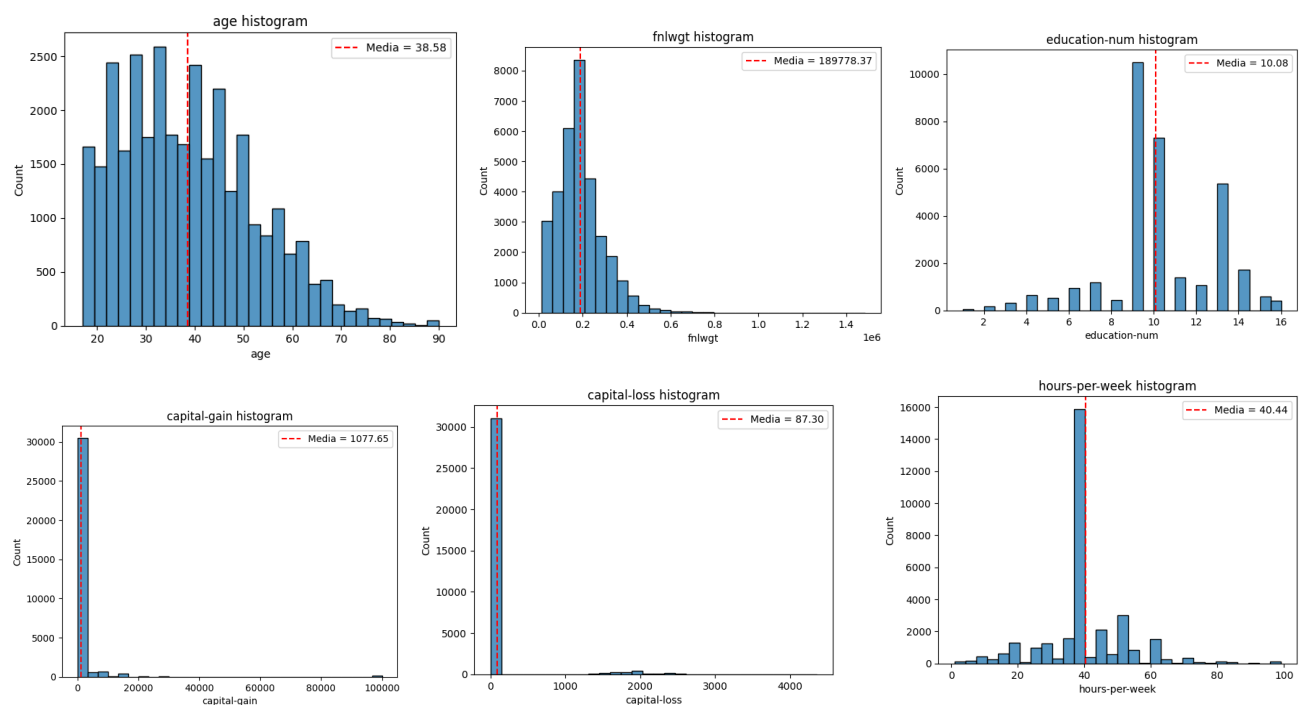


Parcial 2

Decisiones

- Se encontró que el 7% de los datos de las filas presentan “missing values”, sin embargo, se dejaron así dado que no se evidencian problemas y alteraciones al momento de correr el código y obtener los resultados
- Se transformó la variables *income* a un tipo dummy, para el cual se escogió poner 0 a aquellos datos que fueran menores o iguales a 50k, y 1 a aquellos datos que fueran mayores a 50k
- Para la exploración y procesamiento de los datos, se decidió obtener la media por cada variable para identificar el promedio de la edad, horas a la semana, ingreso y otras variables de las personas y saber a qué tipo de personas y características se está realizando el análisis

Procesamiento de Datos



- *age*: cuenta con edades aprox. entre 15 a 90 años, con una media de 38 años
- *workclass*: prevalece el privado con mas de 20,000
- *fnlwgt*: tiene una media aprox de 0.189
- *education-num*: cuenta con datos desde 1 hasta 16, con una media de 10
- *marital-status*: prevalecen las personas casadas sobre los otros estatus
- *relationship*: la mayoría son hombres casados

- *race*: prevalecen las personas blancas con una diferencia de aprox. 20,000 sobre el segundo (negras)
- *sex*: hay el doble de hombres que mujeres
- *capital-gain*: cuenta con una media de 1077
- *capital-loss*: cuenta con una media de 87
- *hours-per-week*: con datos desde 0 hasta 90, tiene una media de 40.44 horas a la semana

Métricas Clasificación Binaria

Entrenamiento

	precision	recall	f1-score	support
0	0.88	0.93	0.91	24720
1	0.74	0.61	0.67	7841
accuracy			0.85	32561
macro avg	0.81	0.77	0.79	32561
weighted avg	0.85	0.85	0.85	32561

Validacion

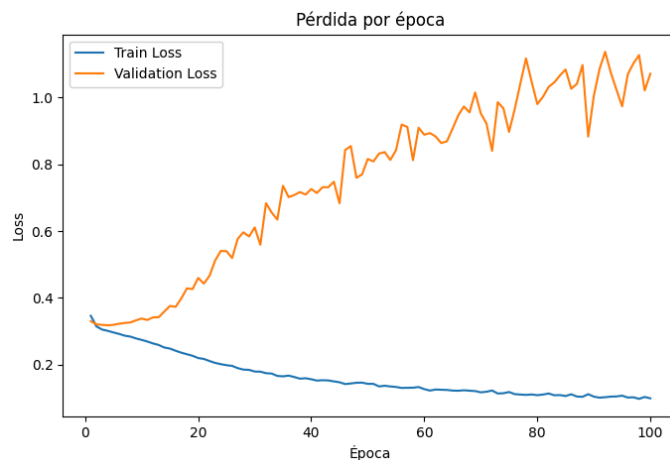
	precision	recall	f1-score	support
0	0.88	0.93	0.90	6231
1	0.71	0.59	0.64	1910
accuracy			0.85	8141
macro avg	0.80	0.76	0.77	8141
weighted avg	0.84	0.85	0.84	8141

Test

	precision	recall	f1-score	support
0	0.88	0.94	0.91	6204
1	0.75	0.61	0.67	1936
accuracy			0.86	8140
macro avg	0.82	0.77	0.79	8140
weighted avg	0.85	0.86	0.85	8140

Se puede evidenciar que en las anteriores tablas que para la clase 0 (es decir las personas que tienen un ingreso igual o menos a 50k) la precisión ronda entre 0.88-0.89 lo que indica que cuando el modelo predice que alguien va a ganar menos de 50k suele acertar, el recall y el F1 están por encima de 0.90 lo que indica que tiene un balance sólido y casi todos los casos reales cuando se ganan 50k o menos se detectan correctamente. Mientras que para la clase 1 (es decir las persona que ganan más de 50k), tiene una precisión entre 0.74-0.75 que indica que cuando se predice que alguien va a ganar mas de 50k es confiable pero puede fallar, el recall y el F1 están por debajo de 0.67 lo que indica que su rendimiento es menor, ya que, deja escapar alrededor del 40% de las personas que realmente ganan mas de 50k aumentando la probabilidad de que hayan falsos positivos. Lo anterior explica que el modelo es bueno para detectar personas con ingresos iguales o menores a 50k, pero no para personas con ingresos mayores a 50k. Con un Accuracy de 85% que significa un buen desempeño global.

Experimento sin Dropout ni Earlystopping



Métricas Entrenamiento:

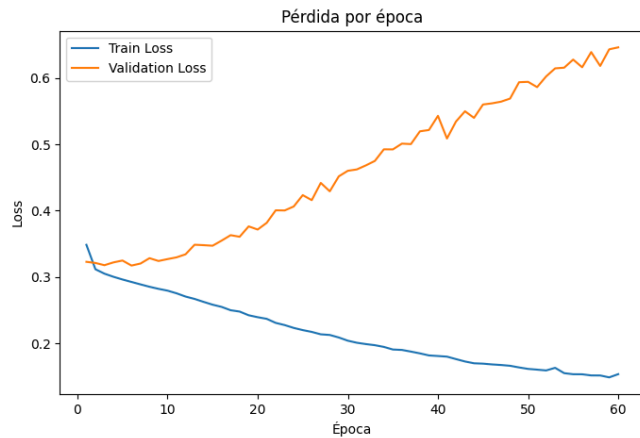
Accuracy: 0.9612, Precision: 0.9367, Recall: 0.8999, F1: 0.9179, ROC-AUC: 0.9403

Métricas de Validación:

Accuracy: 0.8225, Precision: 0.6323, Recall: 0.5817, F1: 0.6059, ROC-AUC: 0.7390

Métricas test:

Accuracy: 0.8262, Precision: 0.6474, Recall: 0.5909, F1: 0.6179, ROC-AUC: 0.7452



Métricas Entrenamiento:

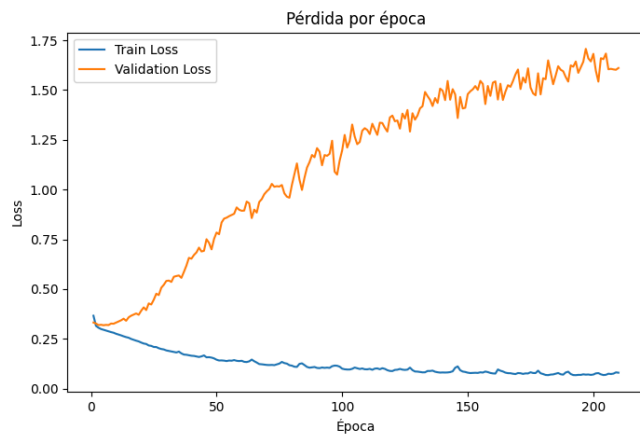
Accuracy: 0.9350, Precision: 0.9234, Recall: 0.7961, F1: 0.8550, ROC-AUC: 0.8876

Métricas de Validación:

Accuracy: 0.8294, Precision: 0.6697, Recall: 0.5382, F1: 0.5968, ROC-AUC: 0.7284

Métricas test:

Accuracy: 0.8313, Precision: 0.6851, Recall: 0.5382, F1: 0.6028, ROC-AUC: 0.7305



Métricas Entrenamiento:

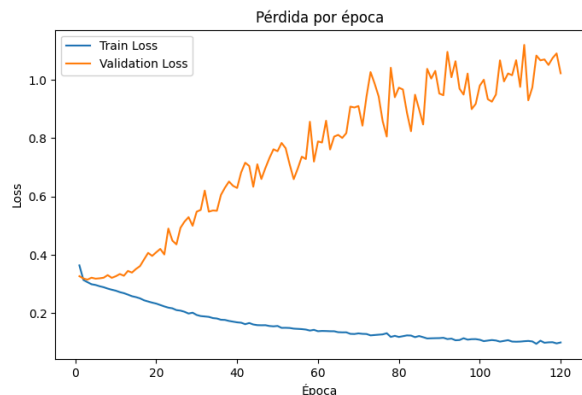
Accuracy: 0.9723, Precision: 0.9399, Recall: 0.9455, F1: 0.9427, ROC-AUC: 0.9632

Métricas de Validación:

Accuracy: 0.8116, Precision: 0.5967, Recall: 0.6073, F1: 0.6020, ROC-AUC: 0.7408

Métricas test:

Accuracy: 0.8141, Precision: 0.6094, Recall: 0.6085, F1: 0.6089, ROC-AUC: 0.7434



Métricas Entrenamiento:

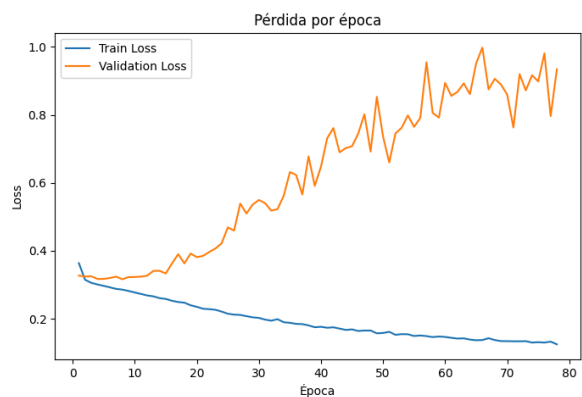
Accuracy: 0.9605, Precision: 0.9207, Recall: 0.9147, F1: 0.9177, ROC-AUC: 0.9448

Métricas de Validación:

Accuracy: 0.8212, Precision: 0.6204, Recall: 0.6126, F1: 0.6164, ROC-AUC: 0.7488

Métricas test:

Accuracy: 0.8254, Precision: 0.6387, Recall: 0.6126, F1: 0.6254, ROC-AUC: 0.7522



Métricas Entrenamiento:

Accuracy: 0.9480, Precision: 0.9076, Recall: 0.8731, F1: 0.8900, ROC-AUC: 0.9225

Métricas de Validación:

Accuracy: 0.8278, Precision: 0.6391, Recall: 0.6110, F1: 0.6247, ROC-AUC: 0.7526

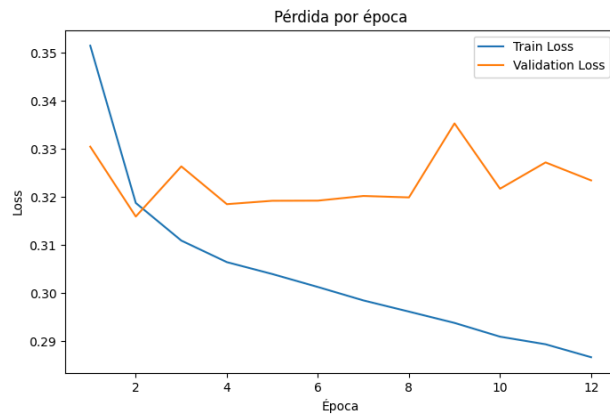
Métricas test:

Accuracy: 0.8271, Precision: 0.6472, Recall: 0.6007, F1: 0.6231, ROC-AUC: 0.7493

En esta primera parte se puede evidenciar que en el experimento 1 el modelo tiene un balance en el entrenamiento, sin embargo, en la validación y test cae significativamente lo que indica que el modelo memoriza pero no generaliza (overfitting). Mismo caso para los modelos 2 y 3, tienen buen rendimiento en el entrenamiento pero en la validación su rendimiento cae (overfitting), siendo el modelo 3 muy fuerte y peor overfitting que el modelo 1. Ahora bien, los modelos 4 y 5 tienen mejor rendimiento que los modelos anteriores pero el modelo 4 sigue teniendo overfitting aunque más estable, y el modelo 5 es más balanceado con una diferencia entre el entrenamiento y validación menor en comparación a los otros modelos. Por lo que, los cinco modelos son parecidos pero el 5 es el que mejor rendimiento tiene en

comparación a los otros y generaliza mejor en la validación y métricas más consistentes, es decir, para este caso es el mejor.

Experimento con Dropout y Earlystopping



Métricas entrenamiento:

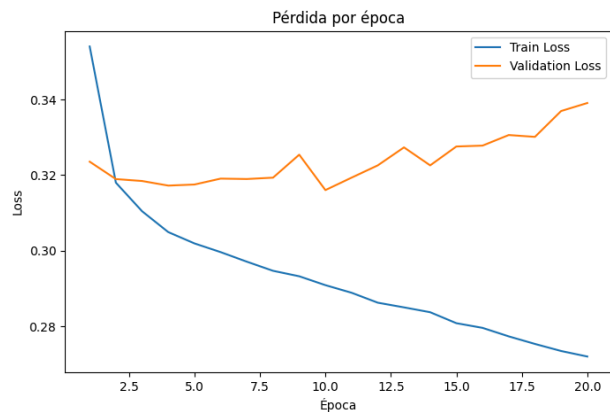
Accuracy: 0.8752, Precision: 0.7955, Recall: 0.6485, F1: 0.7145, ROC-AUC: 0.7978

Métricas de Validación:

Accuracy: 0.8509, Precision: 0.7254, Recall: 0.5864, F1: 0.6485, ROC-AUC: 0.7592

Métricas Test:

Accuracy: 0.8549, Precision: 0.7425, Recall: 0.5971, F1: 0.6619, ROC-AUC: 0.7662



Métricas entrenamiento:

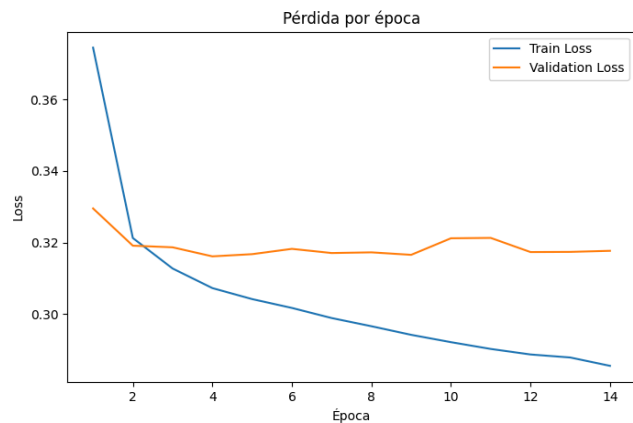
Accuracy: 0.8811, Precision: 0.7834, Recall: 0.6994, F1: 0.7390, ROC-AUC: 0.8190

Métricas de Validación:

Accuracy: 0.8490, Precision: 0.7009, Recall: 0.6220, F1: 0.6591, ROC-AUC: 0.7703

Métricas Test:

Accuracy: 0.8557, Precision: 0.7242, Recall: 0.6348, F1: 0.6766, ROC-AUC: 0.7797



Métricas entrenamiento:

Accuracy: 0.8763, Precision: 0.7829, Recall: 0.6727, F1: 0.7236, ROC-AUC: 0.8068

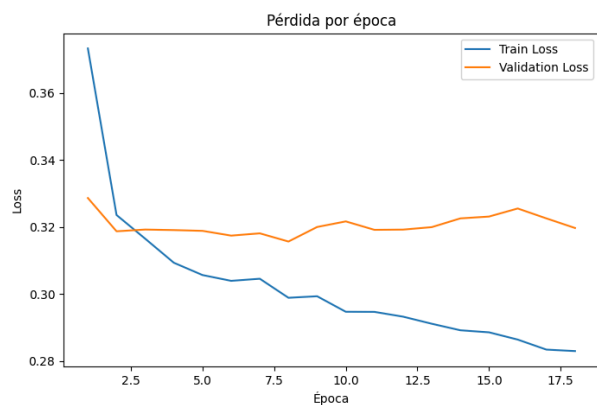
Métricas de Validación:

Accuracy: 0.8500, Precision: 0.7102, Recall: 0.6094, F1: 0.6560, ROC-AUC: 0.7666

Métricas Test:

Accuracy: 0.8548, Precision: 0.7293, Recall: 0.6193, F1: 0.6698, ROC-AUC: 0.7738

Early stopping en época 18



Métricas entrenamiento:

Accuracy: 0.8788, Precision: 0.8042, Recall: 0.6563, F1: 0.7228, ROC-AUC: 0.8028

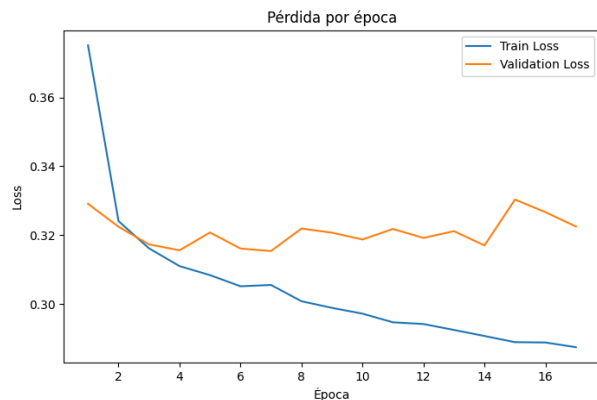
Métricas de Validación:

Accuracy: 0.8477, Precision: 0.7167, Recall: 0.5801, F1: 0.6412, ROC-AUC: 0.7549

Métricas Test:

Accuracy: 0.8563, Precision: 0.7449, Recall: 0.6018, F1: 0.6657, ROC-AUC: 0.7687

Early stopping en época 17



Métricas entrenamiento:

Accuracy: 0.8779, Precision: 0.7793, Recall: 0.6877, F1: 0.7306, ROC-AUC: 0.8129

Métricas de Validación:

Accuracy: 0.8508, Precision: 0.7062, Recall: 0.6230, F1: 0.6620, ROC-AUC: 0.7718

Métricas Test:

Accuracy: 0.8558, Precision: 0.7260, Recall: 0.6322, F1: 0.6759, ROC-AUC: 0.7789

Para esta parte, se realizó los mismos 5 experimentos con los mismos valores de los hiperparametros por cada uno pero se le aplicó la regularización con un Dropout. Por consiguiente se obtuvo que la regularización permitió tener los 5 modelos más estables que sin regularización y no se observa overfitting ni underfitting en ningún modelo. Para los modelos 1 y 4 el balance es mejor pero con una capacidad predictiva no tan alta dado un F1 y ROC menor que los modelos 2 y 3, los cuales tienen un F1 y ROC en la validación y test más altos que los modelos 1 y 4, indicando un buen balance y por ende generaliza mejor. Sin embargo, el modelo 5 es el que tiene un F1 y ROC en el entrenamiento, validación y test superiores que los otros modelos, lo que indica un mejor balance y una capacidad predictiva mejor que los otros modelos. Por lo que, con la aplicación de la regularización el modelo 5 sigue siendo el modelo con mejor rendimiento y generalización mucho más sólida en comparación a los otros modelos, es decir, en general con y sin regularización el modelo con mejor rendimiento es el 5.