PROYECTO MÓDULO 7
DS: ANÁLISIS DE
ASPIRACIONES
PROFESIONALES DE
LA GENERACIÓN Z

CRISTÓBAL YÁÑEZ

#### OBJETIVOS DEL PROYECTO

- Aplicar técnicas avanzadas de ciencia de datos e inteligencia artificial.
- Analizar las aspiraciones profesionales de la Generación Z a partir de un dataset.
- Desarrollar un modelo predictivo (Random Forest) para determinar la carrera aspiracional.
- Implementar una API REST que reciba datos y devuelva la predicción junto con el porcentaje de confianza.
- Exponer la API públicamente mediante Ngrok para facilitar su consulta.

#### PROCESO Y METODOLOGÍA

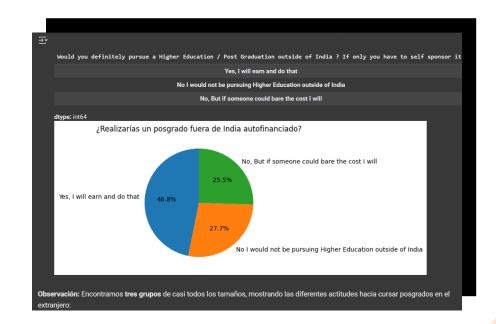
## EDA

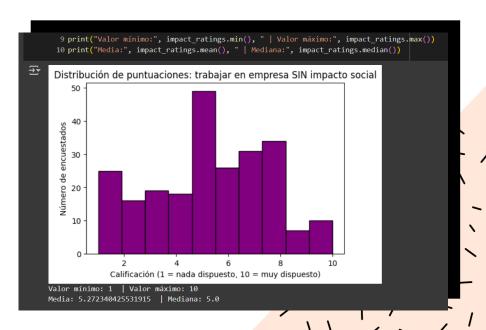
# Modelado

API REST

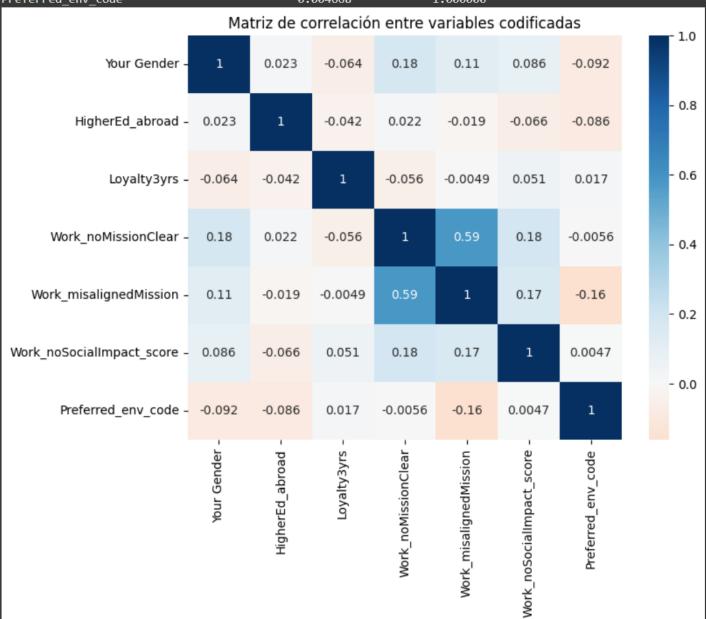
# ANÁLISIS EXPLORATORIO DE DATOS (EDA)

- Limpieza de datos, manejo de valores nulos e inconsistencias.
- Transformación y codificación de variables (One-Hot Encoding, Label Encoding).
- Visualización y análisis de correlaciones para identificar patrones relevantes.

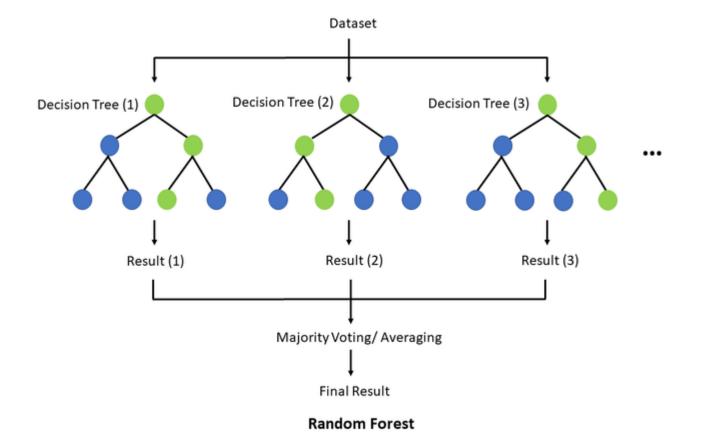




	Matriz de correlación ent	re variables
Preferred_env_code	0.004668	1.000000
Work_noSocialImpact_score	1.000000	0.004668
Work_misalignedMission	0.165004	-0.161593
Work_noMissionClear	0.181731	-0.005608
Loyalty3yrs	0.050563	0.016637
HTRUCI Ed_abi dad	-0.000	-0.003734

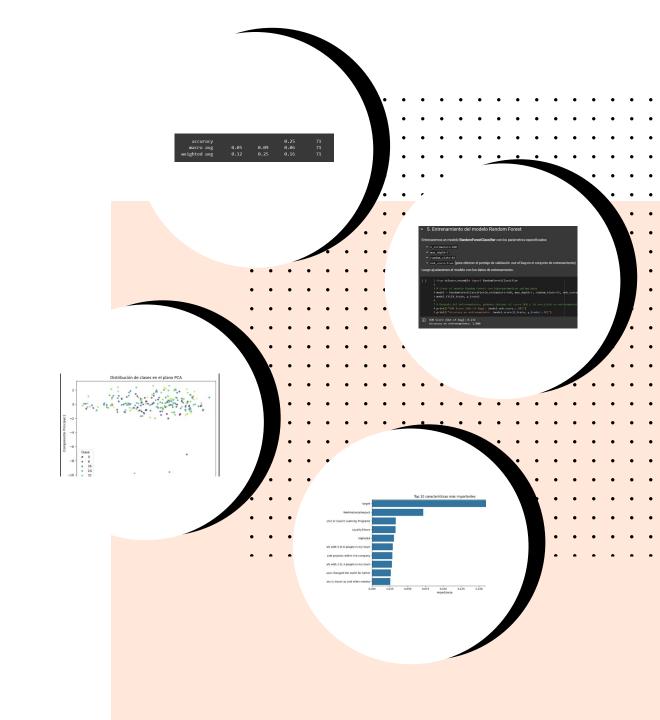


#### MODELADO (RANDOM FOREST)



- Entrenamiento de un modelo Random Forest para clasificación multiclase.
- Evaluación del modelo: métricas

   (accuracy, F1-score, matriz de confusión)
   e importancia de variables.



```
| Discrete | Discrete
```

### DESARROLLO DE LA API REST

- Uso de FastAPI para crear el servicio REST.
- Creación de un endpoint POST que recibe un JSON de entrada y retorna la predicción y confianza.

```
PROBLEMS OUTPUT DEBUG CONSOLE TERMINAL PORTS COMMENTS

ongrok

Goodbye tunnels, hello Agent Endpoints: https://ngrok.com/r/aep

Session Status
Account cyanezg (Plan: Free)
Version 3.20.0
Region South America (sa)
Latency 50ms
Web Interface http://127.0.0.1:4040
Forwarding https://e301-2803-c600-d215-d338-6d93-5dfe-ce0f-b7b8.ngrok-free.app → http://localhost:8000

Connections ttl opn rt1 rt5 p50 p90
```

- Desarrollo y pruebas en Visual Studio Code (entorno local).
- Uso de Ngrok para generar un túnel público y compartir la API.
- Validación de la API mediante Postman y la interfaz Swagger UI.

```
Body Cookies Headers (5) Test Results  

{} JSON  

Preview  

Visualize  

prediction": "Business Operations in any organization, Manage and drive End-to-End Projects or Products, Build and develop a Team",

"confidence": 0.09554870608167572
```

### CONCLUSIONES E IDEAS

- A pesar de limpiar y transformar la data, y ajustar hiper parámetros, optimizando todo para la construcción del modelo, el f1- score del modelo es de mediana confiabilidad (0,25). Una alternativa sería probar modelos más robustos no explorados en proyectos anteriores como XGBoost. Aun así, considerando la diversidad de las preguntas y la complejidad de estas y también de sus respuestas, el modelo es lo suficientemente bueno, y fue testeado y validado con varias entradas, en las que los resultados fueron positivos, y las discrepancias pueden llegar a ser subjetivas y debatibles, dada la naturaleza del estudio.
- La API que predice el empleo según las respuestas de la encuesta, podría alimentarse directamente de ella. Para esto por su puesto habría que tener acceso a la encuesta misma, y no solo al dataset de los resultados. Esto permitiría una interacción más amigable con el cliente, que con la versión actual en la que debe hacer input de las variables mediante un JSON.



. . . . . . . . . . . . .