UNIVERSITY OF HELSINKI

DATA11001

Introduction to Data Science

University of Helsinki

Prepared by: Gayathri Senanayake, Dilusha Senarathna, Kavya Atapattu

Submitted to:

Github link to project

Application link:DS Job Market

Date: 28/10/2024

UNIVERSITY OF HELSINKI

1. Introduction

1.1 Background

The rapid growth of data science has led to an increasing demand for skilled professionals. However, understanding specific job requirements and matching job seekers to relevant roles remains challenging. This project aims to address these challenges by developing a Job Analysis Dashboard tailored to the Helsinki area. Unlike general job boards, this tool focuses on the unique needs of data science students and professionals, offering localized job recommendations, skill assessments, and actionable career insights.

1.2 Problem Statement

Currently, job seekers face difficulties in identifying the most relevant job roles and necessary skills due to a lack of focused, data-driven job analysis tools. The main issues this dashboard addresses include:

- Lack of specialized job recommendations for data science roles.
- Limited tools for identifying the skill gaps necessary for job market readiness.
- Insufficient localized data to reflect real-time demand and trends within the Helsinki job market.

1.3 Objectives

The Job Analysis Dashboard aims to:

- Provide tailored job recommendations based on users' skills and preferences.
- Highlight necessary skill improvements to align with industry standards.
- Enable users to interactively explore job trends and qualifications specific to the Helsinki area.

1.4 Scope of the Report

This report will cover the methodologies and techniques used in data collection, preprocessing, and model development for the dashboard. It also includes evaluations of the machine learning algorithms, ethical considerations, and an analysis of user interaction design. Any limitations related to data availability, geographic focus, and model assumptions will also be discussed.

UNIVERSITY OF HELSINKI

2. Methodology

2.1 Data Collection

Data for this project was gathered using LinkedIn job postings, collected through the Apify LinkedIn job scraper. The scraper retrieved job listings over a two-month period, focusing on data science positions within the Helsinki area. The dataset includes details on job titles, required skills, experience levels, locations, and job descriptions, providing a robust basis for analyzing current trends and skill demands in data science. The inbuilt Apify web scraper was used due to ethical reasons it was not possible to create a customized web scraper without taking permission from linkedin. Considering the time and effort using the Apify LinkedIn job scraper was the best solution.

2.2 Data Preprocessing

To ensure data quality and relevance, several preprocessing steps were applied:

- **Data Cleaning:** Duplicate entries and rows with null values were removed to streamline the dataset.
- **Feature Engineering:** New variables were derived from job descriptions, such as extracted programming languages, job seniority, and company industry. Language translation was also performed to standardize Finnish job descriptions into English, ensuring consistent interpretation.
- **Normalization and Encoding:** Text data, such as job descriptions, was vectorized using TF-IDF and cosine similarity to match user profiles with job listings. Categorical variables were encoded, and numerical data was normalized for model input.

Data processing codes can be accessed through github using Data_Preprocessing.

3. Platform Development and Dashboard

The end product, a user-friendly dashboard, was developed using **Streamlit and Python** to deliver an interactive job recommendation experience. The target audience is Data Science Students and Job seekers in the DS job market. All data is stored in CSV files, supporting easy access and integration with the Streamlit app. The dashboard allows users to view tailored job matches, gain insights into required skills, and interact with dynamic data visualizations.

- **Framework**: Developed using Streamlit and Python, providing a lightweight, web-based interface.
- **Data Storage**: CSV files were used to store job postings and processed user data.
- **User Interaction**: Users can input their CV details, skills, and preferences to receive recommendations, with visualizations highlighting top matches and skill gaps.

UNIVERSITY OF HELSINKI

Our dashboard, accessible through DS Job Market, comprises three tabs: “Key Insights”, “EDA” ,“Job Recommendation” and “About Us”.

Key Insights

This tab provides users with a summary of key job market trends for data science roles in the Helsinki region. It includes high-level statistics on job demand, common required skills, and preferred experience levels, offering users a quick overview of the job market landscape.

EDA (Exploratory Data Analysis)

The EDA tab allows users to explore the dataset in detail, with visualizations that highlight data distributions, correlations, and trends. This helps users understand how factors like job location, title, and skill requirements vary within the Helsinki job market.

Job Recommendation

In this tab, users can input their skills and preferences to receive personalized job recommendations. The system analyzes the input against available job data to present the top matches, along with suggestions for skill enhancements, making this tab highly interactive and tailored to the user’s profile.

4. Limitations and Future Work

The project evolved from its initial plan, with key changes made to enhance usability and performance:

- **Data Complexity**: Managing and preprocessing unstructured text data proved to be a challenge but ultimately allowed us to enhance the recommendation precision.
- **Feedback Mechanism**: Incorporating user feedback for continuous model improvement is a future goal, allowing the dashboard to stay aligned with changing job market demands.
- **Update drop down selection:** To get user inputs there are some selected options given but some users can have skills which are not included in that list. As future work to the list there will be an option “Other” to type whatever skills the user has and the more. More options need to be added to the list.

Reflecting on the project, we recognized the importance of data quality and preprocessing, especially for NLP applications. This experience underscored the value of an iterative development approach, allowing us to adapt and improve based on user needs and technical constraints.

UNIVERSITY OF HELSINKI

5. Conclusion

The "DS Job Analysis Dashboard" provides an innovative solution for job seekers in the data science field, offering tailored insights and recommendations that address both job matching and skill gaps. This project highlights the potential of data science in career services and job market analysis, with further improvements anticipated to enhance the tool’s responsiveness and adaptability to evolving job market trends.