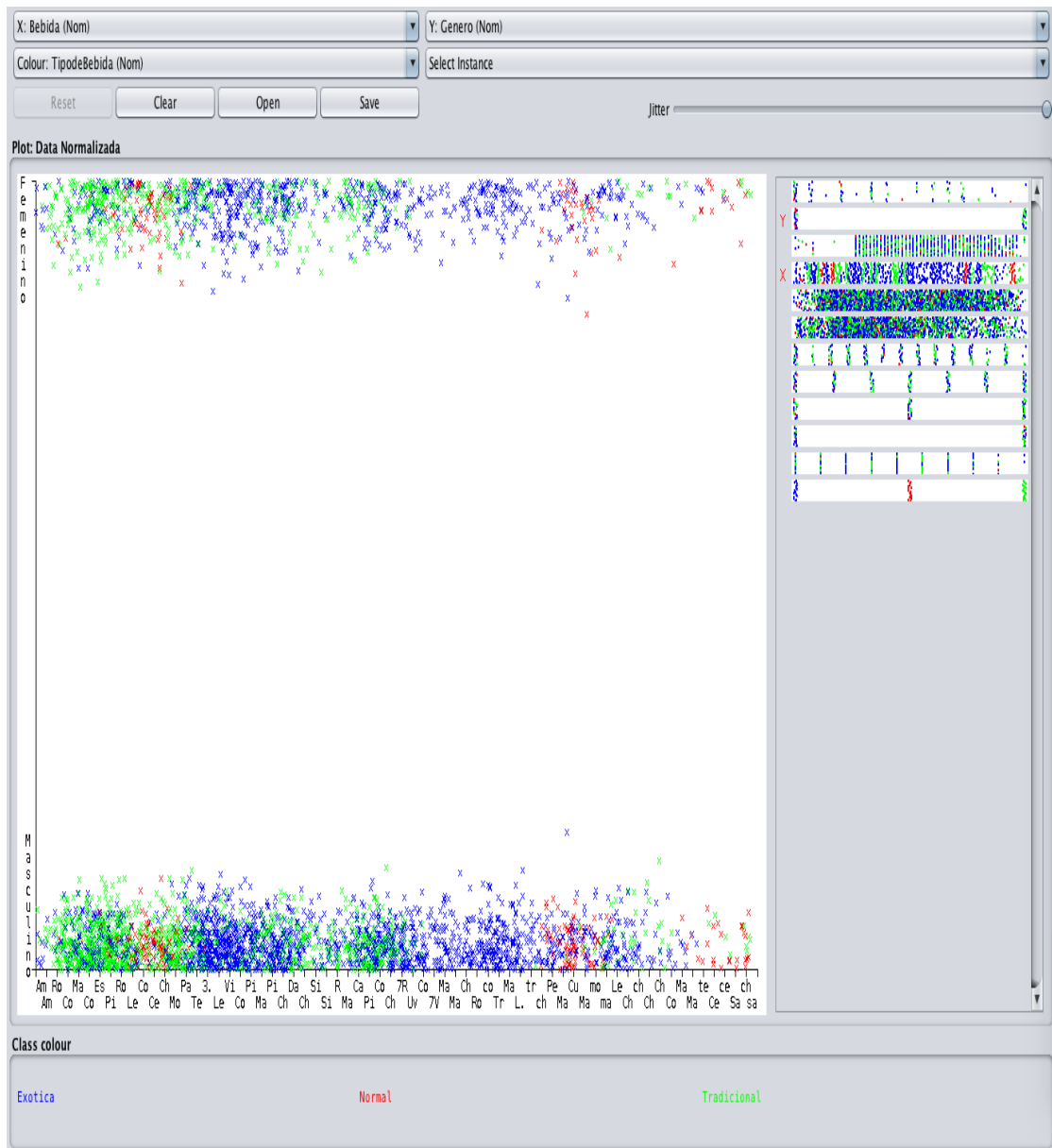


Ilustración 15: Se observa el cruce entre el género y la bebida Autoría propia.

Que el género masculino consume más bebidas exóticas que las tradicionales, el género femenino consume bebidas tradicionales y este es un aporte muy interesante y este afecta a la variable dependiente de la dimensión del proceso de gestión ya que la gerencia puede tomar decisiones en ofrecer a los diferentes géneros ofertas distintas y estas puedan ayudar a tener más clientela.

13. Cuadro cruzado Bebida y cantidad de bebida se muestra en la Ilustración Nº17

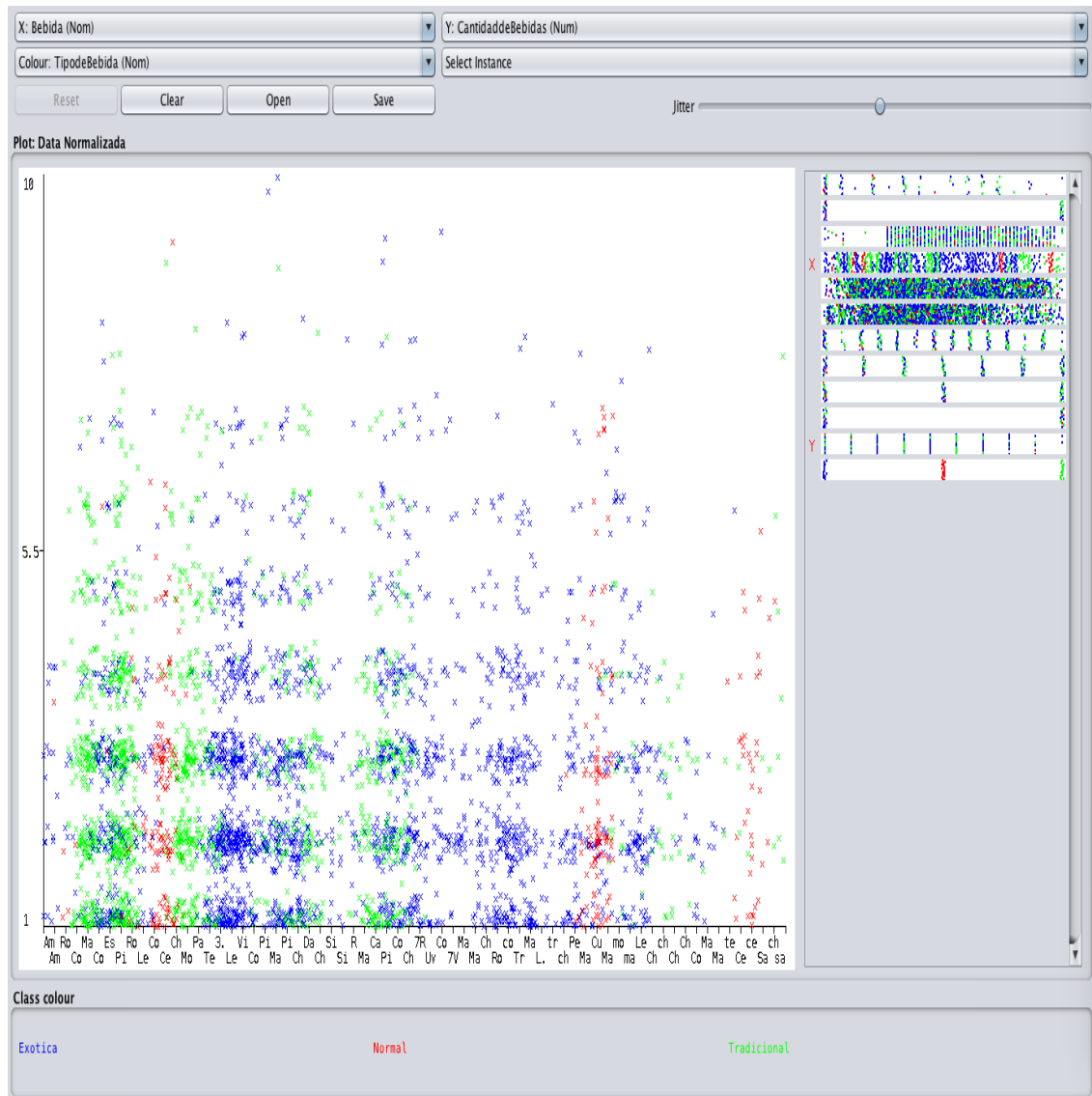


Ilustración 16. Se observa que la bebida y la edad. Autoría propia

En el cuadro podemos ver como se encontró el comportamiento entre la bebida y la cantidad que cada una es consumida, las bebidas exóticas son las más pedidas y con esta información la gerencia de acuerdo a la variable dependiente de la dimensión del proceso económico, operativo y gestión puede ofrecer más trago exótico que trago normal o hacer las promociones adecuadas para que se venda los tragos que no salen mucho y hacer la compra de los respectivos tragos.



14. Cuadro cruzado Música y bebida se muestra en la ilustración N° 18.

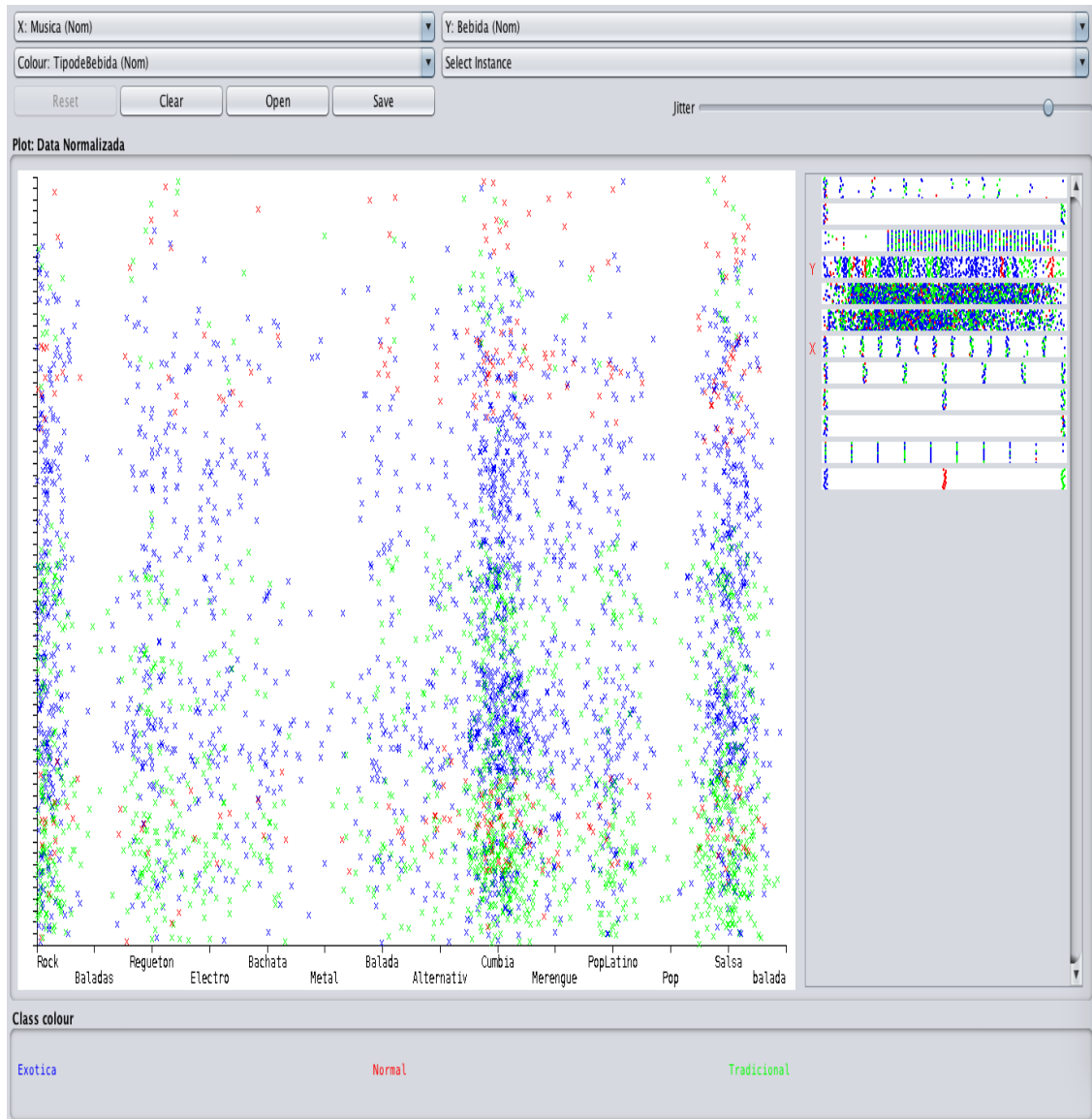


Ilustración 17: Se observa la música y bebida. Autoría propia

En el cuadro podemos observar como la mayoría de nuestros clientes les gusta más el género de rock, cumbia, y salsa son el tipo de música que prefieren nuestros clientes y con los que se vende más los tragos entonces esto afecta a la variable dependiente de la dimensión del proceso de gestión, donde la gerencia tendrá información relevante para poder decidir poner más este tipo de música cosa que traerá más ganancias a la empresa

15. Cuadro cruzado Acompañado y Cantidad de bebidas se muestra en la tabla N° 15.

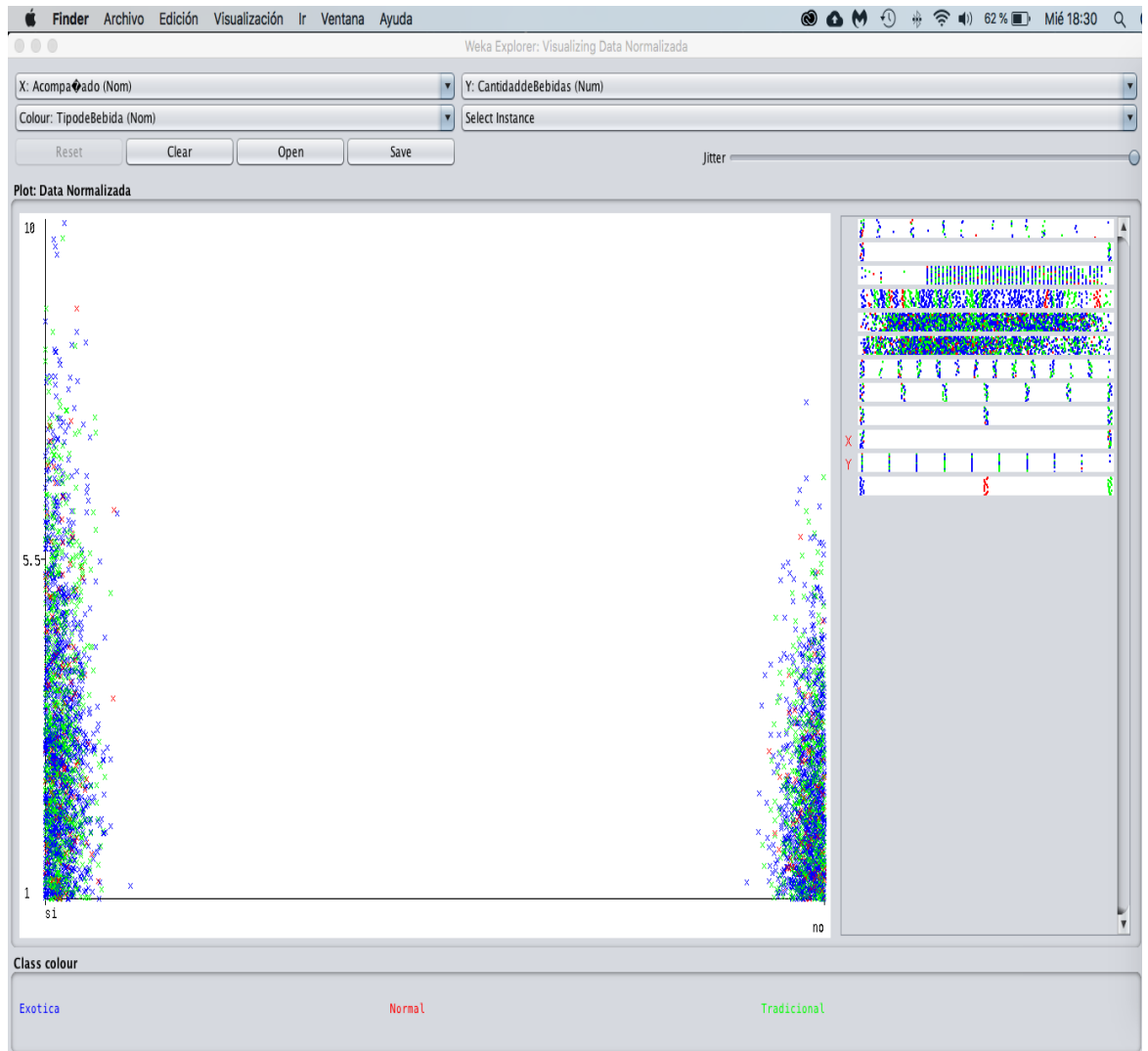
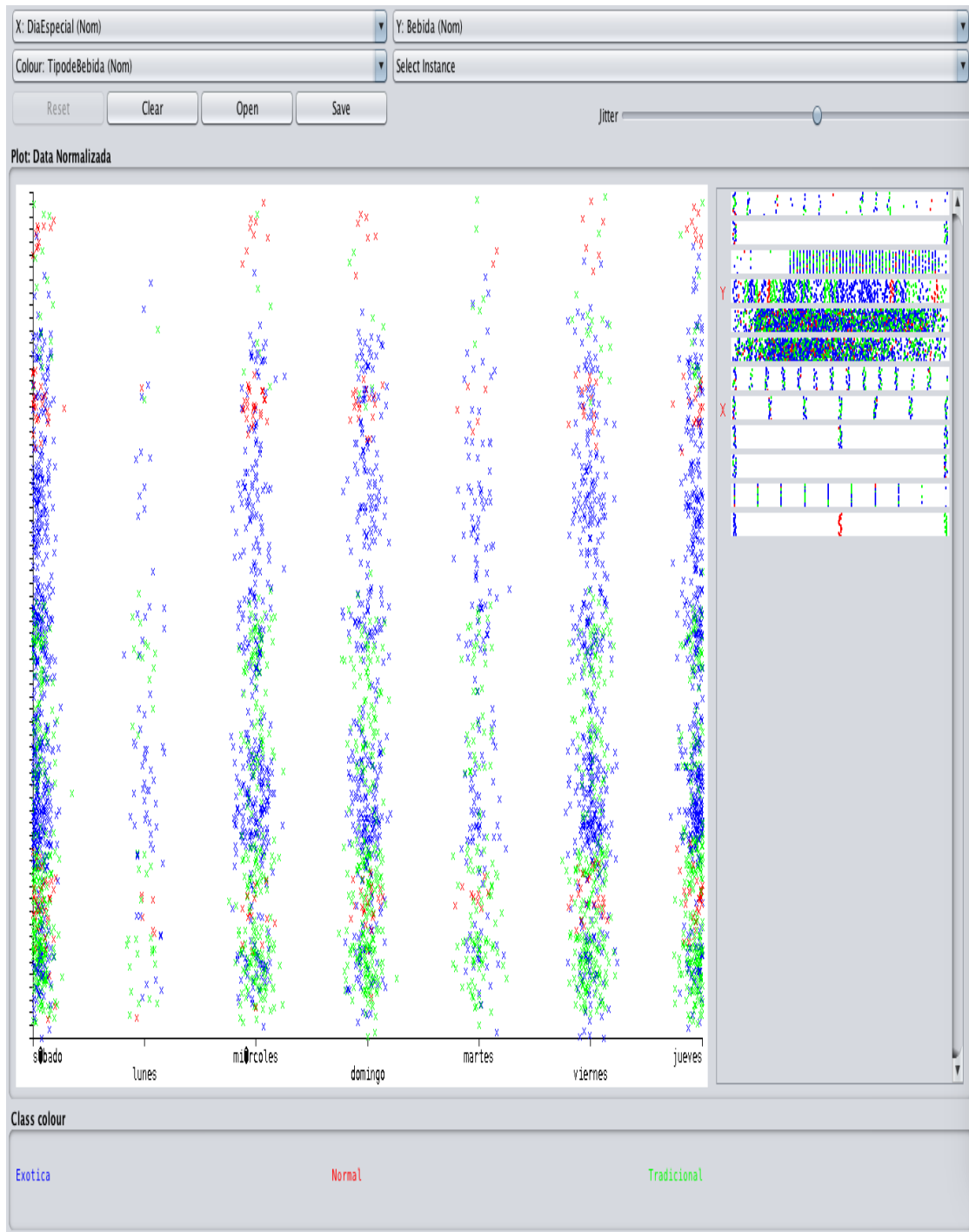


Ilustración 18: Se observa si los clientes vienen acompañados o no y cuanto consumen. Autoría propia

Que los clientes que vienen acompañados consumen más medidas tradicionales a diferencia de los que no que prefieren tomar más bebidas exóticas y esto ayuda que la empresa haga ofertas adecuadas o promociones

16. Cuadro cruzado Día y bebidas se muestra en la tabla N° 16.



*Ilustración 19: Se observa que los clientes comienzan tomando bebidas tradicionales y luego cambian a bebidas exóticas, esto nos ayuda que nuestro clientes les ofertemos los tragos exóticos con ofertas o incentivarlos para que ellos les guste más este tipo de bebida. Autoría propia*



# CÁPITULO IV:

---



## CAPITULO IV: RESULTADOS

### DESARROLLO DEL APLICATIVO FUNCIONAL PARA MOSTRAR EL RESULTADO DE LA INVESTIGACION.

En el siguiente capítulo tiene por objetivo realizar:

- ✓ Análisis del proceso del negocio después de aplicar Machine Learning mediante Patrones de Comportamiento en la Toma de Decisiones
- ✓ Propuesta de mejora del Proceso de Toma de Decisiones
- ✓ Mostrar los datos obtenidos en el Capítulo III, mediante un aplicativo funcional.

#### 4.1. ANÁLISIS DEL PROCESO DEL NEGOCIO DESPUÉS DE APLICAR MACHINE LEARNING MEDIANTE PATRONES DE COMPORTAMIENTO EN LA TOMA DE DECISIONES:

En la empresa Chuchuhuasi S.A. se identificó el proceso de toma de decisiones gerenciales, a través de datos obtenidos de los clientes que visitan al negocio centrándose en las características y gustos personales del cliente. La empresa no cuenta con ningún software específico para la realización de los procesos que realiza. Desde el punto de vista de las TIC (Tecnologías de la Información y las Comunicaciones), su sistema de gestión de información empresarial se basa en:

- ✓ Plantillas en Excel
- ✓ Correo electrónico de la empresa.

#### 4.2. PROPUESTA DE MEJORA DEL PROCESO DE TOMA DE DECISIONES

La propuesta que se ha considerado es la creación de un aplicativo funcional que muestre el resultado del Algoritmo de Vecino más Cercano con la utilización de Machine Learning, para la mejora del proceso de toma de decisiones en la gerencia administrativa de la empresa Chuchuhuasi S.A. Se busca una mejor interoperabilidad, personalización de la información, una mejora en la accesibilidad. El sistema que se hará soportará la creación de la base de datos, en el almacenamiento de archivos que contenga información de los gustos de los clientes, la información de la empresa debe estar debidamente entrelazada y debe sacarse del mismo



lugar. EL trabajo del procesamiento de los datos que se recolectaron debe estar de tal manera que permita el trabajo con esta de forma fácil y en tiempo real para que la gerencia vea el comportamiento de la información que tiene. La información tiene que estar debidamente trabajado para el personal que labora en la empresa. Es necesario establece las políticas necesarias de seguridad y que estas sean robustas. Otro de los objetivos del trabajo es la de analizar, modelar y documentar todos los procesos de ingreso y salida de la PYME basándose en la meotodologia Business Process Management (BPM) que tiene como base el objeto diseñar, modelar, organizar, documentar y optimizar los procesos de negocio de una organización, en busca de mejorar su rendimiento.

**4.3. MOSTRAR LOS DATOS OBTENIDOS EN EL CAPÍTULO III, MEDIANTE UN APLICATIVO FUNCIONAL**

**4.3.1. HERRAMIENTA UTILIZADA. BIZAGI:**

Con esta herramienta para el desarrollo del modelo del proceso del negocio propuesto se utilizar Bizagi modeler BPMS ya que esta nos ofrece diferentes ventajas en nuestro proyecto.

Los modelos generados por la herramienta nos ayudaran a hacer un diagramado de los procesos encontrados relevantes de la empresa en cuestión. Nos ayuda a especificar el trabajo con el objetivo de alimentar algunos de los modelos creados, de igual forma se pueden relacionar la mayoría de los modelos creados. Enfocando nuestro trabajo al Proceso de Toma de Decisiones en función del manejo de una información gerencial.

Algoritmo utilizados por Machine Learning	Herramienta de Uso del Algoritmo (Software).	Uso hasta el año 2019
Algoritmo KNN de búsqueda	Weka, EASA, WebScraping, Shazam, Veronoi	Activo
Algoritmos de Regresión	Weka,	Activo

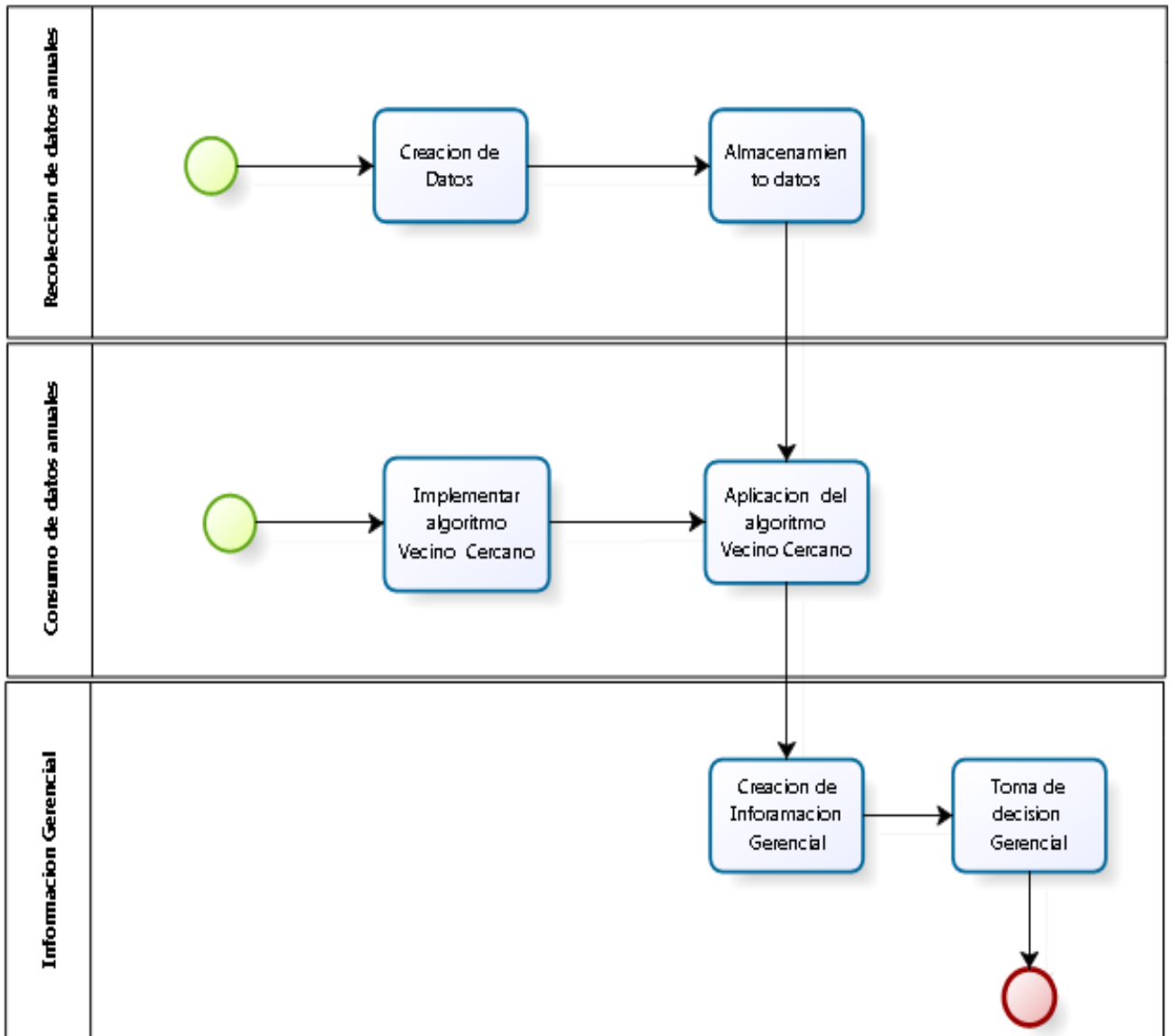


Algoritmos basados en Instancia	Weka	Activo
Algoritmos de Árbol de Decisión.	Weka, EASA	Activo
Algoritmos Bayesianos	Weka, Veronoi	Activo
Algoritmos de Clustering (agrupación)	Weka	Activo
Algoritmos de Redes Neuronales	Weka, EASA	Activo

Mencionar que estos algoritmos son usados por las diferentes herramientas de consumo de datos. Y en todas estas aplicaciones el algoritmo de vecino más cercano KNN, se puede observar es el más utilizado como algoritmo de confianza dentro de su software. Además podemos especificar que el trabajo con Weka es sencillo y a nuestro alcance para la generación de patrones de comportamiento a través de los datos ingresados pertenecientes a la empresa Chuchuhuasi S.A.

**4.3.2. PROCESO DE TOMA DE DECISIONES. BIZAGI:**

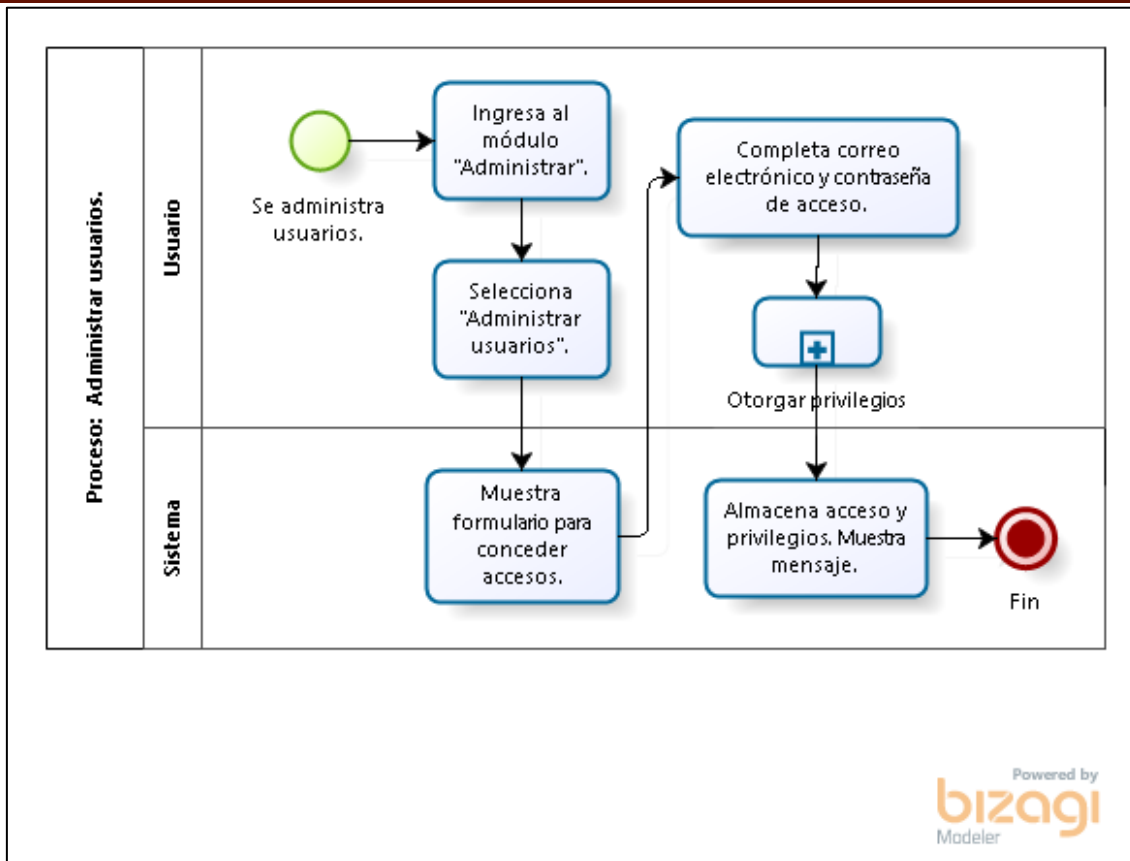
A continuación, se mostrará el mapa del proceso de Toma de Decisiones creado por la herramienta de BPM Bizagi Modeler.



#### 4.3.3. PROCESO DE VALIDACION DE USUARIOS BIZAGI:

A continuación, se muestra el mapa del proceso de validación de usuarios, creado por la herramienta de BPM Bizagi Modeler.





#### 4.3.4. TECNOLOGÍAS UTILIZADAS:

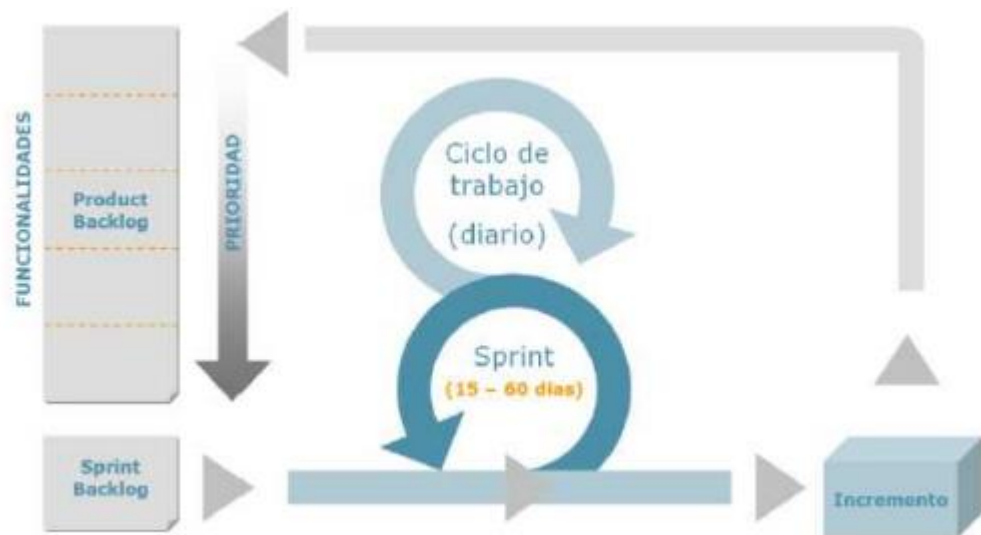
##### a. SCRUM.

La metodología que se utilizara en el desarrollo del proyecto será la técnica denominada SCRUM, es basada en el desarrollo agil de los proyectos de sistemas y esta compuesta por los principales elementos que mostraremos a continuación:

- ✓ **Product Backlog:** Es la lista de procesos funcionales que realiza el cliente (En nuestro caso Chuchuhuasi SA.), se priorizara según las necesidades que este determine. En este proceso se fomentaron varias reuniones con las diversas áreas de la empresa Chuchuhuasi SA., y el grupo que se encargara del desarrollo del mismo, este con la finalidad de conocer las funcionalidades urgentes y necesarias para el desarrollo del proyecto de software. Para la realización del proyecto se urge la necesidad de los objetivos que se quieran conseguir con el producto, conocida y comprendida por todo el

equipo de trabajo y conocida por todo el equipo, y elementos suficientes en el product backlog para llevar a cabo el primer sprint.

- ✓ Sprint Backlog: El sprint backlog es la lista que desarma as funcionalidades del product backlog en las actividades útiles para construir un incremento: una parte completa y operativa del producto. En el sprint backlog se asigna a cada tarea la persona que la va a llevar a cabo, y se indica el tiempo de trabajo que se estima, aún falta para terminarla.
- ✓ Incremento: Parte del sistema desarrollada en un sprint. El aumento es la parte de producto producida en un sprint, y tiene como características que está completamente terminada y operativa: en condiciones de ser entregada al cliente final.



Funcionamiento de Scrum.

ID	PRODUCT BACKLOG	REALIZADA
	Descripción	
1	Creación de una nueva Base de datos.	Si
2	Diseño y desarrollo de las tablas de la Base de datos.	Si
3	Control de acceso al sistema.	Si
4	Creación de las funciones necesarias para dotar al control de acceso de seguridad.	Si
5	Interfaz gráfica Administración	Si
6	Creación de documentos de manera automática	Si



7	Imprimir documentos y etiquetas	Si
8	Repositorio de archivos	Si
9	Validar Tareas	Si
10	Finalizar Tareas	Si

Tabla 5:Product Backlog

4.3.4. LA ARQUITECTURA DEL SISTEMA SE PUEDE DIFERENCIAR EN 3 CAPAS:

a. Nivel de Persistencia

En esta capa se encuentran los datos del sistema. Esta construido mediante el gestor de base de datos el cual realizara toda la gestión de almacenar los datos, el cual recepcione o almacene estas solicitudes de la información desde el nivel lógico. Además de ser un almacenamiento de archivos el cual utilizaran las diferentes partes del sistema el cual lo conforman.

Las empresas en la actualidad manejan un gran volumen de datos los cuales son necesarios almacenarlos para trabajar con ellos y poder convertir esos datos en información utilizando mecanismos profesionales. Sin estos mecanismos sería muy difícil manejar estos datos que maneja la empresa en su totalidad ya que al ser un gran volumen de información se requiere más tiempo de y por ende se tendría una perdida cuantiosa de dinero.

Un ejemplo claro de esto y se asemeja bastante a lo descrito anteriormente sería en la empresa Cuchuhuasi en la cual se está realizando el proyecto.

Lo primero y más importante para la creación de un sistema o aplicación la que nos ayude a la gestión de la base de datos, es sin duda, el modelo y diseño de Base de Datos. Otro de los problemas al crear una base de datos es la creación de las tablas y sus relaciones que tienen que tener para evitar que en el transcurso de la realización del sistema no tenga ningún problema y esté funcionando correctamente es por eso necesario un buen estudio de este. Siempre que se cree la base de datos es muy importante que esta este correctamente hecha la cual nos brinde una eficiencia adecuada y sea escalable en el transcurso del tiempo.



Hay varias características que se tiene que tomar en cuenta al momento de desarrollar el modelo de base de datos, esta tiene que poseer, un acceso veloz, tamaño adecuado para el almacenamiento, tipo de información, un acceso fácil y la forma de sacar la información de esta.

Bueno siguiendo con la estructura de la base de datos esta tiene que tener dos aspectos importantes: eficiente y flexible.

La eficiencia se dice que nos ayudara a que no se duplique los datos en la base de datos y que esta no tenga casillas vacías. Lo flexible seria que nuestra base de datos podemos agregar de manera fácil y sin la necesidad de reemplazar un archivo existente. Y siempre es necesario que la base de datos no tenga redundancia o repeticiones para que sea de fácil acceso rápido y los datos más eficientes.

Como ya se mencionó con anterioridad con referente a las herramientas tecnológicas que se utilizaran en nuestro proyecto, cabe volver a señalar, que la base de datos que sustenta esta aplicación, esta desarrolla mediante MySQL, y que su nombre es Chuchuhuasi.

A continuación se detallan los objetos en un modelo de clases que componen el modelo de negocio:

Diagrama de la Modelo Relacional de la empresa Chuchuhuasi (dbBitacora)

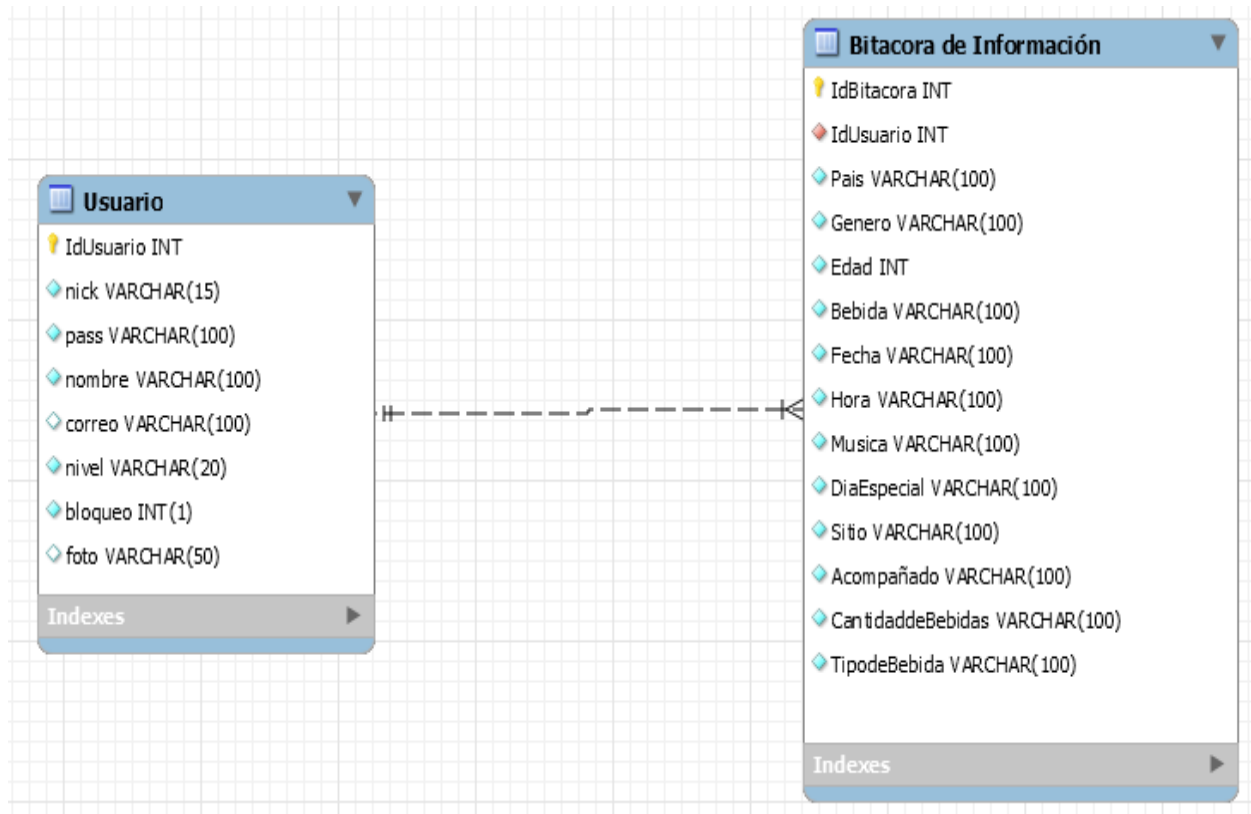


Ilustración 20: Tabla de la base de datos dbBitacora (autoría propia).

**Publicación de la Base de datos empresa Chuchuhuasi en PhpMyadmin:  
Relaciones o tablas: Modelo físico (2 tablas usuario, bitacora de información).**



### b. Nivel Lógico

Esta es una de las capas más importantes de esta arquitectura ya que en estas se describirá las operaciones que necesita la capa de presentación. También llamada como capa de negocio, es en esta en la que se realizaran las reglas de la empresa o negocio que deben de realizarse para su funcionamiento.

Esta capa de nivel lógico es la que se conectara con la capa de presentación la cual recibirá las peticiones de esta y se le muestre lo que necesite o alguna acción, y a su vez esta se conectara con la base de datos llamado también capa de datos o nivel de persistencia, para enviarle una petición y este gestor de datos nos devuelva una respuesta con los datos que se está solicitando para poderlo mostrar en la capa de presentación.

La lógica del sistema está en función de la muestra de los datos consumidos por fechas. Se hace una presentación de gráficos de los datos utilizados en el módulo o página Mostrar de la siguiente forma:

#### **Pagina Mostrar Cuadro de Confusión por preferencia (Musica, Bebidas)**

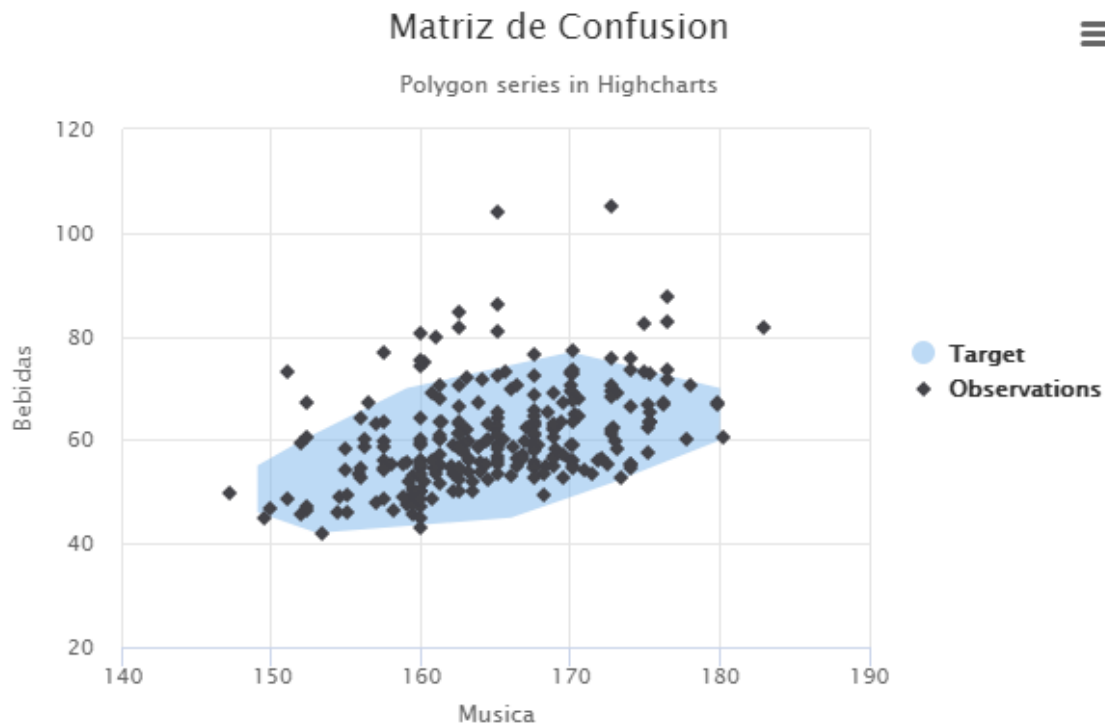


Ilustración 21: Página de Presentación de Matriz de confusión (Autoría Propia)

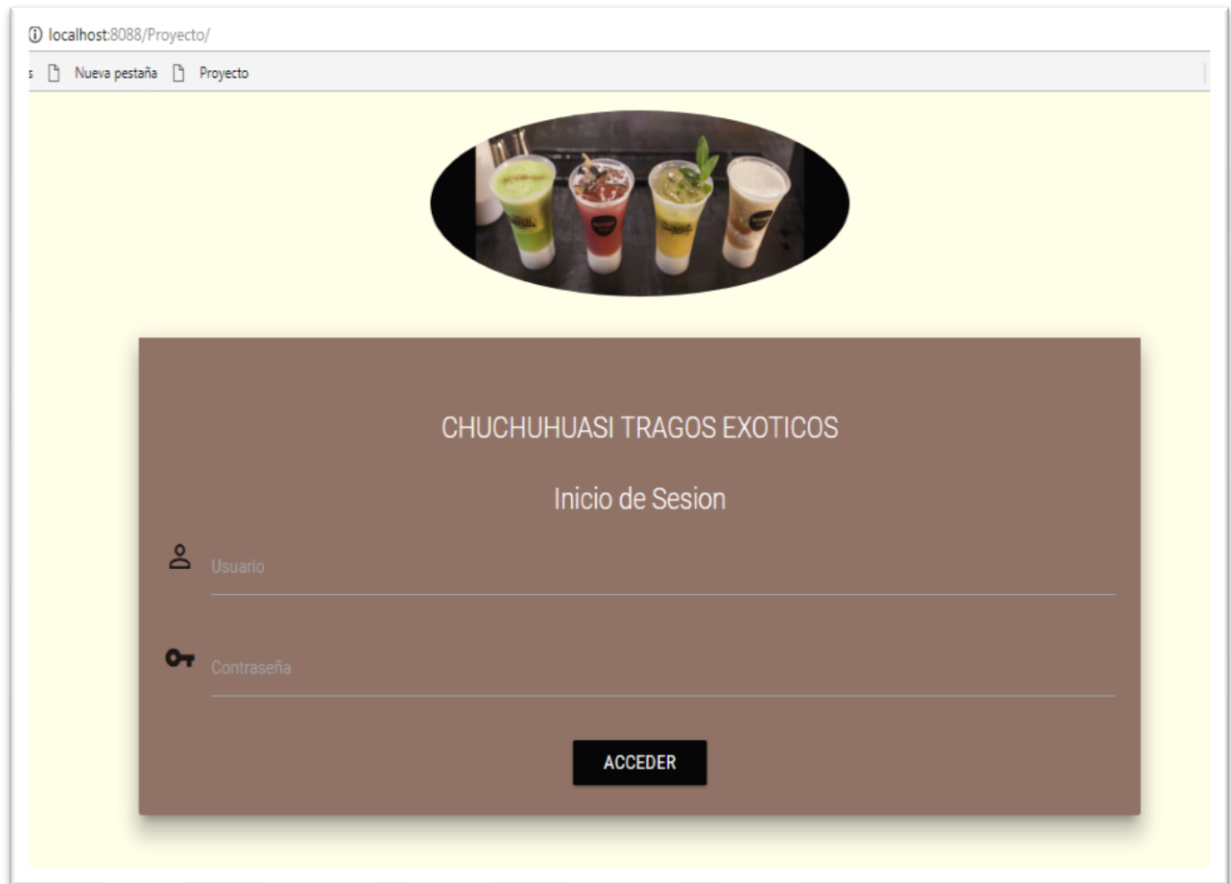
### c. Nivel de Presentación

Esta es la capa de presentación es el que va a tener una interacción con el usuario final, el usuario podrá visualizar el sistema, y el cual le comunicará la información o lo que necesita el usuario. Esta capa solo se podrá relacionar con la capa lógica. Está construida por el conjunto de páginas diseñadas en PHP que permiten el acceso a los usuarios.

Toda esta detalle del sistema lo mostraremos a través de páginas consecutivas denominadas (P1,P2,...Pn).

**P1: Iniciar Sesión.**

*Ilustración 22: Login Usuario (autoría propia)*



*Ilustración 23: Autoría propia*



**Acceso correcto:**

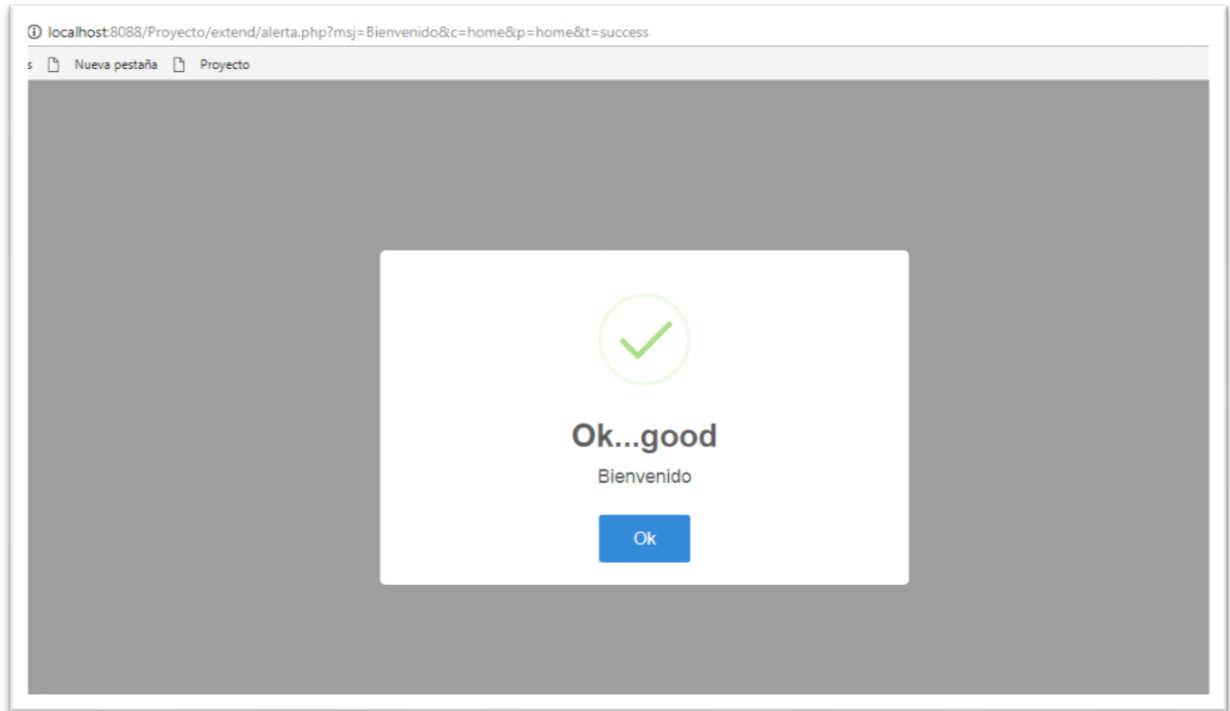


Ilustración 24: Autoría Propia

**Formato de clave:**



Ilustración 25: Autoría Propia

**Error de acceso: mensaje de error.**

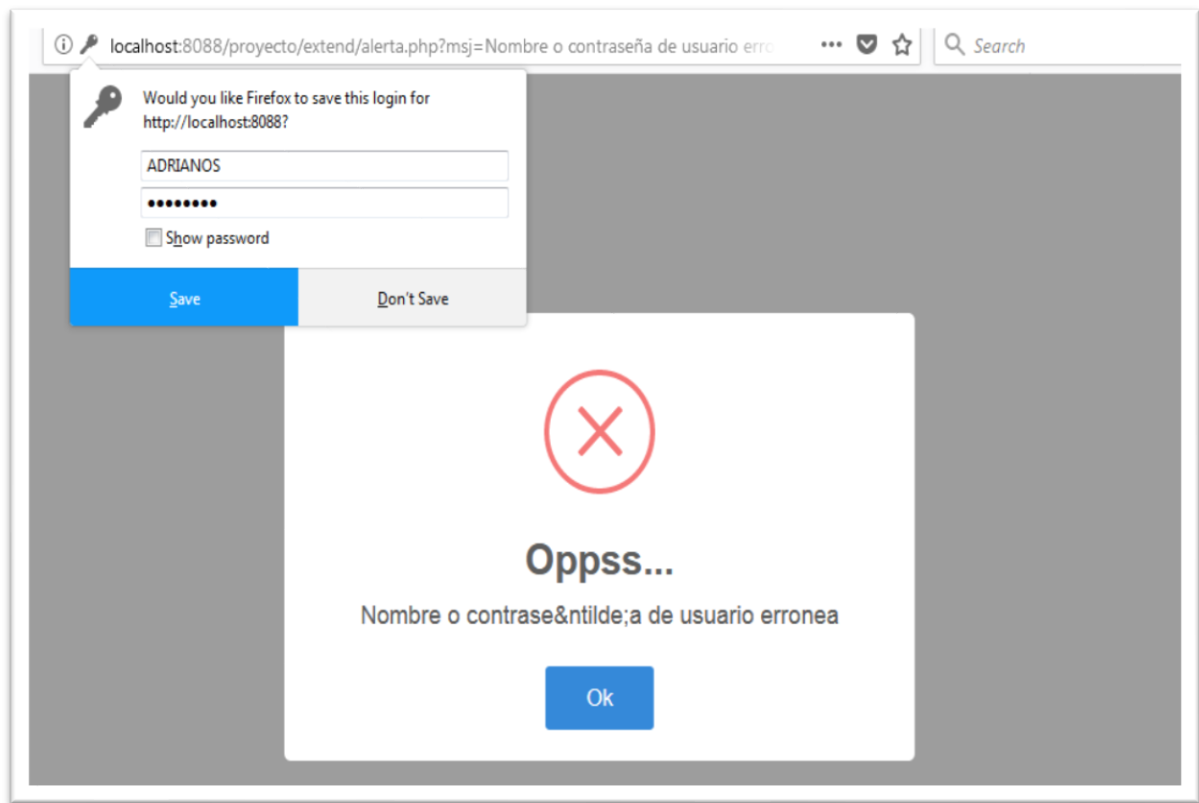


Ilustración 26: Autoría Propia

Iniciar Sesión
Entradas:
<ol style="list-style-type: none"><li>1. Ingresar nombre de usuario y contraseña al sistema de control. Tiene que contar con una clave de ingreso, y ser usuario administrador.</li><li>2. Iniciar sesión.</li></ol>
Resultados esperados:
<ul style="list-style-type: none"><li>• Acceso al sistema de bitácora de información con los datos correctos, que ingresa como administrador 1.</li><li>• Mensaje de logueo correcto</li><li>• Mensaje de error en el ingreso de credenciales del usuario</li></ul>
Condiciones:
<ul style="list-style-type: none"><li>• El usuario 1 concede los accesos y privilegios para otros usuarios.</li></ul>

Tabla 6: (Autoría Propia)

**P2: Alta de Usuarios al sistema:**



<b>Alta de Usuarios al Sistema:</b>
<b>Entradas:</b>
<ol style="list-style-type: none"><li>1. Ingresar los datos correspondientes a la información del administrador del sistema y crear una cuenta.</li><li>2. Guarda foto en la computadora y de no contar con fotografía, se le hace la captura de imagen a la hora del registro.</li><li>3. Modificar datos asignados al propietario o los del dispositivo posteriormente.</li></ol>
<b>Resultados esperados:</b>
<ul style="list-style-type: none"><li>• Creación y almacenamiento de nuevo usuario en la base de datos</li><li>• Cada usuario con su propia cuenta de acceso a la Página Web.</li></ul>
<b>Condiciones:</b>
<ul style="list-style-type: none"><li>• Los campos para completar los datos están validados y se requiere llenar todos los campos.</li><li>• El propietario debe ser el Administrador o funcionario de la empresa.</li></ul>

Tabla 7: Autoría Propia

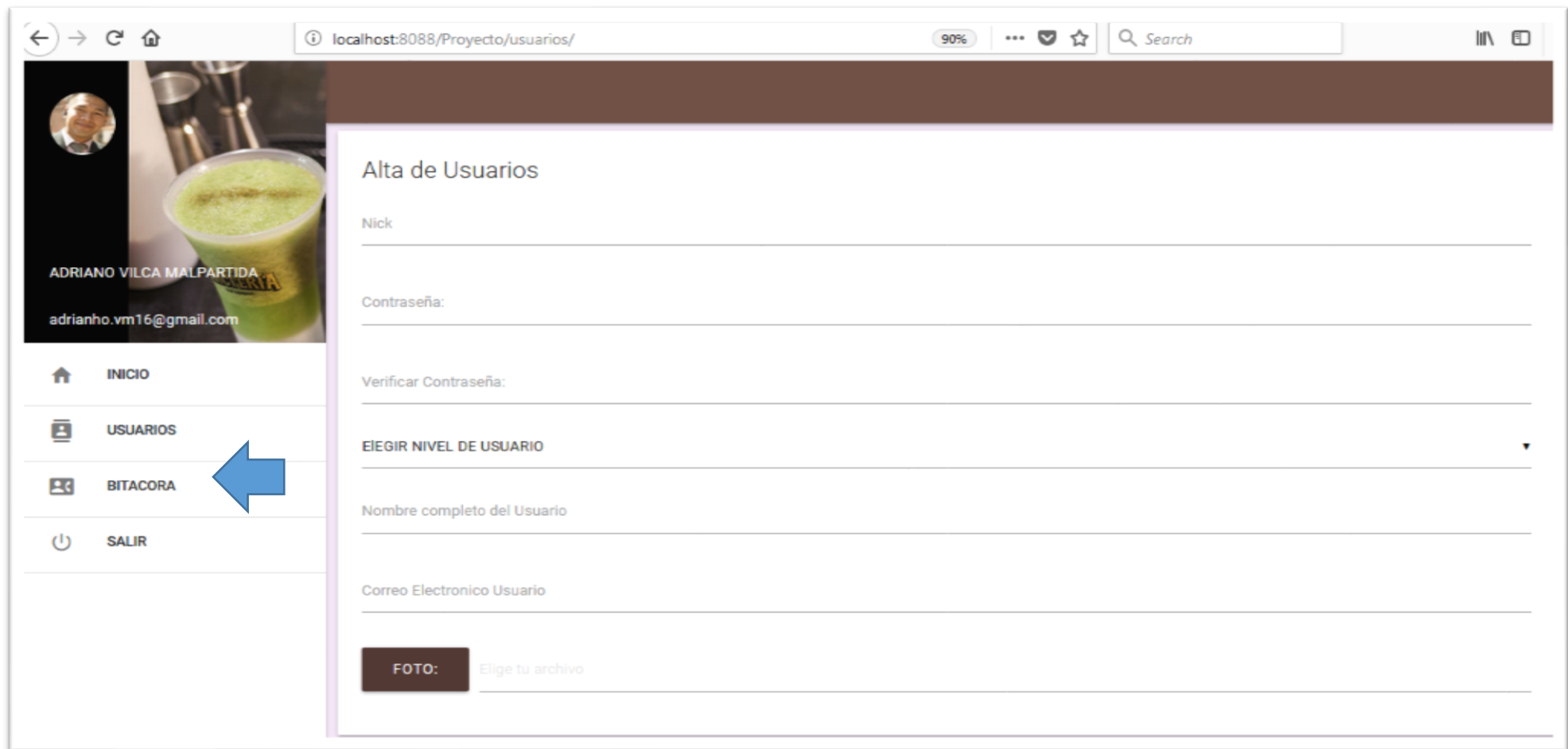


Ilustración 27: Autoría propia



**P3: Muestra de Resultados Bitácora:**

<b>Bitácora :</b>
Entradas:
<ul style="list-style-type: none"><li>• <b>Ingresar a la cuenta del administrador del sistema.</b></li></ul>
Resultados esperados:
<ul style="list-style-type: none"><li>• Muestra los resultado de los patrones predictivos de los datos consumidos obtenidos de la tabla bitácora</li><li>•</li></ul>
Condiciones:
<ul style="list-style-type: none"><li>• Logueo correcto</li><li>• El propietario debe ser el Administrador o funcionario de la empresa.</li></ul>

Tabla 8: Autoría propia

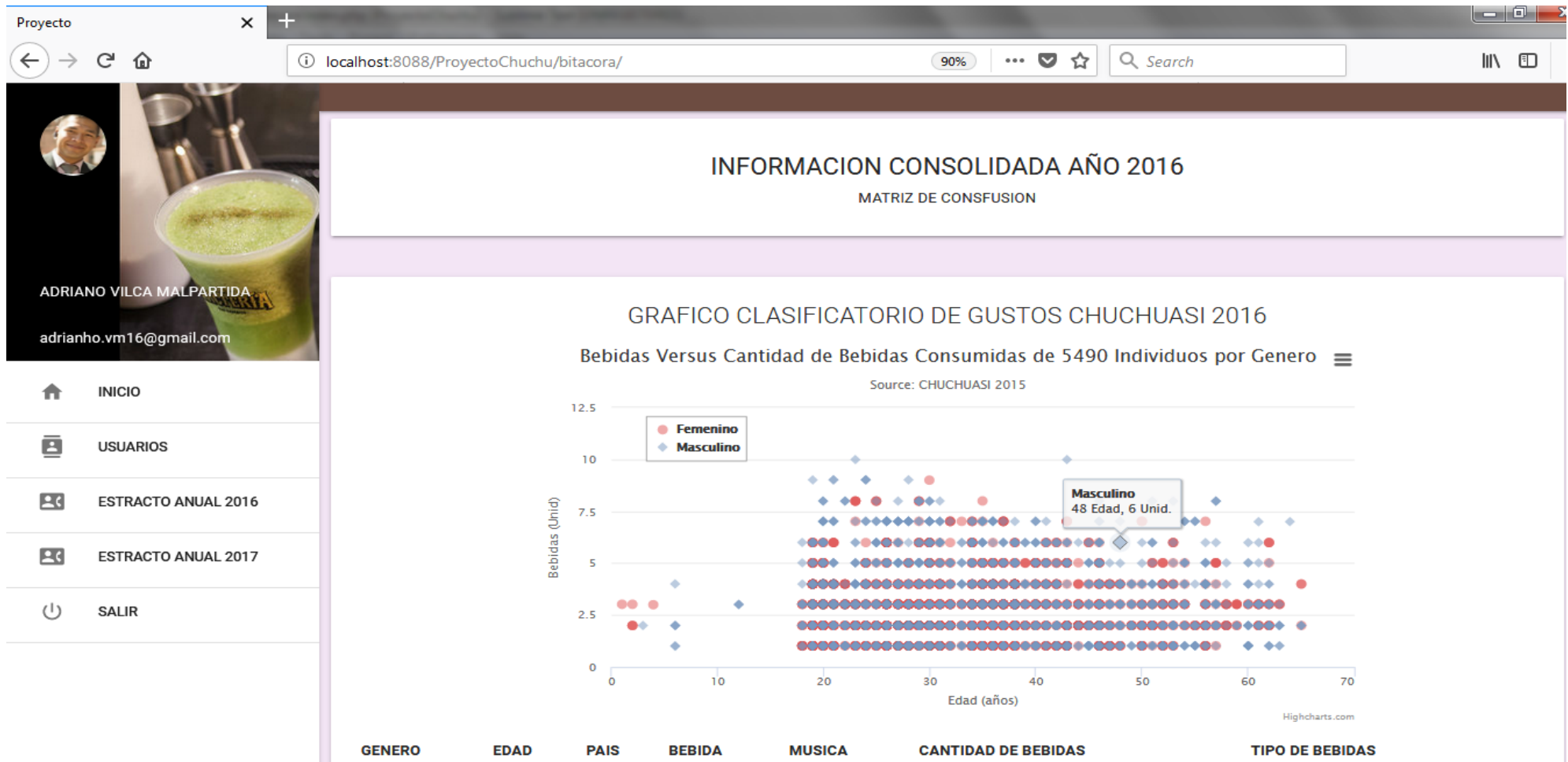
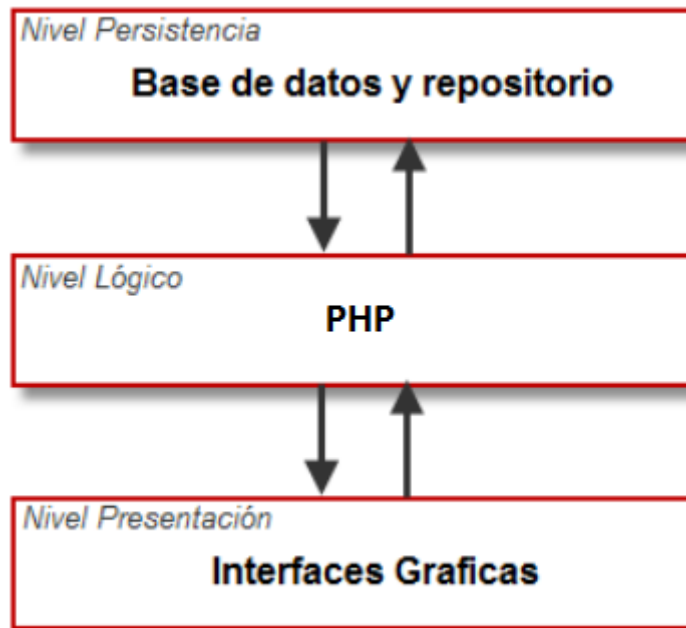


Ilustración 28: Autoría propia.

Ilustración 29: Capas de la arquitectura del sistema.





## CONCLUSIONES

Se ha cumplido con el objetivo principal planteado en nuestra tesis el cual es el Determinar la influencia de Machine Learning mediante patrones de comportamiento, generalizando patrones de comportamiento en los procesos **del negocio**, para recomendar en la **toma de decisiones gerenciales** de la empresa ChuchuHuasi S.A.

Podemos concluir que con la recolección y clasificación de datos de los clientes del Negocio. Al ejecutar varias corridas (iteraciones) de nuestra data tiene un 70% de aceptabilidad con la matriz de confusión la cual nos ayudara a poder clasificar nuestros datos. Aplicando el algoritmo (KVecino más cercano). Esto se verifico en la herramienta de aplicación Weka 3.8. Por ejemplo: El tipo de bebida que pide el cliente.

Concluimos que al implementar y aplicación del algoritmo KNN o vecino más cercano ha generado una herramienta que permite recomendar información sobre el gusto del cliente que visita a la empresa Chuchuhuasi. Específicamente en la variable dependiente de la investigación (Procesos del Negocio). En los cuales se aplicará sobre las dimensiones de (gestión, función, económicos, y operativos).

Concluimos que después de la aplicación de este modelo matemático (KNN o KVecino más Cercanos) se ha generado nueva información que recomendara sobre el consumo de los clientes del negocio (Periodos anuales). Esta información es mostrada a la gerencia a través de un aplicativo en la Web llamado (Chuchuhuasi S.A).





## RECOMENDACIONES

Se recomienda el desarrollo de un modelo matemático dinámico (algoritmo no supervisado), que se actualice de acuerdo al paso del tiempo. Ya que el modelo matemático provisto en este algoritmo y utilizado en esta investigación promueve el manejo de datos estáticos los cuales dificultan el manejo cuando existen grandes volúmenes de información.

El uso del algoritmo KNN nos permitió desarrollar nuestra investigación, pero no abarco la investigación de otros algoritmos similares los cuales también permiten trabajar este tipo modelo predictivo. Ya que nuestro modelo es un modelo supervisado, y puede servir de base para nuevas investigaciones futuras en el ámbito mecanismos computacionales.

Al pertenecer nuestro tema de investigación a un ámbito novedoso que no se abarca con amplitud en el proceso de pregrado en nuestra universidad. Sugerimos la ampliación de los temas de estudio en los cursos de estadística, algoritmos computaciones, investigación operativa, los cuales están enmarcados en la curricula de la Escuela Profesional.



## GLOSARIO

**Acceso:** Operación de identificación de un dato, un registro o una entidad que pertenece a un fichero. Proceso necesario para acceder a una instrucción almacenada en memoria para ejecutarla (Tinaquero, 2018).

**Bienes:** Viene de la palabra Bien; esto quiere decir que es todo lo tangible o intangible, que la persona pueda utilizar de manera útil y este satisfaga sus necesidades de manera directa o indirecta, que también puede ser de manera grupal o individual la cual tenga q ayudar a la mejora de los individuos.

**Bit:** Es la unidad más pequeña de dato que se utiliza en informática, en un dispositivo digital. Con este solo podemos representar dos valores uno o cero, verdadero o falso, etc.

**Byte:** Es la unidad que contiene cuatro dígitos o bits, se dice que 1 byte contiene 8 bits. El icono de un byte es una B, para poder diferenciar del bit el cual tiene un icono b. se utiliza en la informática la cual también es denomina octeto el cual se diferencia del bit.

**Colas:** Las colas son bastante utilizadas en la informática, en los cuales estos son tomados y almacenados de manera que la base de datos la guarde y almacene para su posterior atención de acuerdo al turno generalmente este está en la POO (programación orientada a objetos).

**Embebido:** Es un sistema de cómputo el cual está hecho para realizar una o muchas funciones dedicadas a este, es decir es un sistema de información o de computo que está en ejecución a tiempo real. A la diferencia de los computadores comunes que estos están hechos y adecuados para cubrir una variedad de necesidades del usuario, a diferencia de los sistemas embebidos estos son exclusivamente para cubrir necesidades específicas.

**Formulario:** Son la parte grafica que se le mostrara al usuario y estos son online o de escritorio con validación, estos a su vez son importantes para la realización de registros, reportes, modificaciones, etc. Lo formularios que están hechos en



línea son nombrados como formularios Web y en su mayoría están hechos con código HTML.

**IDE:** Es un entorno de desarrollo integrado o entorno de desarrollo interactivo, en inglés Integrated Development Environment (IDE), es un herramienta informática que brinda servicios integrales para ayudar al programador al desarrollo de software.

**Inventario:** Documento que contiene la lista de las cosas valorables que pertenecen a una persona, entidad u organismo.

**Iteración:** En programación, cuando el bloque de instrucciones de un bucle se ejecuta, se dice que se ha producido una iteración.

**JQuery:** Es una librería que trabaja en multiplataforma de JavaScript, que nos ayuda a facilitar la interacción de los documentos HTML, manejar propiedades, animaciones y lo principal darle interacción a una página web utilizando AJAX.

**Módulo:** Es un sistema que concentra un grupo de subprogramas y estructura de datos. Estos módulos su ventaja es que pueden ser compilados uno indistinto del otro, y esto hace que sea reutilizables y permita a los desarrolladores trabajen en diferentes módulos de manera simultánea, produciendo un buen desempeño en el desarrollo.

**MVC:** Es una arquitectura de software que es implementada para el desarrollo de sistemas modelo vista controlador. Con esto se quiere llegar a la necesidad de crear un sistema que sea robusto y tenga un ciclo de vida correcto, nos ayude el mantenimiento de este, se reutilicé el código y la separación de estos.

**Open source:** Este es un software de distribución libre de código abierto el cual está orientado a compartir el código de manera ética.

**Product owner:** El rol del Product Owner, este es parte del cliente o que trabaje dentro de la misma empresa. No es bueno que el Product Owner sea parte también del equipo de desarrollo, o el Scrum Master mismo, sus intereses se pueden ver enfrentados, pero esto puede variar según el caso. Por defecto, probablemente venga de parte del cliente.



**Product backlog:** El Product Backlog es el conjunto de requisitos funcionales y no funcionales que debe cumplir el producto una vez entregado. No es necesario que el sistema esté terminado al momento de su presentación, basta con definir aquellos requisitos que se conozcan en su momento y alentar a su crecimiento continuo o su modificación.

**Proceso:** Es la unidad de mínima de trabajo al realizar una tarea dicha, este realiza un grupo de instrucciones que ha sido mencionado, en el cual al final se obtiene un resultado deseado.

**Registro:** En base de datos los elementos se les denomina ficheros, estos elementos se dividen y contienen los datos descritos en estos.

**Reporte:** Es la información organizada que viene desde la base de datos. Estos datos son organizados y mostrados por un reporte al usuario de manera atractiva que pueda comprender el usuario.

**SCRUM:** Es una metodología rápida con la que se puede agilizar y disminuir los riesgos en el proceso del proyecto, entre estos nos brinda muchas ventajas en la realización del proyecto, mejorando la comunicación entre los usuarios y los desarrolladores.

**Script:** Un script, archivo de órdenes, archivo de procesamiento por lotes o, cada vez más aceptado en círculos profesionales y académicos, guion es un programa usualmente simple, que por lo regular se almacena en un archivo de texto plano. Los guiones son casi siempre interpretados, pero no todo programa interpretado es considerado un guion. El uso habitual de los guiones es realizar diversas tareas como combinar componentes, interactuar con el sistema operativo o con el usuario. Por este uso es frecuente que los intérpretes de órdenes sean a la vez intérpretes de este tipo de programas.

**SDK:** (Software Development Kit - Kit de desarrollo de software o devkit). Un SDK es el grupo de herramientas que permite a los desarrolladores realizar sus sistemas utilizando paquetes, los cuales le ayudaran a mejorar su aplicativo.



**Servidor:** Es un sistema en el cual se realiza tareas o servicios para el usuario. En otros términos, es una computadora que nos brinda datos o información para que sea consumido por los usuarios o otros ordenadores.

**SGBD:** (Sistema de gestión de base de datos) o en inglés Database management system (DBMS), es un conjunto de herramientas las cuales nos ayudan a modificar una base de datos.

**Sprint:** Es la realización de un grupo de tareas para la realización de un proyecto. Estos sprints generalmente duran entre 15 a 30 días. Cada uno de estos sprint tiene que tener una parte del proyecto para ver su cambio y poder llegar a su objetivo.

**WebService:** Un servicio web (en inglés, web service o web services) es un grupo de servicios los cuales nos pueden ayudar utilizar sus bondades entre diferentes sistemas o aplicativos. Estos servicios pueden ser utilizados en diferentes sistemas desarrollados y sistemas operativos es por eso que este maneja estándares adecuados para poder ser utilizados.



## REFERENCIAS

- Achour, M. (2013). Manual de PHP. E.E.U.U.
- Affiliates, O. C. (2016). *Workbench Reference Manual*. Obtenido de <http://dev.mysql.com/>
- Barrios, O. (2010). Los Negocios Reales, Mejorando procesos. *Los Negocios Reales, Mejorando procesos*, 26.
- Bizagi Limited. (2016). *Bizagi BPM Suite*. Obtenido de Solución para Gestión de procesos de negocio (BPM): <http://www.bizagi.com>
- Cabanillas, A. R. (2010). *Diseño de un prototipo de un Sistema de Trazabilidad de Ganado usando RFID*. Lima - Perú.
- Casillas Santillan, L., Ginesta, M., & Perez Mora, O. (2009). *Bases de Datos en Mysql*. Barcelona: [www.uoc.edu](http://www.uoc.edu).
- Castilla, D. (2014). *A Hybrid Approach Using RUP and Scrum as a Software Development Strategy*. Florida.
- Córdova Forero, J. A. (2014). *Implementación de un Sistema de Matrículas y Pagos para el Centro de Informática de la Universidad César Vallejo*. Lima.
- Delgado, W. M. (2015). *Asp.Net MVC 6 - Una Guía Introductoria*. Obtenido de <https://leanpub.com/aspnetmvc6-unaguaintroductoria>
- Design, M. (2018). <https://materializecss.com/about.html>. Obtenido de <https://materializecss.com/about.html>.
- Diagramas, A. y. (2010). Obtenido de [http://ing.unne.edu.ar/pub/informatica/Alg\\_diag.pdf](http://ing.unne.edu.ar/pub/informatica/Alg_diag.pdf)
- Gauchat, J. D. (2012). *El Gran Libro de libro de HTML5, CSS3 y Javascript*. Barcelona: MARCOMBO S.A.
- Gianni, R. d. (2011). *El libro del BPM 2011*.
- Gonzalez, A. (01 de Julio de 2014). *Clever Data*. Recuperado el 19 de Enero de 2017, de <http://cleverdata.io/que-es-machine-learning-big-data/>
- Group, The PHP. (10 de 05 de 2018). *PHP*. Obtenido de <http://php.net/manual/es/preface.php>
- Harrington, P. (2012). *Machine Learning in action* (Linda Recktenwald ed.). (J. Bleiel, Ed.) Shelter Island, United States of America: Manning Publications Co.
- Harrington, P. (2012). *Machine Learning in Action*. Shelter Island, NY 11964: Manning Publications.
- Hunt, D., & Puglia, A. (2007). *RFID- A Guide To Radio Frecuency Identification*. New Jersey: John Wiley & Sons Inc.



- Jacobson, I., Booch, G., & Rumbaugh, J. (1999). *El proceso unificado de desarrollo de software*. Madrid: Pearson Educación S.A.
- Java - Oracle. (2016). *JAVA*. Obtenido de <https://www.java.com/es/>
- Kniberg, H. (2007). *Scrum y XP desde las trincheras. Como hacemos Scrum*. Nueva York: C4Media, editor de InfoQ.com. .
- Lior Rokach, O. M. (s.f.). *Data Mining With Decision Trees* (2ª Edición ed., Vol. 2). (A. Yun, Ed.) USA: British Library Cataloguing-in-Publication data.
- M., C. M. (2006). *RFID: El código de barras inteligente para Bibliotecas*. Santiago - Chile.
- Martin, A. I. (07 de Julio de 2015). *Machine Learning in Scientometrics*. Recuperado el 20 de Enero de 2017, de Biblioteca Universitaria Politecnica: [http://oa.upm.es/36488/1/ALFONSO\\_IBANEZ\\_MARTIN.pdf](http://oa.upm.es/36488/1/ALFONSO_IBANEZ_MARTIN.pdf)
- Meneses, C. P. (2012). *Diseño de un Sistema de Control de Activos para el Almacén de Electrónica de la Pontificia Universidad Católica del Perú utilizando RFID*. Lima - Perú.
- Microsoft. (2012). *Microsoft Developer Network*. Obtenido de Documentación: [https://msdn.microsoft.com/es-es/library/fx6bk1f4\(v=vs.100\).aspx](https://msdn.microsoft.com/es-es/library/fx6bk1f4(v=vs.100).aspx)
- Microsoft. (2016). *Developer Network*. Obtenido de <https://msdn.microsoft.com/es-es/library/bb545450.aspx>
- Ministerio de Industria, T. y. (2008). *La tecnología RFID: Usos y Oportunidades*. Madrid: red.es.
- Murphey, R. (Agosto de 2013). *Fundamentos de jQuery*. Obtenido de <http://librojquery.com/>
- Ofimática Profesional - Corinne Hervo. (2013). *Microsoft Excel 2013*. Barcelona: Ediciones ENI.
- Oracle. (2016). *NetBeans IDE*. Obtenido de [https://netbeans.org/index\\_es.html](https://netbeans.org/index_es.html)
- Oracle. (2016). *Oracle Documentación*. Obtenido de <http://www.oracle.com/technetwork/es/database/enterprise-edition/documentation/index.html>
- Pal, C. (mayo de 2016). Obtenido de <https://www.cs.waikato.ac.nz/ml/weka/book.html>
- Pal, C. (mayo de 2016). <https://www.cs.waikato.ac.nz/ml/weka/book.html>. Obtenido de <https://www.cs.waikato.ac.nz/ml/weka/book.html>.
- PBWORK. (2008). *Cursos de Ingeniería de Software*. Obtenido de <https://ingsw.pbworks.com/f/Ciclo+de+Vida+del+Software.pdf>
- Percy E. De la Cruz Vélez de Villa, M. R. (2010). Radiofrecuencia de identificación (RFID): Microtecnología de gran impacto. *Revista de Investigación de Sistema e Informática*, 77-86.
- Pérez, J. E. (17 de Diciembre de 2010). *Introducción a AJAX*. Obtenido de [http://www.jesúsda.com/docs/ebooks/introduccion\\_ajax.pdf](http://www.jesúsda.com/docs/ebooks/introduccion_ajax.pdf)



- Pérez, J. E. (17 de Diciembre de 2010). *Introducción a XHTML*. Obtenido de [http://www.jesusda.com/docs/ebooks/introduccion\\_xhtml.pdf](http://www.jesusda.com/docs/ebooks/introduccion_xhtml.pdf)
- Ralph M. Stair, George W. Reynolds. (2001). *Principios de Sistema de Información. Enfoque administrativo*. Madrid: Ediciones Paraninfo.
- Rancel, M. R. (2013). *Aprende a programar*. Obtenido de [ww.aprenderaprogramar.com](http://ww.aprenderaprogramar.com)
- RFID JOURNAL. (2016). *RFID JOURNAL*. Obtenido de <http://www.rfidjournal.com/>
- Roberti, M. (16 de Enero de 2005). *RFID Journal*. Obtenido de RFID Journal: <http://www.rfidjournal.com/articles/view?1338>
- Roberti, M. (16 de Enero de 2005). *The History of RFID Technology*. Obtenido de The History of RFID Technology: <http://www.rfidjournal.com/articles/pdf?1338>
- Román Alexis Baéz Llancafil, J. L. (2013). *Tecnología RFID para el Reconocimiento y Asignación de Pallets y Rollos de Cartulina en Bodega*. Valdivia - Chile.
- Russell, S. J., Norving, P., & Gutierrez, R. B. (1996). *Inteligencia Artificial Un enfoque Moderno. (979), Primera*. Mexico.
- Salazar Caraballo, L. (2013). *Asuntos de la Ingeniería de software*. Obtenido de <http://www.gazafatonarioit.com>
- Sampieri, R. H. (2014). *Metodología de la Investigación*. Mexico: McGRAW-HILL / INTERAMERICANA EDITORES, S.A. DE C.V.
- Sanchez, J. A. (2008). *Sistema de control de acceso con RFID*. México D.F., México.
- SAS, DOINTECH. (2015). *DOINTECH SAS, Automatización, seguridad y control*. Obtenido de <http://www.dointech.com.co/control-de-acceso.html>
- Schwaber, K., & Sutherland, J. (Julio de 2013). *Scrum guides*. Obtenido de La guía de Scrum: <http://www.scrumguides.org/docs/scrumguide/v1/Scrum-Guide-ES.pdf>
- Shai Shalev Shwartz, S. B. (2014). *Understanding Machine Learning from theory to algorithms*. Cambridge: Library of Congress cataloging in publication data. Recuperado el Marzo de 2017
- Sigma Electrónica LTDA. (2016). *Sigma Electrónica LTDA*. Obtenido de <http://www.sigmaelectronica.net/sl500l-p-1971.html>
- SOFTENG. (2016). *SOFTENG*. Obtenido de Metodología Scrum para desarrollo de software - aplicaciones complejas: <https://www.softeng.es/es-es/empresa/metodologias-de-trabajo/metodologia-scrum.html>
- Sparx Systems. (2016). *Sparx Systems*. Obtenido de Enterprise Architect - Herramienta de diseño UML: <http://www.sparxsystems.com.ar/products/ea.html>
- Syed Ahson, M. (2008). *RFID Handbook: Applications, Technology, Security and Privacy*. Estados Unidos: CRC Press.





Telectronica. (2006). *Introduccion a la Identificacion por Radio Frecuencia*. Bueno Aires:  
Telectronica Codificacion S.A.

Tinaquero, D. M. (04 de junio de 2018). *openaccess*. Obtenido de  
<http://openaccess.uoc.edu/webapps/o2/bitstream/10609/81188/8/dmt88TFM0618memoria.pdf>:  
<http://openaccess.uoc.edu/webapps/o2/bitstream/10609/81188/8/dmt88TFM0618memoria.pdf>

Trigas Gallego, M. (s.f.). *Gestión de proyectos Informáticos*. Obtenido de Metodología SCRUM:  
<http://www.scrumguides.org/docs/scrumguide/v1/Scrum-Guide-ES.pdf>

Universidad de Alicante. (2015). *Servicio de Informática ASP.NET MVC 3 Framework*. Obtenido de Modelo vista controlador (MVC): <https://si.ua.es/es/documentacion/asp-net-mvc-3/1-dia/modelo-vista-controlador-mvc.html>

V, O. (s.f.).