

Examen, Semana 12: Aplicaciones Financieras

La base de datos consiste en Empleos, Descripciones de los empleos y Rango salarial. Utilizando al menos estas 3 columnas de los datos, genere un modelo que dado el texto de nuestros datos nos regrese como predicción un salario estimado o un rango salarial estimado.

Utilice algún modelo o variante de Bert para realizar la tarea.

Referencias:

- <https://tfhub.dev/google/collections/bert/1>
- https://tfhub.dev/tensorflow/bert_en_uncased_L-12_H-768_A-12/4 (Bert base)
- <https://towardsdatascience.com/bert-for-dummies-step-by-step-tutorial-fb90890ffe03>

Datos:

https://drive.google.com/file/d/1dY22s182lwwucJirHCTCCD27Blvurcuz/view?usp=share_link

Puntos importantes:

- Cargar alguno de los modelos pre-entrenados BERT
- Llevar a cabo la tokenización del texto
- Convertir dicho vector de tokens en un vector que sea la representación vectorial del texto (usualmente, es lo que nos entrega Bert al final)
- Con los vectores de salida, utilizar una red neuronal o algún otro modelo para las predicciones
- Generar informes del rendimiento del modelo (mean absolute error, etc)

IMPORTANTE: No es necesario que realicen un 'fine tuning' del modelo BERT pre-entrenado, pero si logran hacerlo será tomado en cuenta