**Asignatura: Machine Learning**

**Sección: TIEL26/D-ELEC-N8-PIEI-C107/D Valdivia ELE**

**Nombre del docente:**

**Nombre de los integrantes del grupo:** Margarita Vilches, Sebastian Bascur

**Fecha de entrega**

Informe Nombre del trabajo

Nombre de la unidad de aprendizaje

**Contenido**

[I. Introducción 3](#_Toc471831140)

[II. Objetivo 3](#_Toc471831141)

[III. Desarrollo 3](#_Toc471831142)

[IV. Conclusiones 3](#_Toc471831143)

[V. Referencias bibliográficas 3](#_Toc471831144)

Una vez finalizado el informe, actualiza esta tabla de contenidos, ubicando el mouse sobre ella, y pulsando el botón derecho del mouse. Actualízala en su totalidad y déjela en una página independiente de la Introducción. Finalmente elimina este texto.

**NOTA:** Cada docente DEBE modificar, ajustar o completar los apartados del informe según las características o tipo de informe.

1. Introducción

El uso de modelos de Machine Learning se ha expandido enormemente en la actualidad para resolver problemas en diversas áreas, desde predicción de precios hasta clasificación y detección de fraudes. En este trabajo, abordamos varios problemas utilizando técnicas de aprendizaje supervisado y no supervisado. Exploramos la predicción del precio de una casa en California basándonos en sus características elementales, la clasificación de correos electrónicos en spam o no spam mediante modelos probabilísticos, la recomendación de películas según preferencias de usuarios, y la detección de transacciones fraudulentas empleando máquinas de soporte vectorial (SVM). Además, utilizamos un perceptrón para clasificar flores del dataset Iris. Cada problema se resuelve con métodos estadísticos y métricas de evaluación precisas, buscando modelos efectivos y explicaciones claras de los resultados obtenidos.

1. Objetivo

**El propósito de este informe es evaluar la efectividad de los diferentes modelos de machine learning y las métricas utilizadas para medir su rendimiento.**

1. Desarrollo

**Problema 1: Predicción de precios de viviendas en California**  
Se utiliza un modelo de regresión para predecir el precio de las viviendas en función de sus características clave.  
Desarrollo en el archivo: p1\_housing.ipynb

* **a) Características influyentes en el valor de la vivienda:**  
  La cercanía al océano se identifica como una de las principales características que afectan el valor de la vivienda.

Texto

Descripción generada automáticamente

* **b) Precisión del modelo generado:**  
  El modelo tiene una precisión del 63.69%.

Texto

Descripción generada automáticamente

* **c) Evaluación del modelo con métricas de rendimiento:**
  + **Error Absoluto Medio (EAM):** 51,054.33
  + **Error Cuadrático Medio (ECM):** 4,875,794,585.24
  + **Raíz del Error Cuadrático Medio (RECM):** 69,826.89
  + **Coeficiente de determinación (R²):** 40.45%

**Problema 2: Clasificación de correos electrónicos (Spam vs No Spam)**  
Se clasifica el contenido de los correos electrónicos para determinar si son spam o no.  
Desarrollo en el archivo: p2\_spam.ipynb

* **a) Justificación del modelo:**

La regresión logística es la mejor opción a utilizar por su simplicidad y eficiencia

* **b) Características que afectan la clasificación como spam:**

Las palabras clave como "you", "free", "money", y "credit", junto con símbolos especiales ("!", "(", "$", "#") y enlaces excesivos, son los factores principales para identificar correos como spam.

* **c) Métricas utilizadas para evaluar el modelo:**
  + **Tasa de error**
  + **Exactitud**
  + **Matriz de confusión**
  + **Precisión**
  + **F1-Score**: Esta métrica es la más adecuada debido a su equilibrio entre precisión y recall, lo que resulta crucial para evitar falsos positivos (marcar correos legítimos como spam) y falsos negativos (pasar por alto correos spam reales).

**Problema 3: Recomendación de películas**  
Se desarrolla un sistema de recomendación de películas basado en la similitud entre usuarios y productos.  
Desarrollo en el archivo: p3\_movies.ipynb

* **a) Recomendación de una película:**  
  película de terror
* **b) Recomendación basada en la última película vista (Toy Story):**  
  La recomendación se ajusta en función de la película más recientemente vista por el usuario.
* Texto

  Descripción generada automáticamente

**Problema 4: Detección de fraudes bancarios**  
Se implementa un modelo para detectar transacciones bancarias fraudulentas basadas en patrones históricos.  
Desarrollo en el archivo: p4\_bank.ipynb

* **a) Kernel adecuado para la SVM en este problema:**

El **kernel RBF** es más adecuado ya que captura bien la no linealidad, común en problemas de fraude donde los patrones no son simples.

* **b) Comparación de métricas (Precisión, F1-Score, Recall, Exactitud):**

El **F1-Score** es la métrica más relevante en problemas de fraude, ya que penaliza tanto falsos positivos como falsos negativos, lo que es crítico en la detección de fraudes.

**\*\*EXTRA\*\* PERCEPTRON (5 DÉCIMAS)**

Utilizar un perceptrón para clasificar dos tipos de flores usando el conjunto de datos Iris, simplificado para dos clases y dos características. (Dataset: se encuentra en librería de scikit-learn).

Interfaz de usuario gráfica, Texto, Aplicación, Correo electrónico

Descripción generada automáticamente

1. Conclusiones

A lo largo de este análisis, hemos comprobado cómo distintas técnicas de Machine Learning, como la regresión lineal, Naive Bayes, SVM y sistemas de recomendación, son fundamentales para resolver problemas que van desde la predicción de precios hasta la detección de fraudes. La elección adecuada del modelo y su evaluación mediante métricas como el error absoluto medio, el F1-Score o el coeficiente R² nos permiten obtener una visión clara del desempeño de cada algoritmo en contextos específicos. El uso de datos reales para entrenar estos modelos garantiza soluciones robustas que pueden ser implementadas en escenarios del mundo real, desde el sector inmobiliario hasta el comercio electrónico y la seguridad financiera.