

UNIVERSIDAD DE GUADALAJARA

CENTRO UNIVERSITARIO DE CIENCIAS ECONÓMICO ADMINISTRATIVAS

PROGRAMACIÓN II

MAESTRÍA EN CIENCIA DE LOS DATOS



Avance I Challenge Tesis

P R E S E N T A

José Luis Medrano Medrano

14 de febrero del 2024

jose.medrano7407@alumnos.udg.mx

3314246139

Índice

1. Objetivos	3
1.1. General	3
1.2. Específicos	3
2. Línea Generadora del conocimiento.....	3
3. Alcance.....	4
3.1. Diseño de la metodología.....	4
3.2. Sujeto de estudio	4
3.3. Datos	4
3.4. Diseño	4
3.5. Momento	5
3.6. Herramientas	5
4. Fuente de información.....	6
5. Directores de Tesis	7

1. Objetivos

1.1.General

Proponer un modelo predictivo que determine la probabilidad de rotación temprana de un empleado de producción en empresas manufactureras de Tala, Jalisco.

1.2.Específicos

- Determinar el Perfil del empleado de producción de mayor rotación en empresas manufactureras de Tala, Jalisco.
- Identificar las principales características de los empleados de producción que tienen una mayor rotación en empresas manufactureros de Tala Jalisco.
- Determinar la herramienta de Machine Learning más adecuada para desarrollar el proyecto.

2. Línea Generadora del conocimiento

LGAC2. Smart Data: procesos de gestión y análisis de datos cuantitativos y/o cualitativos para el tratamiento inteligente de información.

A través de esta LGAC se pretende aplicar técnicas de aprendizaje automatizado (del Inglés, Machine Learning) en la solución de problemas en diferentes ámbitos. Para ello es necesario el uso de técnicas para el tratamiento de información, como son la minería de datos, aprendizaje supervisado y no supervisado, y en general, la analítica de datos. Todo lo anterior con el objetivo de apoyar en la solución de problemas de diferente índole social, en proyectos de ciudades sustentables, y en lo general como apoyo a problemas prioritarios nacionales que requieran el tratamiento inteligente de información.

3. Alcance.

3.1.Diseño de la metodología

Esta investigación tiene como finalidad proponer una herramienta que permita predecir la probabilidad de rotación en empresas manufactureras de Tala, Jalisco, por tal motivo, es adecuado trabajar con una perspectiva descriptiva. De acuerdo con Sampieri, et al (2014), dicho enfoque no busca responder a preguntas del porqué se relacionan variables pero sí a recolectar información sobre ellas.

3.2.Sujeto de estudio

El sujeto de estudio serían los postulantes a una empresa manufacturera de Tala Jalisco, rango de edad de 18 a 55 años, género indistinto.

3.3.Datos

Los datos de los que se haría uso para generar el modelo serían de una empresa manufacturera de Tala Jalisco, aplicando todos los protocolos requeridos de confidencialidad para la protección de la información y sin violar la confidencialidad de las personas a las que les pertenece la información.

3.4.Diseño

Este diseño se fundamenta en el estudio de caso que de acuerdo con Sampieri et al (2014), hace uso de distintos procesos de investigación para al profundizar en una unidad holística y responden al planteamiento del problema, probar hipótesis o incluso desarrollar teorías.

Este diseño para este trabajo en particular sería no experimental que como menciona Sampieri, et al (2014) es aquel en el que no se manipulan intencionalmente

variables, encaja con la idea de que los datos que se utilizarán fueron previamente recolectados y en su contexto natural.

Para este trabajo se utilizará el proceso tradicional de ciencia de datos (Tibco, 2023).

1. Recolectar los datos, se recopilan los datos a usarse.
2. Limpieza y preparación de los datos, se preparan todos los datos en un formato utilizable.
3. Análisis exploratorio, se organizan los datos y visualizan.
4. Identificación de variables, se determinan las variables.
5. Procesamiento de datos, se procesan estas variables se prepara la información para el modelado.
6. Preparación del modelo, se desarrolla y ejecuta el modelo.
7. Validación del modelo, se validan la eficiencia del modelo.
8. Resultados, se presentan resultados.

3.5.Momento

Para este estudio se planea utilizar los datos recopilados durante los últimos 5 años por Cool Pak AG, es decir, los datos ya fueron recolectados por lo cuál, de acuerdo con Sampieri et al (2014) se trata de una investigación transeccional o transversal ya que los datos son recolectados en un momento único.

3.6.Herramientas

La herramienta para ejecutar este proyecto sería Python lenguaje de programación que es capaz de facilitar el alcanzar los objetivos planteados.

Se desarrollarán modelos de Machine Learning y Deep Learning en Python 3.8, específicamente para la construcción de las máquinas de soporte vectorial, regresión logística y árboles de decisión.

4. Fuente de información

Los datos por utilizarse en esta investigación serán datos de fuentes secundarias. Antes de la limpieza de los mismo, los datos guardan las siguientes características:

Concepto	Cantidad
Data frames	2
Registros data frame 1	288
Registros data frame 2	514
Variables data frame 1	16
Variables data frame 2	14

5. Directores de Tesis



Alex Guillen Bonilla

Firma

Silvia Ramos Cabral

Firma

