

Introducción a Inteligencia Artificial para ciencias e ingenierías

Primera Entrega

Estudiantes:

Omar Alberto Torres CC 91220873
Carlos Alfredo Pinto Hernández CC 1100953378

Profesor:

Raúl Ramos Pollan

Faculta de ingeniería
Departamento de Ingenieria de Sistemas
Ude@
2023

Desarrollo

(1) Problema predictivo:

Dadas unas características físicas, enfermedades y resultados de laboratorios de una persona se predecirá la probabilidad de presentar enfermedad cerebro vascular o stroke.

(2) Dataset a utilizar

Se utilizará el dataset de kaggle de la competición denominada "Playground Series - Season 3, Episode 2" disponible en el siguiente enlace: <https://www.kaggle.com/competitions/playground-series-s3e2/> el cual contiene más de 15.000 muestras y divididas en 12 columna incluyendo: edad, género, índice de masa corporal, promedio de nivel de glucosa, tipo de residencia, tipo de trabajo, si ha estado casado, si presenta hipertensión o enfermedad cardíaca.

Se obtienen dos archivos uno denominado "train.csv - the training dataset": donde se especifica el target "stroke" en formato binario y el otro documento "test.csv - the test dataset" con el cual se probará el modelo para predecir la probabilidad de presentar un stroke.

(3) Métricas de desempeño

Como métrica de machine learning vamos a utilizar el área bajo la curva ROC entre la probabilidad del modelo vs el resultado (stroke); esta métrica fue definida en la competencia. Como métrica de negocio se podría plantear el incremento del número de personas que ingresan a controles de riesgo cardiovascular detectados por el modelo.

(4) Criterio sobre el cual sería el desempeño deseable en producción

Se espera tener un AUC mayor al 90% dado que son los valores de los modelos de prevención y tamizaje en el ámbito de salud.