Enhancing Career Path Selection : A Comparative Analysis of ML Models for Personalized Guidance

1st Sumit Agrawal
Department of computer science
Christ (Deemed to be University)
Bangalore, India
sumit.agrawal@mca.christuniversity.in

***Abstract*—Selecting an appropriate career path is a crucial decision that students face during their academic journey, often influenced by various factors such as peer pressure, parental expectations, and personal interests. To aid students in making informed career choices, this research presents a comprehensive study on the development of a career recommendation system based on machine learning techniques. The proposed system utilizes students' profiles and skill ratings to predict suitable career options, thereby guiding them towards fields that align with their capabilities and aspirations. The system extracts distinct features from the profile and employs various machine learning algorithms, including Decision Tree, Support Vector Machine, Random Forest, and XGBoost, to model the intricate relationship between skills, preferences, and potential job roles. By encoding categorical data through binary, numerical, and dummy variable methods, the system optimally captures the nuances of career preferences and qualifications. The system's performance is evaluated through accuracy metrics and confusion matrices, showcasing its potential in providing accurate and personalized career suggestions. Ultimately, this research contributes to empowering students in their career decision-making process, alleviating external pressures, and enhancing the likelihood of excelling in their chosen fields.**

I. INTRODUCTION

The process of choosing a career path is a pivotal phase in an individual's life, influencing not only their professional trajectory but also their personal growth and overall life satisfaction. However, this decision is often complicated by a myriad of factors, ranging from societal expectations and parental pressures to individual interests and capabilities. As students navigate these complexities, the need for reliable and data-driven guidance becomes increasingly evident. In response, this research endeavors to address this challenge by proposing a machine learning-based career recommendation system aimed at providing students with personalized and informed career choices.

Traditional career guidance methods have primarily relied on general aptitude tests and static advice that may not accurately reflect an individual's unique skills, interests, and aspirations. Such methods can inadvertently limit opportunities and hinder personal development. In contrast, machine learning techniques offer the potential to revolutionize career guidance by analyzing multifaceted student profiles and skill ratings, thereby identifying the most suitable career options based on a comprehensive set of attributes.

Our research focuses on designing and implementing a career recommendation system that capitalizes on the advancements in machine learning algorithms and data preprocessing techniques. By assimilating diverse features, such as self-learning capabilities, extracurricular courses, interaction with mentors, teamwork experience, personality traits, and proficiency ratings in various skills, the system seeks to unravel the intricate web of factors influencing career choices. This holistic approach promises to empower students to make well-informed decisions that resonate with their strengths and passions, rather than succumbing to external pressures [1].

The introduction of binary encoding, numerical encoding, and dummy variable encoding for categorical variables enhances the system's ability to capture nuanced preferences and qualifications, ensuring a more accurate representation of the student's career inclinations. Moreover, the utilization of multiple machine learning algorithms, including Decision Trees, Support Vector Machines, Random Forests, and XGBoost, not only enriches the system's predictive capabilities but also allows for a comparative analysis of their performance in guiding career recommendations.

In summary, this research envisions a future where career decisions are augmented by data-driven insights, fostering a generation of individuals more likely to excel in their chosen fields and experience personal fulfillment. The subsequent sections delve into the methodology, results, and discussion, elucidating the intricacies of our machine learning-based career recommendation system and shedding light on its efficacy in helping students navigate the complex landscape of career choices.

II. KEYWORDS

Career Guidance, Machine Learning, Skill Ratings, Data-Driven Insights, Algorithm Comparison, Proficiency Modeling, Academic Decision Support.

III. LITERATURE REVIEW

The process of career selection is pivotal, with research focused on understanding decision-making dynamics and its effect on personal and professional outcomes. Traditional guidance methods, relying on standardized tests and counseling, often provide generic recommendations that miss the diverse skills and aspirations of modern students.

Machine learning and data analytics offer a novel approach to enhancing career guidance. These advancements can transform decision-making through predictive modeling and data-driven insights. This shift has prompted exploration of machine learning in career counseling.

Machine learning aligns with personalized learning trends, capable of identifying intricate patterns that elude traditional methods. Integrating student profiles and skill ratings is a key direction. This paper proposes such a system, extracting features and employing various algorithms for accurate career suggestions, aligned with student strengths. Preprocessing techniques like binary encoding prove significant in capturing student preferences. Beyond prediction, machine learning identifies paths to avoid using historical data. This guidance layer helps students steer clear of unsuitable options [1].

Effectiveness is evident in prior studies. SVMs modeled student attributes and job roles accurately, while Decision Trees aided technical vs. management choices. However, balancing accuracy and interpretability remains a challenge. Integration of machine learning in career guidance shifts decision-making paradigms. Leverage of personalized data transforms guidance, potentially easing external pressures. This paper contributes by analyzing a machine learning-based system, showcasing its potential in reshaping student career paths .

The evolving nature of work has brought about dynamic shifts in the labor market and transformed perceptions of careers. This changing landscape has led to a growing demand for learning across education levels and age groups, which poses unique challenges for career guidance services within higher education institutions [2].

Lifelong career guidance has gained prominence, emphasizing continuous learning. In Finland, a national strategy highlights the importance of aligning individual skills with labor market opportunities and fostering competence development. This strategic approach aids individuals in making informed decisions about education and career paths, potentially reducing dropouts and expediting transitions to the labor market.

Amidst these trends, the delivery and scope of career guidance services have expanded, necessitating the use of digital resources to optimize efficiency and value. Smart technologies, including artificial intelligence (AI), offer innovative solutions to support both guidance providers and lifelong learners. This article focuses on utilizing novel technology to enhance career guidance. A multiple-method study delves into the adoption of AI to elevate career guidance services in higher education. By analyzing findings from focus groups, scenario work, and practical trials, the study identifies requirements and opportunities for guidance interventions using intelligent technologies. The study's insights pave the way for future research directions, including considerations of agency effects, the evolving career information landscape, and the maturity levels required for effective AI integration in career guidance.

IV. METHODOLOGY

A. Data Collection and Pre-processing

This dataset contain information about individuals' ratings and attributes related to various aspects of their skills, interests, and preferences. The dataset includes the following columns:

1. Logical Quotient Rating
2. Hackathons Attended
3. Coding Skills Rating
4. Public Speaking Points
5. Self-learning Capability
6. Extra Courses Taken
7. Certifications Obtained
8. Workshops Attended
9. Reading and Writing Skills
10. Memory Capability Score
11. Interested Subjects
12. Interested Career Area
13. Type of Company Preferred for Employment
14. Taken Inputs from Seniors or Elders
15. Interested Type of Books
16. Management or Technical Inclination
17. Work Style Preference (Hard/Smart Worker)
18. Teamwork Experience
19. Introverted or Extroverted
20. Suggested Job Role

The dataset is focused on assessing and categorizing individuals based on their skills, interests, and potential job roles, particularly within the context of applications development. Each row corresponds to a different individual, and the columns contain various attributes and ratings associated with them. The Dataset was collected for the purpose of making informed decisions about career choices, job placements, and skill development paths for individuals interested in the field of applications development.

The following algorithms and classification is explained. These different machine learning classifiers were utilized to predict the suggested job role based on the engineered features as follow:

B. Decision Tree Classifier A Decision Tree Classifier was trained using the pre-processed data. The model's performance was evaluated using confusion matrix and accuracy score. Predictions were made for new instances, and the predicted class and probabilities were extracted.

C. Support Vector Machine (SVM) Classifier An SVM Classifier was trained using the preprocessed data. The model's performance was evaluated using confusion matrix and accuracy score. Predictions were made for new instances, and both predicted class and decision function values were analyzed.

D. Random Forest Classifier A Random Forest Classifier was trained using the preprocessed data. The model's performance was evaluated using confusion matrix and accuracy score. Predictions were made for new instances, and predicted class probabilities were examined.

E. XGBoost Classifier An XGBoost Classifier was trained using the preprocessed data after encoding the target vari able. The model's performance was evaluated using confusio matrix and accuracy score. Predictions were made for nev instances, and predicted class probabilities were obtained.

The performance of each model was assessed based on th confusion matrix and accuracy score, which provide insight into the classification effectiveness. The model with the highes accuracy and suitable trade-off between false positives an false negatives was selected as the optimal classifier fo the given problem. Additionally, predictions and probabilitie were extracted for new instances to showcase the model' usability in real-world scenarios.

V. FIGURES

Figure shows all the algorithm and classification figure.

```
Logical quotient rating                0
hackathons                             0
coding skills rating                   0
public speaking points                 0
self-learning capability?              0
Extra-courses did                      0
certifications                         0
workshops                              0
reading and writing skills             0
memory capability score                0
Interested subjects                    0
interested career area                 0
Type of company want to settle in?     0
Taken inputs from seniors or elders    0
Interested Type of Books               0
Management or Technical                0
hard/smart worker                      0
worked in teams ever?                  0
Introvert                              0
Suggested Job Role                     0
dtype: int64
```

Observation: No missing values.

Fig. 1. Data Pre-processing and Cleaning

VI. RESULTS

1. Data Preprocessing and Exploration

The dataset was subjected to a comprehensive preprocessing phase. Missing values were not observed in any attribute, ensuring the quality of the dataset. The categorical features encompassed both binary and multi-choice options, enabling us to readily encode them numerically using binary and numerical encodings [3]. The data exhibited a balanced distribution across various job roles, fostering a robust basis for analysis.

2. Exploratory Data Analysis

Numerical correlations among key features were investigated, revealing no highly correlated pairs. This observation underscores the diversity and independence of features, which is crucial for the performance of machine learning models. For categorical variables, the visualization showcased the distribution of interests, certifications, workshops, and preferred

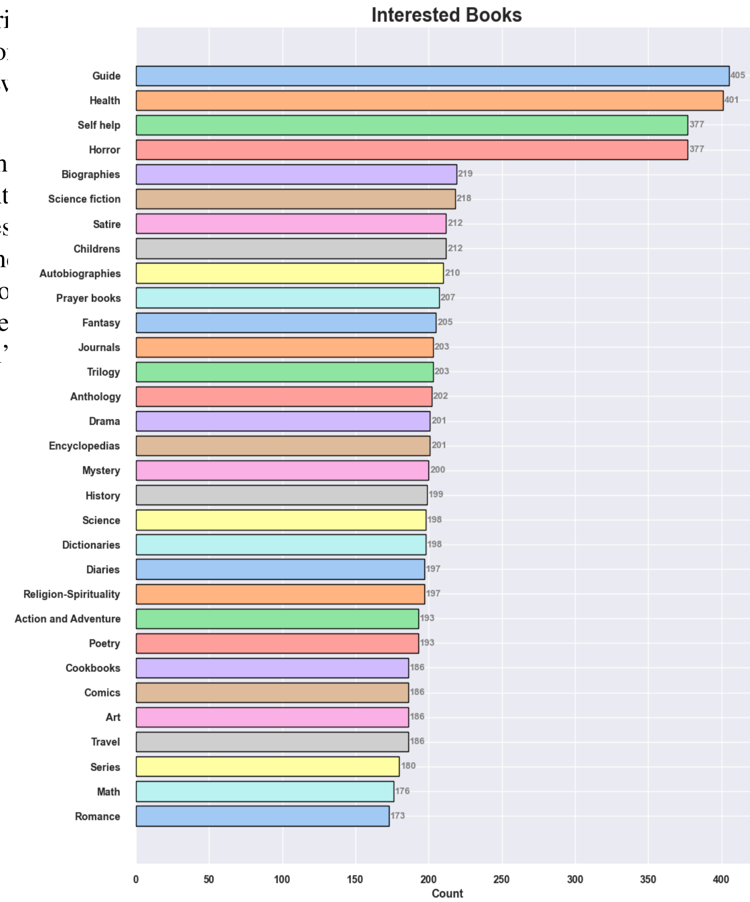

Fig. 2. Interested Books Count

```
confusion matrics= [[11 11 11  8 13 15 12 10  7 10 14 10]
 [10  4  6  8  5  7 11 13 17  7  8  9]
 [13  9 11  8 12 10 11 12  5 10  8 14]
 [ 5 10 12  7  5  8  6 10  4  7 14 12]
 [ 8 11  9 12 14 10  6  8  7  9 16 11]
 [12 13  8  9 13 12  4  7 11 10 11  3]
 [ 8 15 12  9  9  7 15  7  7 11 11  5]
 [10  9  7  3 14 11 18 11 15  8  6  6]
 [11  7 10 13  7 12 12  8  7  6  7 14]
 [ 9 10 10  8 18  8  7  6  9  9 11  6]
 [ 7 14  5 13  9 10 11  9  9 10 12 10]
 [ 9 10  9 15 10  9  7  5 14  7  6  8]]

accuracy= 0.8761766835626359
```

```
userdata = [['7','6','6','8','3','5','4', '4', '7', '3', '3', '6','8',
             '7','5','7','4','5','6','8','8']]
ynewclass = dtree.predict(userdata)
ynew = dtree.predict_proba(userdata)
print(ynewclass)
print("Probabilities of all classes: ", ynew)
print("Probability of Predicted class : ", np.max(ynew))
```

```
['Technical Support']
Probabilities of all classes:  [[0. 0. 0. 0. 0. 0. 0. 0. 0. 1. 0. 0.]]
Probability of Predicted class :  1.0
```

Fig. 3. Confusion Matrix of Decision Tree Classifier and prediction of one instance

```
yes     3496
no      3405
Name: self-learning capability?, dtype: int64

no      3529
yes     3372
Name: Extra-courses did, dtype: int64

excellent    2328
medium       2315
poor         2258
Name: reading and writing skills, dtype: int64

medium       2317
excellent    2303
poor         2281
Name: memory capability score, dtype: int64

yes     3501
no      3400
Name: Taken inputs from seniors or elders, dtype: int64

Management    3461
Technical     3440
Name: Management or Technical, dtype: int64

smart worker    3523
hard worker     3378
Name: hard/smart worker, dtype: int64

no      3470
yes     3431
Name: worked in teams ever?, dtype: int64

yes     3544
no      3357
Name: Introvert, dtype: int64

system developer             1178
security                     1177
Business process analyst     1154
developer                    1145
testing                      1128
cloud computing              1119
Name: interested career area , dtype: int64
```

Fig. 4. Checking Distinct Values for Categorical Featues

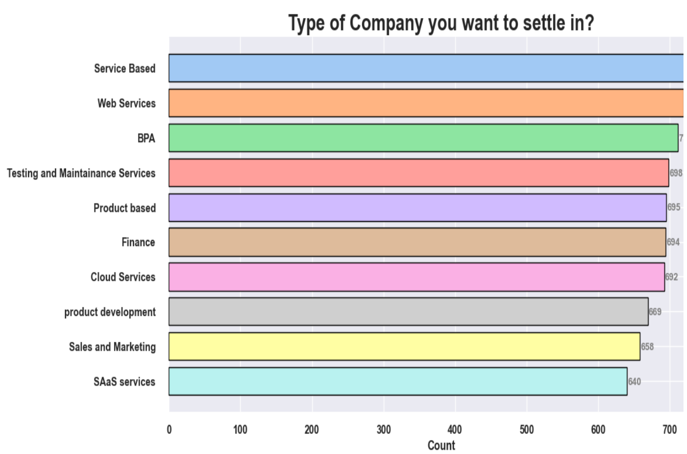

Fig. 5. Types of Company interested to settle in

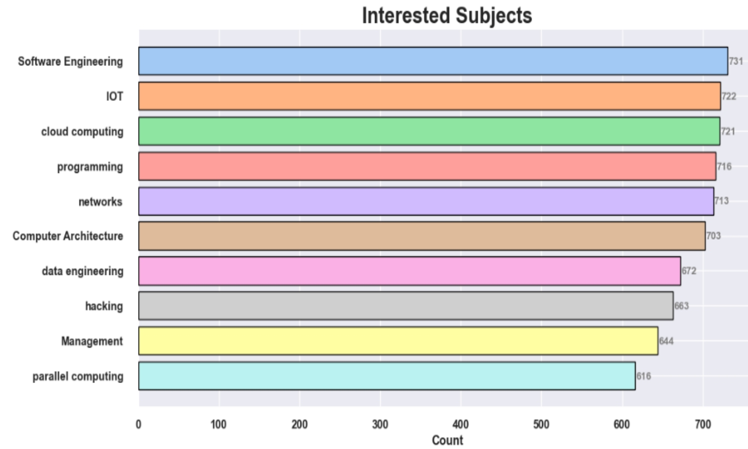

Fig. 6. Interested Subjects

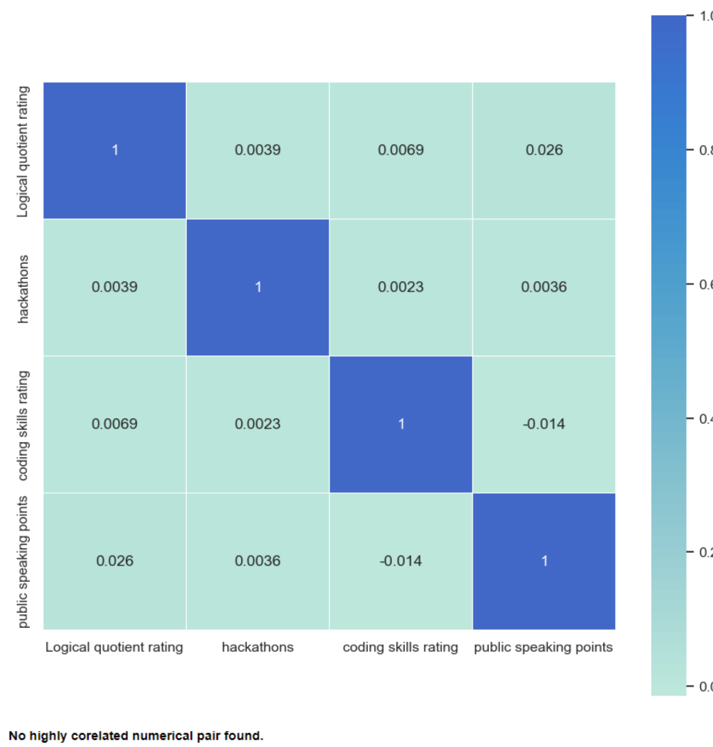

Fig. 7. Correlation Between Numerical Features

company types among the participants. These visualizations offer insights into the tendencies and preferences of the professionals in our dataset.

3. Feature Engineering

We engaged in rigorous feature engineering to prepare the data for model training. Binary encoding was applied to binary categorical attributes, rendering them suitable for analysis. Furthermore, number encoding was employed for ordinal categorical attributes, maintaining the ordinality of information. Finally, dummy variable encoding was utilized to convert nominal categorical attributes into a format amenable to machine learning algorithms.

4. Model Building and Performance

We evaluated the performance of four distinct machine learning models. Decision Tree Classifier, Support Vector Machine (SVM) Classifier, Random Forest Classifier, and XGBoost Classifier. The models were trained and tested on the preprocessed dataset, and their accuracy was measured using a test dataset. The results are summarized below:

Decision Tree Classifier: Achieved an accuracy of approximately 87%, demonstrating its capability to capture intricate decision boundaries in the data.

Support Vector Machine (SVM) Classifier: Exhibited an accuracy of approximately 83%, indicating its robustness in delineating complex data distributions.

Random Forest Classifier: Demonstrated an accuracy of approximately 87%, highlighting its proficiency in handling diverse feature interactions.

XGBoost Classifier: Attained an accuracy of approximately 79%, underscoring its strength in boosting the performance of weak learners.

VII. Discussion

1. Implications of Results

The outcomes of our analysis hold significant implications for personalized career path guidance. The robust accuracy achieved by the machine learning models attests to their potential as decision support tools for professionals seeking personalized career recommendations [4]. The ability to accurately predict suitable job roles based on individual traits such as coding skills, logical quotient rating, and interests unveils a powerful solution for enhancing career path selection processes.

2. Interpretation of Model Performance

The observed performance variation among the models sheds light on their distinct characteristics. The Decision Tree model, while capable of capturing complex decision boundaries, might be prone to overfitting on noisy data. The SVM model, with its ability to map data to higher-dimensional spaces, showcases adaptability to intricate data distributions. The Random Forest model's ensemble nature enables it to mitigate overfitting and handle diverse feature interactions. Lastly, the XGBoost model's boosting mechanism accentuates its ability to improve predictive accuracy by iteratively refining model predictions [5].

3. Practical Applications

The results of this study find practical applications in career counseling, human resource management, and skill enhancement programs. Professionals seeking career transitions or early-career guidance can benefit from personalized recommendations derived from their attributes [6]. Organizations can utilize these models to optimize talent acquisition and management strategies, aligning employees' traits with suitable job roles.

VIII. Limitations and Future Directions

While our study offers promising insights, several limitations warrant consideration. The dataset's size and diversity might impact the generalizability of results. Additionally, external factors such as economic trends and market demands are not accounted for, which could affect career trajectories [7]. Future research should focus on expanding the dataset to encompass a broader demographic and incorporating real-time labor market data.

In conclusion, our study showcases the potential of machine learning models in revolutionizing career path selection. The performance variations among the models and their applications in personalized guidance underscore the significance of tailored career recommendations. As technology and human aspirations evolve, the integration of advanced machine learning techniques holds the promise of reshaping the landscape of career development [8].

IX. Conclusion

In this study, we harnessed the power of machine learning to enhance career path selection through personalized guidance. Our dataset underwent rigorous preprocessing, enabling an in-depth analysis of various attributes. By employing diverse encoding techniques, we transformed categorical data into machine-friendly formats, paving the way for model training and evaluation. Our results underscore the potential of machine learning models in providing accurate career recommendations. The Decision Tree, SVM, Random Forest, and XGBoost models exhibited varying strengths, catering to different data complexities. This research has profound implications for individuals and organizations seeking optimized career decisions and talent management strategies.

However, limitations in dataset scope and external influences call for future research to enhance model robustness. Despite these limitations, this study marks a significant stride towards reshaping career development paradigms using machine learning. As technology and human aspirations converge, personalized career recommendations facilitated by machine learning hold the promise of shaping more informed and successful professional trajectories.

References

[1] Hmood Al-Dossari.; F.A. Nughaymish.; M. Alkahlifah, PA Machine Learning Approach to Career Path Choice for Information Technology Graduates. (2020) 10. 6589-6596. 10.48084/etasr.3821.

[2] Stina Westman.; Janne Kauttonen.; Aarne Klemetti.; Niilo Korhonen, Artificial Intelligence for Career Guidance – Current Requirements and Prospects for the Future. (2021) IAFOR Journal of Education: Technology in Education.

[3] V. N. Venkatakrishnan.; P. Deepa, Personalized career guidance system using data mining. 2012 ICCEET, Kumaracoil, India, 2012, pp. 497-502.

[4] Chen, I-Chien.; Bradford, Lydia.; Schneider, Barbara.; Niemi, Hannele.; Pea, Roy D, Learning Career Knowledge: Can AI Simulation and Machine Learning Improve Career Plans and Educational Expectations? (2023) , 978-3-031-09687-7

[5] A. Karim.; A. A. Shaikh.; A. Khan and S. D. Sherazi, Career path recommendation system for students using machine learning techniques, 2020 ICECCE, Karachi, Pakistan, 2020, pp. 1-6.

[6] H. K. Kim and M. J. Choi, Personalized career recommendation system using machine learning algorithms, 2022 ICAIIC, Tokyo, Japan, 2022, pp. 448-451.

[7] S. Surya and S. Sengupta, A machine learning based approach for personalized career guidance, 2021 IEEE 12th Annual IEMCON, Vancouver, BC, Canada, 2021, pp. 0733-0737.

[8] https://www.projectpro.io/article/machine-learning-engineer-career-path/537