Computer and Information Systems Project

Kyle Sinclair

3538597

COMP 495

26 Nov 2023

Project Proposal Draft

# Introduction

In a time of rapid technological advancement, particularly in machine learning and artificial intelligence, the impact on the economy and job market is significant. This has led to evolving skill sets and knowledge requirements in tech careers. This project aims to capture and analyze hiring data to build a database that tracks these changes in the tech job market. It will not only capture historical trends but identify current trends, including geographical variations within Canada.

Take, for example, the role of a data analyst, which is evolving due to new technologies. This project will explore such changes, examining skills now needed to excel in these roles and regional job market differences, such as between Calgary and Vancouver. Ultimately, this project seeks to provide insights and guidance for professionals adapting to this dynamic field.

# Project Objectives

The primary aim of this project is to analyze the changing technology job market, focusing on how skills and qualifications are evolving due to advancements in areas like machine learning and artificial intelligence. The project involves systematically scraping job postings from Indeed, especially in tech, and creating a scalable database to efficiently manage this data.

A key aspect of the project is developing an intuitive, user-friendly interface to make the data accessible and useful to a wide audience, including job seekers and industry experts. This interface will play a vital role in ensuring the findings are easily understandable.

Overall, this project aims to provide insights into the current and future state of the tech job market, offering a valuable tool for career planning and workforce development in the rapidly changing tech landscape.

# Methods/Technical Strategy

The methods and technical strategies for this project are carefully crafted to analyze the evolving tech job market, focusing on data acquisition through web scraping and in-depth analysis and visualization. The initial phase involves collecting job postings from Indeed using Python and its libraries, ensuring not only data quantity but also quality and relevance. This is followed by data cleaning and preprocessing to transform unstructured web data into a format ready for analysis, utilizing Python’s Pandas library for tasks like text standardization and categorization.

A key component of the project is developing a user-friendly interface, allowing users to interact with and understand the analyzed data. This interface design will be grounded in human-computer interaction principles and user experience design.

Throughout the project, ethical and legal considerations in web scraping, data privacy, and security will be paramount, ensuring responsible research practices. The project aims to blend advanced web scraping, thorough data processing, detailed analysis, and accessible interface design to offer a holistic view of the tech job market's changing trends.

# Required Materials

For the successful execution of the project, a range of essential materials and resources will be utilized, encompassing software tools, database systems, and development technologies.

**Software Tools for Data Collection and Analysis:** Python will be the primary programming language, supported by libraries like BeautifulSoup and Selenium for web scraping, and Pandas for data cleaning and processing. These tools are critical for extracting, transforming, and structuring the data from online job listings.

**Database Management System:** A SQL-based system will be employed for efficient data storage and management. It offers the required scalability and robustness for handling and querying large datasets.

**Data Visualization Tools:** Python’s visualization libraries, Matplotlib and Seaborn, will be utilized to create insightful graphical representations of the analyzed data. These tools are key to understanding and communicating the trends and patterns within the job market data.

**User Interface Development:** The interface will be developed using Python frameworks like Django, coupled with frontend technologies such as HTML, CSS, and JavaScript. This combination ensures a responsive and intuitive user experience.

**Supplementary Learning Resources:** Online courses, tutorials, and forums will be accessed as needed to supplement knowledge in areas such as Python, advanced data analysis, and web development.

# Project Milestones

**Initial Planning and Setup (10 hours):** The project begins with a foundational phase where objectives are solidified, tools are selected, and a comprehensive project plan is developed. This phase also includes a review of ethical and legal considerations relevant to web scraping and data handling.

**Data Collection and Preprocessing (30 hours):** Following the setup, the focus shifts to data collection. Using Python, BeautifulSoup, and Selenium, job postings from Indeed will be scraped. The data will then undergo cleaning and preprocessing using Python's Pandas library to ensure it is analysis-ready. This stage is critical for establishing a robust dataset for subsequent analysis.

**Data Analysis (30 hours):** With the data prepared, an extensive analysis will be conducted. This involves not only exploring and identifying trends but also employing statistical methods to gain deeper insights into the evolving tech job market.

**User Interface Development and Final Reporting (30 hours):** The final phase involves developing a user-friendly interface, ensuring the findings are accessible and understandable. Concurrently, a comprehensive final report will be compiled, encapsulating the entire project from methodology through to conclusions.

**Buffer Time (20 hours):** Additional time is allocated to accommodate unforeseen challenges, further research, or additional refinements, ensuring flexibility and adaptability throughout the project timeline.

# Project Deliverables

**1. Data Collection Scripts:** The project will produce a series of Python scripts used for web scraping. These scripts, developed using libraries like BeautifulSoup and Selenium, will serve as the primary tool for extracting job posting data from Indeed. These scripts are not only crucial for the initial data collection but also serve as a resource for future data scraping endeavors.

**2. Processed Data Set:** A key deliverable is the cleaned and structured dataset, derived from the raw data collected through scraping. This dataset will be processed using Python's Pandas library, ensuring it is in a format suitable for detailed analysis. This dataset will form the foundation for all subsequent analysis and will be a valuable resource for understanding current trends in the tech job market.

**3. User Interface:** An interactive user interface will be developed to allow users to easily engage with the findings of the analysis. This interface will be designed considering principles of human-computer interaction and usability, ensuring that the insights are presented in a user-friendly manner.

**4. Final Project Report:** The comprehensive final report will document the entire project, from the initial planning stages through to the final outcomes. This report will cover the methodologies used, the analysis conducted, the results obtained, and the conclusions drawn. It will serve as a complete record of the project, highlighting its contributions to understanding the tech job market and its relevance to the fields of data science and machine learning.

# Conclusion

In conclusion, this project aims to provide a deep understanding of the evolving technology job market, highlighting shifts in skills and qualifications due to advancements in fields like machine learning and artificial intelligence. By systematically collecting and analyzing job posting data, the project will deliver valuable insights through a range of outputs including data scripts, processed datasets, a user-friendly interface, and a comprehensive final report.

These deliverables will not only fulfill academic objectives but also offer practical guidance for professionals and students in technology fields. The project underscores the practical application of skills from computer science, data analysis, and emphasizes the importance of ethical data handling and legal compliance.

Overall, this endeavor aims to equip individuals with the necessary knowledge to navigate the dynamic tech job market, providing a bridge between academic learning and real-world application. It is anticipated that the outcomes of this project will contribute meaningfully to both academic discussions and career development in the rapidly changing technological landscape.