



**ESCUELA POLITÉCNICA
SUPERIOR DE CÓRDOBA**
Universidad de Córdoba



PETICIÓN DE TEMA DE TRABAJO FIN DE GRADO

Grado en Ingeniería Informática

Tutorial de minería de datos mediante una aplicación web

Autor

Jaime Lorenzo Sánchez

Director(es)

Profesor Nicolas Emilio Garcia Pedrajas

Marzo, 2023



UNIVERSIDAD D CÓRDOBA



Índice general

Introducción	V
Objetivos	VII
Antecedentes	IX
Descripción del proyecto	XI
0.1. Partes del proyecto	XI
0.2. Fases del proyecto	XII
0.2.1. Aprendizaje y familiarización de las tecnologías a emplear en el proyecto	XII
0.2.2. Investigación y planteamiento	XIII
0.2.3. Desarrollo e implementación de la herramienta	XIII
0.2.4. Pruebas	XIII
0.2.5. Documentación	XIII
Distribución temporal del proyecto	XV
Recursos	XVII
0.3. Recursos humanos	XVII
0.4. Recursos Hardware	XVII
0.5. Recursos software	XVII
Bibliografía	XIX

Introducción

Para toda empresa, es imprescindible comprender los comportamientos de los consumidores y los movimientos del mercado, con el objetivo de adaptarse rápidamente y mejorar sus ofertas de productos o servicios.

Por este motivo, se ha producido un auge de la necesidad de las empresas de contratar a ingenieros informáticos con conocimientos tanto de minería de datos como de realización de programas de minería de datos.

En internet se puede encontrar numerosa información de minería de datos, desde documentos hasta vídeos. Sin embargo, si queremos diseñar nuestros propios programas de minería de datos es necesario disponer de un conocimiento sobre lenguajes de programación, métodos de clasificación de minería de datos, etc.

De este modo, cada vez que queramos realizar un programa de minería de datos o detectemos un error en la realización de dicho programa, es necesario realizar una búsqueda en la web hasta dar con aquellos vídeos o documentos que aporten alguna posible solución al error detectado. Dependiendo del nivel de conocimiento de programación del usuario, el tiempo necesario para la solución de dicho error puede volverse bastante largo.

Por tanto, y sobre todo para los usuarios que se inician en este campo de la ingeniería informática, sería de mucha utilidad contar con una herramienta de ayuda de minería de datos, que explique los conocimientos y procedimientos necesarios para analizar los datos contenidos en una base de datos.

Por todo lo anterior, consideramos que el desarrollo del proyecto que proponemos tiene un gran interés.

Para ello, se va a diseñar mediante una aplicación web formada por vídeos explicativos y ejercicios a resolver, un tutorial de minería de datos que abarque los pasos clave en el proceso de minería de datos.

Objetivos

El proyecto a desarrollar trata de la creación de un tutorial de minería de datos para entorno web.

El objetivo principal de este proyecto es diseñar e implementar una herramienta de ayuda que permita analizar, comprender y evaluar un conjunto de datos a nivel de código.

Para ello, se implementará una aplicación en un entorno web que cumpla lo siguiente:

1. Diseño e implementación de un tutorial para el aprendizaje de diversos métodos para recolectar y analizar información a partir de conjuntos de datos (archivos csv, arff, etc).
2. Diseño e implementación de un tutorial para el aprendizaje de diversos métodos para limpiar y transformar los datos al formato deseado por el usuario (normalización de los datos, tratamiento de datos faltantes, etc).
3. Diseño e implementación de un tutorial para el aprendizaje de diversas técnicas de modelado de minería de datos (árboles de decisión, KNN, etc).
4. Diseño e implementación de un tutorial para el aprendizaje de diversos métodos para evaluar los resultados obtenidos (medidas estadísticas, matrices de confusión, precisión, etc).
5. Diseño e implementación de distintos tipos de ejercicios para afinar los conocimientos obtenidos. Dichos ejercicios abarcarán desde completar secciones de código hasta realizar un programa desde cero.
6. La aplicación web podrá ser utilizada por cualquier tipo de usuario ya que estará realizada de tal forma que su uso será sencillo para que el usuario pueda trabajar sin dificultades con ella.

Antecedentes

Si realizamos una búsqueda de tutorial de minería de datos en internet, encontramos desde documentos pdf hasta vídeos, principalmente en youtube.

Sin embargo, es necesario emplear un tiempo considerable en la búsqueda de la información, siendo aún más notorio para aquellos usuarios que no dispongan de ningún conocimiento de programación python, pues deberán de gastar un mayor tiempo en la búsqueda de información sobre este lenguaje de programación.

En este proyecto vamos a realizar una herramienta de ayuda caracterizada por el uso de explicaciones prácticas mediante vídeos, además de aportar diversos ejercicios para que el usuario practique y afine los conocimientos obtenidos.

De este modo, la herramienta a desarrollar será gratuita y accesible para todo el mundo. Además, al disponer de dicha información en una misma aplicación, será de gran utilidad y comodidad para el usuario.

Descripción del proyecto

En este apartado veremos más en profundidad todo lo que se pretende hacer en el proyecto.

La herramienta a desarrollar estará destinada para cualquier tipo de usuario y sin ningún requisito previo de conocimiento (se aportará tanto explicaciones como código de base para la realización de las diversas explicaciones y ejercicios de práctica propuestos).

El proyecto se dividirá en las siguientes secciones:

- **Lenguajes:** Se aportará información de los distintos lenguajes empleados para el desarrollo de la aplicación, así como de enlaces de interés para encontrar información sobre cada lenguaje empleado.
- **Obtención de Datos:** Se aportarán tanto vídeos explicativos como ejemplos de código de diversas formas de obtener distintos conjuntos de datos, así como de visualizar la información de los datos.
- **Limpieza de Datos:** Se aportarán tanto vídeos explicativos como ejemplos de código de diversas formas de detectar y resolver anomalías en conjuntos de datos, para poder explotarlas posteriormente.
- **Transformación de Datos:** Se aportarán tanto vídeos explicativos como ejemplos de código de diversas formas para transformar los datos obtenidos, como por ejemplo, cambiar cadenas de datos en valores numéricos, normalizar los datos, etc.
- **Técnicas de modelado:** Se aportarán tanto vídeos explicativos como ejemplos de código de diversas técnicas de minería de datos.
- **Evaluación de Resultados:** Se aportarán tanto vídeos explicativos como ejemplos de datos de diversas formas de evaluación de los resultados obtenidos.

0.1. Partes del proyecto

Para especificar las distintas partes del proyecto, se ha diseñado una WBS o Estructura de Descomposición del Trabajo [4]:

A continuación, se van a describir cada uno de los bloques de trabajo que descomponen este proyecto:

- **Migración de datos:** Consistirá fundamentalmente en explicar distintos métodos de exportar los conjuntos de datos a un formato que la aplicación pueda emplear, de manera que no se pierda ninguna información.

- Limpieza de datos: Consistirá fundamentalmente en explicar distintos métodos de resolución de anomalías en conjuntos de datos, para poder explotarlos posteriormente.
- Transformación de datos: Consistirá fundamentalmente en explicar distintos métodos para consolidar los datos para que el proceso de minería sea más eficiente y los patrones sean más fáciles de entender.
- Técnicas de Modelado: Consistirá fundamentalmente explicar distintas técnicas encaminadas a identificar patrones implícitos dentro de grandes conjuntos de datos, con el fin de entender sus mecanismos de comportamiento, interrelación y su potencial evolución futura.
- Evaluación de Resultados: Consistirá fundamentalmente en explicar distintos métodos para la evaluación del grado al que el modelo responde a los objetivos buscados.
- Documentación: En este bloque se procede a realizar la memoria del proyecto.

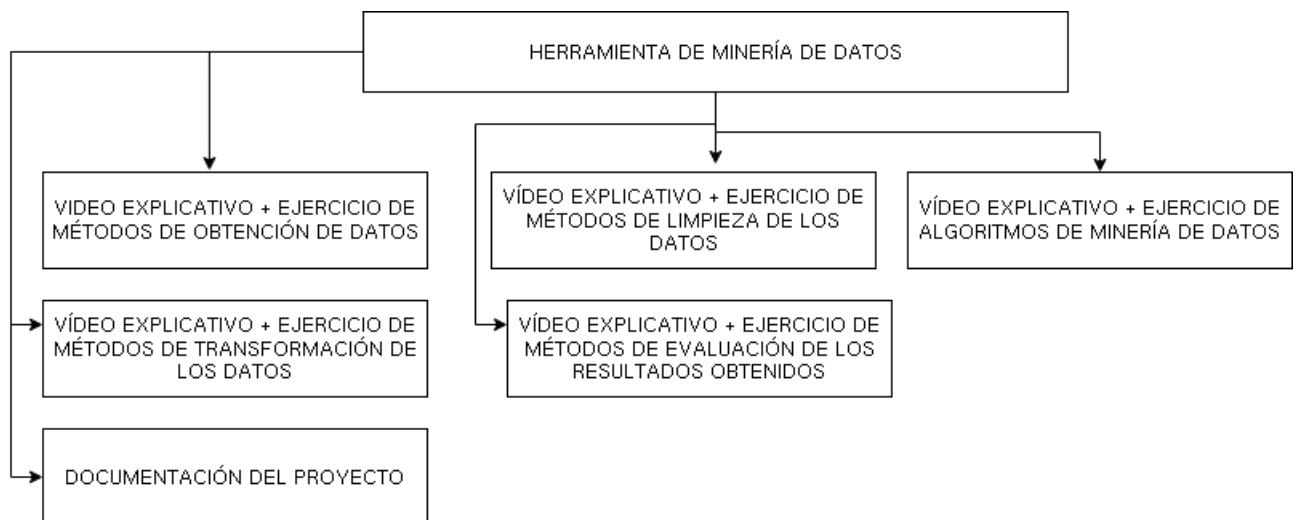


Figura 1: Estructura de Descomposición del Trabajo

0.2. Fases del proyecto

Las fases del proyecto son una división del proyecto orientada a su planificación en el tiempo. Comprenden todas las tareas necesarias para adquirir los conocimientos adecuados para abordar el diseño, desarrollo y ejecución del proyecto.

Las fases del proyecto son las siguientes:

0.2.1. Aprendizaje y familiarización de las tecnologías a emplear en el proyecto

Se utilizarán diversas tecnologías nuevas para el estudiante, por tanto será necesario primero realizar un aprendizaje para familiarizarse con dichas tecnologías.

En este caso, será necesario HTML [5], JavaScript [6], CSS [7], Python [8], Django [9], Bash [10] y PyScript [11].

0.2.2. Investigación y planteamiento

Primer planteamiento de la web aplicando nuevas técnicas y posibilidades actuales.

0.2.3. Desarrollo e implementación de la herramienta

Es la parte más extensa del proyecto, pues abarca tanto el diseño e implementación de las funcionalidades de la herramienta.

0.2.4. Pruebas

Se realizarán todas las pruebas necesarias para garantizar un correcto funcionamiento de la web.

0.2.5. Documentación

Esta fase se desarrollará durante todo el proyecto para la realización de la memoria del proyecto.

Distribución temporal del proyecto

El TFG consta de 12 créditos ECTS y una asignación de 25 horas por crédito en el plan de estudios, por tanto equivale a 300 horas de trabajo. A continuación, se muestra la distribución aproximada y temporal de las principales fases del proyecto.

Tareas	Tiempo (h)
1. Aprendizaje y familiarización de las tecnologías a emplear en el proyecto	35
+ HTML	5
+ JavaScript	5
+ CSS	7
+ Python y PyScript	5
+ Django	8
+ Bash	5
2. Investigación y planteamiento	35
3. Desarrollo e implementación de la página web	150
4. Pruebas	40
5. Documentación	40
Total	300

Cuadro 1: DistribucionTemporalProyecto

Recursos

0.3. Recursos humanos

Esta sección abarca las personas implicadas en la realización del proyecto.

- Directores: Encargados de tutorizar el proyecto, llevar a cabo un seguimiento del mismo, supervisión y orientación en todas las fases del mismo.
 - Profesor Nicolás Emilio García Pedrajas.
- Diseño y desarrollo: Encargado de diseñar y desarrollar la totalidad del sistema, así como cumplir los objetivos de diseño y desarrollo.
 - Jaime Lorenzo Sánchez.

0.4. Recursos Hardware

Se dispondrá de un portátil HP J2I2PN72 con las siguientes especificaciones:

- Procesador: Intel(R) Core(TM) i7-8565U CPU @ 1.80GHz 1.99 GHz.
- RAM: 8,00 GB (7,89 GB usable).

0.5. Recursos software

Los recursos software a utilizar para la realización del proyecto son los siguientes:

- Sistema Operativo: Windows 10 64 Bits
- Lenguajes de programación:
 - HTML.
 - JavaScript.
 - CSS.
 - Python.
- Elaboración de la documentación: Overleaf.
- Editores de código: Windows Visual Studio

Bibliografía

- [1] Orange Data Mining, web consultada el 13-02-2023:
<https://orangedatamining.com/>
- [2] KNIME, web consultada el 12-02-2023:
<https://www.knime.com/>
- [3] SAS, web consultada el 14-02-2023:
https://www.sas.com/es_es/software/enterprise-miner.html
- [4] Estructura WBS, web consultada el 15-02-2023:
<https://biblus.accasoftware.com/es/wbs-workbreakdownstructure-que-es-y-como-se-usa>
- [5] HTML, web consultada el 15-02-2023:
<https://www.w3schools.com/html/>
- [6] JavaScript, web consultada el 15-02-2023:
<https://www.w3schools.com/js/default.asp>
- [7] CSS, web consultada el 15-02-2023:
<https://www.w3schools.com/css/default.asp>
- [8] Python, web consultada el 15-02-2023:
<https://www.w3schools.com/python/default.asp>
- [9] Django, web consultada el 15-02-2023:
<https://docs.djangoproject.com/es/4.1/>
- [10] Bash, web consultada el 15-02-2023:
<https://www.hostinger.es/tutoriales/bash-script-linux>
- [11] PyScript, web consultada el 15-02-2023:
<https://docs.pyscript.net/latest/>