

Reconocimiento de Patrones y Aprendizaje Automatizado

2021-1

Proyecto Final: El Titanic

Daniela Susana Vega Monroy
danisu@ciencias.unam.mx

Fecha de entrega: 3 de Febrero 2021

1 Introducción

La tragedia del Titanic

El hundimiento del RMS Titanic fue una catástrofe marítima ocurrida en la noche del 14 al 15 de abril de 1912, cuando el transatlántico británico Titanic, que realizaba su viaje inaugural de Southampton a Nueva York, chocó contra un iceberg en el océano Atlántico frente a las costas de Terranova. El choque se produjo en el lado de estribor a las 23:40 del 14 de abril, lo que provocó el hundimiento del barco en menos de tres horas, a las 02:20 del 15 de abril.

Fallecieron 1513 personas por golpes diversos, caídas, ahogamiento o hipotermia, lo que convirtió a este naufragio en el más fatal de la época y en una de las mayores tragedias marítimas en tiempo de paz. Numerosas personalidades murieron durante el naufragio, entre las que se encontraban los millonarios Benjamin Guggenheim y John Jacob Astor IV.

Tras el hundimiento, varias comisiones de investigación fueron llevadas a cabo y sus conclusiones se utilizaron para mejorar la seguridad marítima, especialmente a través de nuevos reglamentos. Varios factores se combinaban para explicar tanto el hundimiento como el elevado número de pasajeros que no pudieron sobrevivir. El buque no disponía de suficientes botes salvavidas, y la tripulación nunca había sido entrenada para enfrentarse a estos casos. Como resultado, la evacuación de los pasajeros estuvo mal organizada. El comportamiento del capitán Edward Smith también fue criticado, sobre todo porque había mantenido el barco a una velocidad demasiado alta, dadas las condiciones de navegación. Las circunstancias meteorológicas y climáticas también jugaron un papel determinante.

2 Objetivo

El objetivo del proyecto es entrenar un modelo, para luego crear un nuevo pasajero y poder descubrir su destino a bordo del RMS Titanic.

3 Métodos

Las siguientes referencias me ayudaron a concluir mi trabajo:

<https://www.kaggle.com/tanmayunhale/titanic-top-4-hyperparameter-tuning>

https://youtu.be/rODWw2_1mCI

<http://campus.lakeforest.edu/frank/FILES/MLFfiles/Bio150/Titanic/TitanicMETA.pdf>

<https://www.geeksforgeeks.org/numpy-std-in-python/>

A continuación los puntos clave del proyecto:

1.

```
trainData = trainData.fillna({"Embarked": "S"})
```

```
testData = testData.fillna({"Embarked": "S"})
```

Usado para llenar los valores faltantes de la variable Embarked.

2.

```
totalData = [trainData, testData]
for edades in totalData:
    meanEdad = trainData["Age"].mean()
    stdEdad = testData["Age"].std()
    nullEdad = edades["Age"].isnull().sum()
    edadAlea = np.random.randint(meanEdad - stdEdad, meanEdad + stdEdad, size =
nullEdad)
    divEdad = edades["Age"].copy()
    divEdad[np.isnan(divEdad)] = edadAlea
    edades["Age"] = divEdad
    edades["Age"] = trainData["Age"].astype(int)
trainData["Age"].isnull().sum()
```

Usando 3 valores: Promedio, is_null() y la desviación típica del rango de edades disponible se genera un arreglo de edades para poder llenar los espacios vacíos en el valor Age.

3.

```
genders = {"male": 0, "female": 1}
totalData = [trainData, testData]
for gen in totalData:
    gen['Sex'] = gen['Sex'].map(genders)

ports = {"S": 0, "C": 1, "Q": 2}
totalData = [trainData, testData]
for port in totalData:
    port['Embarked'] = port['Embarked'].map(ports)
```

Se *mapean* valores categóricos a sus equivalentes numéricos.

4.

```
msgSobreviviente = 'Eres de los pocos pasajeros que sobrevivió'
msgVíctima = 'Eres una persona víctima de la tragedia del Titanic'
if rfPrediction == 0:
    print(msgSobreviviente)
else:
    print(msgVíctima)
```

Se imprime el mensaje correspondiente de acuerdo a la predicción del modelo Random Forest.

4 Resultados

Se aplicaron 5 modelos diferentes : Naive Bayes, KNN, Regresión Logística, SVM y Random Forest. Estos fueron sus resultados midiendo su exactitud:

Modelo	Calificación
Naive Bayes	79.70
KNN	76.14
Regresión Logística	80.71
SVM	72.59
Random Forest	81.73

Ahora si creamos a un pasajero de sexo masculino, de primera clase , con 51 años, sin familiares a bordo y que abordó en el puerto Southampton, nuestro modelo Random Forest nos dice lo siguiente:

Eres una persona víctima de la tragedia del Titanic.

6 Conclusiones

Random Forest es un algoritmo de aprendizaje supervisado que es un conjunto de muchos árboles de decisión. Si bien, Decision Trees es un gran algoritmo no siempre nos darán la mejor predicción, ahora imaginemos su resultado si ejecutamos muchos árboles de decisión con diferentes subconjuntos y vemos cuál fue la respuesta más repetida. Eso es un algoritmo Random Forest.

Nuestro dataset tenía claro cuales columnas era extremadamente importantes y cuales no, aparte eliminamos los valores inútiles, haciendo que el campo aleatorio de valores que el algoritmo tenía disponible mucho más pequeño y con la información más importante.

En conclusión, Random Forest es un gran indicador de la importancia que le asigna a nuestras características.

Bibliografía

<https://towardsdatascience.com/why-random-forest-is-my-favorite-machine-learning-model-b97651fa3706>

<https://builtin.com/data-science/random-forest-algorithm>

<https://towardsdatascience.com/predicting-the-survival-of-titanic-passengers-30870ccc7e8>

https://es.wikipedia.org/wiki/RMS_Titanic