



UNIVERSIDAD AUTÓNOMA DEL ESTADO DE HIDALGO
INSTITUTO DE CIENCIAS BÁSICAS E INGENIERÍA
ÁREA ACADÉMICA DE COMPUTACIÓN Y ELECTRÓNICA

DESARROLLO DE PROYECTOS COMPUTACIONALES

**SISTEMA DE PREDICCIÓN DE TRANSACCIONES COMERCIALES PARA
DETECTAR VARIACIONES ATÍPICAS DE UNA EMPRESA EN CDMX
PARA EL AÑO 2023.**

PRESENTAN:

352540 - MARTÍNEZ VEGA ANGEL DE JESÚS

400448 - HERNANDEZ GARCIA KATIA GUADALUPE

ASESORA:

DRA. ANILU FRANCO ARCEGA

UNIVERSIDAD AUTÓNOMA DEL ESTADO DE HIDALGO

CO-ASESOR:

DR. LUIS HERIBERTO GARCÍA ISLAS

UNIVERSIDAD AUTÓNOMA DEL ESTADO DE HIDALGO

Abstract

The next project describes the development of a system capable of predicting transactions of a company through a proposed algorithm. Although there are several documented prediction methods, in this case an algorithm was developed that uses basic functions for a better understanding of the process, in addition to reducing the requirements of the machine to process the huge amount of data and reduce the slowness that machines have with other existing algorithms.

This algorithm is finally linked specifically with the prediction of transactions made in a company to contribute to a greater knowledge in its clientele to satisfy them, in addition to be aware of setbacks.

Resumen

En el siguiente proyecto se describe el desarrollo de un sistema capaz de predecir transacciones de una empresa a través de un algoritmo propuesto. Si bien existen varios métodos de predicción documentados, en este caso se desarrolló un algoritmo que utilice funciones básicas para un mayor entendimiento del proceso, además de aminorar los requerimientos de la máquina destinada a procesar la enorme cantidad de datos y aminorar la lentitud que presentan las máquinas con otros algoritmos existentes.

Este algoritmo finalmente es enlazado específicamente con la predicción de transacciones realizadas en una empresa para contribuir en un mayor conocimiento en su clientela para satisfacerla, además de estar al pendiente de contratiempo.

Índice general

Lista de Imágenes	xI
1. Introducción	1
1.1. Problemática	1
1.2. Solución propuesta	2
1.3. Justificación	2
1.4. Objetivos	2
1.4.1. Objetivo General	2
1.4.2. Objetivos Específicos	3
1.5. Alcances	3
1.6. Limitaciones	3
1.7. Aportaciones	4
2. Fundamento Teórico	5
2.1. Marco Teórico	5
2.1.1. Creación de proyectos informáticos	5
2.1.2. Herramientas	18
2.2. Marco Conceptual	23
2.3. Estado del Arte	24
2.3.1. Modelos de series de tiempo para predecir la inflación en Venezuela	24
2.3.2. Series de tiempo: Una aplicación a registros hidrométricos en una cuenca del Estado de Oaxaca	25
2.3.3. Una evaluación experimental de técnicas de caracteriza- ción de series de tiempo aplicadas a electrocardiogramas .	26
2.3.4. Predicción con series de tiempo y regresión	27

ÍNDICE GENERAL

3. Implementación de la Metodología	29
3.1. Justificación	29
3.2. Aplicación de las Etapas	30
3.2.1. Comunicación:	30
3.2.2. Planeación	31
3.2.3. Modelado	34
4. Desarrollo del proyecto computacional	37
4.1. Construcción	37
4.1.1. Código y Pruebas	37
4.2. Despliegue	43
Validación	45
Conclusiones	49
Trabajo futuro	50
Referencias	52

Lista de Imágenes

2.1.	Figura de los niveles de la metodología de CRISP-DM.	8
2.2.	Fases de la metodología de CRISP-DM.	10
2.3.	Metodología de CRISP-DM.	11
2.4.	Figura que representa la metáfora de las distintas metodologías de Crystal y los minerales a los que se refiere.	12
2.5.	Figura que muestra el espectro de las distintas metodologías de Crystal según los integrantes que necesita.	13
2.6.	Figura que muestra el proceso de Crystal Clean con sus fases y como interactúan los roles junto con las técnicas o estrategias de Crsytal.	14
2.7.	Fases de la metodología de cascada	15
2.8.	Metodología de Kanban: Ejemplo de tablero [23]	17
2.9.	Logotipo de Pandas	20
2.10.	Paquetería de Microsoft Office	21
3.1.	Diagrama de actividades	32
3.2.	Transacciones reales	36
3.3.	Predicciones	36
4.1.	Librerías y etiquetado	38
4.2.	Cargar datos	38
4.3.	Clase graficación	39
4.4.	Normalización	39
4.5.	Algoritmo distancia euclidiana, definicion de valores	40
4.6.	Algoritmo distancia euclidiana, archivo txt	41
4.7.	Algoritmo distancia euclidiana, condicion de valores	42
4.8.	Algoritmo distancia euclidiana, predicción	42
4.9.	Lectura de archivo txt	43

LISTA DE IMÁGENES

4.10. Metodo MLP	43
4.11. Lectura csv	44
4.12. Resultados	46
4.13. Frecuencia por elemento	46
4.14. Composicion total de elementos	47
4.15. Composicion total de elementos, valor porcentual	47
4.16. Grafico, escala de Liker	48



Introducción

En la actualidad, muchas empresas o negocios manejan pagos de todo tipo con sus clientes, ya sean físicos o digitales. Sin embargo, con el paso del tiempo las transacciones se han popularizado tanto debido al alto beneficio que ofrecen, ya que son más prácticas, seguras y rápidas, lo que vuelve a estos modelos de pago una parte vital dentro del desarrollo de las empresas que desean crecer y transportar sus procesos transaccionales al mundo digital.

Por ello, las empresas buscan que sus procesos tengan buenos sistemas informáticos que salvaguarden y aprovechen los datos de los clientes, además de monitorear los procesos que estos realizan, con el propósito de obtener un mejor trato o servicio de las empresas a dichos clientes.

1.1 PROBLEMÁTICA

El equipo de trabajo empresarial labora actualmente con un sistema improvisado basado en la implementación de estadística en sus datos históricos. El sistema busca predecir los picos más bajos sacados del promedio y la media estándar de transacciones en una fecha y hora específica de acuerdo a cada cliente. Este sistema está desarrollado en sql developer, por lo que presenta distintas dificultades al momento de utilizarlo. Una de ellas es que el algoritmo es ineficiente ya que genera alarmas de forma repetitiva que no necesariamente detecta un problema real dentro de los patrones de transacciones de los clientes. Esto hace que el sistema envíe correos electrónicos masivos que generan molestias y falsos positivos dentro de la empresa.

1.2. SOLUCIÓN PROPUESTA

1.2 SOLUCIÓN PROPUESTA

La solución propuesta del presente trabajo es para mejorar los métodos de extracción de datos para la formación de alarmas mediante un historial y un algoritmo que simule una predicción conformada por el mínimo de transacciones que se realizan en la empresa bajo una determinada fecha y hora.

Este sistema utilizara un algoritmo eficiente basado en paradigmas de control y predicción difusa que analice los patrones históricos de las transacciones de los clientes. También contará con interfaces y opciones que hagan más práctica y ágil las tareas que realiza el sistema original.

1.3 JUSTIFICACIÓN

Debido a que el sistema actual de la empresa genera muchos contratiempos, se tiene el proposito de agilizar los procesos que el cliente necesita otorgando una herramienta eficaz y práctica que prediga los patrones y el comportamiento de las transacciones para que genere alarmas de forma precisa. De esta manera el cliente no tendrá problemas con las predicciones, ya que anteriormente la acumulación de datos entregaba problemas y acumulaba alarmas innecesarias. Este sistema ayudará a notificar a la empresa sobre qué clientes presentan una repentina baja de transacciones, con el fin de intervenir rápidamente y detectar una posible problemática. Esto ayudará a la empresa a generar una mejor comunicación con sus clientes, ayudando así a mejorar los servicios para satisfacer mejor las necesidades que estos requieran.

1.4 OBJETIVOS

1.4.1 OBJETIVO GENERAL

Desarrollar un sistema de predicción para identificar variaciones en transacciones comerciales de una empresa mediante la aplicación de algoritmos de series de tiempo.

1.4.2 OBJETIVOS ESPECÍFICOS

1. Preparar el historial de las bases de datos conforme a cada cliente para limpiar e integrar los datos en herramientas digitales mediante las técnicas del proceso KDD.
2. Implementar metodologías que permitan cumplir con tiempo y forma las metas propuestas de análisis, desarrollo, implementación y documentación.
3. Construir un sistema que permita mejorar la relación cliente-empresa otorgando predicciones precisas sobre el comportamiento atípico de las transacciones de clientes preferenciales.
4. Validar la aplicación mediante la norma ISO 9126:2001, realizando la ponderación mediante la escala de Likert, la elaboración de los ítems aplicando la taxonomía de Bloom para determinar el grado de aceptación de la aplicación.

1.5 ALCANCES

Este proyecto principalmente tiene una base de investigación que busca un algoritmo encargado de predicción de datos. Debido a diferentes problemáticas en el uso de algoritmos existentes para predecir datos, surge esta propuesta de predicción desarrollada desde cero, logrando beneficios como ser un algoritmo que perdure un buen tiempo y menor tiempo de compilación. A lo largo de este semestre se desarrollará el algoritmo encargado de la predicción de datos, un algoritmo comparativo ya existente en el mercado y una interfaz que facilite más el acceso a estos dos algoritmos. Se espera que en la primera entrega el algoritmo imprima los valores que obtiene al procesar los datos para un mayor entendimiento del mismo. En un futuro esta opción puede ser opcional o se puede preceder de ella.

1.6 LIMITACIONES

Debido al tiempo estipulado para desarrollar el proyecto, se espera que la primera entrega de este contenga ambos algoritmos funcionales para verificar las mejoras del algoritmo desarrollado en contraste con otros métodos de predicción. Sin embargo, en la entrega del proyecto, el algoritmo aún no será capaz

1.7. APORTACIONES

de ciclarse, no tendrá acceso a la base de datos original funcionando únicamente con datos en formato csv y, por lo tanto, aún no será capaz de lanzar alarmas esperando que esto se encuentre en el trabajo a futuro.

Esto es así, ya que el algoritmo se desarrollara de cero y se someterá incontables pruebas para el manejo de gran cantidad de datos, además de las pruebas de algoritmos existentes que no estén obsoletos.

1.7 APORTACIONES

A lo largo de la investigación en los algoritmos y sistemas de predicción que existen, se notó que la mayoría de documentación existente se encuentra obsoleta, debido a las librerías que usaban. Ya que la mayoría fue actualizada, retirando partes importantes de los algoritmos dejándolo inhabilitado. Otro patrón que se encontró fue la lentitud con la que la mayoría de datos eran procesados, ya que se encontró que un mínimo de 300 datos procesados (sin completar siquiera medio día) tardaba 30 minutos en procesarse.

El algoritmo que se realizó al ser desarrollado desde cero utiliza librerías que a pesar de actualizarse constantemente, jamás se ha registrado el deseo de eliminar las funciones utilizadas, ya que son funciones principales utilizados en innumerables proyectos. De esta manera logramos que el algoritmo perdure un poco más con el tiempo.

Al realizar las pruebas pertinentes con el algoritmo desarrollado se nota una enorme mejora en la velocidad de procesamiento, si bien depende completamente del tipo de procesador que tenga el computador, aun así es posible utilizar el sistema con una pc regular sin necesidad de tarjeta gráfica o componentes muy costosos.



Fundamento Teórico

2.1 MARCO TEÓRICO

Para el desarrollo del proyecto se presenta un marco teórico que abarque conceptos fundamentales durante el desarrollo del proyecto, así como metodologías y herramientas que de igual modo son referentes para el tipo de proyecto que se realizará.

2.1.1 CREACIÓN DE PROYECTOS INFORMÁTICOS

Dentro del desarrollo de proyectos informáticos existen una gran cantidad de metodologías para lograr un desarrollo estructurado, sistemático y fluido, por lo que a continuación se presentan metodologías enfocadas al desarrollo de proyectos informáticos, así como conceptos básicos para comprender mejor dichas metodologías.

Sistema

El concepto sistema aún en la actualidad sigue definiéndose, debido a que se encuentra con diferentes significados de acuerdo al contexto o rama de estudio en el que esté. Además, es uno de los términos más utilizados con diferentes significados. Mirando alrededor se puede notar que el mundo está formado por ellos; siendo colecciones complejas de diferentes elementos relacionados, logrando un desarrollo.

2.1. MARCO TEÓRICO

En otras palabras, un sistema es un grupo de recursos y procesos que logran en conjunto un propósito y satisfacen una necesidad. Éstos elementos pueden ser interactivos o independientes [8].

Mario Bunge señala que el concepto de sistema es dudoso aún y rescata tres definiciones que considera como "populares e incorrectas". La primera definición de sistema que rescata es la de "un sistema es un conjunto de elementos interrelacionados", comentando que solo se podría apoyar en sistemas meramente conceptuales. La segunda definición de sistema comenta que "identifica un sistema con una caja negra con entradas y salidas", en la cual añade que es correcto en casos particulares que no ahondan en el interior del sistema. La tercera definición es una simplificación de la segunda la cual dice "un sistema es una relación binaria", donde Bunge comenta que es otro significado meramente conceptual que no concuerda con los sistemas más concretos[5].

Sistema informático

Grupo de procesos que trabajan de forma colaborativa sobre un conjunto de datos estructurados de forma tal que se acople a las necesidades de la empresa. Recopila, elabora y distribuye los datos de forma selectiva, de ésta manera obtenemos información necesaria para las operaciones que la empresa lleva a cabo. Además en el aspecto organizacional, el sistema es capaz de tener dirección y control, apoyando en los procesos para tomar decisiones con alto nivel de eficiencia [8]. En el libro "Introducción a la gestión de sistemas de información en la empresa" los autores comentan la diferencia entre un sistema de información y un sistema informático argumentando que el sistema informático es un gran conjunto de componentes de hardware y software interconectados que generan salidas con base a entradas y por ello es determinista y formal este tipo de sistema [1].

Software

Información que se almacena en medios digitales. Ésta compuesta de interfaces (medio visual con el que se interactúa) y la lógica abstracta, la cual brinda la capacidad al humano de mantener una comunicación con la máquina, manipulándola a su beneficio. Describiendo el término de forma más técnica. Es un conjunto de programas informáticos, algoritmos, procedimientos y documentación para realizar un sistema que puede cumplir con tareas.[8]

Para Pressman, el software es un producto desarrollado por expertos en la programación, mediante metodos agiles y adaptables a desarrollar un producto de alta calidad y que principalmente satisfaga alguna necesidad del mundo real. Este mismo a lo largo de un periodo de tiempo establecido, se le debe realizar mantenimiento independientemente de que el programa ejecutado este en algún tipo particular de computadora o medio virtual.[18]

Proyecto

Existe una numerosa cantidad de definiciones al respecto, lo más destacable de cada una de ellas es que un proyecto es un plan o esfuerzo integral. El proyecto existe para lograr objetivos y obtener resultados. Conjunto de actividades independientes para un fin específico, con una duración pre-organizada. Cumplir con un proyecto requiere cumplir con todos los objetivos que estuvieron pre-escritos en las especificaciones técnicas, de costo y cumpliendo con el plazo de finalización.[8]

Mark gregory define proyecto como una entidad que en su totalidad es un sistema, dando ejemplos donde un proyecto sigue procedimientos que lo llevan a un fin y solo hasta ese momento es claro su proposito y las consecuencias del mismo [11].

José Ramón Rodríguez argumenta de forma simple que un proyecto es un conjunto o serie de tareas que se realizan en un tiempo determinado bajo un equipo de trabajo con un fin o meta, sinodo asi un proceso con un objetivo y además con un inicio y un fin[20].

Proyectos informáticos

Según Alfons Bataller es un proceso necesario para garantizar el éxito de un proyecto. El PMI dice que una metodología es un sistema en donde se involucran prácticas, técnicas, procedimiento y normas para aplicar una disciplina [2].

José Ramón Rodríguez comenta que un proyecto informático es todo lo que define a un proyecto en general y además, cuentan con un nivel de replicación bastante decente gracias a la existencia de muchas metodologías genrales para realizar proyectos informáticos y a la naturaleza de los software. Tambien asegura que si o si está realizado por especialistas, ya sean informáticos o profesionales que cuentan con las habilidades necesarias dentro de distintas disciplinas o prácticas profesionales para llevar a cabo estas actividades. Por

2.1. MARCO TEÓRICO

último el autor también comenta que los proyectos informáticos cuentan con características como estabilidad, volatilidad y nivel de extensión debido a que el desarrollo tecnológico cambia rápidamente [20].

Metodología Crisp - DM

Esta metodología que es comúnmente usada en proyectos de minería de datos (DM, data mining, en inglés),[16] se define según los autores, como Cross-Industry Standard Process for Data Mining y es un modelo de procesos jerárquicos para la extracción de datos, que se describen en 4 niveles de abstracción, siendo así:

- Fases.
- Tareas genéricas.
- Tareas especializadas.
- Instancias del proceso.

Esto permite a los niveles ir de lo general a lo específico.

Como se puede apreciar ver en la figura 2.1, el primer nivel y segundo nivel que

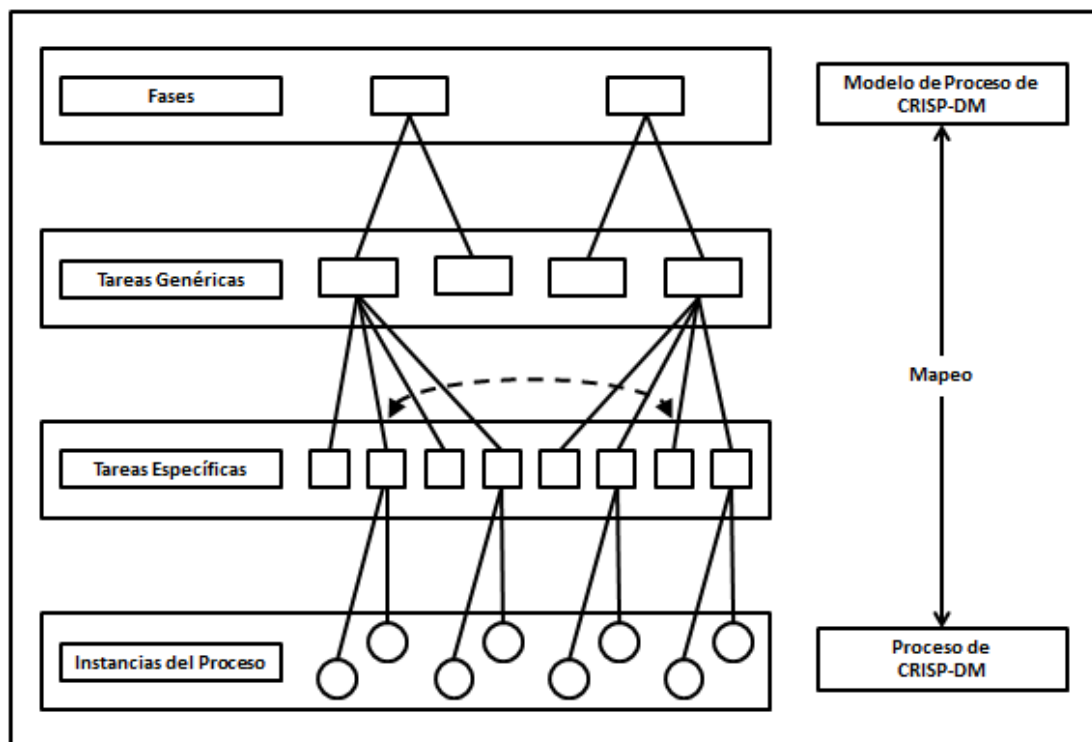


Figura 2.1: Figura de los niveles de la metodología de CRISP-DM.

hacen referencia al proceso de fases y a las tareas genéricas respectivamente, están dentro de lo que los autores denominan modelo genérico CRISP (CRISP Process Model, en inglés). En esta parte, el proceso se enfoca en organizar una serie de fases donde se realizan tareas genéricas que buscan ser completas y estables, es decir, estas tareas deben ser completadas debido a que se desea cubrir un total de variantes posibles sobre procesos existentes dentro de la minería de datos así como las aplicaciones en la misma, y a su vez, estables, porque se requiere que el modelo sea adaptable a desarrollos no pronosticados como otras técnicas de modelización.

La segunda parte denominada como modelo específico (CRISP Process, en inglés) abarca el tercer y cuarto nivel del modelo de Crisp. El tercer nivel trabaja con tareas especializadas, esto quiere decir que se busca definir como las tareas genéricas de segundo nivel son utilizadas para circunstancias diferentes y cómo difieren unas con otras de forma específica. Como ejemplo, se presentan circunstancias donde la limpieza de datos tiene procesos diferentes cuando se trabaja con valores numéricos en comparación con valores categóricos.

El cuarto nivel, es la instancia del proceso, donde hay un registro de acciones, decisiones y resultados que fueron organizados conforme se fueron desarrollando en los niveles superiores para representar lo que ocurrió específicamente en una situación particular.

Como se observa en la figura , entre la parte del modelo genérico CRISP y modelo específico existe una proyección que conecta el nivel genérico con el nivel especializado de tareas. Teniendo en cuenta esto, los autores definen dos tipos de proyecciones en CRISP-DM los cuales son:

Proyección para el presente: Es una proyección que se aplica al proceso genérico para realizar un único proyecto de minería de datos, donde se proyectan las tareas genéricas y sus descripciones únicamente a las tareas específicas del determinado proyecto conforme se vaya necesitando. Esta proyección es útil cuando se trata de un solo uso.

Proyección para el futuro: Es una proyección que especializa al proceso genérico bajo un determinado contexto que sirva para adecuarlo a las necesidades futuras de otros contextos parecidos. La proyección dependerá del contexto específico de la minería de datos y las necesidades a cubrir, pero ambos casos siguen un mismo proceso de proyección para llevar el proceso genérico CRISP

2.1. MARCO TEÓRICO

al modelo de CRISP, que son:

- Analizar el contexto específico del proyecto dentro de la DM y descartar detalles que van fuera de este.
- Añadir detalles que sean parte del contexto del proyecto.
- Especializar las tareas genéricas con base en las características del contexto del proyecto dentro de la DM.
- Homogeneizar las tareas genéricas a formatos más acorde al contexto del proyecto para dar claridad.

Por último, CRISP-DM toma como referencia el modelo actual para los ciclos de vida de la minería de datos adecuada a sí misma en el primer nivel mencionado con anterioridad, debido a que otorga las fases generales de un proyecto con las tareas a realizar y cómo se relacionan unas con otras. Se debe tomar en cuenta que la descripción general no permite identificar todas las relaciones que existen ya que principalmente pueden hallarse relaciones entre cada tarea de minería de datos, ya sean en los objetivos, la problemática o los intereses de los clientes, son contar también las que puede haber existentes en todos los datos. Dentro de este nivel se encuentran seis fases (figura 2.2), Las cuales son:

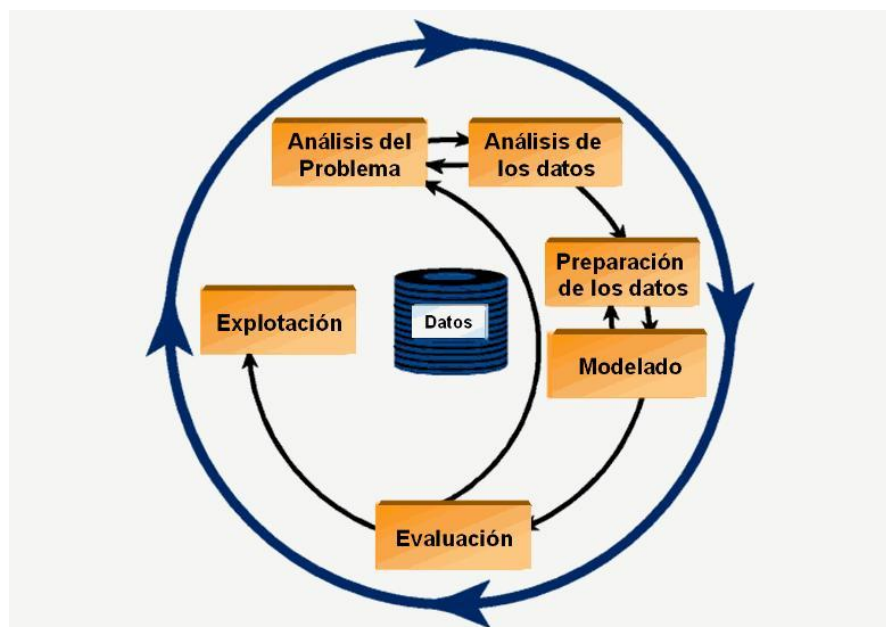


Figura 2.2: Fases de la metodología de CRISP-DM.

- Comprensión del negocio.
- Comprensión de los datos.

- Preparación de los datos.
- Modelado.
- Evaluación.
- Despliegue.

Como se observa en la figura 2.2, las fases están en un constante proceso iterativo por la importancia de pasar de una fase a otra cuando las circunstancias lo ameriten en el flujo de trabajo. También se aprecia que es un proceso cíclico pues esto sucede a menudo al trabajar en proyectos de DM donde el proceso no necesariamente termina al llegar a la fase final.

Business Understanding	Data Understanding	Data Preparation	Modeling	Evaluation	Deployment
Determine Business Objectives <i>Background</i> <i>Business Objectives</i> <i>Business Success Criteria</i>	Collect Initial Data <i>Initial Data Collection Report</i> Describe Data <i>Data Description Report</i>	Select Data <i>Rationale for Inclusion/Exclusion</i> Clean Data <i>Data Cleaning Report</i>	Select Modeling Techniques <i>Modeling Technique</i> <i>Modeling Assumptions</i> Generate Test Design <i>Test Design</i>	Evaluate Results <i>Assessment of Data Mining Results w.r.t. Business Success Criteria</i> <i>Approved Models</i>	Plan Deployment <i>Deployment Plan</i> Plan Monitoring and Maintenance <i>Monitoring and Maintenance Plan</i>
Assess Situation <i>Inventory of Resources</i> <i>Requirements, Assumptions, and Constraints</i> <i>Risks and Contingencies</i> <i>Terminology</i> <i>Costs and Benefits</i>	Explore Data <i>Data Exploration Report</i> Verify Data Quality <i>Data Quality Report</i>	Construct Data <i>Derived Attributes</i> <i>Generated Records</i> Integrate Data <i>Merged Data</i>	Build Model <i>Parameter Settings</i> <i>Models</i> <i>Model Descriptions</i>	Review Process <i>Review of Process</i> Determine Next Steps <i>List of Possible Actions</i> <i>Decision</i>	Produce Final Report <i>Final Report</i> <i>Final Presentation</i>
Determine Data Mining Goals <i>Data Mining Goals</i> <i>Data Mining Success Criteria</i>		Format Data <i>Reformatted Data</i> <i>Dataset</i> <i>Dataset Description</i>	Assess Model <i>Model Assessment</i> <i>Revised Parameter Settings</i>		Review Project Experience <i>Documentation</i>
Produce Project Plan <i>Project Plan</i> <i>Initial Assessment of Tools and Techniques</i>					

Figura 2.3: Metodología de CRISP-DM.

La figura 2.3, muestra cómo se vería la metodología CRISP-DM con ejemplos de sus tareas genéricas en cada fase y las tareas específicas proyectadas por cada tarea genérica.[17]

2.1. MARCO TEÓRICO

Metodología Crystal Clean

Esta metodología en si misma es llamada Crystal y presenta una familia de metodologías diseñadas para adecuarse a varios proyectos dependiendo de los requisitos de los mismos. Alistair Cockburn, el creador de esta metodología se refiere a ello comentando que Crystal Clean, por ejemplo, comparte lo que llama código genético, que indica que la variedad existente de diferentes metodologías Crystal, comparten y apuntan a un desarrollo bastante parecido sustancialmente, ya que como el autor comenta, el nombre de la metodología Crystal proviene de las características de los proyectos que al aumentar de magnitud y la forma critica en la que se desarrollan es bastante parecida a la de los minerales, cambiando de color y dureza. En la figura 2.4; se detalla con más claridad la metáfora, donde mientras más extenso sea un proyecto los niveles y el tipo de Crystal a usar varia de color.

L6	L20	L40	L80
E6	E20	E40	E80
D6	D20	D40	D80
C6	C20	C40	C80
Clear	Yellow	Orange	Red

Figura 2.4: Figura que representa la metáfora de las distintas metodologías de Crystal y los minerales a los que se refiere.

Es así que, Crystal Clear termina siendo una metodología que abarca trabajos de entre 0 y 8 personas donde, según su fundador, Alistair Cockburn, trabajos con esa cantidad de integrantes son por lo regular cortos o sencillos. Dentro de la metodología Crystal y por ende en la metodología Crystal Clear, encontramos partes importantes a tomar en cuenta, siendo así las siguientes:

- **Propiedades:**
En total son 7 propiedades de las cuales Crystal Clear solo requiere las tres primeras, las cuales son:

1. Entrega frecuente
 2. Mejora reflexiva
 3. Comunicación osmótica
- Estrategias y técnicas:
Crystal abarca un gran numero de estrategias y técnicas, de las cuales Crystal Clear puede ocupar un buen conjunto de entre las que existen, por mencionar algunas tenemos las siguientes:
 - Estrategia 1. Exploratoria 360°.
 - Estrategia 2. Victoria temprana.
 - Estrategia 3. Esqueleto andante.
 - Estrategia 4. Re-arquitectura incremental.
 - Estrategia 5. Radiadores de información.
 - Técnica 1. Metodología de la formación .
 - Técnica 2. Taller de reflexión.
 - Técnica 3. Planificación relámpago.
 - Técnica 4. Estimación Delphi mediante la clasificación de expertos.
 - Técnica 5. Reuniones diarias en pie.

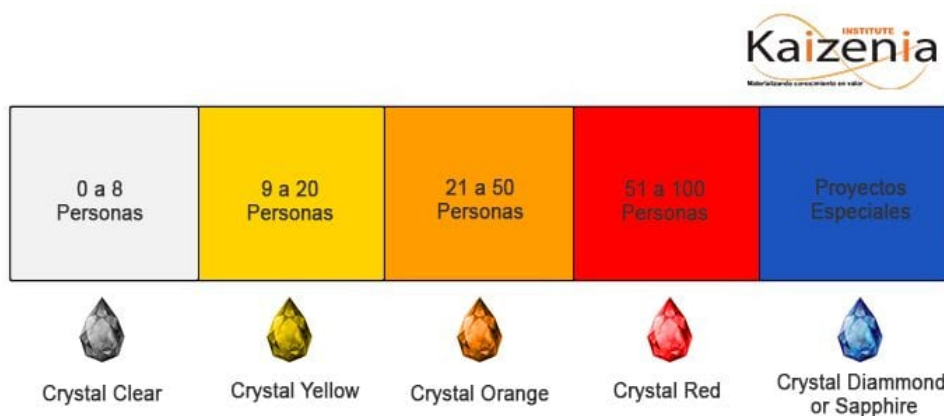


Figura 2.5: Figura que muestra el espectro de las distintas metodologías de Crystal según los integrantes que necesita.

- Fases del proceso de desarrollo:
En particular, Crystal clear lleva a cabo un ciclo de Fases anidadas, se pueden resumir como:
 - El ciclo del proyecto
 - El ciclo de entrega

2.1. MARCO TEÓRICO

- El ciclo de iteración
 - El ciclo de integración
 - La semana y el día
 - El episodio del desarrollo
 - Reflexión sobre el proceso
- Roles:
- El tiempo en cada una de las fases depende de lo que los equipos de trabajo realicen dentro de cada una. Si bien Crystal da importancia a distintos roles de trabajo en el proyecto, en Crystal Clear los mas importantes son:
- Patrocinador o Sponsor
 - Diseñador-programador líder
 - Diseñador-programador
 - Cliente.[7]

En la siguiente figura 2.6 se muestra de forma más gráfica como está integrada Crystal Clean.



Figura 2.6: Figura que muestra el proceso de Crystal Clean con sus fases y como interactúan los roles junto con las técnicas o estrategias de Crsytal.

Metodología de cascada

El modelo de cascada, también llamado ciclo de vida clásico es un modelo clásico. Éste modelo es más efectivo cuando los requerimientos del sistema se comprenden bien, ya que la comunicación y el despliegue fluyen de forma lineal. Suele usarse aún más cuando un sistema que ya existe requiere adaptaciones o mejoras.

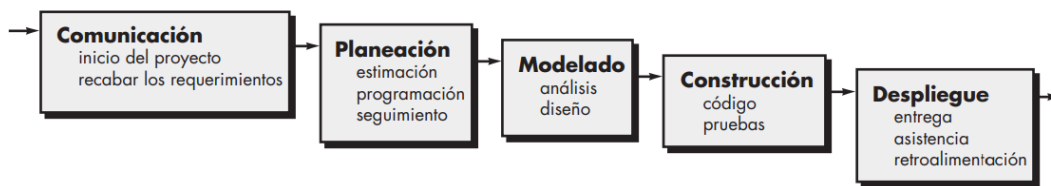


Figura 2.7: Fases de la metodología de cascada

En la figura 2.7 se puede observar sus cinco fases[18].

1. **Comunicación:** En el inicio del proyecto la comunicación es una parte fundamental para que el proyecto tenga éxito, ya que en base a esto el proyecto mismo tendrá éxito. En esta fase el equipo se enfocará en entender los requerimientos del sistema.
2. **Planeación:** Posteriormente llega la planeación, en donde se estima y programa el sistema mismo y las fechas de entrega, de forma en que cada fase sea completada con éxito y sin contratiempo. Además, es fundamental dar seguimiento durante todo el proyecto.
3. **Modelado:** En esta fase se realiza el análisis del proyecto, los requerimientos y la base de datos. Después es momento de diseñar la base de datos, y asimismo la estructura del sistema, siempre cumpliendo con todos los requerimientos.
4. **Construcción:** Cuando el diseño se considera satisfactorio se puede dar paso a la programación del código. Una vez que termina la programación de código se puede abrir pasó a las pruebas, con la finalidad de percatarse de errores y corregirlos antes de hacer la entrega del proyecto.
5. **Despliegue:** Una vez concluido con satisfacción las pruebas, sin encontrar ningún error se puede entregar el proyecto a los clientes, en donde serán capaces de probar el sistema y enviar retroalimentación en caso de ser necesario.

En la actualidad existen innumerable cantidad de softwares que nos ayudan a gestionar proyectos, siendo aplicaciones web, Android o ios- a continuación se nombrarán las más populares:

2.1. MARCO TEÓRICO

- cascade
- Trello
- Jira
- Wrike
- Asana
- ClickUp

Metodología de Kanban

Se basa en la idea de limitar el trabajo en curso, y abrir un nuevo bloque únicamente cuando el bloque de trabajo anterior culmino con éxito o pasó a otra función posterior de la cadena. Esta metodología utiliza un mecanismo de control visual para tener un seguimiento claro durante el trabajo, este proceso puede trabajarse con herramientas como un panel o pizarra con notas adhesivas o un panel electrónico que ayude a gestionar el flujo de trabajo y asignaciones. Aun que se recomienda usar ambos. Este método tiene la ventaja de exponer los cuellos de botella, colas, variabilidad y desperdicios de tiempo. Por lo que existe un mayor control en el rendimiento de la organización, específicamente hablando de la cantidad de trabajo y el ciclo de tiempo requerido para entregarlo. Además, ayuda a los interesados en el proyecto visualizar el desarrollo de éste, además de los efectos de sus acciones o falta de acción, promoviendo la colaboración. Para implementar esta metodología se debe de tener en cuenta estos aspectos:[6]

1. Definir el flujo de trabajo de los proyectos: primero definimos nuestro propio tablero que debe ser visible y accesible a todos los miembros del equipo. Cada columna almacena el estado concreto del flujo de tareas, ayudando a saber en qué situación se encuentra el proyecto. En esta metodología en específico las tareas que se priorizan son las nuevas, dejando las realizadas detrás. No existen fases del ciclo establecidas, pues se establecerán de acuerdo al proyecto.
2. Definir el flujo de trabajo de los proyectos: primero definimos nuestro propio tablero que debe ser visible y accesible a todos los miembros del equipo. Cada columna almacena el estado concreto del flujo de tareas, ayudando a saber en qué situación se encuentra el proyecto. En esta metodología en específico las tareas que se priorizan son las nuevas, dejando las realizadas detrás. No existen fases del ciclo establecidas, pues se establecerán de acuerdo al proyecto.

3. Stop Starting, start finishing: Lema principal de la metodología. De esta manera prioriza el trabajo en curso, en vez de empezar nuevas tareas
4. Control de Flujo: La metodología puede mezclar tareas y proyectos. Lo importante es mantener a los trabajadores con un flujo constante de trabajo, las tareas más importantes en cola para ser desarrolladas y un seguimiento pasivo para no interrumpir al trabajador

Un ejemplo de la metodología es el siguiente:

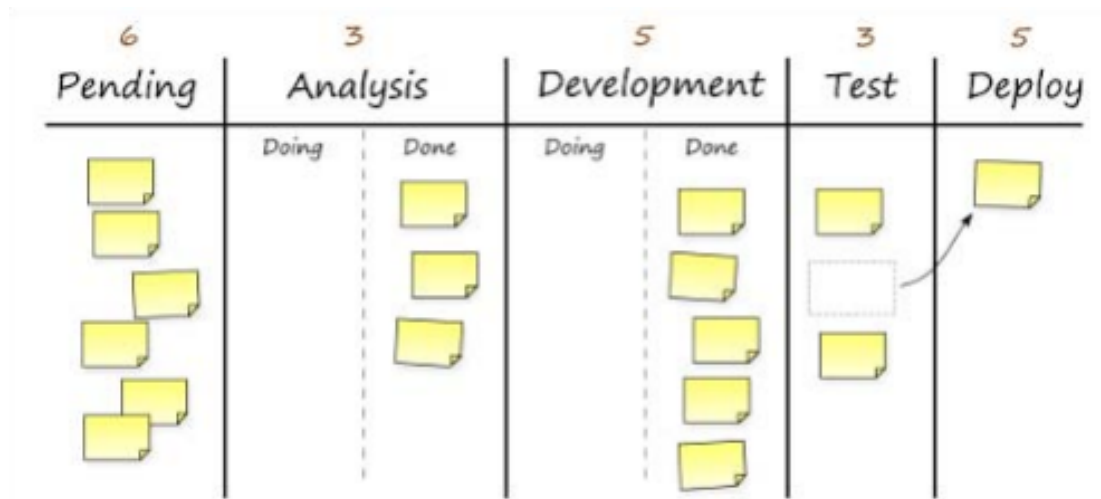


Figura 2.8: Metodología de Kanban: Ejemplo de tablero [23]

Future Driven Development

Esta metodología hace énfasis en cómo se realizan las fases de diseño y construcción. Además, son importantes los aspectos de calidad durante todo el proceso e incluye un monitoreo permanente. Define entregas tangibles y formas de evaluación del progreso del proyecto. Se basa en un ciclo muy corto y repetitivo, nunca superior a dos semanas, siempre orientados a cumplir una lista de Futures que debe tener el software a desarrollar. Un Feature debe cumplir con la siguiente lista de características:[2]

- Debe ser simple y poco costosa de desarrollar, de entre uno y diez días.
- Aportar valor al cliente y relevante.
- Expresar términos de acción, resultado y objeto.

2.1. MARCO TEÓRICO

La metodología tiene cinco fases iterativas.

1. Desarrollo/Modificación de un modelo Global
2. Creación/Modificación de la lista de Features
3. Planificación
4. Diseño
5. Implementación

La metodología presenta ciertas ventajas y desventajas que deben tomarse en cuenta.

Ventajas

- No se malgasta tiempo y dinero con soluciones que en realidad no son un requisito
- Cada componente del producto final fue aprobado
- Rápida respuesta a cambios de requisitos
- Entrega continua y en plazos cortos de software funcional.
- Trabajo conjunto cliente-equipo de desarrollo
- Evita malentendidos.

Deventajas

- Falta de documentación del diseño.
- Fuerte dependencia de las personas. Ya que se evita la documentación y los diseños convencionales, los proyectos ágiles dependen críticamente de las personas.

2.1.2 HERRAMIENTAS

Existen una gran variedad de herramientas para el desarrollo de proyectos informáticos así como en el énfasis de trabajar con técnicas de predicción de series de tiempo, por lo que para trabajar con dichas herramientas es recomendable contar con un buen hardware que acompañe a estas. Dejando en claro esto se dispone de las siguientes herramientas para el desarrollo de proyectos computacionales:

NetBeans

NetBeans es una aplicación que proviene de Apache Software Foundation (ASF). Esta fundación proporciona software de forma gratuita. Dando servicios y soporte, logrando que toda una comunidad colabore. Apache NetBeans es una IDE en la que se pueden desarrollar proyectos de distintos lenguajes y con distintas herramientas que facilitan la realización de proyectos. NetBeans puede desarrollar proyectos con Java, Groovy, php, HTML5, CSS, JavaScript, etc. Las IDE ayudan a los programadores a utilizar todo el potencial en productividad, teniendo un desarrollo continuo sin necesidad de utilizar tantas herramientas. Sin embargo, cada IDE tiene cualidades específicas. En el caso de NetBeans una de las cualidades más útiles es el resaltado de llaves, corchetes, paréntesis y comillas. Logrando identificar fácilmente el inicio y cierre de éstas evitando un problema de dedo. Además, otra cualidad son las sugerencias inteligentes, de ésta manera el código teniendo código con antelación puede ayudar a codificar. Incluso sí en algún momento no se recuerda alguna función, las sugerencias pueden ser útiles al resolver ese acertijo. Otra herramienta practica es la generación de código, gracias a ellas se pueden generar varias construcciones, métodos completos, agregar propiedades de forma automática, logrando de ésta forma mitigar el tedió al momento de programar. También, ésta IDE proporciona sugerencias en los Javadoc, que suelen ser bastante útiles ya sea al agregar nuevo código o editarlo, identificando de forma veloz cual puede ser el problema que genere la ruptura del código, Y así como aquellas herramientas descritas NetBeans tiene muchas más, logrando que la experiencia del programador tenga muchas más facilidades [3].

Pandas

Pandas es una herramienta para manipular y analizar datos, además de ser de código abierto, fue construido sobre el lenguaje Python.

Tiene su inicio en 2008 en AQR Capital Management; para 2009 ya era código abierto y actualmente cuenta con una comunidad amplia de colaboradores activos.

Pandas desde 2015 es patrocinado por NumFOCUS, lo cual ha contribuido a fortalecer el proyecto de código abierto de clase mundial.

Sus principales puntos:

- Un objeto DataFrame, que permite la manipulación de datos con indexa-

2.1. MARCO TEÓRICO



Figura 2.9: Logotipo de Pandas

ción integrada de forma rápida y eficiente.

- Brinda herramientas para leer y escribir datos entre estructuras de datos en memoria y diferentes formatos: CSV y archivos de texto, Microsoft Excel, bases de datos SQL y el formato rápido HDF5.
- Permite la alineación inteligente de datos y manejo integrado de datos faltantes.
- Transformación y rotación flexibles de conjuntos de datos.
- Permite la segmentación inteligente basada en etiquetas , indexación elegante y creación de subconjuntos de grandes conjuntos de datos.
- Podemos insertar y eliminar de las estructuras de datos para la mutabilidad del tamaño.
- La indexación de ejes jerárquicos brinda una forma intuitiva de trabajar con datos de alta dimensión en una estructura de datos de menor dimensión.
- Funcionalidad de series temporales: generación de rango de fechas y conversión de frecuencia, estadísticas de ventanas móviles, cambio de fecha y retraso.
- Python con pandas se usa en una gran variedad de dominios académicos y comerciales , que incluyen finanzas, neurociencia, economía, estadísticas, publicidad, análisis web y más.

La misión de Panda es ser el bloque de construcción fundamental de alto nivel para realizar análisis prácticos de datos del mundo real en Python.

Si visión es crear un mundo donde el software de manipulación y análisis de datos sea accesible para todos, gratis para que los usuarios lo usen y modifiquen, flexible, rápido y fácil de usar.

Para Panda es fundamental el respeto y los valores, por lo que brinda un espacio

acogedor para sus usuarios y colaboradores, sin importar la experiencia, género, identidad y expresión de género, orientación sexual, discapacidad, apariencia personal, religión, raza, etnia o nacionalidad[24] [14].

Microsoft Office



Figura 2.10: Paquetería de Microsoft Office

Si bien las herramientas para desarrollar software son parte fundamental a la hora de trabajar en los proyectos computacionales, no podemos dejar de lado la parte de documentación, la cual es pilar para llevar a cabo de forma ordenada y metodológica todo el flujo de trabajo de el proyecto en cuestión. Para cumplir con esta labor de documentar los proyectos, existe una variedad de herramientas que de entre las más destacadas tenemos la paquetería completa de Microsoft Office, la cual actualmente ya cuenta con servicio en la nube llamado Microsoft 365, pero no deja de lado la utilidad que presenta la paquetería office, la cual cuenta con los elementos necesarios para una buena documentación. para entrar en contexto, Microsoft Corporation es una empresa tecnológica multinacional , fundada por Bill Gates y Paul Allen el 4 de abril de 1975 en Albuquerque , Nuevo México en Estados Unidos. Su misión es empoderar a cada persona y cada organización en el planeta para lograr sus objetivos y metas planteadas. Además Microsoft permite la transformación digital para la era de una nube inteligente y un borde inteligente. La paquetería de Microsoft Office cuenta con recursos destacables como:

- Microsoft Word: excelente para editar y documentar

2.1. MARCO TEÓRICO

- Microsoft Power point: Bueno en general para elaborar presentaciones que aporten en el transcurso del desarrollo del proyecto
- Microsoft Excel: Útil para generar representaciones graficas con base en datos numéricos o estadísticos

En general, la paquetería de Office es eficiente con un correcto uso a la hora de documentar proyectos.[15]

Weka

Weka es un software desarrollado por la universidad de Waikato (Nueva Zelanda), tiene una alta colección de algoritmos de Maquinas de conocimiento que pueden ser aplicados sobre datos mediante varias interfaces que ofrece. Esta aplicación también tiene la capacidad de realizar transformaciones sobre datos, realizar clasificación, regresión, clustering, asociación y visualización. Una gran ventaja que proporciona Weka es su diseño, ya que es una herramienta extensible y añadir nuevas funcionalidades no es complicado en la ecuación.[9]

Weka trabaja con el formato .arff, acronimo de Attribute-Relation File Format. Este archivo se estructura en tres partes:

1. *Cabecera*: Nombre de la relación, expresada con datos tipo String, por lo que si se requieren espacios en blanco es necesario agregar el entrecomillado @relation <nombre-de-la-relacion>
2. *Declaraciones de atributos*: Atributos que tendrá el archivo junto a su tipo de dato. El nombre del atributo será tipo String. Serán con 5 tipos de datos, ya sea *Numeric* en donde nos encontraremos con números reales, *Integer*, en donde toparemos con números enteros
3. *Sección de datos*: Datos que componen la relación separando entre comas los atributos y con saltos de línea las relaciones.

GitHub

GitHub es una aplicación gratuita creada por programadores y para ellos, tiene la capacidad de conectar aplicaciones y desarrollar automatizaciones basadas en programación de eventos. En la mayoría de los casos GitHub es la herramienta perfecta para tener como repositorio, trabajar conjuntamente con un equipo y compartir tus proyectos. [10]

A continuación se nombrarán algunas herramientas que GitHub tiene para trabajar:

- Acciones: Códigos prediseñados útiles para realizar operaciones en un flujo de trabajo. Estas acciones pueden hacer hasta 500 integraciones
- Fuentes de eventos: Desencadenan flujos de trabajo. Cuando un evento ocurre en una aplicación de terceros, se utiliza una fuente de eventos.
- Flujo de trabajo: Peticiones que se ejecutan de forma automática. Secuencias de pasos o programación diseñada con un código personalizado. Esta programación desencadena eventos.

2.2 MARCO CONCEPTUAL

Sistemas de Trabajo del Conocimiento (STC)

Sistemas de información que permiten manejar la información en los procesos de crear e integrar nuevos conocimientos para la empresa, normalmente este conocimiento o información ayuda a la empresa a tomar mejores decisiones.[8]

Knowledge Discovery in Databases (KDD)

Término creado en 1989, se refiere al proceso de extracción de información desde una base de datos, procesarlos y por medio de ello obtener nueva información que servirá para el futuro. El proceso KDD está compuesto por varios pasos, para que de esta forma se complete de forma certera, siendo selección de datos, pre procesamiento y limpieza de los mismos, minería de datos e interpretación y evaluación de los mismos. Es importante recalcar la importancia de tener un proceso iterativo, siempre acompañado de un experto que sea capaz de comunicarnos cómo manejar la información en cualquier circunstancia que se nos presente.[19]

El concepto KDD es relativamente nuevo por lo que su desarrollo permanece constante hoy en día, entrelazándose con áreas como base de datos, aprendizaje automático, reconocimiento de patrones, estadística, teoría de la información, inteligencia artificial, razonamiento con incertidumbre, visualización de datos y computación de altas presentaciones. Una vista general sobre los sistemas KDD se pueden encontrar en la siguiente referencia[13]

2.3. ESTADO DEL ARTE

Modelo Predictivo

También referido como aprendizaje supervisado. Ya que en el momento de capacitación los datos son presentados con datos de entrada y de salida. Estos algoritmos tienen dos fases. La primera fase, llamada fase de entrenamiento o supervisión trabaja con un conjunto de datos que servirán para entrenamiento, posteriormente los parámetros internos se ajustarán de forma que se minimice el error de predicción de la variable dependiente. En la siguiente fase, llamada fase de prueba se aplica la fase de prueba en la cual se realizará la estimación del error en el modelo de prueba, (no en el de supervisión). El error encontrado será una aproximación más cercana. El objetivo de este algoritmo es encontrar modelos que minimicen el error de predicción. En la siguiente figura encontraremos un gráfico en donde se muestra el comportamiento de los errores de entrenamiento en diferentes complejidades [25]

Series de tiempo

Es una serie o secuencia de datos ordenados y distanciados con valor cronológico sobre sus características, observados en diferentes momentos. Los datos pueden adoptar diferentes comportamientos de acuerdo al tiempo, siendo una tendencia, ciclo, etc. Las series de tiempo suelen estar enfocadas en predicciones y éstas predicciones se basan en el comportamiento pasado de la variable de interés.[4]

2.3 ESTADO DEL ARTE

El estado del arte es una recopilación de proyectos que presentan una similitud con el trabajo que se está realizando, por ello en este apartado se muestra una recopilación de trabajos que cuentan con modelados y sus respectivas técnicas de predicción de series de tiempo.

2.3.1 MODELOS DE SERIES DE TIEMPO PARA PREDECIR LA INFLACIÓN EN VENEZUELA

El equipo del proyecto presento varios modelos continuos para realizar finalmente las series de tiempos oportunas que buscaban manejar un pronóstico de la evolución que tenía la inflación en Venezuela en ese entonces. Se busco

hacer estimaciones en cortos y medianos periodos de tiempo de la tendencia de los precios en el país, quedando como principal insumo esta variable y su evolución respecto al tiempo.

Los modelos funcionaron de la siguiente manera respectivamente: El primer modelo fue realizado con procedimientos de suavización exponencial, donde el equipo de trabajo comenta que se realiza mediante la obtención de la ecuación que describe el proceso estocástico, esto para obtener el promedio de los niveles experimentados bajo la última variable más reciente de precios; El segundo modelo corresponde a los resultados bajo especificaciones de ARIMA que buscan encontrar dependencias de una variable específica respecto al historial de cambios a través de la integración de resultados pasados de la misma variable, así como algún evento importante a lo largo del lapso de tiempo del historial; El último modelo fue uno de función de transferencia con técnicas de regresión con series de tiempo para obtener los datos importantes sobre la inflación en Venezuela bajo un modelado llamado ARMA, esto, incluyendo también (para más eficacia) una combinación de ecuaciones que se puedan integrar a las técnicas series de tiempo.

Como conclusión, el equipo de trabajo de este proyecto comenta que el hecho de proyectar series de tiempo en corto plazo, implico que no se necesitará de una gran cantidad de datos en comparación con los análisis de regresión múltiple. Además de esto, advierten que el sistema solo debe usarse como alternativa para predecir la inflación ya que solo está modelado bajo una variable que es la de la última actualización de los precios y a pesar de ser funcional no está fundamentado en un sistema económico como tal, lo que implica menos rigurosidad en cómo es el proceso de proyección pues no toma en cuenta posibles variables determinantes de una inflación. Por otro lado, comentan la eficacia de predicción que tiene tomar en cuenta el historial de cambios de la propia inflación para obtener un resultado de cambios presentes o futuros.[12]

2.3.2 SERIES DE TIEMPO: UNA APLICACIÓN A REGISTROS HIDROMÉTRICOS EN UNA CUENCA DEL ESTADO DE OAXACA

La autora de esta tesis presenta un proyecto con el fin de utilizar series de tiempo en la obtención de registros hidrométricos en una cuenca del rio Verde en Paso Ancho, Oaxaca. Por lo que decidió aplicar un modelo pertinente para esta tarea analizando series de tiempo para encontrar el modelo que pueda obtener

2.3. ESTADO DEL ARTE

los mejores resultados de un historial anual instantáneo de la planta hidroeléctrica y así conseguir la información esencial para determinar conductas futuras que ayudaran a planear un proceso de obtención de energía eléctrica, prevención de desastres naturales o ver la reserva de agua en abasto de la cuenca.

El proceso fue mediante la variable de caso de uso de gastos mínimos anuales instantáneos en metros cúbicos por segundo modelando el periodo que comprende desde el año 1957 al 2005 y utilizando una función de autocorrelación parcial muestral, obteniendo así que el mejor modelado será el de ARIMA, la cual al realizarse las oportunas validaciones se realizó la predicción mediante el enfoque de Box-Jenkins, obteniendo así resultados que si bien fueron óptimos y tuvieron errores cuadráticos mínimos, aun así no fueron los más deseados.

Esto último lleva a la conclusión de que debido a que diferían con la ingeniería, la autora comenta que puede ser debido a que la ciencia no ve procesos naturales como un proceso general en sí mismo, pero que esta diferencia de perspectiva puede llevar a un desarrollo de nuevas investigaciones que apunten a un análisis más subjetivo que pueda explicar de mejor manera el modelo con las variaciones pico que estas presentan.

2.3.3 UNA EVALUACIÓN EXPERIMENTAL DE TÉCNICAS DE CARACTERIZACIÓN DE SERIES DE TIEMPO APLICADAS A ELECTROCARDIOGRAMAS

se han incorporado numerosos algoritmos capaces de procesar datos para encontrar información sumamente valiosa. Sin embargo, con el continuo crecimiento en la rama de datos incrementan las opciones que tiene un analista al momento de procesar datos. El objetivo de este proyecto es encontrar los algoritmos con mayor valor de precisión y descubrir que factores pueden afectar estos datos para obtener un resultado más o menos preciso. Una característica importante de los electrocardiogramas es que pueden tratarse como una serie de tiempo, más específicamente una serie de valores que miden un evento a lo largo del tiempo. Es importante la mención de la transformación de los datos, para que éstos tengan una representación indexable. Solo con los datos tratados pueden procesarse con los algoritmos y obtener un resultado más preciso. Como primer paso se trabajó con el término patrón de interés o motif según Mueen et ál. [mueen] Éste es una sub secuencia de valores en la serie de tiempo, en donde se detecta que la semejanza a su vecino más cercano es mayor a la semejanza en otra sub secuencia en la serie de tiempo con sus vecinos más cercanos. Es

importante generar un vector característico adecuado sin perder información relevante, ya que puede causar falsos positivos si no se realiza adecuadamente. En éste proyecto se utilizan vectores de 8, 16 y 32 dimensiones para valorar el grado de pérdida de información. Además de diferentes algoritmos de transformación. Siendo Clipped y FAX. En cuanto a las funciones de distancia primeramente se destaca que en éste sistema la semejanza entre dos patrones es inversamente proporcional a su distancia, entonces asumimos que entre menor sea ésta distancia, mayor será la semejanza. En éste proyecto se utilizaron 3 algoritmos diferentes para hacer la comparativa, en donde la base es un algoritmo de distancia, pero con algunas características únicas. [22] Para este trabajo se utilizaron registros de electrocardiogramas obtenidos en el Instituto Tecnológico de Massachusetts disponible a través del sitio de Internet <http://www.physionet.org>.

- Las búsquedas más eficientes para las técnicas APCA y SAC es cuando la separación mínima entre secuencias es de 8 y 16 posiciones.
- Cuando el espacio de separación entre patrones incrementa a 32 y 64 unidades, DFT introduce un menor número de falsos positivos.
- Las técnicas PAA y Clipped produjeron consistentemente el mayor número de falsos positivos.

2.3.4 PREDICCIÓN CON SERIES DE TIEMPO Y REGRESIÓN

El siguiente proyecto presenta modelos básicos de pronóstico con base en series cronológicas, regresión lineal, exponencial y parabólica. Esto con el objetivo de conocer la evolución y tendencia del mercado y negocios. Ya que para empresas de mediana y pequeña estatura aumentaría la posibilidad de éxito logrando planificar estrategias de producción. El trabajo presentó la teoría básica de los modelos de regresión, pero para ello fue necesario tener los datos de la empresa como ejemplo, posteriormente se registran las ventas realizadas de acuerdo a cierto periodo de tiempo. Primero se realiza una comprensión de naturaleza probabilística obteniendo un modelo de regresión lineal. Obteniendo los resultados de éste modelo se obtiene una tendencia secular, es decir, un patrón de la serie de tiempo que puede analizarse y obtener predicciones a través de ésta. Posteriormente se modifica la ecuación utilizada de tal forma que las predicciones sean más específicas de acuerdo a las tendencias del mes en específico en que se trata el problema. Este procedimiento logra eliminar

2.3. ESTADO DEL ARTE

la variación por tendencia, es decir, los valores reales ahora son expresados en porcentajes del valor que ésta en tendencia cambiando de acuerdo al mes. Para mayor precisión en las predicciones también se determina la razón promedio de cada mes (rpm), los valores desestacionados mensuales (VD) Finalmente, con la ecuación de tendencia mensual y el mes en que se desee estimar se obtiene el valor de tendencia del mes deseado. Si este valor es multiplicado por el índice estacional se obtiene la predicción. Así como se explicó el método de regresión lineal, también existen los métodos exponenciales y parabólicos que se aplican de acuerdo a las características específicas de los datos, tomando el método más conveniente. Éstas metodologías están basadas en la estadística y son sumamente usadas gracias a su fácil comprensión e implementación. El proyecto trabajó con diferentes herramientas de fácil adquisición siendo Excel, con el fin de evaluar las posibles metodologías y ponerlas al alcance de los datos de microempresas que no cuenten con muchos recursos para la compra de software especializados. Si bien estas herramientas son útiles, es sumamente necesario asegurarse en que la información que trabajas es confiable [21].

3

Implementación de la Metodología

En el desarrollo de proyectos informáticos existe un proceso a seguir que está basado en alguna metodología especializada para estos trabajos. Teniendo esto en cuenta existe un patrón sustancial que considera fases o etapas dentro de todo el proceso que no pueden faltar en alguna metodología existente, un análisis y un diseño e implementación, así como fases que ayuden en la evaluación y sostenibilidad de los proyectos a lo largo del tiempo, que pueden ir variando respecto a la metodología en cuestión. En este proyecto informático se hará uso de la metodología cascada por lo que en este capítulo se dará a conocer la razón, más de la metodología y la implementación de cada una de sus etapas dentro del contexto de este proyecto informático.

3.1 JUSTIFICACIÓN

El desarrollo del proyecto computacional se verá desarrollado bajo la metodología de cascada, un método tradicional pero bastante efectivo para la naturaleza del trabajo en cuestión, ya que se busca mejorar y adaptar a un nuevo software uno ya existente además de que dicho proyecto se realiza en un periodo de tiempo moderadamente corto. Gracias a la estructura de sus fases que permite gestionar bien el inicio del proyecto para un flujo de trabajo más eficaz en cuanto a la construcción futura del software final, por lo que la metodología cascada es una elección sensata para llevar a cabo este proyecto.

3.2 APLICACIÓN DE LAS ETAPAS

Ya que se sabe las razones por las cuales se eligió la metodología de cascada, así como saber más sobre el funcionamiento y proceso de la misma, es prudente dar a conocer como es que estas etapas se implementaran bajo el contexto del proyecto computacional que se aborda, por lo que a continuación se presenta de manera breve como se llevaran a cabo dentro de las distintas fases las actividades que demanda el proyecto desde su inicio hasta su implementación y retroalimentación.

3.2.1 COMUNICACIÓN:

Dentro de la fase inicial del desarrollo, al elegir el proyecto en cuestión y dar inicio al trabajo, consecuentemente, se realizó la recopilación de requisitos y requerimientos que necesitaba el sistema a través del equipo de trabajo con el cliente, formulando preguntas al mismo sobre la naturaleza del sistema ya existente, así como analizando este último verificando sus fallos, aciertos y su funcionalidad en cuanto a lo general así como su algoritmo de predicción para determinar que requisitos necesitará y que requerimientos tanto actuales como nuevos se encontrarán en el nuevo sistema informático.

La fase de comunicación en la metodología cascada sirve como el periodo de arranque dentro del proyecto ya que abarca el inicio del proyecto así como la recopilación de requerimientos que necesita el sistema en cuestión y cómo estos requerimientos se manejan y estructuran dentro del sistema actual y se estructurarán en el nuevo sistema informático.

Inicio y recopilación de requerimientos

Para comprender mejor la naturaleza de este proyecto se considera prudente hablar del análisis de requerimientos recabados con anterioridad y que comprenden desde los requerimientos funcionales, los no funcionales y los de dominio. Para verlo de forma estructurada y práctica se enlistaran a continuación los mas importantes de cada tipo de requerimiento a tomar en cuenta:

- **Requerimientos funcionales:**

- El sistema:

- Recibe un histórico de una empresa/cliente para crear un patrón inicial.

- Produce predicciones desde el patrón inicial del histórico real cada 5 minutos..
 - Evalúa las predicciones con el histórico real de ese momento del tiempo.
 - Genera una alerta si la predicción no se cumple durante 3 o más momentos en el tiempo.
 - Manda un correo electrónico si se genera una alerta.
 - Realiza el mismo proceso para otra empresa/cliente.
- Requerimientos no funcionales:
- El sistema:
- Trabaja de forma local.
 - Tendrá una interfaz simple.
 - Se iniciará cada que el propietario desee realizar las predicciones.
 - Presentara predicciones simples que pueden tener un gran margen de error.
 - Manejara un punto de respuesta largo o moderado dependiendo de los datos a procesar.
 - Solo usara máximo de 2 a 5 registros de las BD de empresas/cliente.
- Requerimientos de dominio:
- El sistema:
- Tendrá datos normalizados de 0 a 1 para poder realizar el proceso de predicción.
 - El proceso de predicción utilizara distancia euclidiana para evaluar los puntos en determinado espacio-tiempo mediante steps.
 - Usara un modelo forecasting para las series de tiempo.

3.2.2 PLANEACIÓN

En esta fase se presenta todo el organigrama del proyecto donde se desglosa la planificación de actividades consideradas, así como con su tiempo determinado por periodos cortos dentro de todo el tiempo estimado para el desarrollo de este trabajo.

3.2. APLICACIÓN DE LAS ETAPAS

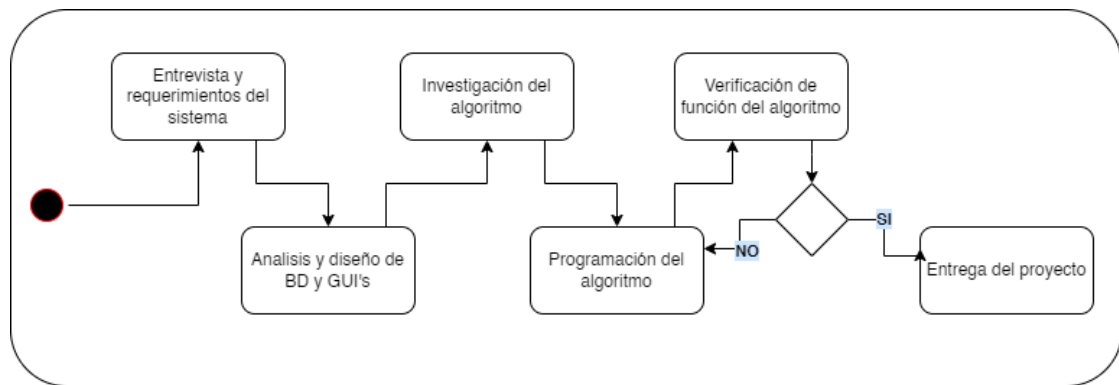


Figura 3.1: Diagrama de actividades

Se realizaron estimaciones de lo recabado en la fase de comunicación para verificar que requisitos además de los requerimientos que serán los principales y secundarios para que en el lapso de tiempo disponible para el trabajo puedan realizarse mediante actividades planificadas y programadas con sus propios tiempos estimados y destinados para completar al final el proyecto computacional como lo esperado, a su vez de realizar un monitoreo de cada actividad que se va iniciando, desarrollando y terminando durante todo el ciclo de vida del proyecto.

Estimación, Programación y Seguimiento

En la siguiente tabla se presenta el organigrama de actividades a seguir durante todo el periodo estimado para el proyecto, así como la planificación de la documentación del mismo:

CAPÍTULO 3. IMPLEMENTACIÓN DE LA METODOLOGÍA

CRONOGRAMA DE ACTIVIDADES						
Actividad	Jun	Jul	Ago	Sep	Oct	Nov
FASE: COMUNICACIÓN						
Búsqueda de proyecto						
Selección de proyecto						
Búsqueda de asesora de tesis						
Obtención de requerimientos						
FASE: PLANEACIÓN						
Organigrama de actividades						
FASE: MODELADO						
Análisis						
Diseño						
FASE: CONSTRUCCIÓN						
Código						
Pruebas						
FASE: DESPLIEGUE						
Entrega						
Aistencia						
Retroalimentación						

3.2. APLICACIÓN DE LAS ETAPAS

CRONOGRAMA DE ACTIVIDADES						
Actividad	Jun	Jul	Ago	Sep	Oct	Nov
DOCUMENTACIÓN						
Portada						
Resumen y abstract						
Índice						
Introducción						
Capítulo I Y II						
Capítulo III y IV						
Apartado de validación						
Trabajo futuro						
Referencias bibliográficas						
Anexos						
Apéndices						

3.2.3 MODELADO

Como ya se mencionó anteriormente, los proyectos informáticos cuentan con una estructura primordial que dependiendo de la metodología está enfocada a iniciar preferentemente con un análisis y diseño antes de comenzar cualquier construcción de código. Esto debido a que el análisis y diseño funcionan para proporcionar bases solidas en el proyecto pues de aquí se parte para un desarrollo solido en cuanto a todo el ciclo de vida de dicho trabajo y como se realizara y trabajaran las distintas áreas dentro del mismo.

En esta fase, se ha llegado al inicio de lo que vendrá siendo toda la estructura del sistema, pues abarca el análisis y diseño de lo que anteriormente se recabo, contando con como se va a estructurar el sistema internamente junto con su algoritmo de predicción y como se visualizara esto en el exterior, todo con base a los requerimientos de funcionalidad que se piden analizando cada uno de ellos y diseñando los modelos preliminares para empezar a continuación la etapa de desarrollo.

Análisis y Diseño

En este apartado se encuentra el análisis y el diseño del proyecto enfocado a los requerimientos funcionales, no funcionales y de dominio, basados en el trabajo en cuestión y qué de acuerdo a dicho proyecto informático cuenta con varias propiedades a tomar en cuenta sobre como trabaja interna y externamente. También se presenta el análisis de la base de datos y la estructura dentro del modelo relacional. Presentando también un diseño de como trabajara el proyecto informático mediante un diagrama de actividades.

Análisis

Para el análisis del proyecto es fundamental tener en cuenta el tipo de predicciones que se encuentran hoy, la elección del método a utilizar, qué beneficios tiene, su capacidad de adaptación en cuanto al tipo de serie que se introduzca, asegurarse que siga siendo funcional por un largo periodo de tiempo.

Análisis de la Base de datos

Al realizar el análisis debemos siempre tomar en cuenta que las entidades no deben ser repetitivas. En este caso, este sistema tiene como objetivo ser predictivo. Por esta razón la importancia de tener el registro real del número de transacciones cada 5 minutos y tener el número predictivo con los mismos datos. De esta manera se logra asegurar que constantemente use los datos históricos más viejos y recientes para obtener un algoritmo óptimo. El sistema contará con 3 entidades: Empresa, fundamental ya que el patrón de cada empresa varía, por lo que el algoritmo se implementará en cada una. Transacciones reales, registrando las transacciones en tiempo real y transaccionesPrediccion, lugar donde se registrará el resultado del algoritmo. De esta manera, el sistema será capaz de comparar los resultados reales con las predicciones, alzando una alarma en caso de ser necesario y con la capacidad de mejorar las predicciones con los nuevos datos.

3.2. APLICACIÓN DE LAS ETAPAS

Modelo Relacional

MITDBA.MON_TOEMPRESAS_REAL	
P *	FECHAHORA DATE
P *	CD_EMPRESA VARCHAR2 (4 BYTE)
	TRXS NUMBER
	APROBADAS NUMBER
PK_MON_TOEMPRESAS_REAL (FECHAHORA, CD_EMPRESA)	
PK_MON_TOEMPRESAS_REAL (FECHAHORA, CD_EMPRESA)	

Figura 3.2: Transacciones reales

En la figura 3.2 se registran el número de transacciones realizadas en tiempo real. Además, es posible que en algunas transacciones existan fallos, por lo que es necesario tener en cuenta las transacciones aprobadas, para validar el número final de transacciones por fecha, hora, minuto y segundo.

MITDBA.MON_TOEMPRESAS_WARNING	
P *	FECHAHORA DATE
P *	CD_EMPRESA VARCHAR2 (4 BYTE)
	TRXS NUMBER
	APROBADAS NUMBER
	TRXS_MIN NUMBER
	TRXS_MAX NUMBER
	TRXS_AVG NUMBER
	TRXS_MIN_ACEPTADO NUMBER
	WARNING VARCHAR2 (2 BYTE)
PK_MON_TOEMPRESAS_WARNING (FECHAHORA, CD_EMPRESA)	
PK_MON_TOEMPRESAS_WARNING (FECHAHORA, CD_EMPRESA)	

Figura 3.3: Predicciones

La figura 3.3 es una tabla que también ocupará nuestro sistema, ya que de esta manera será capaz de registrar las transacciones reales mínimas, es decir el promedio mínimo de transacciones que llega a tener alguna empresa, las transacciones reales máximas, el registro de la predicción y a través de ésta predicción el valor mínimo aceptable de transacciones. Si las transacciones reales no alcanzan el mínimo aceptable según el algoritmo lanzará un warning y éste será registrado en la tabla.

4

Desarrollo del proyecto computacional

4.1 CONSTRUCCIÓN

El desarrollo del algoritmo fue en base al conocimiento previo de predicciones en series de tiempo y los patrones que la empresa de transacciones tiene. Se seleccionó la distancia euclidiana como base para encontrar diferencias entre cada serie de grupos. Además como metodo de comparación se añadió un algoritmo basado en IA para predicciones de datos.

4.1.1 CÓDIGO Y PRUEBAS

El proyecto requiere de varias bibliotecas importantes para realizar el algoritmo y graficación de datos. A continuacion en la figura 4.1 se muestran las librerías ocupadas para el desarrollo del algoritmo, procesamiento y graficación de datos.

4.1. CONSTRUCCIÓN

```
1  #Manipulacion y tratamiento de datos
2  from tokenize import PlainToken
3  import numpy as np
4  import pandas as pd
5
6  #Visualización de Dato
7  import matplotlib.pyplot as plt
8  import seaborn as sb
9
10 #Normalizar datos
11 from locale import normalize
12 from sklearn import preprocessing
13 from sklearn.metrics.pairwise import euclidean_distances
14
15 #Biblioteca de Distancia Euclidiana
16 from scipy.spatial import distance
17
```

Figura 4.1: Librerías y etiquetado

Las librerías se usan para 4 funciones importantes. La primera aporta el uso de series de tiempo, principalmente utilizada en las predicciones íntimamente relacionadas con el tiempo. Posteriormente, se encuentra la librería encargada de trabajar con matrices y series numéricas para el desarrollo de las operaciones creando grupos de datos. Después otro elemento relevante es la graficación de datos que aporta principalmente en el entendimiento de estos datos de acuerdo a los patrones encontrados. Por último se encuentran las librerías que realizan las operaciones como la distancia euclidiana.

El sistema utilizó una función para la lectura de los datos como se muestra en la figura 4.2, a través de esta librería se realiza la lectura de los datos.

```
datos= pd.read_csv("prueba3.csv", sep=',')
#d1 = Graficacion(datos)
#d1.imprimirGrafico()
```

Figura 4.2: Cargar datos

Como se observa en la figura 4.3 trabaja a través de clases por las diferentes ventajas que esto conlleva, además de facilitar el ciclado del mismo algoritmo con diferentes datos. Gracias a ésto es posible observar la clase destinada a graficar los datos

```

class Graficacion:
    def __init__(self, datos):
        self.datos = datos

    def imprimirGrafico(self):
        plt.figure(figsize=(22,6))
        sb.lineplot(x=self.datos['fechaHora'], y=self.datos['trxs'])
        plt.title('Transacciones por 5 minutos')
        plt.show()

```

Figura 4.3: Clase graficación

La distancia euclidiana requiere la normalización de los datos para evitar la pérdida de estos. Para ello existe la librería NumPy es capaz de transformar un conjunto de datos en una serie y la librería sklearn tiene la capacidad de normalizar los datos como se muestra en la figura 4.4. Normalizar los datos ayuda a controlar datos de diferentes escalas, evitando la distorsión y pérdida de información.

```

class NormalizarDatos:
    def __init__(self, datos):
        self.datos = datos

    def normalizar(self):
        trx = self.datos['trxs']
        d = pd.DataFrame(trx)

        tran = d.to_numpy()
        scaler = preprocessing.MinMaxScaler()
        normalizedlist=scaler.fit_transform(tran)

        return normalizedlist

```

Figura 4.4: Normalización

Como se muestra en la figura 4.5, el algoritmo recibe los datos que va a procesar y el tamaño de las agrupaciones apuntando a mantener un valor variable. Ya que distintos patrones de transacciones de acuerdo a "x" tiempo puede requerir distintos números de agrupaciones. Logrando que cada caso de predicción sea mucho más específico.

Estas dos entradas de datos suman a nuevas variables para el desarrollo del algoritmo, "muestra1" siendo la agrupación seleccionada para ser comparada con el resto de agrupaciones buscando encontrar los datos más similares. La variable "muestra2" es la encargada de representar las demás agrupaciones en el ciclo for, para lograr encontrar la distancia euclidiana entre ambas agrupaciones.

4.1. CONSTRUCCIÓN

En esta sección del algoritmo además se crean otras variables de apoyo para el ciclo y recorrido de datos.

```
class DistEuclidiana:
    def __init__(self, datos, longMuestra):
        self.datos=datos
        n1 = NormalizarDatos(self.datos)
        self.normalized_trx = n1.normalizar()
        self.longIterador =longMuestra+1
        self.longMuestra = longMuestra
        self.distancias=[]
        self.f = open("archivo.txt", "w")

    def distancia(self):
        #Definir valores
        trx = self.datos['trxs']
        d = pd.DataFrame(trx)

        tran = d.to_numpy()
        indice=len(self.normalized_trx)-self.longMuestra
        muestra1=self.normalized_trx[indice:indice+self.longMuestra]
        muestra1 =np.concatenate(muestra1)
        #Crear archivo y agregar Datos Normalizados
        self.f = open("archivo.txt", "a")

        muestra2 =[]
        iterador=len(self.normalized_trx)-self.longIterador
        longResta =1
        distancias=[]
        cumplen=[]
```

Figura 4.5: Algoritmo distancia euclidiana, definicion de valores

El algoritmo tiene la capacidad de crear y editar archivos tipo txt con los resultados que arroja, así mismo agrega los datos de la agrupación seleccionada para comparar(muestra1), la agrupación que se recorrerá con cada ciclo(muestra2), la posición en la que se encuentra (iterador). Esto con el propósito de una mayor comprensión del algoritmo, visualización de qué datos se están procesando y concebir los resultados de forma más ordenada. Esto puede observarse en la figura 4.6.


```

#Inicio de for
for i in range(self.longMuestra+1,len(self.normalized_trx),longResta):

    muestra2=self.normalized_trx[iterador:iterador+self.longIterador-longResta]
    muestra2 =np.concatenate(muestra2)
    #Agregar muestra 2
    self.f = open("archivo.txt", "a")
    self.f.write("Muestra1: ")
    self.f = open("archivo.txt", "a")
    self.f.write(str(muestra1))
    self.f = open("archivo.txt", "a")
    self.f.write("          Iterador:")
    self.f = open("archivo.txt", "a")
    self.f.write(str(iterador))
    self.f = open("archivo.txt", "a")
    self.f.write("          Muestra 2:")
    self.f = open("archivo.txt", "a")

    self.f.write(str(muestra2))

```

Figura 4.6: Algoritmo distancia euclidiana, archivo txt

Posteriormente, se realiza la operación de distancia euclidiana con una librería encargada de realizar esta operación de forma directa, logrando mayor precisión sin pérdida de datos. Para evaluar la cercanía de grupo a grupo se resta 1 menos el valor obtenido de la distancia. Se tomaron en cuenta únicamente los valores mayores a .9, pues son los más cercanos al grupo seleccionado. Finalmente, estos resultados son agregados al archivo txt como se muestra en la figura 4.7.

Estos resultados se agregan en un arreglo llamado "distancias" la segunda agrupación se arrastra un lugar con ayuda de las variables iterador y "longResta" logrando una nueva agrupación de datos en la variable "muestra2".

4.1. CONSTRUCCIÓN

```
#Operacion euclidiana
dist = distance.euclidean(muestra1,muestra2)
distancia2 = 1- dist
self.f.write("      Distancia: ")
self.f = open("archivo.txt", "a")
self.f.write(str(distancia2))
if distancia2 >= 0.9:
    self.f.write("      Si ")

    cumplen.append(tran[iterador+1])
    self.f.write("\n")

else:
    self.f.write("      No ")
    self.f.write("\n")

#Agrega resultado a un array
distancias.append(dist)
#Se recorre el segundo valor, 1 punto
self.longIterador=self.longIterador+1
longResta = longResta+1
iterador=len(self.normalized_trx)-self.longIterador
self.distancias=distancias
```

Figura 4.7: Algoritmo distancia euclidiana, condicion de valores

Todos los datos que cumplen con la condición de ser mayor a .9, son datos que tienen una enorme similitud y gracias a estos se realiza la predicción, ya que con el historial de estos patrones se logra suponer cuál será el patrón siguiente. Al final del archivo, todas las transacciones que cumplieron con la condición se guardaron en un array llamado `cumplen`".

En la figura 4.8 se agrega este array al final del archivo txt y se promedian todos los valores que pertenecen a este array obteniendo de esta manera la predicción.

```
cumplen = np.concatenate(cumplen)
self.f.write(" Datos que cumplen: ")
self.f = open("archivo.txt", "a")
self.f.write(str(cumplen))
suma = 0

for valor in cumplen:
    suma = suma + valor
longCumplen = len(cumplen)
promedio = suma /longCumplen
self.f.write("\n")
self.f.write("Predicción: ")
self.f = open("archivo.txt", "a")
self.f.write(str(promedio))
return distancias
```

Figura 4.8: Algoritmo distancia euclidiana, predicción

Finalmente, se encuentra la clase capaz de leer el archivo txt como se observa en la figura 4.9, donde se almacena la información. Es importante mencionar que este archivo editara y remplazará los valores de acuerdo a los resultados que el algoritmo obtenga y datos que procese.

```
def leerArchivo(self):
    self.f = open("archivo.txt", "r")
    print(self.f.read())
```

Figura 4.9: Lectura de archivo txt

4.2 DESPLIEGUE

Al iniciar se encuentra la interfaz con cuatro pestañas diferentes como se muestra en la figura 4.11, la primera pestaña destinada a realizar la predicción con un algoritmo comparativo llamado Método MLP: Multi-Layered Perceptron. Esta interfaz permite introducir el csv para procesar los datos, generar los datos muestra y reiniciar el proceso. Es importante mencionar que el diseño de la pestaña "Supervisadoz .Agrupamiento.es el mismo, la única diferencia radica en el algoritmo que se utiliza para procesar los datos.

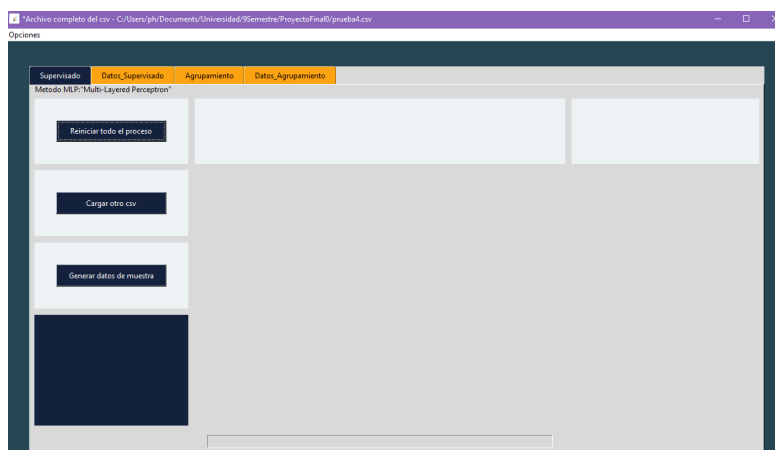


Figura 4.10: Metodo MLP

En cambio, en la pestaña llamada "DatosSupervisados" se pueden observar los valores del csv como se muestra en la figura. Además, la pestaña

4.2. DESPLIEGUE

"DatosAgrupamiento" tiene la misma función que esta, solo que es capaz de leer los datos introducidos en la pestaña ".Agrupamiento".

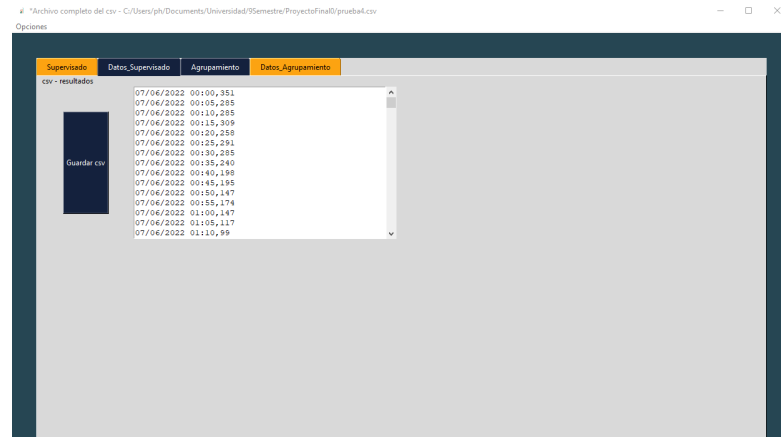


Figura 4.11: Lectura csv

Validación

El sistema pasó por un proceso de validación, los enunciados seleccionados son los siguientes

- Las instrucciones para el uso del sistema son adecuadas
- La secuencia de uso para las opciones del sistema es simple y ordenada.
- Las ventanas emergentes de graficos muestran los datos claramente.
- Los mensajes de error son entendibles y oportunos.
- La conexión y ejecución del archivo de datos es rápido y simple.
- Las alertas enviadas por el sistema son oportunas.
- El sistema resuelve el problema de forma precisa a la realidad.
- Las gráficas son adecuadas para el entendimiento de los datos
- Las opciones del sistema abarcan un uso justo y completo.
- Los resultados cuentan con una fácil interpretación.
- La carga del archivo de datos se realizo sin ningún problema.
- Las ventanas emergentes de graficos se generan adecuadamente
- El sistema inicia sin ningún problema
- Al cerrar el sistema no hay ningún inconveniente
- La velocidad de procesamiento de los datos no demora más de 10 minutos

Estos enunciados serán evaluados de acuerdo a los siguientes valores

- Muy en desacuerdo (Valor:1)
- En desacuerdo (Valor:2)
- Indiferente (Valor:3)

4.2. DESPLIEGUE

- De acuerdo (Valor:4)
- Muy de acuerdo (Valor:5)

Después de realizarse la validación se obtuvieron los resultados que se muestran en la figura 4.12, en la última columna se realiza una sumatoria de todos los valores.

A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R
Marca temporal	¿Cuál es tu nombre?	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	Valor por sujeto
8/12/2022 19:47:25	Brenda	5	5	5	5	5	5	5	5	5	5	5	5	4	5	5	74
8/12/2022 20:02:15	Orlando De La Cruz Her	5	5	4	4	5	5	5	5	5	5	5	5	5	5	5	73
8/12/2022 20:04:19	Atenea Buenfil	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	75
8/12/2022 20:08:38	Elena Jiménez	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	75
8/12/2022 20:09:53	Rebecca Hernández	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	75
8/12/2022 20:11:24	Alicia De la Torre	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	75
8/12/2022 20:13:12	Keyla Ramírez	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	75
8/12/2022 20:13:55	Shany Espinoza	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	75
8/12/2022 20:15:21	Alexa Márquez	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	75
8/12/2022 20:23:10	Paola Santillan Viveros	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	75
8/12/2022 20:25:27	Vanessa	4	5	5	4	4	4	4	4	5	4	5	5	4	5	5	67
8/12/2022 20:29:55	Xcaret	5	5	5	4	4	4	4	5	5	4	5	4	5	5	5	69
8/12/2022 20:31:46	Schecel Rosales	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	75
8/12/2022 20:34:56	Johan López del Toro	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	75
8/12/2022 20:42:11	Emiliano Candelas	5	5	5	3	5	5	4	3	5	5	5	5	5	5	5	70
8/12/2022 20:43:33	GUSTAVO TADEO PARC	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	75
8/12/2022 20:47:13	Sergio González	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	75
8/12/2022 20:57:18	Paola Alejandra Macías	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	75
8/12/2022 21:04:41	Paty	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	75

Figura 4.12: Resultados

Posteriormente, con ayuda de la fórmula CONTAR.SI() se obtuvo la frecuencia por elemento en cada afirmación, los resultados fueron los mostrados en la figura 4.13.

Elemento	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
Muy de acuerdo	19	19	19	15	18	18	17	17	20	18	19	18	17	20	20
De acuerdo	1	1	1	4	2	2	3	2	0	2	1	2	3	0	0
Indiferente	0	0	0	1	0	0	0	1	0	0	0	0	0	0	0
En desacuerdo	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Muy en desacuerdo	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0

Figura 4.13: Frecuencia por elemento

Para obtener el valor absoluto en la escala de Likert se realizó una sumatoria del total de valores frecuentes por elemento, como se visualiza en la figura 4.14.

Elemento	Total
Muy de acuerdo	274
De acuerdo	24
Indiferente	2
En desacuerdo	0
Muy en desacuerdo	0
Total	300

Figura 4.14: Composicion total de elementos

Posteriormente estos datos se dividen entre el total para encontrar el valor porcentual, como se muestra en la figura 4.15.

Elemento	Valor porcentual
Muy de acuerdo	91,33%
De acuerdo	8,00%
Indiferente	0,67%
En desacuerdo	0,00%
Muy en desacuerdo	0,00%
Total	100,00%

Figura 4.15: Composicion total de elementos, valor porcentual

Finalmente, se obtuvo un grafico con está informacion. De acuerdo con la escala de liker la validacion presenta satisfacci3n, ya que la grafica que se muestra en la figura 4.16 presenta una satisfaccion mayor a 81

4.2. DESPLIEGUE

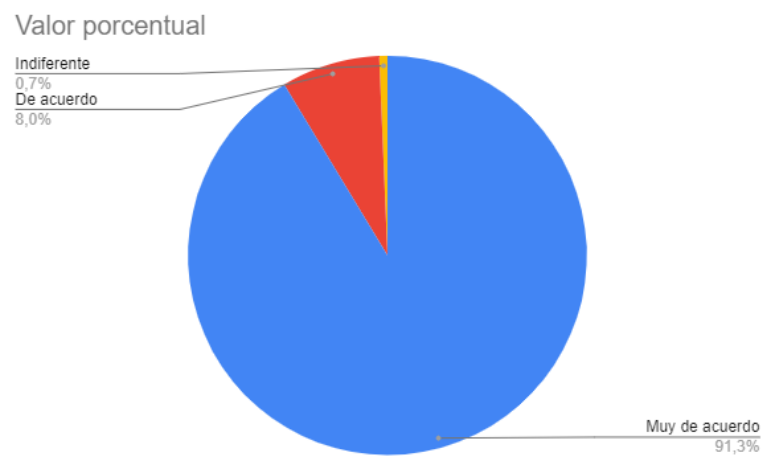


Figura 4.16: Grafico, escala de Liker

Conclusiones

Es indiscutible que la tecnología avanza de forma acelerada, aportando mejores herramientas para el desarrollo del ser humano. Sin embargo, así como trae beneficios y agiliza procesos, también supone una continua actualización para los desarrolladores, ya que las tecnologías antiguas se inhabilitan rápidamente. Por esta razón concluimos la importancia de apegarse a funciones que no dependan del tiempo, colocándole más tiempo de función y uso.

Actualmente, se ha descubierto la importancia de los datos y sus distintos usos. Logrando que sistemas predictivos e inteligentes sean una herramienta fundamental en diferentes ramas. Estos sistemas se vuelven sumamente valiosos en cada rama en la que se desarrollen, aportando un mayor conocimiento, detectando con mayor velocidad irregularidades o indicios de estas.

En este caso se aplicó en la rama empresarial; sin embargo, las predicciones pueden aportar información relevante en otras ramas como la medicina, música, economía, sociedad, etc.

Trabajo futuro

Posteriormente, el sistema será capaz de ciclar las clases, obteniendo la predicción en cada agrupación de los datos seleccionados, de esta manera se obtendrán las predicciones. Además, se conectará directamente a la base de datos, por lo que será capaz de predecir en tiempo real con los datos más actualizados.

De esta manera, el usuario del sistema tendrá la capacidad de seleccionar que datos se utilizarán para las próximas predicciones. Una vez que el sistema obtenga y registre los datos en tiempo real y realice la comparativa con las predicciones realizadas, el sistema emitirá una alarma si se detectan datos más bajos a la predicción, aportando un mayor conocimiento en los datos, clientes y mejorando el servicio de la compañía.

Referencias

- [1] Rafael Lapiedra Alcamí y col. «Introducción a la gestión de sistemas de información en la empresa». En: *Tema 2. Aspectos fundamentales de los sistemas de información en la empresa: 2.1. Concepto de sistema de información*. 1.^a ed. Universitat Jaume I, Servei de Comunicació i Publicacions, 2011, pág. 14.
- [2] RICART J. E. ANDREU R. y J. VALOR. *Estrategia y Sistemas de Información*. Madrid: Mc Graw-Hill, 1991.
- [3] Apache. 2017. URL: <https://netbeans.apache.org/>.
- [4] P.J. Brocwell y R.A. Davis. *Introduction on Time Series and Forecasting*. 2.^a ed. Springer, 2002.
- [5] Mario Bunge. «Ontología 2: Un mundo de sistemas: Volumen IV Tratado de Filosofía». En: *Tratado de Filosofía*. 1.^a ed. Vol. 4. Gedisa Mexicana, 2017, pág. 25.
- [6] F. Chia JoiLin. *Knowledge kanban system for virtual research and development. Robotic and Computer-Integrated Manufacturing*. 2012, págs. 119-134.
- [7] Alistair. Cockburn. *Crystal clear a human-powered methodology for small teams*. 1998-2004. URL: https://www.researchgate.net/publication/234820806_Crystal_clear_a_human-powered_methodology_for_small_teams.
- [8] WARD J. EDWARDS C. y A BYTHEWAY. *Fundamentos de Sistemas de Información*. Madrid: Mc Graw-Hill, 1998.
- [9] D. Garcia Morate. *Manual de Weka. Doctorado Formación en la Sociedad del Conocimiento*. URL: <https://knowledgesociety.usal.es/sites/default/files/MANUAL%20WEKA.pdf>.

REFERENCIAS

- [10] GITHUB. URL: https://github.com/PipedreamHQ/pipedream/tree/master/components/github?gclid=Cj0KCQjwxveXBhDDARIsAI0Q0x0C0tC1S_9XWlmY9fjdLsFmfUk0cSnIYMjMyslBdZRVfjVqlT5jJR8aAty3EALw_wcB#github-api-integration-platform.
- [11] Mark. Gregory. *Integrating project management and systems analysis*. <https://www.research.com/publication/10.25080/Majora-92bf1922-00a>. 2005.
- [12] José Guerra, Gustavo Sánchez y Belkis Reyes. *Modelos de series de tiempo para predecir la inflación en Venezuela*. Banco Central de Venezuela, 2001.
- [13] M. J. Ramírez-Quintana J. Hernández-Orallo y C. Ferri. *Introducción a la Minería de Datos*. Prentice Hall / Addison-Wesley, 2004.
- [14] Wes McKinney. «Data Structures for Statistical Computing in Python». En: *Proceedings of the 9th Python in Science Conference*. Ed. por Stéfan van der Walt y Jarrod Millman. 2010, págs. 56-61. DOI: 10.25080/Majora-92bf1922-00a.
- [15] Microsoft. *about Microsoft*. Ver. latest. 2022. URL: <https://www.microsoft.com/es-mx/microsoft-365/microsoft-office>.
- [16] J. Clinton (SPSS) P. Chapman (NCR) y col. *CRISP-DM 1.0 Step-by-step data mining guide*. U.S.A, 1999-2000, pág. 3.
- [17] J. Clinton (SPSS) P. Chapman (NCR) y col. *CRISP-DM 1.0 Step-by-step data mining guide*. U.S.A, 1999-2000, págs. 9-15.
- [18] Roger S. Pressman. *Ingeniería de software: un enfoque práctico*. México: McGraw-Hill, 2010, pág. 1.
- [19] Hernández-Arteaga PTimarán-Pereira (S. R.), T. Khabaza (SPSS) Caicedo-Zambrano (S. J.) y J. C. Hidalgo-Troya Alvarado Pérez. *El proceso de descubrimiento de conocimiento en bases de datos. En Descubrimiento de patrones de desempeño académico con árboles de decisión en las competencias genéricas de la formación profesional*. Colombia: Ediciones Universidad Cooperativa de Colombia, 2016, págs. 63-86.
- [20] Jose Ramon Alvarez Rodriguez. «Gestion de proyectos informaticos/ Information Project Management: Metodos, herramientas y casos/ Methods, Tools and Cases». En: *1. Qué es un proyecto*. 1.^a ed. Editorial UOC, 2007, págs. 32-34.

- [21] Ines F.Vega Lopez S. L. Lara Devora y Danuel Ernesto Lopez Barron. *Predicción con series de tiempo y regresión*. 2013. URL: <https://doi.org/10.15765/pnrm.v2i4.262>.
- [22] Ines F.Vega Lopez S. L. Lara Devora y Danuel Ernesto Lopez Barron. *Una Evaluación Experimental de Técnicas de Caracterización de Series de Tiempo Aplicadas a Electrocardiogramas*. 2010. URL: https://www.researchgate.net/publication/322701951_Una_Evaluacion_Experimental_de_Tecnicas_de_Caracterizacion_de_Series_de_Tiempo_Aplicadas_a_Electrocardiogramas.
- [23] Marvin G. Soto. «La metodología Kanban» URL: <https://marvin-soto.medium.com/la-metodolog%7B%5C'i%7Da-kanban-6ab002502831>.
- [24] The pandas development team. *pandas-dev/pandas: Pandas*. Ver. latest. Feb. de 2020. DOI: 10.5281/zenodo.3509134. URL: <https://doi.org/10.5281/zenodo.3509134>.
- [25] IBM Developer Works. 2012. URL: <https://www.ibm.com/developerworks/ssa/industry/library/ba-predictiveanalytics1/#ibm-pcon>.