University of Ibadan Journal of Science and Logics in ICT Research (UIJSLICTR)

ISSN: 2714-3627

A Journal of the Faculty of Computing, University of Ibadan, Ibadan, Nigeria

Volume 17 No. 1, June 2026

journals.ui.edu.ng/uijslictr
http://uijslictr.org.ng/
uijslictr@gmail.com

University of Ibadan
Journal of Science and Logics in ICT Research

Predicting Information Technology Career Paths Using Stacking Ensemble Learning Model

[1]✉ Udosen A. A., [2]Okoro A., [3]Oduba T., [4]Awodele S., [5]Alimi S. and [6]Nteziryayo, D.

[1,2,3,4]Department of Computer Science, Babcock University, Ilishan-Remo, Ogun State, Nigeria.
[5]Blue Oak Technologies Ltd., Lagos State, Nigeria.
[6]Department of Computer Science, Adventist University of Central Africa (AUCA), Kigali, Rwanda
Emails of authors: udosena@babcock.edu.ng, okoro6473@student.babcock.edu.ng, oduba9852@student.babcock.edu.ng, awodeles@babcock.edu.ng,[5]sheriff.alimi@blueoaktechnologiae.com, deogratias.nteziryayo@auca.ac.rw

Abstract

Selecting a profession route in the fast-changing Information Technology (IT) industry is now an even more complicated issue among students and job applicants, and this has resulted to job dissatisfaction and skill gaps. The conventional counselling practices are often based/ on subjectivity and are not capable of examining the large number of modern tech positions and the individual ability of a person. This paper outlines the creation of an IT Career Prediction System that was developed to deliver recommendations based on personal data, using technical skills, including Database Fundamentals and AI/ML, and psychological traits, including Openness and Conscientiousness. A stacking technique was employed to implement an ensemble machine learning approach in order to achieve high predictive reliability by combining four base classifiers, such as Support Vector Machine (SVM), Decision Tree (DT), Random Forest Classifier (RFC), and Naive Bayes (NB). The developed ensemble learning method combined the computational powers of the four based classifiers used with a performance accuracy of 99.1%, and was effective by applying interpretability and balancing the risk of overfitting when giving practical advice. Hence, the paper concludes that the automated system provides a scalable, objective, and accurate alternative to conventional career guidance, which ends up enhancing more effective career alignment and long-term professional success in the IT industry.

Keywords: *Career prediction, Ensemble learning, Machine learning, Stacking, Skill assessment, Support Vector Machine*

1. Introduction

Students and job seekers in the modern competitive and dynamic job market are overwhelmed by a plethora of career choices, and the amount of the choices offered makes the process of decision-making difficult, but making informed choices regarding their careers is vital to the individuals involved. Traditional career guidance methods, including counselling, often are not capable of capturing an individual's unique strengths, interests, and potentials. Categorically, these approaches typically depend on standardized tests and general advice, which may not help with the specific aspirations and capabilities of each individual [1].

Udosen A. A., Okoro A., Oduba T., [4]Awodele S., Alimi S. and Nteziryayo, D. (2026). Predicting Information Technology Career Paths Using Stacking Ensemble Learning Model. *University of Ibadan Journal of Science and Logics in ICT Research (UIJSLICTR),* Vol. 17 No.1, pp. 22 - 30

Career paths have a historical background around the 20th century when the introduction of psychological and vocational tests was made in order to assist individuals choose the right career paths among their interests, talents, and personality traits [2]. These tests were meant to match individuals with careers that would result in job satisfaction and productivity. But the success of such traditional practices has been subjected to examination, and research indicates that a high level of job dissatisfaction, low performance, and career mismatches is prevalent as a result of the old or simplistic methods of assessment [3][4]. Conversely, this developed system is targeted to youths and IT professionals who are unsure of the career path to follow; resulting to a cycle of unhappiness and poor performance.

The limitations to traditional career guidance include that career counsellors may rely on personal judgment, stereotypes, or limited exposure, leading to biased recommendations. Furthermore, this traditional guidance depends on interviews, observations, and paper-based assessments, which cannot process large amounts of data [5][6]. However, the limitations of traditional career guidance have brought about the development and adoption of digital technologies, particularly big data analytics, which revolutionize the field of career prediction by providing more objective insights and personalized recommendations based on extensive data analysis

In order to make proper career predictions, various relevant factors were put into consideration. A Mendeley dataset was used to train the model to be able to determine the best career match of a person depending on their unique profile gotten through a skill assessment form that measures the level of proficiency in various fields of the user. Basically, these career path prediction models are based on data analytics, machine learning, and artificial intelligence that shows the future of a person based on their education, skills, past experience, and current market trends [5][6].

The dataset, CareerMap - Mapping Tech Roles with Personality & Skills, used for the model training was published on April 3, 2023. It is a complete resource that is intended to be used by AI based career guidance systems. With around 9,180 records and 28 features, including technical abilities such as Database Fundamentals and AI/ML, and psychological features such as Openness and Conscientiousness, the dataset helps in researchers in finding the right career path in tech by matching personal skills and personality with the job description through the development of machine learning models that reduces the dimensions of the dataset to the most important attributes needed to obtain an accurate prediction.

Hence, this study aims at developing an elaborate IT career path prediction tool which can guide the users in the most appropriate careers that not only best fit their skills, but also give them job satisfaction, and these systems are likely to be embraced more by individuals, institutions, and organizations that are concerned with human capacity development. By leveraging data-driven techniques and personalized skill assessments, the system will offer tailored recommendations that go beyond traditional methods, best suited for the user and ultimately increasing job satisfaction and long-term career success.

The use of the system is likely to continue to expand as educational institutions and organizations begin to see the benefits of these innovative tools on an individual basis, impacting the future of career development and decision making. Hence, the system developed in this study will offer tailored recommendation that go beyond traditional methods with respect to career path prediction.

2. Related Works

Career forecasting has been a domain where the technology of Machine Learning (ML) has been used to move away from once simple and conventional rule-based methodology to the more advanced predictive system. In this part, the modern systems are reviewed to identify the technical gaps which the proposed stacking ensemble learning model (SELM) aims to overcome.

A. Comparative Analysis of the Existing Career Prediction Models

There have been a number of studies that examine how to predict career path by combining student attributes with algorithmic prediction. Among those, one is Deshpande et al. [7] which had proposed a guidance system through aligning career path with the interests and abilities of students. Their work highlighted the need to examine in detail the career choice potentials, but the study did not provide information on the provenance of the data and the requisite preprocessing steps. But it is hard to judge the reliability of the results due to the lack of data transparency, which is solved by the current study, since it uses the well-documented data set.

VidyaShreeram and Muthukumaravel [8] and Holland [2] compared the performances of the general guidance systems. Their research showed that the ML techniques are more effective than the conventional questionnaire-based techniques and that the highest accuracy was obtained by the Random Forest (RF) classifiers (93%). However, their model's features, such as parental education and family economy focused heavily on socio-economic status rather than the specific technical competencies required for IT roles. Furthermore, while their study highlighted the efficiency of single-model classifiers, it suggested a need for more robust architectures to handle complex, non-linear career data.

To fill the gap of model clarity, Liu et al. [9] and Locke [3] investigated Automated Machine Learning (AutoML) to predict STEM careers. Their approach applied the Logistic Regression model due to its interpretability, which enables the researcher to determine the weight of each predictive factor. Even though their system was superior in the area of feature generation and optimization, it was not externally tested on independent data. Also, the use of a single simple model such as Logistic Regression could not reflect the high-dimensional trends of the current tech skill sets.

B. Identification of the Research Gap

A review of related literature showed a consistency in trade-offs and they include:

1. Interpretability vs. Accuracy: Under certain models such as Liu et al. [9] and Locke [3], interpretability is given priority at the expense of high accuracy as in ensemble learning methods [8] [2].
2. Feature Scope: The current systems concentrated on either socio-economic aspects or the macro-level academic achievement, often ignoring the combination of psychological attributes (the Big Five personality traits) and technical abilities (proficiency in the use of structure query language (SQL) and Machine Learning algorithms).

The architecture suggested in this study will cover these gaps and will be used to implement a stacking ensemble learning architecture. Through the use of the four different base classifiers to develop an ensemble learning model, this study attained high levels of predictive accuracy and still have the interpretability needed to make actionable student career counselling.

3. Methodology

This stage outlines the methods and methodologies adopted for this study.

A. Study Design

Figure 1 is a diagrammatic representation of the system flowchart, which shows the system workflow from data collection, preprocessing, and the model training and testing. The dataset was pre-processed to handle missing values and encoded using categorical variables, creating data consistency and compatibility with machine learning algorithms.

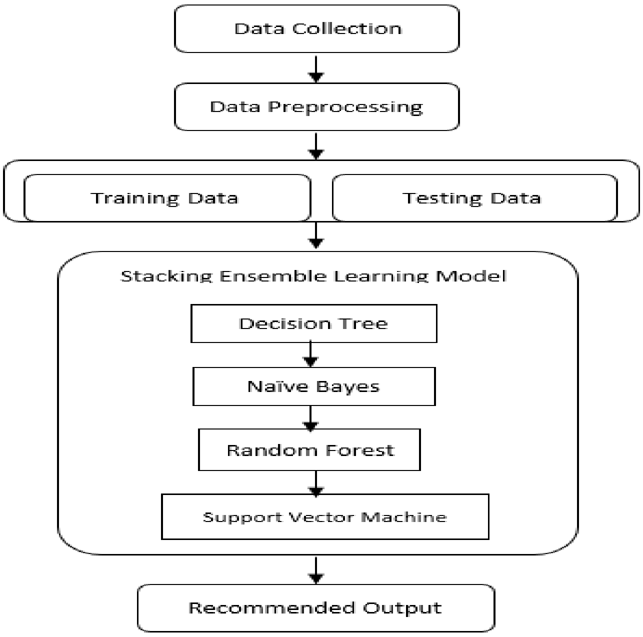

Figure 1. System Flowchart

Using the dataset, five various machine learning algorithms including Support Vector Machine (SVM), Naive Bayes (NB), Random Forest (RF), and Decision Trees (DT) were used as the meta-models for the developed ensemble learning model to train and evaluate the model in comparison to the performance of the individual models used based on four major metrics, such as accuracy, precision, recall, and F1-score.

B. Dataset

A Mendeley dataset, CareerMap - Mapping Tech Roles With Personality Skills, created by Pimplikar, Pati, Singh, and Shah [10], was used for this study. The dataset consists of about around 9,180 records with 28 features. It covers different dimensions in technology jobs as well as the personal user expertise in those areas. With this dataset, this study was able to establish the patterns and relationship between skills, personality and appropriate careers.

C. Data Preprocessing

The dataset used for this research underwent a set of preprocessing steps that would guarantee the integrity and its appropriateness for predictive modelling. The missing values were systematically dealt with to prevent potential bias and to preserve the quality of the dataset. Categorical variables are the skills, personality traits and career related variables that are encoded using numerate representation to utilize them within machine learning algorithms. In addition, the use of min-max scaling was done to normalise the features needed to improve the model’s convergence and performance. The cross validation was stratified by a factor of 10. A good classification accuracy will be achieved with the use of 90% for training and 10% for testing each fold. These preprocessing measures did not only improve the validity of the dataset, but also contributed to the development of a robust system.

D. Model Performance and Justification

Before building the predictive machine based on Pearson Correlation Analysis as in the following figure (Figure 2) which presents a Pearson correlation matrix of the various skills and traits, the dataset had to be understood.

E. Pearson Correlation Analysis

The heat map shown in Figure 2 not only points to numbers, it is an actual representation of the natural groupings of the tech world. As an example, the dark red areas (Positive Correlations) can demonstrate that such skills as Database Fundamentals and SQL tend to move in one direction almost always, similarly to how Artificial Intelligence/Machine Learning knowledge is frequently accompanied by strong python skills.

This correlation was made possible through:

1. Positive Correlations: Instead of considering the individual skills in isolation, the heat map enabled us to understand which particular competencies such as Software Engineering or Analytical Thinking actually drivers a student in a given direction. Finding the needles of the compass in the data; it is these attributes that will most likely influence where a student will get the greatest success.

2. Predictive Insights: Our ability to analyse such relationship meant that we were able to identify which bits of information were unnecessary. In a case where two skills would never contradict each other about the potential of a student to succeed in the academic world, we would be in a position to simplify the system. This is why the last model is far more rapid and remains focused on the characteristics that do make a successful career match, as opposed to being lost in the background noise.

3. Feature Selection: This was more of a quality than quantity step. Machine learning models automatically simplified the dataset by detecting, selection, and combining the features with a high correlation. This not only renders the system more computationally efficient but opens up the way to allow the algorithms to concentrate on the most distinct and powerful predictors resulting in a far more dependable outcome.

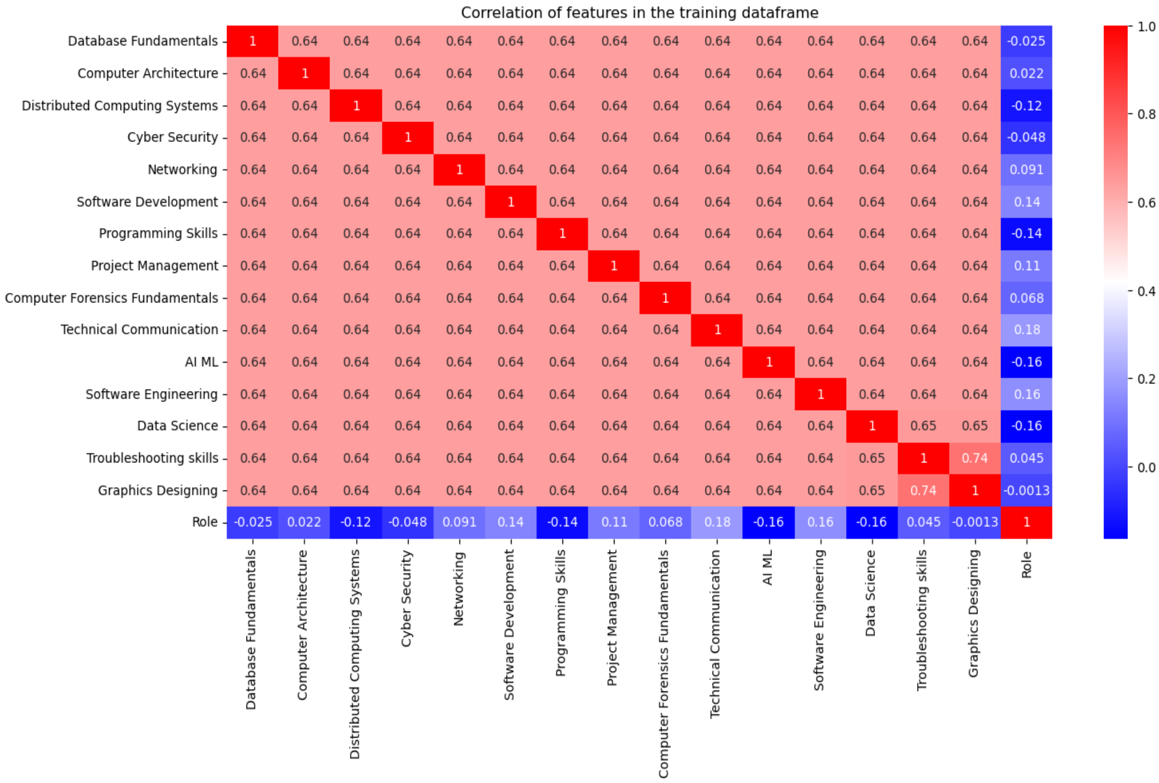

Figure 2. The heat map of the dataset

F. *Algorithmic Framework*

The developed ensemble learning model is based on a multi-model approach where different algorithms bring their own strengths to the final prediction. The justification behind the choice of these particular models is as follows:

1) Support Vector Machine (SVM): It was selected based on its outstanding better performance in high-dimensional data. It effectively differentiates the complex career categories by building the best hyperplanes.
2) Decision Tree (DT): It was selected because of its interpretable and rule-based structure. This explicability is critical in gaining user confidence in a career guidance environment.
3) Random Forest Classifier (RFC): This mitigates the errors of overfitting and enhances the generalisation of the heterogeneous dataset by combining several decision trees.
4) Naive Bayes (NB): It was proven to be very efficient in terms of its ability to handle categorical inputs, such as personality traits, because the algorithm is characterised by its probabilistic nature and relative simplicity of implementation.

The ensemble learning model for this study was developed as a stacking classifier that combines the predictive power of four different base learners, such as Random Forest, Naive Bayes, Support Vector Machine, and Decision Tree. The output from a model was passed as an input for the next model. This was done to get an optimized output from the model as the succeeding model caters to the weaknesses of the preceding model.

This specific model was trained, validated, and then serialized using python pickle library for deployment. The developed stacking ensemble learning model (SELM) is used to guarantee that the system not only identifies a career path but that it does so with a degree of accuracy and interpretability that is even greater than any single-model methods.

G. *Model Evaluation Metrics*

The performance of the models was compared under the conventional classification metrics which consisted of accuracy, precision, recall, F1-score and standard deviation (SD). These measures give a thorough analysis of the predictive ability of the models as it is given in Table 1. These metrics give a comprehensive assessment of predictive ability, reliability, and generalization for different data samples.

4. Discussion of Results

Although Support Vector Machine (SVM) was the best model with highest accuracy (98.1 %) with precision (97.2 %) and recall (98 %), its best performance shows its ability in processing high dimensional features of the skill. In completing the assessment form, and in defining clear decision lines between career categories. In terms of the performance of the Decision Tree (DT), its accuracy was reported as 97.2%, which showed high interpretability and transparency, while it was slightly inferior in terms of precision and recall when compared to other classifiers such as SVM. Also, the rule-based structure of DT makes it more useful for creating user-friendly career-recommendation rules that are easy to understand and trust. Moreover, the training of the Random Forest Classifier (RFC) gave an accuracy of 97.3%, which is better than the Decision Trees algorithm in terms of robustness and stability due to the use of ensemble learning to reduce overfitting.

It aggregated several decision trees to ensure more accurate predictions, and was appropriate for complex and varied dataset, Naïve Bayes (NB) classifier was also used, although it made the assumption of the features independence, it achieved a competitive accuracy of 96.1% with consistent precision and recall (96.4% and 98.2% respectively) in independence. It performs well on the categorical inputs like personality traits and preferences, showing that it can be useful on large data sets.

The stacking ensemble learning model (SELM) combined these base classifiers' advantages. As illustrated in the table below, the accuracy of the developed model was 96.1% and the mean values of precision and recall were 97.8% and 96.8% respectively, thus it proved to be more stable than the other base models used. The system had a more or less stable performance for all the classifiers and the prediction accuracy was not very different. This suggests that the models were well calibrated and provided good performance.

A salient feature is the overall robustness of the developed model, that is, the developed model was less prone to biases or errors from any one of the single models used by the combination of the computational power of the base classifiers.

The developed offer a unique advantage over the others in that they have been created to include interpretability features as well as prediction capabilities. Rather than just suggesting careers, these machine learning algorithms provide clarity on why a certain career was suggested, describe the skills and attributes that are compatible with that career, and provide steps on how to grow a career. The accuracy, transparency, and actionable guidance make this system a transformative tool in the realm of career counselling, moving beyond data-driven predictions to human-centric support. The use case diagram depicted in figure 3 illustrates the process of making a prediction.

This system was developed with Streamlit, a data library capable of transforming data scripts into interactive web applications, and PostgreSQL. Its database was used to store structured data like user profiles, career information and prediction logs.

5.0 System Implementation

Figure 4 shows a snapshot of the index page of the Career path Prediction System.

Table 1: Performance Classification Models

MODEL	ACCURACY	PRECISION	RECALL	F1 SCORE	STD DEV
SVM	98.1%	97.2%	98%	97.6%	1
DT	97.2%	96.3%	97.1%	96.7%	1.2
NB	96.1%	95.4%	96%	95.7%	1.5
RFC	97.3%	96.4%	98.2%	97.3%	1.1
SELM	99.1%	97.8%	96.9%	93.7%	0.9

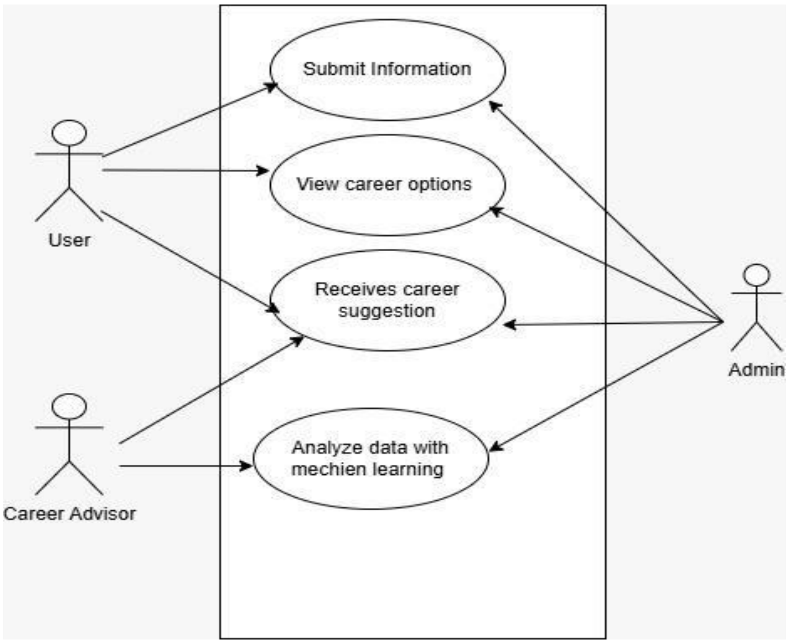

Figure 3. Use Case Diagram of the Career Path Prediction System

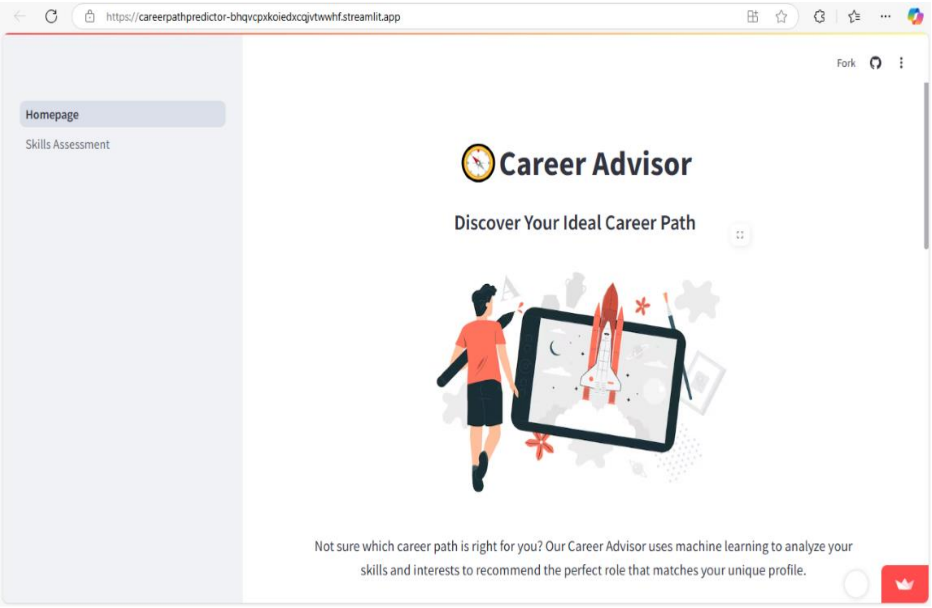

Figure 4. Homepage

The figure 5 shows a section of the skill assessment form that is to be completed by the system users.

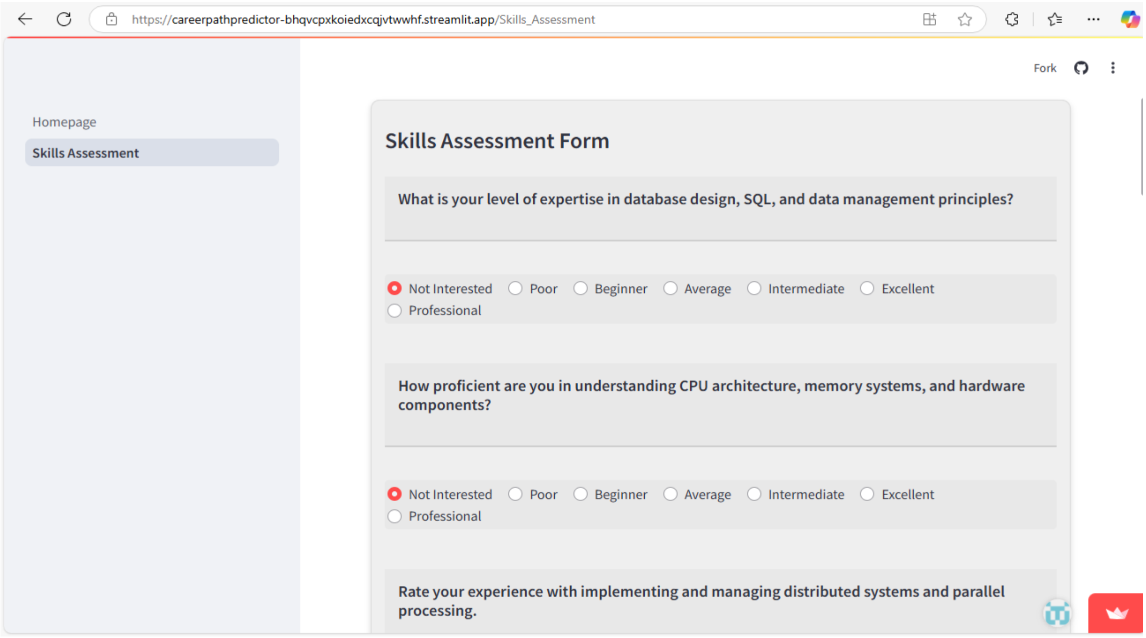

Figure 5. Skill assessment form

Figure 6 illustrates the Recommendation Page, showcasing how the system presents the user's predicted career path.

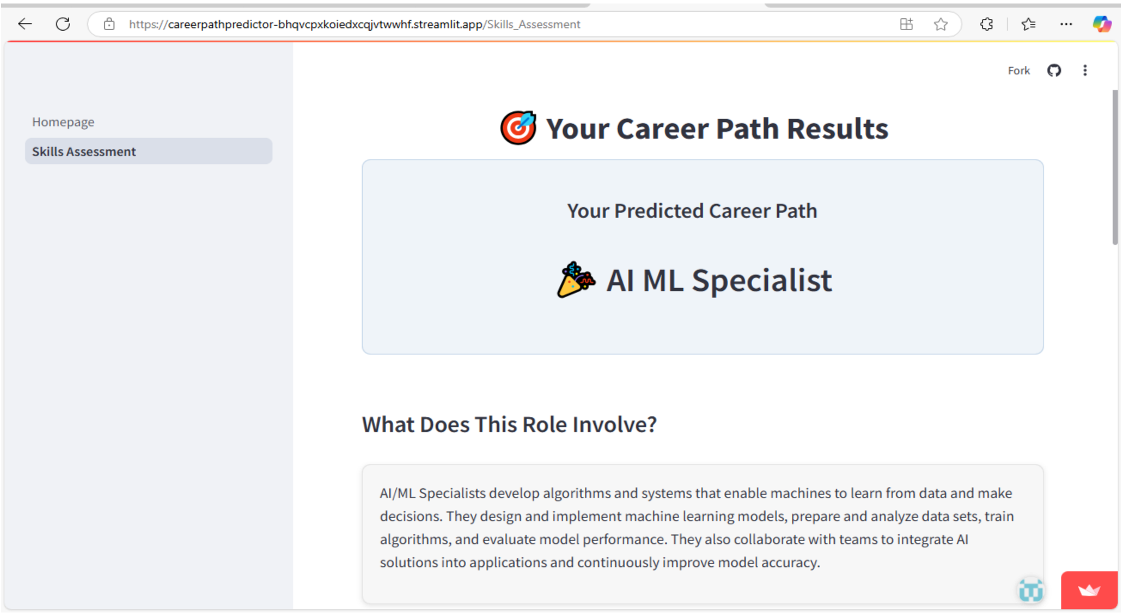

Figure 6. Recommendation Page

5. Conclusion

The introduction of a model for predicting career paths has greatly aided users in making more informed and better career choices. It has It made decision making in career path easy, convenient and accurate which makes people more confident in their future career and has a higher success rate in their future career. When fully developed, the system will help to reduce human errors and provide a cost-effective solution that will be easily accessible for individuals.

References

[1] N. C. Gysbers and P. Henderson, Career Guidance and Counseling in the 21st Century: An International Perspective. Routledge, 2014.

[2] J. L. Holland, Making Vocational Choices: A Theory of Vocational Personalities and Work Environments, 3rd ed. Odessa, FL, USA: Psychological Assessment Resources, 1997.

[3] E. A. Locke, “The effects of self-efficacy, goals, and task strategies on task performance,” J. Appl. Psychol., vol. 61, no. 3, pp. 267–276, 1976.

[4] D. E. Super, “A life-span, life-space approach to career development,” in Career Choice and Development, D. Brown and L. Brooks, Eds. Jossey-Bass, 1990, pp. 197–262.

[5] Y. R. Shrestha, J. Latoo, and U. Farooq, “A survey on big data analytics in career prediction systems,” Int. J. Comput. Appl., vol. 975, pp. 29–33, 2020.

[6] A. Gupta and S. Singh, “Machine learning for career prediction: A systematic review,” Int. J. Comput. Appl., vol. 141, no. 8, pp. 1–7, 2016.

[7] A. Deshpande, A. Dubey, A. Dhavale, U. Navatre, A. Gurav, and A. K. Chanchal, “Implementation of an NLP-driven chatbot and ML algorithms for career counselling,” in Proceedings of the 2024 International Conference on Inventive Computation Technologies (ICICT), Lalitpur, Nepal, 2024, pp. 853–859.

[8] N. VidyaShreeram and A. Muthukumaravel, “Student career prediction using machine learning approaches,” EAI Endorsed Transactions, 2021. [Online]. Available: https://doi.org/10.4108/eai.7-6-2021.2308642

[9] Y. Liu, L. Zhang, L. Nie, Y. Yan, and D. S. Rosenblum, “Fortune teller: Predicting your career path,” 2022.

[10] A. Pimplikar, R. Pati, A. Singh, and A. Shah, “CareerMap - Mapping tech roles with personality & skills,” Mendeley Data, vol. 1, 2023. [Online]. Available: https://doi.org/10.17632/5z68cvxssn.1