Date: 05/10/2024

PROJ-CS781: Mid-term Review

Project Title: Employability Prediction Based on Personality Traits and Cognitive Analysis obtained from Social Media Profiles	Group: 26

Team (Roll / Name):
    1. Pinaki Banerjee (13000121003)
    2. Debanjan Konar (13000121006)
    3. Anuvab Chakravarty (13000121036)
    4. Soumyajit Dey Sarkar (13000221080)
	Guide: Mrs. Nairanjana Chowdhury

Questions:

1. Justify the uniqueness of your project.

Ans: The hiring process, regardless of an organization's size, is a subject that continuously sparks intense debate and scrutiny. Determining "employability"—or more specifically, identifying the qualities that make an individual truly "employable"—remains a complex and challenging endeavor. Despite the widespread use of modern tools like online resumes, LinkedIn profiles, virtual interviews, and take-home assignments, the overall hiring process is still inherently flawed and riddled with inefficiencies. Human biases and errors in judgment often lead to less-than-optimal hiring decisions.

This project seeks to address these issues by streamlining the hiring process through the introduction of an intelligent prediction tool. The tool is designed to narrow the search space for recruiters and provide data-driven insights, ultimately making the process more efficient and reducing the likelihood of errors. By leveraging advanced analytics and personality trait predictions, this tool aims to serve as a supplementary aid that enhances decision-making in hiring, helping organizations identify candidates who are a better fit for specific roles.

2. What are the expected benefits from your project?

Ans: The expected benefits from our project are plentiful. The most notable among them are:

- **Improved Hiring Accuracy**: The tool uses data-driven models based on personality traits and cognitive abilities, reducing human bias and improving the accuracy of hiring decisions by matching candidates more precisely with job roles.

- **Streamlined Hiring Process**: By leveraging social media profile data and automating the personality analysis process, the tool significantly reduces the time and effort spent by HR teams in evaluating candidates.
- **Enhanced Candidate-Role Fit**: The prediction tool helps map individuals' personality traits and cognitive abilities to the requirements of specific job roles, increasing the likelihood of better long-term job performance and satisfaction.
- **Reduction in Hiring Costs**: Automation and the use of predictive analytics reduce the need for extensive manpower in entry-level and bulk hiring processes, thus lowering operational costs for organizations.
- **Unbiased Hiring Decisions**: By focusing on objective data such as personality and cognitive skills, the tool minimizes unconscious biases, leading to fairer and more inclusive hiring practices.
- **Data-Driven Career Counselling**: The tool can be used by career counsellors to provide students and professionals with personalized career recommendations based on their social media activity and personality analysis.
- **Scalability Across Industries**: The hiring metric can be customized for different industries, job roles, and skill sets, making the tool adaptable for various sectors and job types, from technical roles to management.
- **Improved Job Satisfaction and Retention**: By ensuring better alignment between personality traits and job requirements, the tool can lead to higher employee engagement and retention rates, reducing turnover costs.
- **Continuous Learning and Development**: The tool can identify gaps in personality or cognitive skills, enabling employers to channel their training programs more effectively and employees to focus on areas for self-improvement.
- **Future-Proof Hiring**: As personality and cognitive skills remain relevant even as technical skills evolve, this tool ensures that organizations are hiring adaptable and long-term fit candidates who can grow within the company.

3. a) State the findings from your analysis till date. Show your analysis documentation in PRD till date following the template.

Ans: **Find attached PRD document for analysis documentation.** The findings from the analysis are as follows:

The analysis phase has progressed with the successful completion of key functional components, including automated social media data collection, post preprocessing, and machine learning model integration for personality trait predictions. The core browser extension has also been developed, enabling seamless data capture from platforms like Facebook. However, work remains on developing the hiring metric and linking cognitive abilities to employability. The non-functional requirements such as scalability and

response time optimizations are currently underway. Overall, the project is approximately 30% complete, with major foundational features implemented.

b) What is the % completion progress of analysis? Explain the calculation logic.

Ans: **Currently the analysis phase is completed by 30%.**

Hiring Metric Formulation and **Linking Cognitive Abilities:** Together, these tasks represent **66%** of the project scope. The remaining tasks (data collection, preprocessing, model integration, API integration, etc.) constitute **34%** of the project scope.

Further details can be found in the attached Project Plan document.

4. a) State the findings from your design till date. Show your design documentation in PRD till date following the template.

Ans: **The design documentation has been attached to this submission. It can also be found in the PRD document.**
The findings from the design of the employability prediction tool to date highlight key advancements in integrating social media data collection and personality trait prediction. The design includes automated post collection using browser automation tools like Selenium, preprocessing of collected data for analysis, and the use of models like the Big 5 and MBTI for personality classification. Additionally, the tool is structured as a SaaS platform to support employers in employability assessments. Security and data privacy are prioritized, and future enhancements focus on scalability and accuracy improvements.

b) What is the % completion progress of design? Explain the calculation logic.

Ans: **Currently the design phase is complete by at least 50%.**

The currently working prototype can already successfully extract content from a social media in an automated fashion and perform personality classification based on MBTI. Moreover, we have a high-level design that can relate employability to personality types which remains to be implemented. The only perspective that remains is to successfully relate employability to cognitive ability and formulate a customizable hiring metric based on that.

5. State the tools you are using for analysis and design; apply Software Engineering concepts.

Ans: Some of the major tools that we are using for analysis and design are as follows:

- **StarUML**: Used for creating **UML diagrams** like class, use case, and sequence diagrams to model system architecture and design.

- **VS Code**: An IDE used for **coding, debugging, and refactoring**, ensuring alignment with design patterns and software engineering principles.
- **Google Colab**: For cloud-based execution when working with larger datasets or GPU-based model training.
- **Dia**: Utilized for creating **flowcharts and DFDs** to represent data flow and process structure in the system design.
- **Microsoft Project**: Applied for **project management**, task scheduling, and resource allocation using Gantt charts and Agile planning techniques.

6. Are you foreseeing any risk in completing your project?

Ans: These are some identifiable risks that present themselves:

- **Data Privacy and Legal Concerns**: Analyzing social media profiles for employability can raise privacy and ethical issues, especially concerning user consent and compliance with regulations like GDPR and CCPA. Failure to handle data securely could lead to legal consequences.
- **Bias in Predictions**: The machine learning models may inherit biases from the training data, leading to unfair or skewed predictions for certain demographics or personality traits. Ensuring the model is free from bias is crucial for fairness.
- **Accuracy and Over-Simplification**: Personality traits alone may not provide a complete picture of a candidate's qualifications, potentially overlooking important factors like soft skills or professional experience. This could lead to inaccurate or overly simplified assessments.
- **Data Availability and Quality**: Relying on social media data for personality analysis could be problematic if the data is incomplete, curated, or if users aren't active on social platforms, leading to inaccurate predictions.
- **Adoption and Trust**: Employers might be hesitant to adopt or trust the tool fully, as they may be skeptical about relying on automated predictions for hiring decisions. Demonstrating the tool's accuracy and effectiveness will be essential for market acceptance.

7. State the study references you have used in this semester (after Synopsis preparation)

Ans: Some of the major references that we have used in this semester for research purposes in the design phase are as follows:

- [1] Personality Classification Model of Social Network Profiles based on their Activities and Contents by Mervat Ragab Bakry, Mona Mohamed Nasr, Fahad Kamal Alsheref on International Journal of Advanced Computer Science and Applications,

Vol. 13, No. 7, 2022, https://thesai.org/Downloads/Volume13No7/Paper_3-Personality_Classification_Model_of_Social_Network.pdf

- [2] Souri, A., Hosseinpour, S. & Rahmani, A.M. Personality classification based on profiles of social networks' users and the five-factor model of personality. Hum. Cent. Comput. Inf. Sci. 8, 24 (2018). https://doi.org/10.1186/s13673-018-0147-4
- [3] Hernandez, Rayne, Knight, Ian Scott – "Predicting Myers-Briggs Type Indicator with Text Classification", https://web.stanford.edu/class/archive/cs/cs224n/cs224n.1184/reports/6839354.pdf
- [4] Öngöre, Özgür. (2014). A Study of Relationship between Personality Traits and Job Engagement. Procedia - Social and Behavioral Sciences. 141. 1315-1319. 10.1016/j.sbspro.2014.05.226. https://www.researchgate.net/publication/265336122_A_Study_of_Relationship_between_Personality_Traits_and_Job_Engagement
- [5] GRADUATE EMPLOYABILITY QUALITIES AND PERSONALITY PREFERENCE AS DETERMINANTS OF JOB PERFORMANCE IN NIGERIA by Nwogu, Momoh on European Scientific Journal September 2015 edition vol.11, No.25 , https://core.ac.uk/download/pdf/236411658.pdf

8. Submission of RM (excel file) and PP (Microsoft Project Plan) as separate attachments following the given templates. Show relevant parts of PRD prepared till date following the template .

Ans: **Requirements Analysis Matrix (RM) and Project Plan (PP) prepared using MS Projects have been included in the submission. PRD document has also been included with all the relevant sections filled.**

9. What is the % completion progress of the prototype (refer target for 7th Semester as set by Guide)? Explain the calculation logic.

Ans: **We would put the completion status of the prototype at 25%.**

The currently working prototype can successfully extract content from a social media in an automated fashion and perform personality classification based on MBTI. However, in order to devise a solution that can relate employability to personality types, we need to train a model to relate various job roles to their corresponding personality types. The latter section of the project would also deal with relating cognitive abilities with employability for which the design hasn't been finalized yet.

10. Additionally, Guides should ask questions (what's, how-to's) on understanding of the target system and expected functions.

Ans: The what and how-to questions put forward by our guide and their answers are as follows:

1. What are the main goals of the prediction tool for employability?

- **Answer:** The main goal of the prediction tool is to reduce the complexity of the hiring process by analyzing social media posts to predict an individual's personality traits. By leveraging models such as the Big 5 and MBTI, the tool will aid in determining employability and cognitive fit for specific job roles, helping employers make more accurate hiring decisions.

2. How will the social media posts be collected and processed for analysis?

- **Answer:** The system will use browser automation tools (like Selenium or Puppeteer) to capture posts from social media profiles, such as Facebook. After collection, the posts will be automatically preprocessed (removing stopwords, performing lemmatization, etc.) to make the data suitable for personality trait predictions. These posts will then be passed to trained machine learning models for analysis.