

Maestría en Inteligencia Analítica para la Toma de Decisiones

Deep Learning y Redes Neuronales

Proyecto Kaggle | Kobe Bryant Shot Selection

Integrantes:

Carmona Lozano, Silverio | 199618115 Oñate Vega, Lady D. | 200814451

Quispe Castilla, Carlos J. | 201627368

Suarez Acero, Johan D. | 200713010





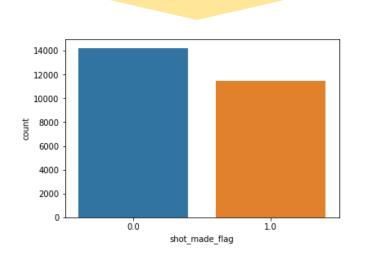
Descripción del problema seleccionado

Using 20 years of data on Kobe's swishes and misses, can you predict which shots will find the bottom of the net?

Clasificación:

- 25 variables
- 25.7 mil registros para entrenar/validar
- 5 mil registros por predecir
- Datos numéricos y categóricos
- Datos de cada partido:
- Clase de tiro
- Ubicación en el campo
- Tiempo del partido
- Fecha/temporada del partido
- Partido local/visitante
- Oponente



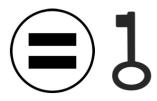




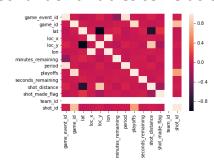


Metodología de abordaje del problema

Registros iguales o ID

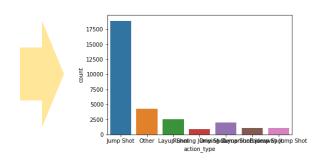


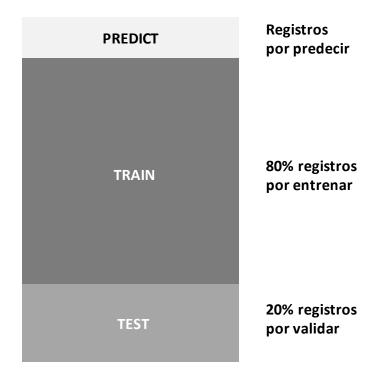
Retirando variables correlacionadas



Agrupando Clases Pareto











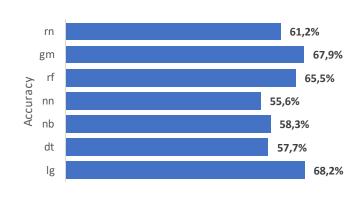
Metodología de abordaje del problema

MODELAMIENTO DE DATOS

```
models = {'lg': LogisticRegression(),
          'dt': DecisionTreeClassifier(),
          'nb': GaussianNB(),
          'nn': KNeighborsClassifier(),
          'rf': RandomForestClassifier(),
          'gm': LogisticGAM(),
          'rn': MLPClassifier()}
X_train, X_test, y_train, y_test = train_test_split(X, y, random_state=1)
for model in models.keys():
   models[model].fit(X_train, y_train)
# predict test for each model
y pred = pd.DataFrame(index=X test.index, columns=models.keys())
for model in models.keys():
   y pred[model] = models[model].predict(X test)
for model in models.keys():
   print('accuracy '+ model, metrics.accuracy_score(y_test, y_pred[model]))
   print('f1_score '+ model, metrics.f1_score(y_test, y_pred[model]))
   print('f1 beta score ' + model, metrics.fbeta_score(y_test, y_pred[model],10))
```

7 modelos

SELECCIÓN DE MODELOS



Accuracy | F1 Score | Beta Score

PREDICCIÓN



Registros por predecir

