**Tarea 3 Algoritmos de Aprendizaje Supervisado**

Tutor

Universidad Nacional Abierta y a Distancia-UNAD

Escuela de Ciencias Básicas, Tecnología e Ingeniería ECBTI, Ingeniería de Sistemas

Curso

de abril de 2024

**Introducción**

En el vasto campo del aprendizaje automático, los modelos de aprendizaje supervisado son esenciales para predecir y clasificar datos basados en ejemplos etiquetados. En esta aventura de análisis y modelado, utilizaremos el poderoso lenguaje Python y la plataforma interactiva Jupyter Notebooks. Este entorno nos permite explorar, preprocesar y entrenar modelos de Regresión Lineal, Regresión Logística y Árboles de Decisión con datasets reales descargados desde Kaggle.

Además nos sumergiremos en tres datasets distintos, cada uno con su propio objetivo y desafíos. El primero, el "Vehicle dataset", nos permitirá predecir el precio de los automóviles en función de sus atributos técnicos y de mercado. Por otro lado, el "Heart Disease Cleveland UCI dataset" nos desafiará a clasificar la presencia de enfermedad cardíaca basada en datos médicos de pacientes. Y finalmente, el "Red Wine Quality dataset" nos llevará al mundo del vino, clasificando su calidad en 10 clases.

A través de pasos cuidadosamente diseñados, exploraremos los datos, limpiaremos cualquier ruido o valores faltantes, seleccionaremos las características más relevantes y entrenaremos nuestros modelos. Estaremos atentos a métricas como precisión, recall, F1-score y R cuadrado para evaluar el rendimiento de nuestros modelos en conjuntos de datos de prueba y utilizaremos gráficos para visualizar y comprender mejor los resultados obtenidos.

Al final de esta fase, no solo habremos diseñado y entrenado modelos de aprendizaje automático, sino que también habremos aprendido a interpretar y analizar los resultados obtenidos. Finalmente, compartiremos nuestro trabajo en GitHub para que otros puedan explorar y aprender de nuestras experiencias y resultados.

**Objetivos**

* Leer y analizar la guía de actividades.
* Estar atentos a la web conferencia creada por los tutores explicando es desarrollo de esta fase paso a paso.

**DESARROLLO DE LA ACTIVIDAD**

**Elaborar un cuadro sinóptico sobre los diferentes modelos de Aprendizaje Supervisado que incluya definición, casos de uso, ventajas y desventajas**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Modelo** | **Definición** | **Casos de Uso** | **Ventajas** | **Desventajas** |
| Regresión Lineal | Modelo que busca establecer una relación lineal entre las variables independientes y la variable objetivo. | Predicción de valores numéricos, como precios de casas, ingresos, etc. | Fácil interpretación, rápido entrenamiento. | Sensible a outliers, no captura relaciones no lineales. |
| Regresión Logística | Utilizado para problemas de clasificación binaria, estima la probabilidad de que una observación pertenezca a una clase. | Clasificación de correos electrónicos como spam/no spam, diagnóstico médico, etc. | Buen desempeño en problemas lineales. | No es adecuado para clasificaciones multiclase. |
| Árboles de Decisión | Modelo de representación gráfica de posibles decisiones basadas en ciertas condiciones, dividido en nodos y hojas. | Clasificación y regresión. | Fácil interpretación, manejo automático de features. | Propenso al sobreajuste si no se controla su crecimiento. |
| Support Vector Machines | Modelo que busca encontrar el hiperplano que mejor separa las clases en un espacio dimensional superior. | Clasificación y regresión. | Buen desempeño en espacios de alta dimensionalidad. | Sensible a la elección de su función de kernel. |
| K-Nearest Neighbors | Clasifica un punto basado en la clase predominante de sus 'k' vecinos más cercanos, donde 'k' es un número definido por el usuario. | Clasificación. | Fácil de entender e implementar. | Lento en la predicción con grandes conjuntos de datos. |
| Random Forest | Conjunto de árboles de decisión que trabajan de forma conjunta, el resultado se obtiene por. votación | Clasificación y regresión. | Buen desempeño en la mayoría de los casos. | Menos interpretable que un solo árbol de decisión. |
| Gradient Boosting | Construye árboles de decisión en serie, cada uno corrigiendo los errores del anterior, obteniendo un modelo más robusto. | Clasificación y regresión. | Buen desempeño en la mayoría de los casos. | Sensible a overfitting si no se ajustan los parámetros. |

**Referencias Bibliográficas:**

* Carlos Véliz. (2020). [*Aprendizaje automático. Introducción al aprendizaje profundo*](https://bibliotecavirtual.unad.edu.co/login?url=https://search.ebscohost.com/login.aspx?direct=true&db=nlebk&AN=2600876&lang=es&site=eds-live&scope=site&ebv=EB&ppid=pp_I). El Fondo Editorial de la Pontificia Universidad Católica del Perú.https://bibliotecavirtual.unad.edu.co/login?url=https://search.ebscohost.com/login.aspx?direct=true&db=nlebk&AN=2600876&lang=es&site=eds-live&scope=site&ebv=EB&ppid=pp\_I Cap 3, 4, 5 y 6
* David Julian. (2016). [*Designing Machine Learning Systems with Python*](https://bibliotecavirtual.unad.edu.co/login?url=https://search.ebscohost.com/login.aspx?direct=true&db=nlebk&AN=1218065&lang=es&site=eds-live&scope=site&ebv=EB&ppid=pp_Cover). Packt Publishing. https://bibliotecavirtual.unad.edu.co/login?url=https://search.ebscohost.com/login.aspx?direct=true&db=nlebk&AN=1218065&lang=es&site=eds-live&scope=site&ebv=EB&ppid=pp\_Cover. Cap 2
* Giuseppe Bonaccorso. (2018). [*Machine Learning Algorithms*](https://bibliotecavirtual.unad.edu.co/login?url=https://search.ebscohost.com/login.aspx?direct=true&db=nlebk&AN=1881497&lang=es&site=eds-live&scope=site&ebv=EB&ppid=pp_Cover): Popular Algorithms for Data Science and Machine Learning, 2nd Edition: Vol. 2nd ed. Packt Publishing. https://bibliotecavirtual.unad.edu.co/login?url=https://search.ebscohost.com/login.aspx?direct=true&db=nlebk&AN=1881497&lang=es&site=eds-live&scope=site&ebv=EB&ppid=pp\_Cover Cap 3, 4, 5 y 8
* Minguillón, J. Casas, J. y Minguillón, J. (2017). [*Minería de datos: modelos y algoritmos*](https://elibro-net.bibliotecavirtual.unad.edu.co/es/ereader/unad/58656). Editorial UOC. https://elibro-net.bibliotecavirtual.unad.edu.co/es/ereader/unad/58656. Cap 4, 5, 13
* Pratap Dangeti. (2017). [*Statistics for Machine Learning*](https://bibliotecavirtual.unad.edu.co/login?url=https://search.ebscohost.com/login.aspx?direct=true&db=nlebk&AN=1560931&lang=es&site=eds-live&scope=site&ebv=EB&ppid=pp_Cover) : Build Supervised, Unsupervised, and Reinforcement Learning Models Using Both Python and R. Packt Publishing. https://bibliotecavirtual.unad.edu.co/login?url=https://search.ebscohost.com/login.aspx?direct=true&db=nlebk&AN=1560931&lang=es&site=eds-live&scope=site&ebv=EB&ppid=pp\_Cover Cap 2, 3 y 4
* Romero Villafranca, R. y Zúnica Ramajo, L. (2020). [*Métodos estadísticos para ingenieros*](https://elibro-net.bibliotecavirtual.unad.edu.co/es/ereader/unad/129644). Editorial de la Universidad Politécnica de Valencia. https://elibro-net.bibliotecavirtual.unad.edu.co/es/ereader/unad/129644. Cap 12