



Cas Kaggle

Adrià Fernández Ortiz
Eduard José García Mendelev



1. Cas Kaggle Elegido

- League of Legends SOLO-Q Ranked Games
- Objetivo: Clasificar cuál de los 2 equipos ha ganado la partida.
- Dataset con 25000 partidas.
- El dataset contiene 59 atributos correspondientes en su mayoría con datos de cada partida desde el punto de vista del equipo azul.



2. Tratamiento de datos

- Comprobación de NaNs y filas duplicadas
- Comprobación de distribución del dataset
- Normalización con MiniMax de los datos no categóricos

False

Nombre de filas duplicadas:
0

Distribució de la classe objectiu:

hasWon

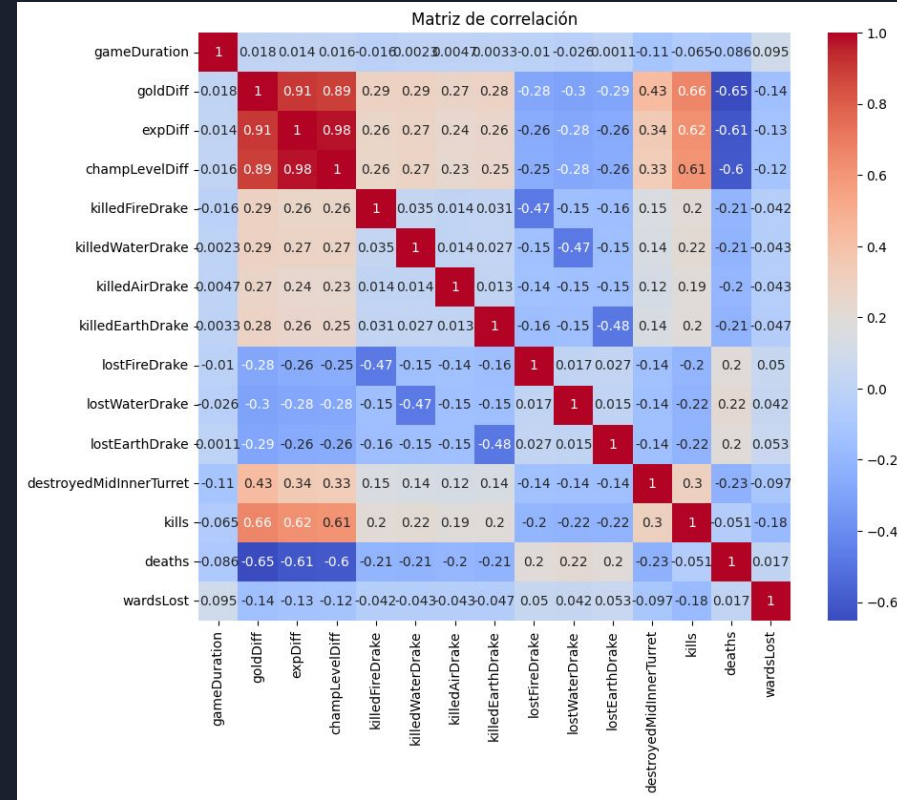
0 121740

1 120832

Name: count, dtype: int64

3. Selección de Características

- Estudio preliminar con Random Forest
- Eliminación de atributos menos importantes
- Estudio más exhaustivo con RFE (Recursive Feature Elimination) para quedarnos con las 15 más importantes.
- Buscamos correlaciones para evitar atributos redundantes.





4. Entrenamiento de modelos

- Entrenamos 4 modelos (KNN, Random Forest, Regresión Logística, SVM)

Presició del model amb KNN: 0.81

Presició del model amb Regressió Logística: 0.83

Presició del model amb SVM: 0.83

Presició del model amb Random Forest: 0.83



4.1 Mejores hiperparámetros para cada modelo

- Buscando los mejores hiperparámetros para cada modelo con Randomized Search.

Millors hiperparàmetres trobats:

```
{'n_estimators': 300, 'max_depth': 30, 'bootstrap': True}
```

Precisió del model RF optimitzat: 0.83

Millors hiperparàmetres trobats:

```
{'kernel': 'linear', 'gamma': 0.001, 'degree': 3, 'C': 0.1}
```

Precisió del model SVM optimitzat: 0.83

Millors hiperparàmetres trobats:

```
{'weights': 'distance', 'p': 1, 'n_neighbors': 11, 'metric': 'minkowski'}
```

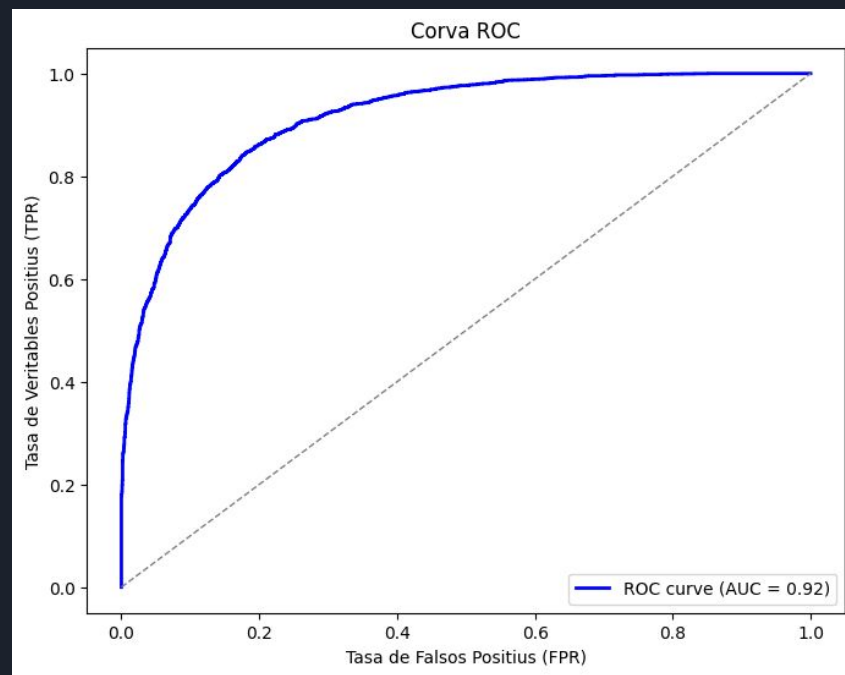
Precisió del model KNN optimitzat: 0.82

Millors hiperparàmetres trobats:

```
{'solver': 'liblinear', 'penalty': 'l1', 'C': 0.1}
```

Precisió del model de Regressió Logística optimitzat: 0.83

5. Visualización de Resultados





5. Conclusiones

Los resultados obtenidos muestran que los cuatro modelos implementados (KNN, Regresión Logística, Random Forest y SVM) ofrecen un rendimiento muy similar, con precisiones en torno al 82-83% incluso después de optimizar sus hiperparámetros.

Se puede concluir que cualquiera de los modelos analizados sería adecuado para la taberna propuesta, ya que todos logran un rendimiento muy similar en términos de accuracy. Esto refleja que los modelos implementados han sido bien ajustados y que los datos disponibles son suficientes para alcanzar ese nivel de precisión.