

ENTREGA 1
PROYECTO NCAA 's Clairvoyance.

PROFESOR
RAUL RAMOS POLLAN

INTEGRANTES

ORIANA MEJÍA CARDONA
VIVIANA SALAZAR PINEDA
ANDRES RAMÍREZ ARISTIZABAL

CONTENIDO

PROBLEMA A RESOLVER:

En este proyecto, el problema a resolver que se tratara, será acorde a predecir el resultado del torneo de la liga de basketball NCAA. Se harán predicciones en acorde a los ganadores y perdedores del mismo.

DATASET A UTILIZAR:

La base de datos que se utilizara para este proyecto va a adquirirse a través de una competencia kaggle titulada “March Machine Learning Mania 2017 ”:

<https://www.kaggle.com/competitions/march-machine-learning-mania-2017/overview/evaluation>

donde se usará las tablas necesarias de las 10 distintas que se encuentran en el dataset, siendo la principal la tabla de TourneyDetailedResults, en donde se encuentran un total de 914 registros distribuidos a lo largo de **34 columnas** nombradas como: **Season, Daynum, Wteam, Wscore, Lteam, Lscore, Wloc, Numot, Wfgm, Wfga, Wfgm3, Wfga3, Wftm, Wfta, Wor, Wdr, Wast, Wto, Wstl, Wblk, Wpf, Lfgm, Lfga, Lfgm3, Lfga3, Lftm, Lfta, Lor, Ldr, Last, Lto, Lstl, Lblk, Lpf**

MÉTRICAS DE DESEMPEÑO:

La métrica de desempeño a utilizarse ser la de logistic loss proporcionada por la siguiente fórmula:

$$-\frac{1}{n} \sum_{i=1}^n [y_i \log(\hat{y}_i) + (1 - y_i) \log(1 - \hat{y}_i)]$$

en donde más bajo sea el valor, significa que la predicción sera mas acertada en dicho caso

PRIMER CRITERIO A TENER EN CUENTA:

Si la fiabilidad de predicción de ganador/perdedor baja de un 70%, no suena efectivo poner el modelo en producción, porque generaría un fallo en el crecimiento de la efectividad del mismo en entornos como apuestas.