

PROYECTO FINAL – DEEP LEARNING ANN CLASIFICACIÓN BINARIA

PREVISIÓN LLUVIA

Iván Pinar Domínguez ©

Enunciado del Caso

En este proyecto final vamos a intentar realizar la predicción de lluvia para el día siguiente en base al dataset histórico con información de clima recopilado en 10 años en Australia. Este dataset se denomina “weatherAUS” y es un fichero CSV adjunto en este caso práctico.

La variable objetivo es “RainTomorrow”, es decir, la predicción de si habrá o no lluvia mañana.

NOTA: Utilice el script solución que ha visto durante el curso del caso de uso de clasificación binaria como base para realizar este proyecto final.

Para ello siga los siguientes pasos:

- 1) **Realice la importación de las librerías básicas de cálculo y visualización así como el dataset en un dataframe df:**

Resultado:

- 2) **Realice la limpieza del dataframe teniendo en cuenta eliminar aquellas columnas que superan un 10% de valores erróneos,**

posteriormente elimine todos aquellos registros que contienen algún valor NaN:

Resultado:

- 3) **Realice la visualización de la variable objetivo “RainTomorrow” para ver si el dataset está balanceado, adicionalmente, cree 2 nuevas columnas para reemplazar la columna “RainToday” por “Lluvia Hoy” y “RainTomorrow” por “Predicción Lluvia” de tal manera que las etiquetas “No” / “Yes” pasen a ser “0” / “1” (puede usar el siguiente código):**

```
df["Lluvia Hoy"] = np.where(df["RainToday"]=="Yes",1,0)
df["Predicción Lluvia"] = np.where(df["RainTomorrow"]=="Yes",1,0)
df=df.drop(["RainTomorrow","RainToday"],axis=1)
```

Por ultimo visualice la correlación entre las variables, ¿cuáles son las variables más correlacionadas con la variable objetivo?:

Resultado:

- 4) Realice la limpieza de columnas que estén poco correladas (por debajo de 0,1 en valor absoluto):

Resultado:

- 5) Finalmente en este paso EDA, convierta las variables categóricas con la técnica OneHotEncoder de las variables ['Location','WindGustDir','WindDir9am','WindDir3pm'], las variables de "Lluvia Hoy" y "Predicción Lluvia" ya se han convertido previamente de manera manual:

Resultado:

6) Train / Test con una división 75% / 25% respectivamente:

Resultado:

7) Escala los datos para que queden entre 0 y 1:

Resultado:

8) Creación del modelo de clasificación binaria, se recomienda utilizar 2 capas densas añadiendo 1 dropout tras cada 1 y finalmente una capa de clasificación binaria con 1 neurona:

Resultado:

9) Entrene el modelo definiendo un objeto EarlyStopping con patience = 5 y utilizando para el entrenamiento 30 epochs como máximo, ¿Cuál es el valor de pérdidas en el conjunto de validación de la última

iteración del entrenamiento? Visualice la evolución de las pérdidas durante el entrenamiento:

Resultado:

- 10) **Evaluar el modelo sobre el conjunto de test, realice la predicción de clases y la matriz de confusión, ¿cuál es el f1-score para cada categoría binaria?:**

Resultado: