

Reporte metodología de propuesta de proyecto:

MACHINE LEARNING SUPERVISADO:

El aprendizaje supervisado es una rama de la maquina de aprendizaje, la cual esta conformada por un conjunto de algoritmos que buscan aprender patrones de los datos sujetas a una variable objetivo, la cual puede ser continua o discreta. Hay diferentes tipos de algoritmos, a continuación se brinda una descripción de algunos de estos:

Arboles de Decisión:

El algoritmo de árbol de decisión funciona a partir de un conjunto de datos donde se generan diagramas de construcciones lógicas, estos diagramas se pueden ver como un sistema de decisiones basados en reglas lógicas, las cuales pueden servir para representar y categorizar una serie de reglas que permitan resolver no solo el problema, sino también de poder replicar en nuevos datos para generar una predicción.

Clasificación Naïve Bayes:

El algoritmo de Naive es un clasificador probabilístico fundamentado en el teorema de Bayes, adicional y con base a la información o los datos el algoritmo usa supuestos adicionales, los cuales relacionan las variables con la variable dependiente o la variable objetivo para poder estimar las probabilidades.

Regresión por Mínimos Cuadrados:

Este algoritmo de mínimos cuadrados, busca minimizar los errores al cuadrado entre la variable objetivo y una estimación de coeficientes que permitan aproximar un conjunto de variables independientes a la variable objetivo. Al optimizar los coeficientes, estos eran los que reduzcan la diferencia entre la variable objetivo y la variable estimada. La regresión por mínimos cuadrados ajusta variables continuas.

Regresión Logística:

La regresión logística a diferencia de la regresión por mínimos cuadrados busca ajustar variables categóricas, y bajo la transformación logit hace la estimación de probabilidades en las cuales puede ser clasificada una variable dentro de alguna categoría.

MACHINE LEARNING NO-SUPERVISADO:

Por otro lado machine learning no supervisado tiene como objetivo encontrar patrones en un conjunto de datos, a diferencia de machine learning supervisado, no se necesita una variable objetivo o dependiente que permita guiar al algoritmo. Estos algoritmos permiten encontrar patrones en los datos o características en las variables.

Algoritmos de Cluster:

Los algoritmos de cluster permiten encontrar grupos entre un conjunto de variables, estos grupos se forman de similitudes que encuentran los algoritmos entre las variables o conjunto de datos.

Análisis de Componentes Principales:

El análisis de componentes principales, permite encontrar las variables o conjunto de datos que permitan explicar la mayor varianza en la información.