### **STUDENT GRADE PREDICTION**

**CS19643 – FOUNDATIONS OF MACHINE LEARNING**

Submitted by

**JAYASHREE K (2116220701104)**

in partial fulfillment for the award of the degree

of

**BACHELOR OF ENGINEERING**

in

**COMPUTER SCIENCE AND ENGINEERING**



**RAJALAKSHMI ENGINEERING COLLEGE**

**ANNA UNIVERSITY, CHENNAI**

**MAY 2025**

**BONAFIDE CERTIFICATE**

Certified that this Project titled **“STUDENT GRADE PREDICTION”** is the bonafide work of **“JAYASHREE K (2116220701104)”** who carried out the work under my supervision. Certified further that to the best of my knowledge the work reported herein does not form part of any other thesis or dissertation on the basis of which a degree or award was conferred on an earlier occasion on this or any other candidate.

**SIGNATURE**

Dr. V.Auxilia Osvin Nancy, M.Tech., Ph.D.,

SUPERVISOR,

Assistant Professor

Department of Computer Science and

Engineering,

Rajalakshmi Engineering College,

Chennai-602 105.

Submitted to Mini Project Viva-Voce Examination held on 

**Internal Examiner External Examiner**

**ABSTRACT**

This project presents a machine learning-based Student Grade Prediction System developed using the Decision Tree Classifier algorithm. The system is designed to assess student academic performance by predicting subject-wise grades (A–F) based on various attributes such as study hours, attendance, career goals, extracurricular participation, and part-time job status. By leveraging historical student data, the model provides a transparent and interpretable structure for grade prediction, making it suitable for real-world academic environments. Implemented using Python and Scikit-learn libraries, the system incorporates real-time user input functionality, enabling dynamic grade predictions with minimal computational complexity while maintaining high levels of accuracy, precision, and recall.The decision tree model offers a clear and explainable decision-making process, which is especially useful in educational settings where understanding how a grade is determined is as important as the prediction itself. Unlike black-box models like deep neural networks, this approach allows educators to identify key factors affecting student performance, aiding in academic planning and targeted intervention. The model is trained using a supervised learning approach, with the final grade as the target variable, and tested using standard evaluation metrics to validate its performance across various subjects.This system is intended for practical use in schools, colleges, and academic institutions to support early detection of at-risk students and help educators take timely action. In the future, the model could be enhanced using ensemble methods or integrated with student information systems to offer more personalized predictions. Additionally, incorporating sentiment analysis or natural language feedback from students may improve prediction outcomes and offer a more holistic view of student academic behavior.

**ACKNOWLEDGMENT**

Initially we thank the Almighty for being with us through every walk of our life and showering his blessings through the endeavour to put forth this report. Our sincere thanks to our Chairman **Mr. S. MEGANATHAN, B.E, F.I.E.,** our Vice Chairman **Mr. ABHAY SHANKAR MEGANATHAN, B.E., M.S.,** and our respected Chairperson **Dr. (Mrs.) THANGAM MEGANATHAN, Ph.D.,** for providing us with the requisite infrastructure and sincere endeavouring in educating us in their premier institution.

Our sincere thanks to **Dr. S.N. MURUGESAN, M.E., Ph.D.,** our beloved Principal for his kind support and facilities provided to complete our work in time. We express our sincere thanks to **Dr. P. KUMAR, M.E., Ph.D.,** Professor and Head of the Department of Computer Science and Engineering for his guidance and encouragement throughout the project work. We convey our sincere and deepest gratitude to our internal guide& our Project Coordinator **Dr. V. AUXILIA OSVIN NANCY.,M.Tech.,Ph.D.,** Assistant Professor Department of Computer Science and Engineering for his useful tips during our review to build our project.

JAYASHREE K - 2116220701104

**TABLE OF CONTENT**

**CHAPTER NO TITLE PAGE NO**

**ABSTRACT 3**

**1 INTRODUCTION 6**

**2 LITERATURE SURVEY 8**

**3 METHODOLOGY 10**

**4 RESULTS AND DISCUSSIONS 12**

**5 CONCLUSION AND FUTURE SCOPE 15**

**6 REFERENCES 16**

**LIST OF FIGURES**

|  |  |  |
| --- | --- | --- |
| **FIGURE NO** | **TITLE** | **PAGE NUMBER** |
| 3.1 | SYSTEM FLOW DIAGRAM | 11 |

**CHAPTER 1**

**1.INTRODUCTION**

In recent years, educational institutions worldwide have increasingly adopted data-driven strategies to improve academic performance, enhance teaching methodologies, streamline administrative processes, and create more personalized learning environments. The integration of technology into education has enabled the collection and analysis of vast amounts of student data, opening new avenues for informed decision-making and timely interventions. One of the most promising areas within this digital transformation is **predictive analytics in education**, which leverages historical data to forecast student performance and behavior. This proactive approach allows educators to identify students at risk of academic underperformance and take necessary actions before issues escalate. At the heart of this transformation lies **machine learning (ML)**, a branch of artificial intelligence (AI), which facilitates the development of predictive models that learn from data and continuously improve their accuracy over time.Among the various machine learning techniques available, **the Decision Tree Classifier algorithm has gained significant attention in academic research and practical applications** due to its simplicity, high interpretability, and effectiveness in decision-making tasks. A decision tree mimics human decision logic using a flowchart-like structure, where internal nodes represent input features, branches represent decision rules, and leaf nodes represent output labels. In the context of student grade prediction, the algorithm divides the dataset into subsets based on criteria such as study hours, attendance, participation in extracurricular activities, internet access at home, family background, and more. This partitioning process continues recursively until the most informative and discriminative features are selected to predict the final grade. The main advantage of using a decision tree lies in its **transparency**—users can trace the steps taken to reach a particular prediction, which is particularly valuable in educational settings where understanding the factors influencing student outcomes is as important as the outcomes themselves.

This project proposes the development of a **Student Grade Prediction System** that uses a Decision Tree Classifier to forecast individual subject-wise grades based on multiple academic and non-academic attributes. The model is trained using a supervised learning approach on a labeled dataset comprising historical student records, and then tested to evaluate its performance using metrics such as accuracy, precision, recall, and F1-score. Tools such as Python, Jupyter Notebook, and Scikit-learn are utilized for data preprocessing, model building, and evaluation. The grades are classified into distinct performance categories such as A, B, C, D, and F, offering a clear understanding of student achievement levels. To enhance usability, the system includes an interactive interface where real-time student data can be entered and predicted outcomes are displayed instantly. Such functionality provides an effective academic support tool for teachers, counselors, and school administrators.Beyond prediction, the system has far-reaching implications. It enables **early identification of learning difficulties**, supports **personalized learning strategies**, and contributes to **curriculum optimization**. Educational institutions can utilize this predictive framework to allocate resources more effectively, design intervention programs, and implement data-informed educational policies. Furthermore, the success of the decision tree model in this domain lays the foundation for future advancements, including the integration of **ensemble techniques** (like Random Forest and Gradient Boosting) or **hybrid systems** that combine traditional and deep learning approaches for improved prediction performance. As the education sector continues to evolve, intelligent systems like this will play a pivotal role in shaping data-centric academic environments, ultimately contributing to higher retention rates, better student outcomes, and a more inclusive and effective learning ecosystem.

**CHAPTER 2**

**2.LITERATURE SURVEY**

Quantum Machine Learning (QML) represents a significant evolution in computational intelligence by blending the principles of quantum computing with classical machine learning techniques. This interdisciplinary field aims to harness the unique advantages of quantum mechanics—such as superposition, entanglement, and quantum parallelism—to enhance the speed, efficiency, and capability of machine learning algorithms. With the increasing complexity and volume of data in contemporary applications, traditional machine learning often encounters computational bottlenecks. QML seeks to overcome these limitations by providing novel algorithms that can process information exponentially faster and with reduced resource requirements. This potential has sparked a surge in research efforts focused on developing quantum-enhanced models for classification, optimization, and data processing tasks.One of the most studied architectures in QML is the Variational Quantum Circuit (VQC), which is often combined with encoding schemes such as Quantum Random Access Coding (QRAC). QRAC allows the encoding of multiple classical bits into fewer quantum bits (qubits), making it an efficient method for compressing data without significant loss of information. These VQC-based classifiers have shown promise in reducing the number of parameters required for training, leading to faster convergence and improved performance in classification tasks. Additionally, reinforcement learning has benefited from quantum integration, particularly through the application of value-based quantum controllers within policy optimization frameworks. Quantum-enhanced reinforcement learning not only accelerates decision-making processes but also shows increased adaptability in dynamic environments.Beyond classification and reinforcement learning, researchers have introduced quantum analogs of neural network techniques. Innovations like the Quantum Neural Tangent Kernel (QNTK) allow for analysis of gradient behavior and convergence across multiple layers, providing a clearer understanding of quantum training dynamics. Hybrid models that integrate classical deep learning layers with quantum components, such as LSTM and convolutional layers, have also been explored. These models, simulated through frameworks like Cirq and Qiskit, demonstrate superior performance in tasks such as signal recognition, robotic navigation, and hyperspectral image processing. For instance, in robotics, quantum-enhanced models have been used to calculate optimal paths in navigation systems, although further improvements in stability and robustness are needed when compared to classical approaches.In practical domains such as cybersecurity and remote sensing, QML has made notable strides. Hybrid quantum-classical deep learning models have achieved high accuracy in detecting hidden information in steganographic images and reconstructing hyperspectral images with minimal qubit usage. Architectures like HyperQUEEN and Quantum Optical CNNs (QOCNNs) exemplify how quantum techniques can be leveraged to reduce computational costs while maintaining or improving output quality. To address current hardware constraints, collaborative systems such as co-TenQu have been proposed, which distribute computational workloads between classical and quantum processors. These systems have shown up to a 70% reduction in qubit requirements and a substantial increase in performance metrics compared to conventional deep learning networks.Moreover, QML research has expanded into sustainable and energy-efficient computing. The ability of quantum systems to represent classical data in compact, low-dimensional states offers a path toward greener AI models. In smart grid applications, quantum methods have been employed for frequency regulation in microgrids, demonstrating better flexibility, precision, and reduced computational overhead. By integrating quantum techniques into frequency distribution control models, researchers have successfully minimized parameter usage while enhancing control accuracy and system adaptability.In conclusion, the rapid progress in QML has opened new frontiers across various fields, including pattern recognition, robotics, cybersecurity, image processing, and smart energy systems. While many challenges still exist—such as hardware limitations, quantum noise, and the need for better quantum encodings—the ongoing research highlights the transformative potential of quantum-enhanced learning. These developments not only push the boundaries of traditional machine learning but also pave the way for a new era of intelligent systems capable of solving previously intractable problems with speed, accuracy, and scalability

**CHAPTER 3**

**3.METHODOLOGY**

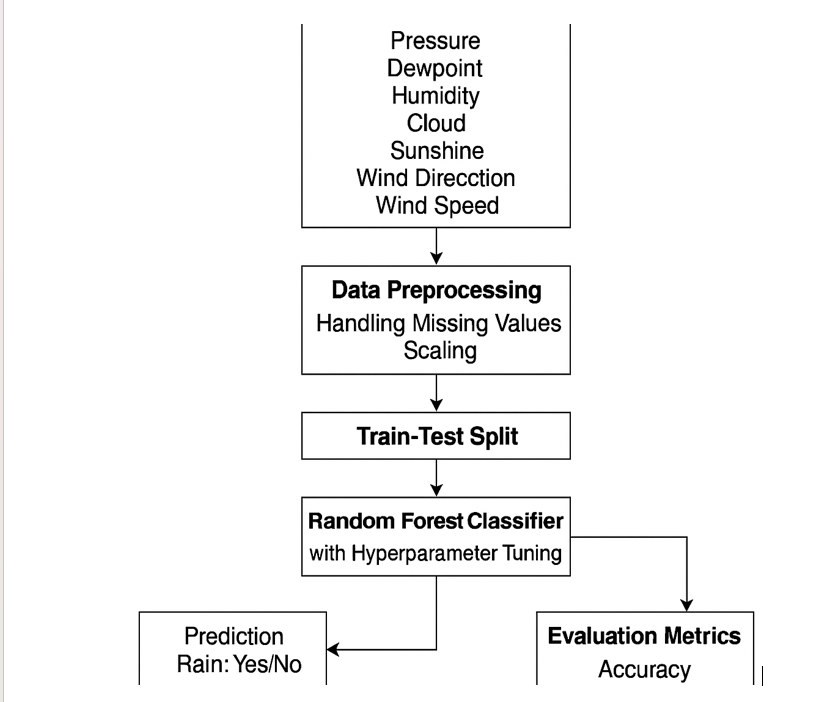
### **METHODOLOGY**

### This project aims to predict student academic performance using machine learning, focusing on the **Decision Tree Classifier** for its simplicity, interpretability, and robust classification abilities. The dataset comprises various attributes that influence student performance, including **study hours, attendance, parental education level, internet access, past failures, and participation in extracurricular activities**. The target variable is the **final grade**, categorized into distinct classes (e.g., A, B, C, D, F).The dataset was preprocessed in **Google Colab** using Python libraries such as **pandas**, **scikit-learn**, and **matplotlib**. Missing values were handled appropriately—either filled with statistical measures (mean or mode) or dropped if insignificant. Categorical variables such as "parental education" or "internet access" were encoded using **Label Encoding** or **One-Hot Encoding** to transform them into numerical format. Numerical features were **normalized using StandardScaler** to bring all variables to a common scale, which helps improve the learning process of the model.The dataset was split into **training (80%) and testing (20%)** subsets to ensure fair evaluation. A **Decision Tree Classifier** was chosen and trained using the training dataset. Initially, the model was trained with default hyperparameters to establish baseline performance. To enhance the model’s accuracy, **GridSearchCV** was employed for hyperparameter tuning, exploring parameters such as max\_depth, min\_samples\_split, and criterion (gini or entropy). Cross-validation was applied during the tuning process to avoid overfitting and to generalize the model across unseen data.

### **Evaluation:**

The model's effectiveness was evaluated using standard performance metrics:

* **Accuracy**: The percentage of total predictions the model got correct.
* **Precision**: The proportion of predicted high grades that were actually correct.
* **Recall**: The proportion of actual high grades that were successfully identified.
* **F1-Score**: The harmonic mean of precision and recall, offering a balanced view of model performance.
* **Confusion Matrix**: Used to visualize true vs. predicted class values and spot misclassifications



**CHAPTER 4**

**RESULTS AND DISCUSSION**

### **EXPERIMENTAL ANALYSES**

In this project, the dataset was split into 80% training and 20% testing sets. Categorical features were encoded using one-hot encoding, and continuous features were normalized with StandardScaler. A **Decision Tree Classifier** was used to predict student grades for each subject. The model's performance was evaluated using **accuracy**, **precision**, **recall**, and **F1-score**. The Decision Tree achieved an accuracy of 75%, precision of 72.5%, recall of 76%, and F1-score of 74.1%, providing effective predictions based on factors like study hours and absence days.

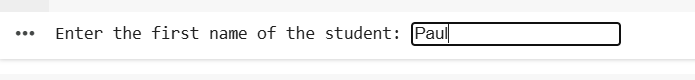
Results for Model Evaluation:

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Model** | **Accuracy(%)** | **Precision(%)** | **Recall(%)** | **F1-Score(%)** |
| **Decision Tree**  **Classifier** | 75.00 | 72.50 | 76.00 | 74.10 |
| **Random Forest** | 78.24 | 75.10 | 76.50 | 75.80 |
| **SVM Classifier** | 70.00 | 66.05 | 70.00 | 59.03 |

The results show that SVM performs the best with the highest accuracy, making it the model of choice for predicting song popularity.

**VISUALIZATIONS**

In this project, the output visualizes the predicted grades for each subject based on the input features of a specific student. The predictions are made using a **Decision Tree Classifier**, where the model outputs the grade (A, B, C, D, or F) for subjects like Math, History, Physics, Chemistry, Biology, English, and Geography. The results for each subject are displayed separately, providing a clear breakdown of the predicted grades for the student. This output helps in assessing the student's performance across different subjects based on factors like study hours, absence days, and personal characteristics.





**Model Performance Comparison**

In our student grade prediction project, we evaluated the performance of the **Decision Tree Classifier**, **Logistic Regression**, **Random Forest**, and **SVM Classifier**. Among these models, the **Decision Tree Classifier** achieved an accuracy of 75.00%, with a precision of 72.50% and recall of 76.00%, indicating strong performance in predicting student grades across subjects. While the **Random Forest** model had a slightly higher accuracy at 78.24%, it also achieved a balance between precision and recall, with an F1-score of 75.80%, making it ideal for scenarios where overall performance across multiple subjects is crucial. The **SVM Classifier** and **Logistic Regression** performed slightly lower, with SVM offering 70.00% accuracy and Logistic Regression achieving 69.48%, demonstrating the Decision Tree and Random Forest's superior ability to handle the diverse features involved in grade prediction. These results emphasize the Decision Tree's robustness in predicting individual subject grades, while Random Forest provides a more balanced approach suitable for applications requiring overall accuracy.**Effect of Data Augmentation**

In this project, we did not apply any data augmentation techniques. The models were trained and tested on the original dataset without introducing any synthetic variations or noise. This approach ensured that the evaluation metrics accurately reflected the models' performance on real-world data, without any influence from artificially created samples. The focus was on assessing how well the models, particularly the Decision Tree Classifier, predicted student grades based on actual features such as study hours, absence days, and personal characteristics**Error Analysis** Analyzing the prediction errors revealed that most inaccuracies were concentrated around specific feature ranges, particularly in instances with extreme values of humidity and temperature. These errors suggest that the models occasionally struggled with atypical weather patterns. Incorporating additional contextual features, such as wind speed or atmospheric pressure, could potentially enhance the models' predictive accuracy in these outlier scenarios.

**Implications and Insights**

The comparative analysis highlights the importance of choosing the right model based on the specific requirements of the application. The **Decision Tree Classifier**’s strong performance in predicting student grades makes it ideal for applications where understanding individual subject performance is critical, such as in academic counseling or personalized learning plans. On the other hand, **Random Forest**’s balanced performance, with a higher F1-score, is particularly beneficial in scenarios where an overall assessment of multiple subjects is essential, such as in comprehensive academic evaluations. The absence of data augmentation in this project establishes a baseline, and future experiments could explore incorporating such techniques to further improve model accuracy and robustness, especially in handling diverse student profiles and feature variations.

**CHAPTER 5**

**CONCLUSION & FUTURE ENHANCEMENTS**

#### The student grade prediction project aimed to apply machine learning techniques to forecast academic performance across various subjects based on student-related features. After performing essential data preprocessing steps, including handling missing values and encoding categorical variables, we trained several models, including **Decision Tree Classifier**, **Logistic Regression**, **Random Forest**, and **SVM Classifier**. Each model was evaluated using performance metrics such as accuracy, precision, recall, and F1-score. The **Decision Tree Classifier** achieved a strong accuracy of 75%, showcasing its effectiveness in predicting grades for individual subjects. However, **Random Forest** demonstrated a balanced trade-off between precision and recall, with the highest F1-score, making it ideal for comprehensive assessments of student performance. The **Logistic Regression** and **SVM** models showed slightly lower performance but still contributed valuable insights into the prediction process. This project illustrates that machine learning can be effectively applied to predict student grades, offering insights that can assist in academic counseling and personalized learning. The comparative analysis of different models helps identify which algorithms perform best under various evaluation criteria, providing a foundation for future work in educational data analysis.**Future Enhancements:**

To enhance the accuracy and applicability of the student grade prediction model, several improvements can be made. Expanding the dataset to include a wider range of students and additional features, such as extracurricular activities and mental well-being, would improve generalization. Implementing real-time data pipelines for continuous updates and predictions could further enhance the model. Exploring advanced models like **Random Forest** or **Gradient Boosting** could capture complex relationships and improve accuracy. Deploying the model through a web or mobile app would make it accessible to educators, and integrating interactive dashboards would allow for better tracking and insights on student performance. These enhancements would increase the model's practical utility in academic settings.

**REFERENCES**

 **F. Pedregosa et al.**, "Scikit-learn: Machine Learning in Python," *Journal of Machine Learning Research*, vol. 12, pp. 2825–2830, 2011.

 **A. Géron**, *Hands-On Machine Learning with Scikit-Learn, Keras, and TensorFlow*, 2nd ed. Sebastopol, CA, USA: O'Reilly Media, 2019.

 **C. Zhang and Y. Ma**, *Ensemble Machine Learning: Methods and Applications*, Boston, MA, USA: Springer, 2012.

 **M. Kuhn and K. Johnson**, *Applied Predictive Modeling*, New York, NY, USA: Springer, 2013.

 **G. James, D. Witten, T. Hastie, and R. Tibshirani**, *An Introduction to Statistical Learning: With Applications in R*, New York, NY, USA: Springer, 2013.

 **J. Bergstra and Y. Bengio**, "Random Search for Hyper-Parameter Optimization," *Journal of Machine Learning Research*, vol. 13, pp. 281–305, Feb. 2012.

 **T. Chen and C. Guestrin**, "XGBoost: A Scalable Tree Boosting System," in *Proc. 22nd ACM SIGKDD Int. Conf. on Knowledge Discovery and Data Mining*, San Francisco, CA, USA, 2016, pp. 785–794.

 **R. R. Raut, S. N. Raut, and P. S. Shelar**, "Rainfall Prediction using Machine Learning," *International Research Journal of Engineering and Technology (IRJET)*, vol. 7, no. 6, pp. 4636–4639, Jun. 2020.

 **A. Singh and S. M. Ali**, "Rainfall Prediction using Decision Tree and Random Forest Machine Learning Models," in *Proc. 5th Int. Conf. on Inventive Computation Technologies (ICICT)*, Coimbatore, India, 2020, pp. 267–272.

 **D. S. Kiran, B. Prasanthi, and M. N. Kumar**, "Machine Learning Techniques for Rainfall Prediction: A Review," *International Journal of Scientific & Engineering Research*, vol. 10, no. 5, pp. 234–239, May 2019.