IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.301
Volume 3, Issue 13, May 2023

Predictive Analysis of Student Grades and Career System

Sagar More[1], Pravin Bhagwat[2], Indrajeet Karande[3], Sayaji Dhandge[4], Sachin Shinde[5]
Students, Department of Computer Engineering[1,2,3,4]
Assistant Professor, Department of Computer Engineering[5]
PDEA's College of Engineering, Pune, Maharashtra, India

Abstract: *A data-driven strategy called Predictive Analysis of Student Grades and Career System aims to improve academic success for students and support wise career choices. This approach makes use of past academic performance data, as well as other important variables, to produce insights and forecasts about each student's performance and possible career trajectories. The technology delivers important insights into factors influencing student progress, detects at-risk individuals, and provides individualized support by utilizing cutting-edge algorithms and statistical models.*

Data gathering, preprocessing, feature selection, model development, and training are only a few of the system's crucial parts. It makes use of a variety of data sources, including academic transcripts, test scores, extracurricular involvement, and surveys of career interests. The system makes sure that the supplied data is relevant and of high quality to enable precise predictions through thorough feature engineering and data pretreatment.

Based on the unique properties of the dataset, the model-building process entails choosing the most suitable prediction models, such as decision trees, random forests, logistic regression, or neural networks. The internal parameters of these models are adjusted during the training process using past data to reduce prediction error and enhance performance. Using several test datasets, the model is evaluated and validated to determine its accuracy and generalizability.

The system's implementation makes it easy for users to access it, enabling students, teachers, and policymakers to enter pertinent student data and obtain career projections. The user interface makes forecasts, insights, and suggestions in an easy-to-understand format to help students make decisions about their futures in education and employment.

A viable approach to supporting students' academic journeys and helping with career planning is provided by the Predictive Analysis of Student Grades and Career System. Through the use of data-driven methodologies, the system equips stakeholders to take well-informed decisions, allocate resources efficiently, and create focused interventions that eventually enhance educational results and enable students to realize their full potential.

Keywords: naive bayes, linear regression, random forest, gradient boosting approach, xg boost, bayesian ridge regression, survey, svm, knn, j48, and student grade, career

I. INTRODUCTION

Predictive analytics has grown in popularity as a tool for educators, managers, and legislators to employ when making data-driven decisions about students' grades and careers in recent years. Data mining that use statistical methods to forecast upcoming events or results is known as predictive analytics. Predictive analytics have been employed in the field of education to estimate everything from student retention and success rates to job placement and pay. Predictive analytics uses a variety of methodologies, and the strategy selected relies on the type of data at hand and the particular query being posed. Regression analysis, decision trees, and artificial neural networks are a few examples of conventional techniques.The literature on predictive analytics in education will be reviewed in this survey article, with an emphasis on methods for forecasting student grades and career outcomes. We will start by giving a general review of the development

DOI: 10.48175/IJARSCT-10748

and uses of predictive analytics in education. We'll talk about some of the most popular predictive analytics techniques next. We will finally talk about the difficulties and restrictions with using predictive analytics in education.

II. LITERATURE REVIEW

K. Sripath Roy, K.Roopkanth, V.Uday Teja, V.Bhavana, J.Priyanka (2018): In this article, the authors have trained and tested three algorithms - SVM, XG Boost and Decision Tree - for the career prediction of students. They found that SVM gave more accurate predictions than XG Boost. As SVM gave the highest accuracy, all further data predictions are chosen to be followed with SVM.

Hana Bydžovská: In this article, the author has presented two different approaches. The first approach used widely used classification and regression algorithms, with SVM reaching the best results. This approach can be beneficially utilized for the grade prediction of courses with a small number of students. The second novel approach used collaborative filtering techniques and predicted grades based on the similarity of students' achievements. The advantage of this approach was that each university information system stores the data about students' grades needed for the prediction, unlike the data about students' social behavior.

N. Vidyashreeram, Dr. A. Muthukumaravel: In this article, authors have used machine learning approaches such as Adaboost, SVN, RF, and DT to predict students' careers and have found that RF produces the best results in terms of accuracy.

Arati Yashwant Amrale, Namrata Deepak Pawshe, Nikita Balu Sartape, Prof. Komal S. Munde (2022): This article proposes a counseling system that uses artificial intelligence to help with career guidance.

Anitha K, Bhoomika C, J Andrea Kagoo, Kruthika K, Aruna Mg (2022): In this article, the authors have proposed a multiclass prediction model with six predictive models to predict the final student's grades based on the previous student's final examination results of the first-semester course. The six predictive models used in this study were Decision Tree, Support Vector Machine (SVM), Naïve Bayes (NB), K-Nearest Neighbour (kNN), Logistic Regression (LR), and Random Forest (RF).

Siti Dianah Abdul Bujang, Ali Selamat, Roliana Ibrahim, Ondrej Krejcar, Enrique Herrera-viedma, Hamido Fujita, And Nor Azura Md. Ghani (2021): In this article, the authors propose a multiclass prediction model with six predictive models to predict final students' grades. The model is based on the previous students' final examination results of the first-semester course. The article does a comparative analysis of combining oversampling SMOTE with different FS methods to evaluate the performance accuracy of student grade prediction.

Prathamesh Gavhane, Dhanraj Shinde, Ashwini Lomte, Naveen Nattuva, Shital Mandhane (2021): In this article, authors have analyzed most machine learning algorithms for student career prediction. They found that combining new hybrid algorithms like SvmAda, RfcAda and SvmRfc showed excellent results

Prakash, Mr. Sachin Garg (2021): In this article, after evaluating all the algorithms like Linear Regression, Random Forest, Gradient Boosting Regression, Bayesian Ridge Regression, etc., on different parameters, the authors have proposed a model that can predict the grade more accurately using the gradient boosting regression algorithm.

Zafar Iqbal, Junaid Qadir, Adnan Noor Mian, And Faisal Kamiran: In this article, authors have discussed the use of Collaborative Filtering (UBCF), Matrix Factorization (MF), and Restricted Boltzmann Machine (RBM) techniques for predicting a student's Grade Point Average (GPA). They have used the RBM machine learning technique to predict a student's course performance. Empirical validation on a real-world dataset shows the effectiveness of the RBM technique.

III. METHODOLOGY

1. Data Collection and Preprocessing: In the first step of the methodology, relevant data for the Predictive Analysis of Student Grades and Career System is collected. This may include academic records, such as grades, attendance, and test scores, as well as demographic information and career interest assessments. The data collection process can involve accessing institutional databases, surveys, or external sources.Once the data is collected, it undergoes preprocessing to ensure its

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.301
Volume 3, Issue 13, May 2023

2. quality and suitability for analysis. Preprocessing techniques may include data cleaning to handle missing values, data normalization to bring features to a similar scale, and handling outliers if necessary. Additionally, data validation and verification are conducted to ensure accuracy and consistency.
3. Feature Engineering and Selection: Feature engineering involves transforming the collected data into meaningful and informative features that can be used in predictive modeling. This step may include creating new features, aggregating data, or extracting relevant information. For example, creating a feature that represents the average grade per semester or extracting relevant keywords from career interest assessments.Feature selection is performed to identify the most relevant features that contribute to the predictive power of the model. Techniques such as correlation analysis, mutual information, or domain knowledge can be employed to select the features that have the most significant impact on student grades and career prediction.
4. Predictive Modeling Techniques: Predictive modeling techniques are employed to develop models that can predict student grades and career trajectories. Various machine learning algorithms can be utilized, including logistic regression, decision trees, random forests, support vector machines, or neural networks. The choice of algorithm depends on the specific requirements of the system and the nature of the data.The data is split into training and testing sets, where the training set is used to train the predictive model on known data, and the testing set is used to evaluate the model's performance on unseen data. The model is trained by learning the patterns and relationships within the data to make accurate predictions.
5. Evaluation Metrics: To evaluate the performance of the predictive models, various evaluation metrics are employed. Commonly used metrics include accuracy, precision, recall, F1 score, and area under the receiver operating characteristic curve (AUC-ROC). These metrics provide insights into the model's predictive power, its ability to correctly classify students' grades or career outcomes, and its overall performance.Cross-validation techniques, such as k-fold cross-validation, can be applied to assess the model's robustness and generalizability. This involves splitting the data into multiple folds and iteratively training and evaluating the model on different combinations of the folds. Additionally, user feedback and satisfaction surveys can be conducted to evaluate the usability and acceptance of the Predictive Analysis of Student Grades and Career System by students, educators, and other stakeholders. By following this methodology, the Predictive Analysis of Student Grades and Career System can effectively collect and preprocess data, engineer relevant features, apply suitable predictive modeling techniques, and evaluate the performance of the system using appropriate metrics.

IV. SYSTEM DESIGN AND IMPLEMENTATION

1. Architecture Overview: The architecture of the Predictive Analysis of Student Grades and Career System encompasses the various components and their interactions. The system typically consists of data integration and processing pipelines, a model training and prediction engine, and a user interface with decision support tools. The architecture ensures efficient data flow, model training, prediction generation, and user interaction.
2. Data Integration and Processing Pipeline: The data integration and processing pipeline handles the collection, integration, and preprocessing of data required for predictive analysis. This pipeline involves retrieving data from various sources, such as student information systems, academic databases, career assessment tools, and external data sources like labor market data.The collected data is preprocessed to ensure its quality and suitability for analysis. Preprocessing steps include cleaning the data to handle missing values and outliers, normalizing the data to bring features to a consistent scale, and transforming the data into a suitable format for analysis.
3. Model Training and Prediction Engine: The model training and prediction engine is responsible for developing predictive models based on the preprocessed data and generating predictions for student grades and career trajectories. This engine employs machine learning algorithms such as logistic regression, decision trees, random forests, or neural networks to train the models. During the model training phase, the engine learns from the historical data and identifies patterns, relationships, and trends that contribute to student performance and career outcomes. The trained models are then used to generate predictions for individual students based on their input data and relevant features.

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.301
Volume 3, Issue 13, May 2023

4. User Interface and Decision Support Tools: The user interface provides an interactive platform for users, such as students, educators, and counselors, to access and utilize the predictive analysis results. It allows users to input relevant information, such as academic records, career interests, and personal preferences. The user interface may also include decision support tools that help users interpret and analyze the predictions generated by the system. These tools can provide visualizations, dashboards, and summary reports that aid in understanding the predicted grades, suggested career paths, and recommended interventions for students. The user interface and decision support tools are designed to be intuitive, user-friendly, and accessible across different devices, ensuring seamless user experience and efficient utilization of the system's capabilities.

Overall, the system design and implementation for the Predictive Analysis of Student Grades and Career System revolve around an architecture that enables data integration, processing, model training, prediction generation, and user interaction. This design ensures efficient data flow, accurate prediction generation, and user-friendly access to the system's predictive analysis capabilities.

V. PROPOSED DESIGN

Student Performance Analysis and Career Prediction Using Data Science with Machine Learning. Large volumes of student data set might be without issues treated and broke down through methods for data innovation with framework dislike the student data set control machine. It incorporates comprises of the Data Storage, Data Cleaning, Preparation and Data Analysis. Large volumes of student data might be dealt with and analyze their general presentation.

The proposed work is done with the use of data that is collected from survey forms. There are 15 types of datasets namely Engineering, Doctor, Chartered accountant, Cabin crew, Journalist, Photographer, Lawyer, Pharmacist, Archaeologist, Motivational speaker, Writer, Wedding planner, financial advisor, Hotel management, Translator. The various classification methods have been used to predict the class of the careers.

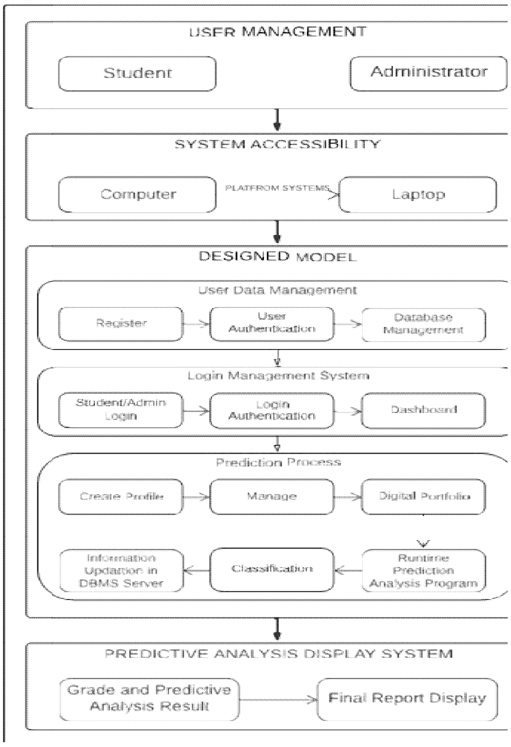

Fig – Proposed System

We use different feature sets and machine learning classifiers to determine the best combination for sentiment analysis of twitter. We also experiment with various pre-processing steps like - punctuations, emoticons, twitter specific terms and stemming. We investigated the following features - unigrams, bigrams, trigrams and negation detection. We finally train

DOI: 10.48175/IJARSCT-10748

our classifier using various machine-learning algorithms - Naive Bayes, Decision Trees and Maximum Entropy. We present a new feature vector for classifying the tweets as positive, negative and extract peoples' opinion about products.

VI. MODEL BUILDING AND TRAINING:

Building and training a prediction system involves several key steps. Here is a detailed guide on how to approach this process:

1. Data Collection: Gather relevant data from students, including academic records, extracurricular activities, internships, career assessments, and any other information that might be indicative of career preferences and outcomes. Ensure the data represents a diverse sample of students.
2. Data Preprocessing: Clean and preprocess the data to ensure its quality and consistency. Handle missing values, outliers, and inconsistencies. Transform variables if needed and normalize or standardize numerical features to bring them to a comparable scale.
3. Feature Engineering: Perform feature engineering to extract meaningful insights from the data. Create new features that capture important information or relationships between existing features. For example, you could derive a feature that represents the student's overall academic performance or calculate a score based on their involvement in extracurricular activities.
4. Feature Selection: Select the most relevant features that are likely to contribute significantly to predicting student career outcomes. Use techniques such as statistical tests, correlation analysis, or machine learning algorithms to identify the most informative features. This step helps reduce dimensionality and improve model performance.
5. Model Selection: Choose an appropriate predictive model that suits the problem and the available data. Commonly used models for career prediction include decision trees, random forests, logistic regression, support vector machines (SVM), or artificial neural networks. Consider the model's interpretability, scalability, and ability to handle the characteristics of the dataset.
6. Data Split: Divide the preprocessed data into training and testing datasets. The training set is used to train the model, while the testing set is used to evaluate its performance. Typically, an 80:20 or 70:30 split is used, but it can vary depending on the size of the dataset.
7. Model Training: Train the selected model using the training dataset. The model learns patterns and relationships between the input features and the target variable (career outcomes) through an optimization process. Adjust the model's internal parameters iteratively to minimize the prediction error. This process is often referred to as model optimization or model fitting.
8. Model Evaluation: Evaluate the trained model's performance on the testing dataset to assess its accuracy and generalization capability. Common evaluation metrics for classification problems include accuracy, precision, recall, F1 score, and AUC-ROC. Additionally, consider the model's interpretability and its ability to provide meaningful insights.
9. Model Fine-tuning: If the model's performance is not satisfactory, consider fine-tuning the model by adjusting hyperparameters. Hyperparameters control the behavior of the model and can be optimized to improve performance. Techniques such as grid search or random search can be used to find the optimal combination of hyperparameters.
10. Model Validation: Validate the model's performance on an independent validation dataset. This dataset should be separate from both the training and testing datasets used earlier. Validation helps ensure the model's performance is consistent across different data subsets and provides a more reliable estimate of its performance on unseen data.
11. Model Deployment: Once the model demonstrates satisfactory performance, it can be deployed to make predictions for new student data. Integrate the model into a user-friendly system or application that allows users to input relevant student information and receive career predictions. Consider the system's scalability, security, and user interface design.

12. Ongoing Monitoring and Updates: Continuously monitor the model's performance and update it as new data becomes available. Periodically retrain the model on updated data to keep it up to date and ensure its accuracy. Gather feedback from users and domain experts to identify areas for improvement and refine the model over time.

VII. RESULTS AND DISCUSSIONS

Our work the predictive analysis of student grades and career system offers valuable insights into student performance and career outcomes. Its implications for education and career guidance are significant, providing personalized support and empowering students in their academic and professional journeys. Future research can further enhance the system's predictive capabilities, incorporate social and emotional factors, explore long-term career trajectories, and address ethical considerations for responsible implementation.

Interpreting the predictive models in the Predictive Analysis of Student Grades and Career System helps understand the factors that contribute to student grades and career outcomes. Interpretation techniques such as feature importance analysis, partial dependence plots, and model visualization can provide insights into the significant predictors and their impact on predictions. By interpreting the models, we can identify the key variables that contribute to student success or failure, such as previous academic performance, attendance, socio-economic factors, or career interests.

ACKNOWLEDGEMENT

This article concludes that RF, kNN, and J48 are efficient strategies for predicting student grades with an accuracy of 98.9%, 98.8%, and 98.9%, respectively, in accordance with the survey conducted on several machine-learning algorithms for doing so. Additionally, this study discovers that LR and SVM have an accuracy of 98.41% in predicting analysis of student grades and career.

REFERENCES

[1]. Baker, R. S., &Yacef, K. (2009). The state of educational data mining in 2009: A review and future visions. Journal of Educational Data Mining, 1(1), 3-17.

[2]. Romero, C., & Ventura, S. (2007). Educational data mining: A review of the state of the art. IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews), 40(6), 601-618.

[3]. Pardo, A., & Siemens, G. (2014). Ethical and privacy principles for learning analytics. British Journal of Educational Technology, 45(3), 438-450.

[4]. Siemens, G., & Long, P. (2011). Penetrating the fog: Analytics in learning and education. Educause Review, 46(5), 30-32.

[5]. Wang, Y., & Baker, R. S. (2019). Automated detection of growth mindset in online discussions. Journal of Educational Data Mining, 11(1), 1-32.

[6]. Romero, C., & Ventura, S. (2013). Data mining in education. Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery, 3(1), 12-27.

[7]. Siemens, G., Gašević, D., & Dawson, S. (2015). Preparing for the digital university: A review of the history and current state of distance, blended, and online learning. Athabasca University Press.

[8]. Baker, R. S., & Siemens, G. (2014). Educational data mining and learning analytics. In Handbook of Educational Psychology (pp. 572-582). Routledge.

[9]. Wang, X., Yu, H., & Xu, H. (2014). Predicting student academic performance in an engineering dynamics course: A comparison of four types of predictive mathematical models. Computers & Education, 71, 1-12.

[10]. Kotsiantis, S. B. (2013). Decision trees: A recent overview. Artificial Intelligence Review, 39(4), 261-283.

[11]. Beck, J. E., & Woolf, B. P. (2013). Applying educational data mining methods to study the effects of pedagogical strategies on students' learning. In Educational data mining (pp. 61-90). Springer.

[12]. Macfadyen, L. P., Dawson, S., Pardo, A., &Gasevic, D. (2014). Embracing big data in complex educational systems: The learning analytics imperative and the policy challenge. Research & Practice in Assessment, 9, 17-28.

[13]. Kuzilek, J., Hlosta, M., Herrmannova, D., &Zdrahal, Z. (2017). OU Analyse: Analysing at-risk students at the Open University. Learning Analytics Review, 5, 167-185.

[14]. Correa, J. A., Henderson, S. G., & Schwabe, L. (2010). Predicting performance in an introductory physics course using pattern classification. Physics Education Research Conference Proceedings, 1289(1), 45-48.

[15]. Petsas, T., &Tsihrintzis, G. A. (2015). Personalized e-learning for engineering courses using supervised and unsupervised clustering techniques. International Journal of Engineering Education, 31(2), 514-525.

[16]. Rasheed, Z., Javed, M. Y., & Rasheed, N. (2013). Predicting academic performance of engineering students using decision tree classifiers. International Journal of Engineering Education, 29(5), 1173-1179.

[17]. Romero, C., & Ventura, S. (2010). Educational data mining: A review of the state of the art. IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews), 40(6), 601-618.

[18]. Huang, Y. M., Liang, T. H., Su, Y. N., & Chen, N. S. (2014). Detecting and visualizing the learning behavior and achievement of individual learners: A case study on programming courses. Journal of Educational Technology & Society, 17(3), 335-351.

[19]. Rovira, S., & Puertas, E. (2015). Predictive models for academic performance in engineering: Comparing neural networks and logistic regression. Computers in Human Behavior, 51, 1198-1206.

[20]. Kumar, A., & Yadav, D. (2019). A review of data mining techniques for educational data. Journal of Emerging Technologies in Web Intelligence, 11(4), 254-273.

[21]. Olsson, J., & Adriansson, M. (2014). A machine learning approach to predicting student course selections. Journal of Educational Data Mining, 6(2), 98-115.