Team Project: Job Role Recommendation App

Project Overview

In this project, you will develop a web application that analyses job holder profiles and provides profile-based recommendation of job roles in the Data Science and Machine Learning (DSML) sector.

Please note that details are subject to changes if there is a need. Changes will be communicated to you in the class or via Canvas Announcements.

Team Composition

You will work in a team of **maximum 3 students** for the project.

As the project work involves a range of skillsets (data analysis, web development, report / documentation, presentation, leadership), you should try to have team members who can complement each other's strengths in these areas.

Development Framework

You will use **Streamlit** to develop the web app for this project.

Streamlit is an open-source Python library that is very easy to use and allows you to create and deploy data apps quickly. You can find step-by-step guides on the Streamlit documentation page.

You may decide on your preferred development / deployment platform. For milestone deliverables, you will be asked to upload the **zip file containing your code base** for submission on Canvas.

For example:

- You may use Streamlit Community Cloud linked to your GitHub repository. If you do this, please make sure to **keep your GitHub repo private**, and be aware of the limitations.
- You may use Anaconda and deploy your app locally.

Dataset

This project will use combined datasets from *Kaggle's Annual Machine Learning and Data Science Survey* from the past three years (2020-2022).

Download **all three datasets** from their respective sources below. You will need a Kaggle account and may need to click "Join the competition" to access the data.

- 2022 survey
- 2021 survey
- 2020 survey

Your team should determine and perform **any required data cleaning and preparation** to use these datasets for your app.

Project Tasks

There are **four milestones** for this project.

The following elaboration provides you with certain specifications to meet, but also leaves certain project parameters for you to decide on. There are no absolute "correct answers" for these parameters. You are expected to explore the dataset and decide on parameters that provide meaningful insights into the data.

(1) Exploratory data analysis

Build the first version of your app that allows users to visualize and explore the data to derive insights.

Use Streamlit's multipage app feature to offer 2 visualization functionalities:

(1.1) Trend over time

Show the trend of *<x>* over the years 2020-2022, where *<x>* is selectable by users.

- Provide users with **at least 2 options** for *<x>*, which are features of your choice taken from the dataset. They should offer meaningful insights into the data. (You may check out the Code section in the Kaggle dataset sources for inspiration.)
- You may define a sensible scope (e.g., by country, gender, age, ...) relevant to *<x>* for the visualization. In other words, it is not a must to capture all survey participants for every visualization. This scope (if any) must be made explicitly clear in the visualization.

(1.2) Statistical tendencies

Show [statistics] of *<y>* in *<period>*, where *<y>* and *<period>* are selectable by users.

- Provide users with **at least 2 options** for *<y>*, which are features that you select from the dataset. These may or may not be the same features as *<x>* in (1.1).
- Users shall be able to set *<period>* to:
  - Any single year (e.g., 2020);
  - Any two consecutive years (e.g., 2020-2021);
  - All three years within the dataset.
- You may decide on the statistical measure and the scope (see point (1.1)) that will offer meaningful insight.

You will be asked to discuss the insights from these visualizations in your final report and presentation (Milestones 3 and 4).

You will also give a live demo of the app in class by this milestone.

(2) Recommendation system

Add a recommendation functionality to your app.

- Allow app users to input their personal / professional profiles, and show recommended job roles based on the similarity of their profiles with those in the dataset.
- Your team shall determine what features to collect via user input, to perform the recommendation with reasonable accuracy.

You should well document your recommendation methodology for inclusion in your final report and presentation (Milestones 3 and 4).

You will also give a live demo of the app in class by this milestone.

(3) Presentation (recording)

Prepare a video recording of your project presentation. Your team may choose one or more team members to present.

The presentation is essentially **an articulation of your final report** and should cover:

- Dataset and preprocessing
- Exploratory analysis insights (Milestone 1)
- Approach for the job role recommendation task (Milestone 2)
- Reflection and conclusion

Visualizations from the app could be used to support the points you discuss in the presentation. However, the presentation **should not be about your app development process nor a demo of your app functionalities**. *Separate opportunities will be arranged for the app demo.*

The video duration should not exceed 10 minutes.

(4) Final report

The report should be a single PDF document not exceeding 8 pages.

The report should **present your data analysis methodology and not the app development**. Follow this outline:

- 1 cover page with project title and listing of team members (name and student ID)
- 1 page describing the dataset and preprocessing
- 2 pages describing exploratory analysis insights from Milestone 1
- 2 pages describing your approach for the recommendation task from Milestone 2
- 1 page of reflection and conclusion
- 1 page of references

Do not include screenshots of the code base in the report. If you need to include app screenshots, do so **sparingly**.

Deliverables and Deadlines

Milestone / Tasks	Deliverables	Due Date
0: Form project teams	Signed up for project groups on *Canvas > People > Groups*	27 Aug 2023 23:59
1: Exploratory data analysis	**The zip file of your code base**, implementing Milestone 1 specifications, submitted on *Canvas > Assignment > Project Milestone 1*	01 Oct 2023 23:59 *Demo:* 5 Oct 2023 18:30
2: Recommendation system	**The zip file of your code base**, implementing Milestone 2 specifications, submitted on *Canvas > Assignment > Project Milestone 2*	19 Nov 2023 23:59 *Demo:* 17 Nov 2023 18:30
3: Presentation	**The video recording of your presentation** submitted on *Canvas > Assignment > Project Milestone 3*	19 Nov 2023 23:59
4: Final report	**Final report in PDF format** submitted on *Canvas > Assignment > Project Milestone 4*	19 Nov 2023 23:59

Grading Components

Component	Weightage
Working web app with data visualization interface (Milestone 1)	5%
Data visualization quality (Milestone 1)	5%
Working web app with recommendation interface (Milestone 2)	5%
Recommendation quality (Milestone 2)	5%
Presentation (Milestone 3)	5%
Final report (Milestone 4)	5%
Total	**30%**

All the best!