[DATATHON PROJECT REPORT]

[project by the Datamind team]

Contents

- Datasets used
- How the datasets were used(relationships)
- Inner mechanics
- Future features

Members List:
mohammed Bakhit(mb2107875@qu.edu.qa)
Nima Abdi (nabdi@hbku.edu.qa)
Rama Dalqamouni
(ramadalgamouni41@gmail.com)
Noor Almousa (noorma377@gmail.com)
Laila Juma (200771065@qu.edu.qa)

*DATASETS INDEX

NPC DATASETS

LABOR FORCE

- 2_labour_force_2023_AE
- Annual_Bulletin_Labour_force_2018_AE
- Bulletin_labor_force_2022
- LF_Q4_2020_AE

Datasets From other

- Courses from Coursera 2024 (Kaggle.com)

How the DATASETS Were Used(relationships):

- The number of employees in an occupation (gender or nationality total) from the various years are used in order to check the demand for an occupation by looking at the trend (by comparing the values to previous year(s) we can see if an occupation has increased in demand then we can show occupations / jobs that match the student and the have a high demand
- The Average salary was taken from each occupation then checking if it increased or decreased then getting an average helps getting a realistic output from the AI model which motivates users

Inner Mechanics

- Azure was Used in order to prove that the using an AI that is capable of being modified without breaking the structure and cloud storage(for the database) for future proofing is possible
- Open AI was used as a proof of concept to how the AI will react to answers given the data
- AI Communicates with the datasets and is given specific instructions via “prompt engineering” to Follow the datasets give Closely and given the context of it being an educational tool, this is done so that the amount of outputs given are from a set of pools and it allows the AI to logically profile the user’s personality (+other attributes) to match with the perfect jobs

Future Features

- Integrate it with school DB’s so that it can take the users test results, academic performance, teacher inputs then integrate them into the (let’s get better function)
- Integrating it with big organizations such as google, Microsoft or amazon so that the courses can be free for students
- Creating an ML model that can predict the future salaries & demand for occupation by working with a larger more detailed dataset provided by the government.
- Finetuning The AI model Based off of user feedback.
- As a proof of concept for future integration (using azure) an ml model that predicts the occupation based off scores (in the personality test + test score)

ML model description:

K-Nearest Neighbors (KNN) model predicts the best job category for a user based on their personality traits. It helps match job seekers to careers where they are most likely to succeed using AI.

How It Works:

1. Dataset:

- We created a sample dataset with personality traits (Openness, Conscientiousness, Extraversion, Agreeableness, Neuroticism).
- Each row represents a person's trait scores and their best-matched job category (e.g., Marketing, Software Engineer, Data Scientist)

Prediction Process:

- A new user enters their personality scores (from a test).
- The KNN model finds similar users in the dataset.
- It recommends a job category based on the most common

Github Repo: