Proyecto Kaggle – Otto Group
Maestría en Inteligencia Analítica para la Toma de Decisiones
Deep Learning
Intersemestral

Presentado por:

John Pablo Calvo

Edgar Andrés García

Camilo Alejandro Rodriguez

Juan Carlos Eraso

Descripción del problema

- ***Otto Group*** es una compañía multinacional dedicada al e-commerce, con subsidiarias en mas de 20 países, vende millones de productos de diversas líneas vía web a nivel mundial.

- Debido a la diversa infraestructura global y a la similitud en las características de los productos, tiene ***problemas*** en la clasificación adecuada de estos en sus respectivas categorías.

- Para realizar un buen análisis de producto, es indispensable tener máxima ***precisión*** al momento de clasificar los productos en sus respectivas categorías.

- Se identifica como un ***problema de clasificación Multiclase***

Metodología de Abordaje del Problema

- La base contiene **93 variables**, incluyendo un código ID y la variable de desempeño que para este caso es nominada ***Target***
- La variable de desempeño contiene ***9 clases***, y las variables predictoras son de tipo discreto.
- La competencia contiene dos bases: una de entrenamiento (149,369 registros) y una de pruebas (61,789 registros).
- Ninguna de las dos bases contenía ***missing values*** ni ***outliers*** en sus variables.
- Se deben escoger métodos adecuados para la predicción de variables categóricas multiclase.

Metodología de Abordaje del Problema

- Se utilizan tres ***métodos de clasificación***:
  - ***Random Forest:***

    Se basa en la creación de un conjunto de árboles de decisión seleccionados aleatoriamente. Se busca reducir el efecto del ruido, brindando resultados con el mejor precisión.
  - ***Regresión Logística:***

    Los modelos de regresión logística multinomiales son ampliamente utilizados para predecir variables categóricas que pueden tomar mas de dos modalidades diferentes, determinando que ocurra o no el evento.
  - ***Redes Neuronales:***

    Los modelos de redes neuronales han demostrado su gran poder de predicción y su baja tasa de errores en múltiples competencias, llevando a su uso extendido pesar de su baja interpretabilidad.