

Proyecto: Minería de datos

Juan Pablo Echeagaray González, Emily Rebeca Méndez Cruz, Grace Aviance Silva Arostegui

Resumen—Referencia perrona de este libro [1]

Index Terms—Data Science, Machine Learning, Data Analysis

I. INTRODUCCIÓN

II. CRÉDITOS

- Juan Pablo Echeagaray González - A00830646
- Emily Rebeca Méndez Cruz
- Grace Aviance Silva Arostegui

III. MODELOS DE MACHINE LEARNING

III-A. *Árbol de decisión*

[2]

III-B. *Support Vector Machine (SVM) Grace Aviance Silva Arostegui*

Support vector machine (SVM) es un algoritmo de aprendizaje supervisado que se utiliza en muchos problemas de clasificación y regresión, e incluso para la detección de valores atípicos [1]. Este modelo de aprendizaje automático se basa en una separación de diferentes clases a través de un hiperplano en un espacio de dimensión superior [3].

Dado un conjunto de muestras (ejemplos de entrenamiento) se etiquetan clases y se entrena una SVM para construir un modelo que prediga la clase de una nueva muestra.

Para la base de datos que tenemos y el análisis que queremos llevar a cabo, en donde el enfoque es considerar en cada uno de los registros de las mujeres cual es el riesgo que tienen para desarrollar cáncer cervical. Dado que en la base de datos el rango de riesgos va de 0 a 4, tenemos así 5 clases de las cuales buscaríamos 5 hiperplanos que tengan el margen lo más ancho posible entre las clases, para así poder entregar lo mejor posible al algoritmo y hacer mejores clasificaciones futuras. Cabe mencionar que utilizamos un 80/20 de los datos para entrenamiento y prueba respectivamente.

III-C. *Red Neuronal*

Después de inspeccionar el mapa generado hemos notado que hay algunos puntos que parecen tener datos geográficos erróneos, descartar la entrega a estos clientes es algo inaceptable, así que una de las siguientes tareas en el proyecto será desarrollar un método de limpieza efectivo que ayude a mejorar la información geográfica que obtengamos de cada punto.

III-D. *Regresión Logística*

IV. RESULTADOS

V. CONCLUSIONES

V-A. *Áreas de mejora*

V-B. *Modelo seleccionado*

VI. REFLEXIONES

APÉNDICE A DATOS

Los datos usados en este proyecto pueden descargarse aquí

APÉNDICE B CÓDIGO

El código desarrollado se encuentra en el siguiente repositorio

APÉNDICE C

EVIDENCIAS DE TRABAJO EN EQUIPO

REFERENCIAS

- [1] A. Géron, *Hands-On Machine Learning with Scikit-Learn, Keras, and TensorFlow*. Culemborg, Netherlands: Van Duuren Media, 2019.
- [2] Sci-kit Learn, “sklearn.tree.DecisionTreeClassifier.” [Online]. Available: <https://scikit-learn.org/stable/modules/generated/sklearn.tree.DecisionTreeClassifier.html>
- [3] M. P. Deisenroth, A. A. Faisal, and C. S. Ong, *Mathematics for machine learning*. Cambridge University Press, 2020.