

Primera entrega de proyecto

POR:

Claudia Yaneth Giraldo Vergara

Sebastian Zuluaga

MATERIA:

Introducción a la inteligencia artificial

PROFESOR:

Raúl Ramos Pollan



UNIVERSIDAD DE ANTIOQUIA

FACULTAD DE INGENIERÍA

MEDELLÍN 2023

1. Planteamiento del problema

Estados Unidos es el país de las oportunidades, cuenta con un buen nivel de desarrollo, una economía fuerte y una gran estabilidad, esto hace que la demanda laboral sea alta, por lo que sus salarios y beneficios son, por lo general, más competitivos que los que ofrece una empresa por fuera de los Estados Unidos.

Estas oportunidades laborales hacen que muchos trabajadores quieran emigrar a este país, sin embargo, para poder ejercer de manera legal en los Estados Unidos es necesario tener una visa de trabajo, la cual puede ser difícil de conseguir, puesto que depende de muchos factores, desde el título profesional hasta el código postal de la compañía para la que se va a laborar, inclusive, algunas veces parece cuestión de suerte.

Nuestro proyecto consiste en la predicción del resultado a la aplicación de una visa de trabajo en Estados Unidos, dadas las características laborales de un aplicante (título profesional, compañía en la cual va a trabajar, número de empleados de la compañía, salario, etc.).

2. Dataset

El dataset a utilizar para este proyecto es este, proviene de Kaggle LINK (<https://www.kaggle.com/datasets/jonamjar/h1b-data-set-2017>) en la cual se encuentran aproximadamente 700 mil datos de aplicaciones para la visa de trabajo en el año 2017, proporcionados por la Oficina de Certificación Laboral Extranjera, del Departamento del Trabajo de Estados Unidos de América. El dataset está compuesto por un solo archivo .csv que proporciona los datos pedidos en una aplicación para la visa, consta de 53 variables (columnas), las cuales contienen información como:

- Estado del caso: aprobados, en estudio, retirados.
- Nombre del empleador: empresa a la cual se presenta el postulante.
- Nombre del cargo
- Título requerido
- Trabajos de tiempo completo: si/no
- Salario
- Año de la postulación
- Sitio de trabajo

3. Métricas

Para el análisis del dataset utilizaremos el Métricas de regresión en aprendizaje automático: Error absoluto medio (MAE)

En MAE, el error se calcula como un promedio de diferencias absolutas entre los valores objetivo y las predicciones. El MAE es una puntuación lineal, lo que significa que todas las diferencias individuales se ponderan por igual en el promedio. Por ejemplo, la diferencia entre 10 y 0 será el doble de la diferencia entre 5 y 0.

Se calcula mediante la siguiente formula:

$$MAE = \frac{1}{N} \sum_{i=1}^N |y_i - \hat{y}_i|$$

4. Desempeño

Nuestro modelo de predicción aplicado una empresa de asesoramiento para trámites de visado debería tener un porcentaje de acierto de al menos el 65%, ya que el resultado de la predicción definirá si un cliente realiza el proceso de aplicación para una visa laboral. Si el porcentaje es menor, daría cavidad a predicciones erróneas, por ejemplo, arrojaría una predicción positiva (es decir: aprobada) a un cliente que desea tener una visa H1B, pero cuando realice la aplicación, este se le sea negada. Esto sería contraproducente para la empresa ya que no sería confiable ante sus posibles clientes que desean contratar sus servicios de asesoramiento.

5. Bibliografía

- H1B Data Set 2017 – Manoj | Kaggle (2017). Recuperado el 8 de marzo de 2023, de https://www.kaggle.com/datasets/jonamjar/h1b-data-set-2017?select=H-1B_Disclosure_Data_FY17.csv