



Trabajo Práctico 1

Reservas de hotel - Resumen checkpoint 4

[75.06] Organización de datos
Primer cuatrimestre 2023
Grupo 19: Sudanalytics

Integrantes

Nombre	Padrón	Mail
Re, Adrián Leandro	105025	are@fi.uba.ar
Lorenzo, Luciano Andrés	108951	llorenzo@fi.uba.ar
Toulouse, Alan	105343	atoulouse@fi.uba.ar
Tonizzo, Nicolas	107820	ntonizzo@fi.uba.ar

Fecha de entrega: 25/05/2023

Índice

1. Introducción	2
2. Resultados	2
3. Conclusiones	2
4. Referencias	2

1. Introducción

El objetivo de esta instancia fue entrenar una red neuronal secuencial para realizar predicciones de manera similar a cómo lo veníamos haciendo en los checkpoints anteriores.

2. Resultados

Aunque las redes neuronales son conocidas por su capacidad para capturar patrones complejos, en este caso específico, parece que no fueron la mejor opción. Las features con las que trabajamos en este dataset pueden no haber sido completamente capturadas por las redes neuronales que entrenamos. En cambio como vimos en el checkpoint3, modelos como XGBoost o Random Forest, parecen obtener muchos mejores resultados para este tipo de datasets.

Creemos que fue un poco incomodo utilizar random search con keras y el KerasClassifier. Así que para el segundo notebook decidimos no utilizarlo, y buscar los hiperparametros manualmente. En este intentamos variar y jugar un poco con las dense layers de la red, para ver si podíamos encontrar un mejor patrón en los datos.

Al añadir muchas dense layers el modelo overfitteo mucho el dataset, obteniendo un score mucho mayor para el conjunto de entrenamiento que para el de validación. Por eso en las siguientes redes que creamos, utilizamos los parámetros de regularización de las dense layers, para que el modelo no se ajuste tanto a los datos de entrenamiento. Luego de esto logramos 'emparejar' bastante el score de entrenamiento con respecto al de validación. A pesar de que este ultimo no haya subido mucho, al menos nos aseguramos de que el modelo puede generalizar bien.

3. Conclusiones

Llegamos a la conclusión de que las redes neuronales son altamente versátiles, lo cual es una ventaja para poder cambiar mucho el modelo y poder encontrar el verdadero patrón de los datos, pero también es muy ir complejo ir cambiandolo para encontrarlos y tambien es muy propenso al overfitting.

Concluimos que XGBoost o Random Forest son mejores opciones para este tipo de problemas.

4. Referencias

- Paper "Hotel booking demand datasets" proporcionado por la cátedra.