ISSN No:-2456-2165

Intelligent Career Guidance Systems

[1]Rohith Kanathur, [2]Rakshita Akhilesh, [3]Soumya A, [4]Manasa Kanathur
R V College of Engineering

Abstract:- This study focuses on conducting a cost-benefit analysis to determine the value of pursuing a master's degree in different fields of study and from different universities. While previous research has explored the impact of higher education on career prospects and income potential, there is still a gap in providing individuals with personalized recommendations to make informed decisions. The objective of this research is to address this gap by developing a recommendation system that considers various factors such as academic scores, intended major, location preferences, and budget. By weighing the costs associated with obtaining a master's degree against the potential benefits, individuals will have the necessary information to assess whether pursuing a master's degree is worth the time and financial investment.

The results of this study will provide valuable insights into the value of pursuing a master's degree in different fields and from different universities. By quantifying the costs and benefits, individuals will be able to make informed decisions about their educational and career paths. The analysis of the results will include comparisons with existing literature to validate the findings and assess the novelty of the personalized recommendation system. By examining the data and interpreting the results, this research will provide individuals with a clear understanding of the potential returns they can expect from pursuing a master's degree in their chosen field of study, thereby empowering them to make well-informed decisions about their education and future career prospects.

I. INTRODUCTION

Recent advancements in evaluating the value of pursuing a master's degree and providing personalized recommendations have seen significant developments in the utilization of data analytics and machine learning techniques. These advancements have allowed for more accurate and data-driven assessments of the costs and benefits associated with obtaining a master's degree, as well as enhanced guidance for individuals in making informed decisions about their educational paths.

One notable area of progress lies in the analysis of large datasets to evaluate the outcomes of graduates from different universities and fields of study. Researchers have leveraged these datasets to examine factors such as employment rates, salary growth, and career trajectories. By analyzing this wealth of information, studies have been able to provide insights into the long-term value of a master's degree from specific institutions or in particular disciplines.

Furthermore, advancements in machine learning algorithms have paved the way for the development of personalized recommendation systems. These systems consider a multitude of factors, including academic performance, interests, location preferences, and financial constraints. By leveraging these factors, the recommendation systems generate tailored suggestions for individuals seeking to pursue a master's degree.

In addition to data analytics and machine learning, the integration of interactive web platforms and online resources has revolutionized the accessibility of information related to universities and programs. These platforms provide comprehensive data on various aspects, including tuition fees, living expenses, financial aid options, employment statistics, and alumni outcomes. By centralizing this information and making it easily accessible, prospective students can conduct their own research and make informed decisions based on reliable and up-to-date data.

In summary, recent advancements in the field have harnessed the power of data analytics, machine learning, and interactive web platforms to provide individuals with more accurate assessments of the value of pursuing a master's degree. These advancements have also facilitated the development of personalized recommendation systems, enabling tailored suggestions that consider a wide range of factors. Through these developments, individuals are now better equipped to make informed decisions about their educational and career paths, ultimately leading to more rewarding and successful outcomes.

II. LITERATURE REVIEW

In [1], the system tailors course recommendations for post-graduate students based on their academic history, interests, and career aspirations. A web application allows users to input their preferences and generates personalized course suggestions to aid informed decision-making.

According to Daohu et.al. [2], PaUtilizing historical data like grades and enrollment, this model predicts academic performance within a university. It identifies trends and patterns, helping understand factors influencing success and progression.

By considering learner preferences and interactions [3], this system suggests relevant online educational resources in International Finance. Learners benefit from a tailored learning experience and improved knowledge acquisition.

ISSN No:-2456-2165

This method in [4] employs deep learning to recommend courses based on learner preferences, enhancing engagement and learning outcomes. Incorporating external factors and addressing model interpretability are avenues for future development.

The study in [5] reveals that comprehensive pre-college career guidance positively affects student persistence and performance in university, suggesting a potential strategy for improving student outcomes.

According to Mafufah Hastin et al. [6], the study showcases the positive impact of guidance and counseling services on students' career maturity. It helps students gain clarity, self-awareness, and decision-making skills related to their career paths.

[7] involves utilizing technology and AI, this system offers personalized career guidance by analyzing students' skills and interests. However, limited human interaction and potential lack of context are considerations.

The system described in [8] is a web-based system that streamlines career development, offering tools for goal-setting and networking. Potential limitations include underrepresentation of diverse paths and limited support for career transitions.

According to the study done by Areej Kamal et al. [9], it explores emotional intelligence and job seeker attitudes in career decision-making. The system employs adaptive learning algorithms to customize recommendations based on user progress and preferences.

The paper by Kasem Seng et al. [10], E-learning's role in career counseling is discussed, highlighting features such as self-assessment tools and job market information. Challenges include capturing subjective career decision-making and ensuring accurate resources.

System described in [11] connects students with professionals for career guidance, aiming to address the scarcity of such advice. However, ethical and privacy concerns should be considered.

The research paper in [12] is an AI-driven web-based expert system that offers career counseling for college students. The agent-based approach enhances adaptability and learning from student interactions.

III. PROBLEM DEFINITION

The problem domain of the project falls under data science. The project involves gathering, processing, and analyzing data on universities and student performance to create a personalized recommendation system. It involves the application of machine learning algorithms to identify patterns in the data and generate accurate predictions about which universities are most likely to meet the individual's needs. Additionally, the project requires the use of data visualization techniques to present the data in a meaningful way to users.

The problem statement is to determine the value of pursuing a master's degree in different fields of study and from different universities by conducting a cost-benefit analysis. It provides a personalized recommendation system that helps individuals make informed decisions about which universities to apply to and attend.

IV. METHODOLOGY

Gather the datasets from relevant sources: This step involves collecting the necessary datasets that contain information about universities, master's degree programs, tuition fees, living expenses, financial aid options, employment statistics, and other relevant factors. These datasets can be obtained from sources such as university databases, government reports, educational websites, and surveys conducted by educational institutions or organizations.

Perform data cleaning and exploratory data analysis: Once the datasets are gathered, the next step is to clean the data by removing any errors, inconsistencies, or missing values. Data cleaning ensures that the dataset is accurate and reliable. Exploratory data analysis involves examining the dataset to gain insights, identify patterns, and understand the distribution and relationships between different variables. This analysis helps in understanding the data and making informed decisions about data preprocessing and feature engineering.

Split the dataset into a training set and a testing set: To evaluate the performance of the machine learning model, the dataset needs to be divided into two subsets: a training set and a testing set. The training set is used to train the model, while the testing set is used to assess how well the trained model performs on unseen data. Typically, the dataset is randomly split, allocating a certain percentage (e.g., 70-80%) for training and the remaining portion for testing.

Choose an appropriate ML model and train it on the training set: Based on the nature of the problem and the available data, an appropriate machine learning model is selected. Models such as logistic regression, decision trees, random forests, or SVM may be suitable for this project.

Evaluate the performance of the trained model on the testing set: Once the model is trained, its performance is evaluated using the testing set. Performance metrics such as accuracy, precision, recall, and F1 score are computed to assess how well the model predicts the target variable based on the input features. These metrics provide insights into the model's predictive capabilities and help determine its effectiveness in making accurate recommendations.

Fine-tune the model to improve its performance: After evaluating the model's initial performance, fine-tuning techniques are applied to improve its predictive accuracy. This may involve adjusting hyperparameters, such as regularization strength or tree depth, using techniques like grid search or random search.

V. IMPLEMENTATION

Python is a favored language for machine learning projects due to its simplicity, extensive libraries, and active community support. It offers tools and frameworks designed for machine learning tasks. Python's advantages include:

Ease of Learning and Readability: Python has an intuitive syntax that is easy to grasp, aiding both beginners and experienced programmers. Its readability accelerates development and maintenance of machine learning projects.

Abundant Libraries: Python boasts specialized libraries for data manipulation, analysis, and machine learning. Notable examples include NumPy, Pandas, Scikit-learn, TensorFlow, and PyTorch. These libraries provide functions and algorithms to simplify complex tasks and speed up development.

Strong Community and Documentation: Python has a large and active community of developers and data scientists. This ensures access to numerous resources, tutorials, and code samples, facilitating problem-solving and learning. Python's extensive documentation enhances understanding and usage of its libraries and frameworks.

Data Manipulation and Analysis: Python's Pandas library offers robust tools for data manipulation and analysis. It enables data loading, cleaning, preprocessing, handling missing values, and exploratory data analysis. Pandas simplifies data transformation, merging, and insight extraction.

Machine Learning Frameworks: Python provides popular machine learning frameworks like Scikit-learn, TensorFlow, and PyTorch. Scikit-learn covers various algorithms for tasks such as regression, classification, clustering, and dimensionality reduction. TensorFlow and PyTorch are deep learning frameworks for complex neural network creation and training.

Visualization Capabilities: Libraries like Matplotlib and Seaborn in Python facilitate rich data visualizations for better data understanding and result communication. These libraries offer diverse plots, charts, and graphs to present outcomes effectively.

Integration and Extensibility: Python easily integrates with languages like C/C++ and Java, enabling utilization of existing code and libraries. This flexibility supports specialized libraries or optimization of resource-intensive project parts. Python's integration with databases, web frameworks, and APIs suits varied project needs.

Deployment Options: Python provides multiple options to deploy machine learning models. You can integrate models into web apps via frameworks like Flask or Django. Alternatively, cloud platforms like AWS Lambda, Google Cloud Functions, or Microsoft Azure Functions allow creation of serverless APIs for inference.

➢ *Key Libraries Mentioned:*

- NumPy: Supports efficient numerical operations and multi-dimensional arrays for handling data in machine learning projects.
- Pandas: Offers tools for data manipulation and analysis, easing data loading, cleaning, preprocessing, and transformation.
- Scikit-learn: A versatile machine learning library with various algorithms for regression, classification, and model evaluation.
- Matplotlib: Enables data visualization through a variety of plots and charts.

When organizing a project, adhere to conventions such as meaningful variable and function names, proper commenting, and a well-structured file organization. Use docstrings to provide comprehensive documentation for files, modules, classes, functions, and methods. The choice of tools like Jupyter Notebook, Flask for web development, Visual Studio Code for code editing, and Git for version control enhances project development and collaboration.

VI. TESTING AND MAINTENANCE

In a machine learning project, testing is crucial to ensure the correctness and robustness of your code and system. Different levels of testing include:

Unit Testing : Unit testing is done using Pytest and involves writing tests to verify the correctness of individual components. Key features of Pytest include test discovery, fixture support for creating reusable test environments, assertions for checking results, parameterization for testing multiple scenarios, and test coverage measurement.

Unit tests can be written for input data validation, model prediction accuracy, handling invalid inputs, and testing edge cases.

Integration Testing : Integration testing checks if different parts of the system work together seamlessly. It involves scenarios like data preprocessing integration, model training and evaluation integration, input data integration, API integration, deployment and infrastructure integration, and end-to-end testing.

Integration tests ensure the interaction between components and data flow are handled properly.

System Testing : System testing evaluates the performance and functionality of the entire system, including the machine learning model, data processing, and user interface.

Test data preparation, defining test scenarios, determining performance metrics, test execution, and error handling are important aspects of system testing.

Hypothesis testing, on the other hand, is a statistical method used to assess claims or hypotheses about a population based on sample data. It involves formulating a null hypothesis (H0) that assumes no effect or difference, and an alternative hypothesis (H1 or Ha) that represents the claim you're investigating. The two hypotheses are mutually exclusive and help you make informed inferences about the population based on the sample data.

When dealing with classification models, several relevant metrics help assess their performance. Accuracy, the proportion of correct predictions to total predictions, is commonly used, though its adequacy can be limited for imbalanced classes. Precision gauges correctly predict positive instances out of all predicted positives, useful when false positives are costly. Recall (also known as Sensitivity or True Positive Rate) measures correctly predicted positives out of all actual positives, important when false negatives are costly. F1 Score combines precision and recall into a balanced metric. Area Under the ROC Curve (AUC-ROC) evaluates overall model performance across different thresholds, particularly beneficial for varying operating points. For regression tasks, Mean Squared Error (MSE) calculates the average squared difference between predicted and actual values. Root Mean Squared Error (RMSE) and Mean Absolute Error (MAE) present errors in the original unit of the target variable. R-squared (Coefficient of Determination) shows how well the model explains variance in the target variable. Mean Absolute Percentage Error (MAPE) indicates average percentage difference between predicted and actual values.

The dataset, sourced from Levels.fyi, contains worker attributes such as company, salary, experience, education, and location. It facilitates analysis of whether a Master's degree is worth pursuing, especially in STEM roles. The dataset dimensions are 62642 by 29, with attributes including company, level, title, compensation, location, years of experience, and base salary. Another dataset for predicting Graduate Admissions features GRE and TOEFL scores, university rating, statement of purpose and recommendation strength, undergraduate CGPA, research experience, and chance of admit. The dataset size is 400 by 9, and sources include ets.org.

Key evaluation metrics encompass Accuracy (total correct predictions over all predictions), Precision (correctly predicted positive instances over all predicted positives), Recall (correctly predicted positive instances over actual positives), and F1 Score (a balanced combination of precision and recall). For regression tasks, Root Mean Squared Error (RMSE) offers an interpretable metric in the original unit, while Mean Absolute Error (MAE) captures average magnitude of errors. R-squared indicates the proportion of variance explained by the model, with a perfect fit at 1.

Table 1 Performance Analysis

Model	Accuracy	Precision	Recall	F1 Score
Logistic Regression	0.95	0.78	0.91	0.84
SVM	0.98	0.91	0.91	0.91
GNB	0.93	0.70	0.90	0.82
Decision Tree	0.92	0.75	0.75	0.75
Random Forest	0.94	0.77	0.83	0.80
KNN	0.93	0.68	0.91	0.78

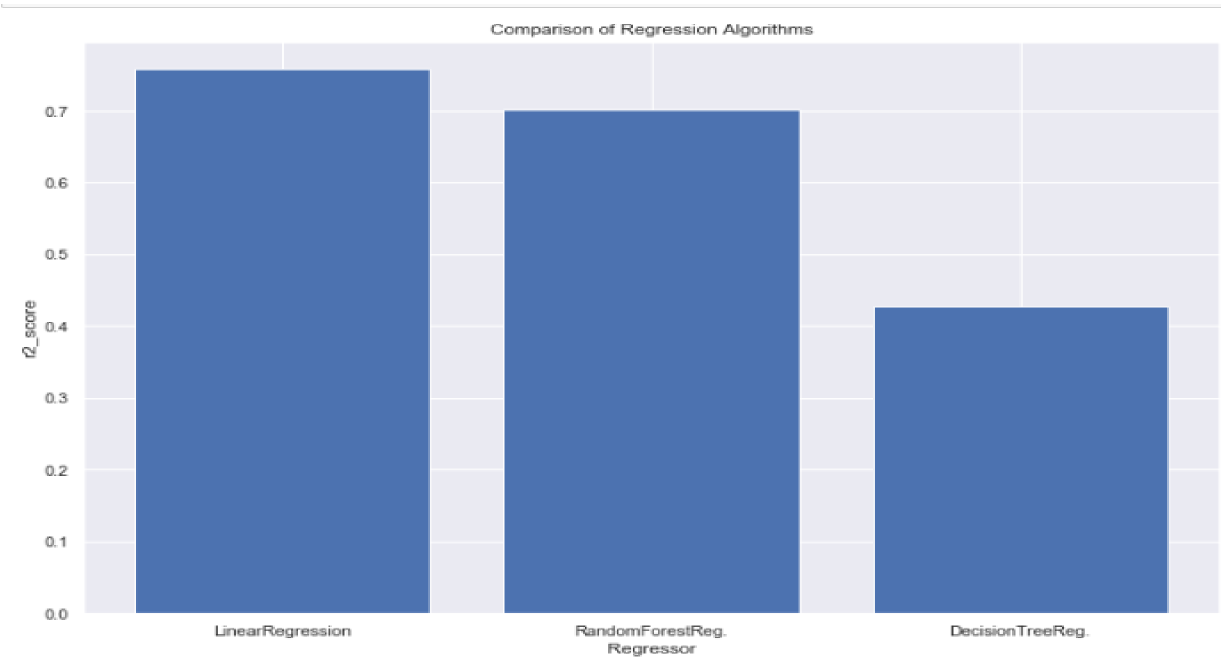

Fig 1 Comparison of Regression Algorithms

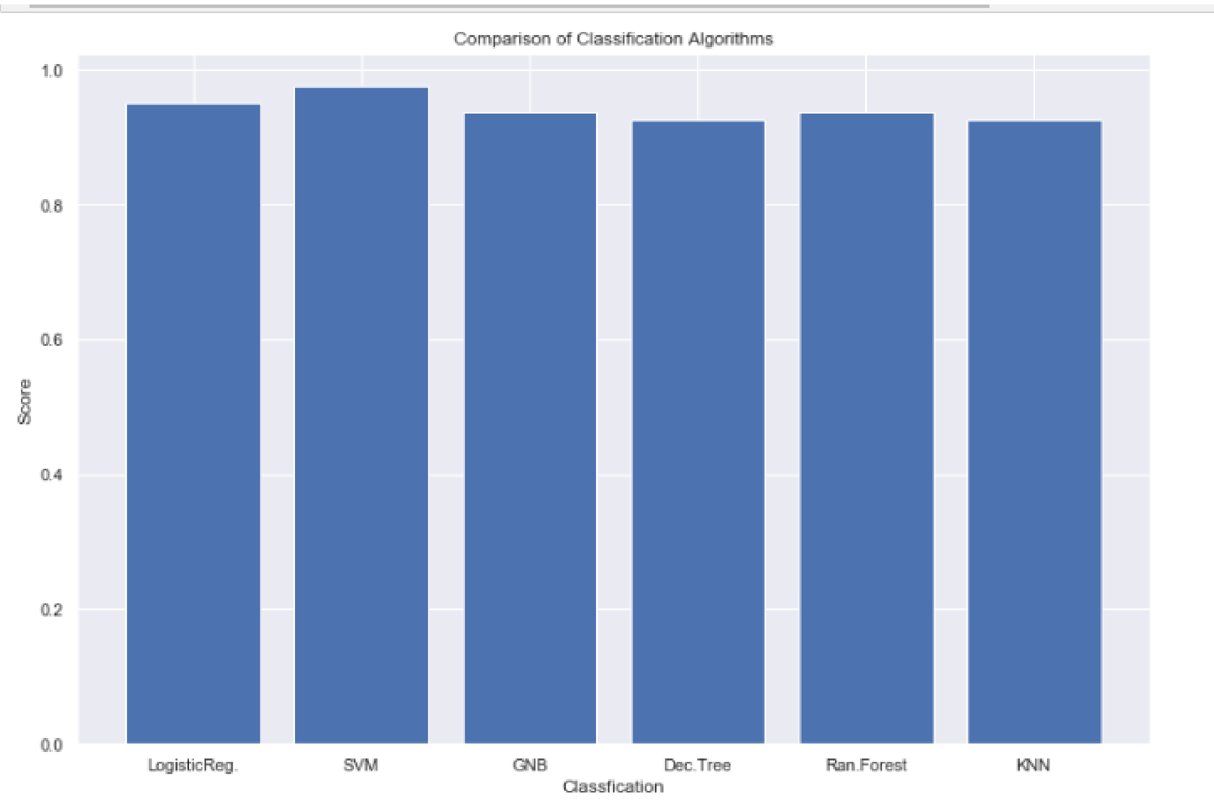

Fig 2 Comparison of Classification Algorithms

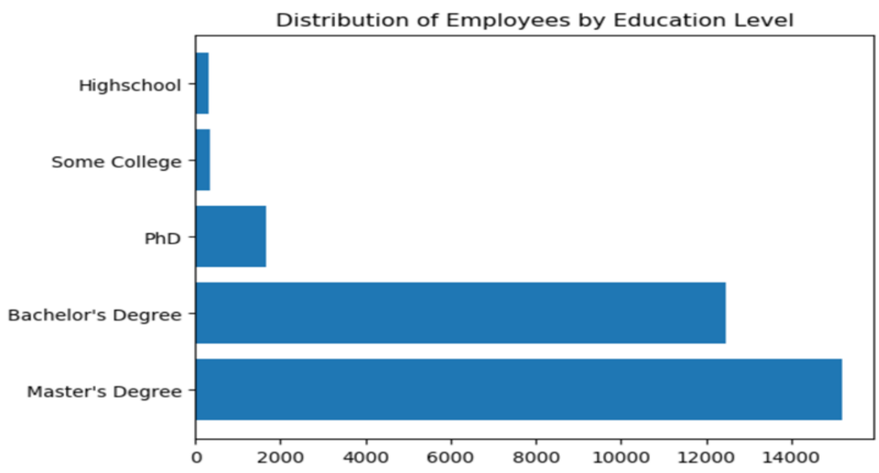

Fig 3 Distribution of Employees by Education

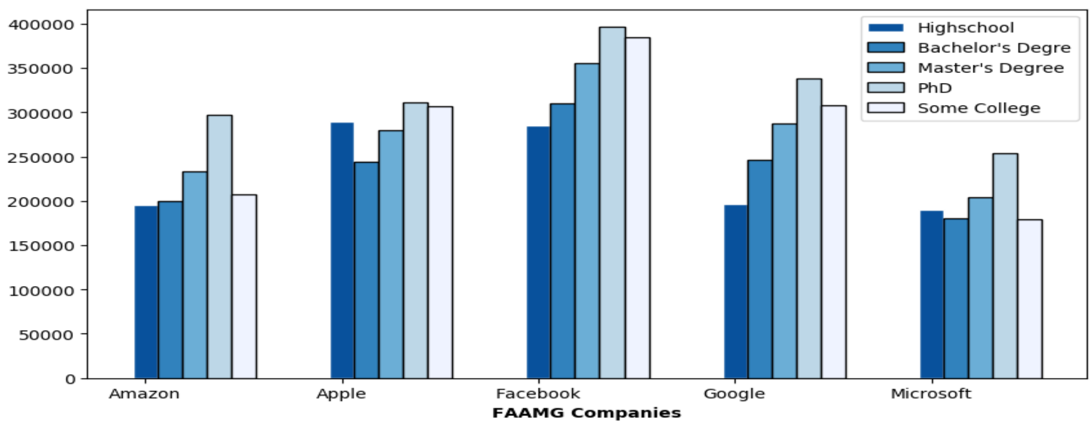

Fig 4 Salary Distribution v/s Company

VII. CONCLUSION

Predicting whether pursuing a master's degree is worth it or not using machine learning techniques is a relevant and valuable application in the field of education and career planning. Informed Decision-Making: Many individuals face the decision of whether to pursue a master's degree, which entails significant investment in terms of time, effort, and finances. By developing a machine learning model to predict the worthiness of pursuing a master's degree, individuals can make more informed decisions based on data-driven insights. This helps them assess the potential benefits and drawbacks of obtaining a master's degree in their specific circumstances. Machine learning models can take into account various factors such as the cost of tuition, potential salary increments, employment opportunities, and industry trends. By considering these factors, the model can provide an estimation of the return on investment (ROI) associated with pursuing a master's degree. This information empowers individuals to make cost-benefit analyses and weigh the potential advantages and disadvantages before committing to further education. Machine learning models can be trained on historical data that includes information about individuals who pursued a master's degree and their subsequent career outcomes. By analyzing patterns and correlations within the data, the model can provide personalized recommendations to individuals based on their unique profiles, academic background, work experience, and career aspirations. These recommendations can guide individuals in making decisions aligned with their goals and increase their chances of success.

The machine learning project involves predicting the chance of admission based on various factors such as GRE scores, TOEFL scores, letter of recommendation, SOP, and CGPA. The project aims to develop a model that can accurately predict the likelihood of an applicant being admitted to a particular educational institution.

The project utilizes a dataset that includes information on these factors for a set of individuals. The dataset is used to train and evaluate the machine learning model. Various evaluation metrics such as accuracy, precision, recall, F1 score, and regression metrics like MSE and RMSE are used to assess the model's performance.

Through the project, the team explores different aspects of the data, conducts feature engineering, and selects an appropriate algorithm for the prediction task. The model is then evaluated using cross-validation techniques and fine-tuned to achieve the best possible performance.

In conclusion, this machine learning project provides a solution for predicting the chance of admission, which can be valuable for both applicants and educational institutions. By leveraging the provided dataset and implementing an appropriate machine learning model, the project offers a tool to assess an applicant's likelihood of admission based on their academic achievements, recommendation letters, and other factors.

The project's accuracy and performance metrics indicate the model's effectiveness in predicting admission outcomes. However, it is important to recognize the limitations and potential biases inherent in the dataset and model. Factors such as sample bias, limited features, and changing admission policies can impact the model's performance and generalizability.

This machine learning project holds significant relevance in the context of college admissions. It provides a data-driven approach to support decision-making for both applicants and educational institutions. For applicants, the model offers insights into their chances of admission and can guide them in making informed choices about their educational paths. For educational institutions, the model can help streamline the admission process and identify promising candidates.

Additionally, the project showcases the application of machine learning techniques in the field of education. It demonstrates how data analysis and predictive modeling can be used to gain insights and improve decision-making in the admissions process. The project's findings and methodologies can serve as a foundation for further research and development in the field of educational data analysis and predictive modeling.

Overall, this machine learning project provides a valuable tool for predicting admission chances, facilitating decision-making, and promoting fairness and transparency in the college admissions process.

REFERENCES

[1]. Shivani Pal, Anjali Shahi, Shubham Rai, Reshma Gulwani, "Course Recommendation System for Post-graduate (Masters in Science) Aspirants," in Proceedings of International Conference on Computational Intelligence and Data Engineering, January 2021, DOI: 10.1007/978-981-15-8767-2_31.
[2]. Daohua Yu, Xin Zhou, Yu Pan, "University Academic Performance Development Prediction Based on TDA," in Entropy, vol. 25, no. 1, pp. 24, December 2022, DOI: 10.3390/e25010024, License: CC BY 4.0.
[3]. Daifu Qiao, "Online Education Resource Recommendation System of International Finance Course Based on Preference Data Collection," in e-Learning, e-Education, and Online Training, March 2023, DOI: 10.1007/978-3-031-21164-5_4.
[4]. Jingbin Zhang, "Course Recommendation Method and System of Education Platform Based on Deep Learning," in Application of Big Data, Blockchain, and Internet of Things for Education Informatization, Second EAI International Conference, BigIoT-EDU 2022, Virtual Event, July 29–31, 2022, Proceedings, Part I, January 2023, DOI: 10.1007/978-3-031-23950-2_44.

[5]. Raymond D. Carson, Philip A. Reed, "Pre-College Career Guidance on Student Persistence and Performance at a Small Private University," in Career and Technical Education Research, vol. 40, no. 2, pp. 99-112, October 2019, DOI: 10.5328/cter40.2.99.

[6]. Mafufah Hastin, Najlatun Naqiyah, Eko Darminto, "Guidance and Counseling Services to Develop Student Career Maturity," November 2022, DOI: 10.46245/ijorer.v3i6.261, License: CC BY-SA 4.0.

[7]. Manoj Kumar H, Sufiyan S A, Vinay L, "STUDENT CAREER GUIDANCE," in International Research Journal of Computer Science, vol. 9, no. 8, pp. 312-315, August 2022, DOI: 10.26562/irjcs.2022.v0908.30.

[8]. Ae-Kyoung SUN, "Multi-level Analysis on Factors Influencing Career maturity of High School Students," May 2022, DOI: 10.24299/kier.2022.352.139.

[9]. Areej Kamal, Batul Naushad, Hadia Rafiq, Shahab Tahzeeb, "Smart Career Guidance System," in Proceedings of 2021 9th International Conference on Computer and Communications Security (ICCCI), October 2021, DOI: 10.1109/ICCIS54243.2021.9676408.

[10]. Kasem Seng, Akram M. Zeki, "Career Guidance and Employment Management System," in Proceedings of 2019 International Conference on Advanced Computer Science and Applications (ACSAT), December 2019, DOI: 10.1109/ACSAT.2019.20.

[11]. Milen Todorov, Gergana Avramova-Todorova, "Generalized network model for career guidance using emotional intelligence, self-assessment and attitudes toward the employer," in Proceedings of 2021 IEEE Bulgarian Control Days (BdKCSE), May 2021, DOI: 10.1109/BdKCSE53180.2021.9627315.

[12]. Kazi Fakir Mohammed, Sushopti Gawade, "Online career counseling system using adaptive e-learning," in Proceedings of 2017 IEEE International Conference on Recent Innovations in Signal Processing and Embedded Systems (RISE), December 2017, DOI: 10.1109/RISE.2017.8378222.

[13]. Ishaan Dutta, Atharv Shah, Ms. M Safa, "Building a Recommendation System for Career Advice for Students from Professionals," in Proceedings of 2022 5th International Conference on Communication Systems and Network Technologies (ICCCSNT), November 2022, DOI: 10.1109/ICCCI54379.2022.9740837.

[14]. Ying Cao, Lei Zhang, "Research about the college students career counseling expert system based on agent," in Proceedings of 2011 International Conference on Advanced Information Management and Service (IMS), December 2011, DOI: 10.1109/AIMSEC.2011.6010332.

[15]. S Vignesh, C Shivani Priyanka, H Shree Manju, "An Intelligent Career Guidance System using Machine Learning," in Proceedings of 2021 12th International Conference on Advanced Computing and Communication Systems (ICACCS), May 2021, DOI: 10.1109/ICACCS51430.2021.9441978.

[16]. Raza, S., Al-Ebadi, R., & Luo, Y. (2020). An Intelligent Career Guidance System for University Students. IEEE Access, 8, 74951-74962. DOI: 10.1109/ACCESS.2020.2991741.

[17]. Oladimeji, P., Nwachukwu, C., & Oyerinde, D. (2020). Development of an Intelligent Career Guidance System using Rule-based Reasoning. Journal of Intelligent Systems, 29(1), 179-194. DOI: 10.1515/jisys-2018-0402.

[18]. Kim, S., Kim, G. J., & An, H. (2018). An Intelligent Career Guidance System based on Learning Analytics. IEEE Transactions on Education, 61(2), 138-146. DOI: 10.1109/TE.2017.2786078.

[19]. Shi, J., Wu, H., & Jiang, H. (2019). An Intelligent Career Guidance System for College Students Based on Semantic Similarity Computing. IEEE Access, 7, 45989-46001. DOI: 10.1109/ACCESS.2019.2903013.

[20]. Chen, Y., Cheng, L., & Li, Y. (2018). An Intelligent Career Guidance System for College Students based on Big Data and Data Mining. In Proceedings of the 2018 International Conference on Artificial Intelligence and Big Data (ICAIBD) (pp. 274-279). DOI: 10.1109/ICAIBD.2018.8393690.

[21]. Zhang, Y., Gu, Q., Guo, Y., & Ma, Y. (2019). An Intelligent Career Guidance System Based on Multi-source Data Analysis. Journal of Computer Applications, 39(2), 544-549. DOI: 10.11772/j.issn.1001-9081.2018.01.221.

[22]. Majumdar, D., & Mukherjee, A. (2018). An Intelligent Career Guidance System based on Personality and Learning Styles using Fuzzy Logic. In Proceedings of the 2018 International Conference on Advanced Computing and Communication Systems (ICACCS) (pp. 1-6). DOI: 10.1109/ICACCS.2018.8722895.

[23]. Ren, H., Zhang, J., & Cheng, H. (2018). An Intelligent Career Guidance System for College Students based on Personalized Recommendation. Journal of Intelligent & Fuzzy Systems, 35(6), 6433-6441. DOI: 10.3233/JIFS-171102.

[24]. Bhanu, B. V., Sudheer, V., & Poonia, R. C. (2018). An Intelligent Career Guidance System for Effective Decision Making. In Proceedings of the 2018 International Conference on Advances in Computing, Communication Control and Networking (ICACCCN) (pp.299-304).DOI: 10.1109/ICACCCN.2018.8668124.

[25]. Prasad, A., Kumar, V., & Jain, M. (2017). An Intelligent Career Guidance System using Machine Learning Techniques. In Proceedings of the 2017 2nd International Conference on Signal Processing and Integrated Networks (SPIN) (pp. 94-98). DOI: 10.1109/SPIN.2017.8340100.