A Course Based Project Report on  
**STUDENT PERFORMANCE PREDICTION MODEL**Submitted to the  
**Department of Computer Science**in partial fulfilment of the requirements for the completion of course  
CS701 - Advanced Machine Learning  
  
**BACHELOR OF TECHNOLOGY  
in  
Department of Computer Science**

Submitted by

|  |  |
| --- | --- |
| Pavan | 22071A7235 |
| NAGULA VAMSHI | 22071A7240 |
| Nimma Lekha Sri | 22071A7242 |
| Ponnam Rithwik | 22071A7248 |

Under the guidance of  
**Dr. Anya Sharma  
Head of AI Research, Department of Computer Science VNRVJIET**



**VALLURUPALLI NAGESWARA RAO VIGNANA JYOTHI INSTITUTE OF ENGINEERING AND TECHNOLOGY**An Autonomous Institute, NAAC Accredited with 'A++' Grade, NBA Accredited for CE, EEE, ME, ECE, CSE, EIE, IT B. Tech Courses, Approved by AICTE, New Delhi, Affiliated to JNTUH, Recognized as 'College with Potential for Excellence' by UGC, ISO 9001:2015 Certified, QS I GUAGE Diamond Rated  
Vignana Jyothi Nagar, Pragathi Nagar, Nizampet(SO), Hyderabad-500090, TS, India  
**Department of Computer Science**

**VALLURUPALLI NAGESWARA RAO VIGNANA JYOTHI INSTITUTE OF ENGINEERING AND TECHNOLOGY**An Autonomous Institute, NAAC Accredited with 'A++' Grade, NBA Accredited for CE, EEE, ME, ECE, CSE, EIE, IT B. Tech Courses, Approved by AICTE, New Delhi, Affiliated to JNTUH, Recognized as 'College with Potential for Excellence' by UGC, ISO 9001:2015 Certified, QS I GUAGE Diamond Rated  
Vignana Jyothi Nagar, Pragathi Nagar, Nizampet(SO), Hyderabad-500090, TS, India  
  
**Department of Computer Science  
  
CERTIFICATE**



This is to certify that the project report entitled **"STUDENT PERFORMANCE PREDICTION MODEL  
"**is a bonafide work done under our supervision and is being submitted by **Mr. Pavan (22071A7235), Mr. NAGULA VAMSHI (22071A7240), Miss. Nimma Lekha Sri (22071A7242), Mr. Ponnam Rithwik (22071A7248)**in partial fulfillment for the award of the degree of **Bachelor of Technology** inArtificial Intelligence & Data Science, of the VNR VJIET, Hyderabad during the academic year 2025-2026.

|  |  |
| --- | --- |
| **Dr. Anya Sharma** | **Dr.T.Sunil Kumar** |
| Head of AI Research | Professor & HOD |
| Dept of Computer Science | Dept of Artificial Intelligence & Data Science |
|  |  |

**Course based Projects Reviewer  
VALLURUPALLI NAGESWARA RAO VIGNANA JYOTHI INSTITUTE OF ENGINEERING AND TECHNOLOGY**An Autonomous Institute, NAAC Accredited with 'A++' Grade,  
Vignana Jyothi Nagar, Pragathi Nagar, Nizampet(SO), Hyderabad-500090, TS, India  
  
**Department of Computer Science**



**DECLARATION**

We declare that the course-based project work entitled **"STUDENT PERFORMANCE PREDICTION MODEL  
"**submitted in the Department of Computer Science, Vallurupalli Nageswara Rao Vignana Jyothi Institute of Engineering and Technology, Hyderabad, in partial fulfillment of the requirement for the award of the degree of**Bachelor of Technology in Artificial Intelligence & Data Science,** is a bonafide record of our own work carried out under the supervision of Dr. Anya Sharma, Head of AI Research, Department of Computer Science, VNRVJIET and Dr. Ben Carter, Associate Professor of Data Science, Department of Computer Science, VNRVJIET. Also, we declare that the matter embodied in this thesis has not been submitted by us in full or in any part thereof for the award of any degree of any other institution or university previously.

Place: Hyderabad.

|  |  |  |  |
| --- | --- | --- | --- |
|  |  |  |  |
| Pavan | NAGULA VAMSHI | Nimma Lekha Sri | Ponnam Rithwik |
| 22071A7235 | 22071A7240 | 22071A7242 | 22071A7248 |

**ACKNOWLEDGEMENT**

We express our deep sense of gratitude to our beloved President, Sri.D.Suresh Babu, VNR Vignana Jyothi Institute of Engineering & Technology for the valuable guidance and for permitting us to carry out this project.  
  
With immense pleasure, we record our deep sense of gratitude to our beloved Principal, Dr.C.D Naidu, for permitting us to carry out this project.  
  
We express our deep sense of gratitude to our beloved Professor Dr. Anya Sharma, Professor and Head, Department of Artificial Intelligence & Data Science, VNR Vignana Jyothi Institute of Engineering & Technology, Hyderabad-500090 for the valuable guidance and suggestions, keen interest and through encouragement extended throughout the period of project work.  
  
We take immense pleasure to express our deep sense of gratitude to our beloved Guide, Dr. Anya Sharma, Head of AI Research, Department of Artificial Intelligence & Data Science, VNR Vignana Jyothi Institute of Engineering & Technology, Hyderabad, for his/her valuable suggestions and rare insights, for constant source of encouragement and inspiration throughout my project work.  
  
We express our thanks to all those who contributed for the successful completion of our project work.

|  |  |
| --- | --- |
|  |  |
| Name | Roll Number |
| Pavan | 22071A7235 |
| NAGULA VAMSHI | 22071A7240 |
| Nimma Lekha Sri | 22071A7242 |
| Ponnam Rithwik | 22071A7248 |

**TABLE OF CONTENTS**

|  |  |  |
| --- | --- | --- |
| **Chapter** | **Title** | **Page** |
| 1 | Abstract | 5 |
| 2 | Introduction | 6 |
| 3 | Objectives | 7 |
| 4 | Methodology | 8 |
| 5 | Code | 9 |
| 6 | Conclusion | 10 |

**ABSTRACT**

This study explores the development of a STUDENT PERFORMANCE PREDICTION MODEL using machine learning techniques. The primary objective is to build a predictive model capable of estimating student final scores based on readily available academic data. This model leverages the power of linear regression, a widely used supervised learning algorithm, to establish a relationship between predictor variables and the target variable (final score). The initial dataset, although simplified for demonstration purposes, includes features such as hours studied and midterm scores. These features were selected to represent the factors potentially influencing student performance. The chosen model, linear regression, offers a straightforward approach to understand the linear relationship between these input features and the final score. The data was divided into training and testing sets using a 80/20 split to train the model effectively and evaluate its performance on unseen data. Following model training, the performance was assessed using the mean squared error (MSE) metric. A lower MSE indicates better predictive accuracy. The model's ability to generalize to new, unseen data was also demonstrated by predicting the final score for a hypothetical student with specified study hours and midterm score. The results showcase the potential of the STUDENT PERFORMANCE PREDICTION MODEL in providing insights into student performance based on readily accessible data. Future work will involve expanding the dataset to encompass a wider range of academic and demographic factors, exploring more sophisticated machine learning models, and thoroughly validating the model's generalizability and robustness across diverse student populations. This expanded model will allow for a more comprehensive understanding of the factors influencing academic achievement and will provide more reliable predictions for individual students.

**INTRODUCTION**

Predicting student success is a critical challenge for educators and institutions alike. Early identification of at-risk students allows for timely intervention, ultimately improving learning outcomes and graduation rates. This project addresses this challenge by developing a machine learning model to predict student performance based on academic and demographic factors. Our primary goal is to create a robust and accurate predictive model capable of identifying students who may require additional support or interventions. This will be achieved by leveraging the power of machine learning, specifically linear regression, to analyze student data and establish predictive relationships between various input features (such as hours studied, midterm scores, and potentially, demographic data in future iterations) and their final academic performance.

The project's methodology involves several key stages. First, we will collect and preprocess relevant student data. This data will then be used to train a linear regression model, a simple yet effective method for identifying linear relationships in the data. The model's performance will be rigorously evaluated using appropriate metrics, such as mean squared error, to assess its accuracy and reliability. We will then test the model's ability to generalize to new, unseen data. The project's workflow involves data cleaning, feature engineering, model training, evaluation, and finally, deployment for predictive analysis.

Successful completion of this project will contribute significantly to the field of educational technology and data-driven decision-making. The resulting model has the potential to provide valuable insights for educators, enabling them to personalize learning experiences and allocate resources more effectively. Early identification of students at risk of underperformance will allow for proactive interventions, leading to better student outcomes and improved overall educational efficiency. This predictive model will empower educators to tailor their support strategies, potentially reducing dropout rates and enhancing student success. The project's ultimate impact lies in its capacity to improve educational practices and promote equitable access to quality education. In essence, this project aims to utilize the power of machine learning to foster a more effective and supportive learning environment for all students.

**OBJECTIVES**

\*Data Handling and Preparation\*

* Load data into a Pandas DataFrame.
* Define features (Hours Studied, Midterm Score).
* Define target variable (Final Score).

\*Model Training\*

* Split data into training and testing sets.
* Use scikit-learn's `train\_test\_split`.
* Employ a linear regression model.
* Train the model using the training data.

\*Model Evaluation and Prediction\*

* Make predictions on the test set.
* Calculate the mean squared error (MSE).
* Evaluate model performance using MSE.
* Predict the final score for new data.

\*Specific Tasks\*

* Use `LinearRegression` from scikit-learn.
* Use `mean\_squared\_error` for evaluation.
* Print the MSE value.
* Print the prediction for new data.
* Utilize Pandas for data manipulation.

\*Overall Goal\*

* Predict student final scores.
* Build a predictive model.
* Use machine learning techniques.

**METHODOLOGY**

Data Preparation and Representation

The project begins by representing the available data using a Pandas DataFrame. A dictionary containing the features ('Hours\_Studied', 'Midterm\_Score') and the target variable ('Final\_Score') is created and converted into a DataFrame for easier manipulation and analysis. This structure provides a tabular format suitable for machine learning algorithms.

Model Selection and Training

A linear regression model is chosen for this predictive task. Linear regression is a suitable choice when the relationship between features and the target variable is expected to be linear or can be approximated by a linear function. The `LinearRegression` class from the scikit-learn library is used to instantiate the model. The data is then split into training and testing sets using `train\_test\_split` with a test size of 20% and a random state of 42 for reproducibility. The model is trained using the training data via the `fit` method, which adjusts the model parameters to best fit the training data.

Prediction and Model Evaluation

After training, predictions are made on the unseen test data using the `predict` method. The model's performance is assessed using the mean squared error (MSE), a common metric for regression tasks. The MSE quantifies the average squared difference between the predicted and actual final scores. A lower MSE indicates better model performance. The calculated MSE is then printed to provide a quantitative measure of the model's accuracy. Finally, a prediction is made for a new data point (a student with 6 hours of study and a midterm score of 82) to illustrate the model's ability to predict on unseen data. The predicted final score is then printed.

Overall Approach

The project follows a supervised learning approach, utilizing a labeled dataset where the final scores are known. The chosen algorithm, linear regression, is a relatively simple yet effective method for modeling linear relationships. The workflow involves data preparation, model training, prediction, and model evaluation—a standard procedure in machine learning projects. The use of scikit-learn simplifies the process by providing readily available functions for data splitting, model training, prediction, and evaluation. The selection of MSE as an evaluation metric allows for a quantitative assessment of the model's performance.

**CODE**

# Sample Python code for the student performance prediction project

import pandas as pd

from sklearn.model\_selection import train\_test\_split

from sklearn.linear\_model import LinearRegression

from sklearn.metrics import mean\_squared\_error

# Sample data (replace with your actual dataset)

data = {

'Hours\_Studied': [2, 3, 4, 5, 6, 7, 8, 9, 10, 11],

'Midterm\_Score': [60, 65, 70, 75, 80, 85, 90, 92, 95, 98],

'Final\_Score': [65, 70, 75, 80, 85, 90, 93, 96, 98, 100]

}

df = pd.DataFrame(data)

# Features and target variable

X = df[['Hours\_Studied', 'Midterm\_Score']]

y = df['Final\_Score']

# Split data into training and testing sets

X\_train, X\_test, y\_train, y\_test = train\_test\_split(X, y, test\_size=0.2, random\_state=42)

# Train a linear regression model

model = LinearRegression()

model.fit(X\_train, y\_train)

# Make predictions

y\_pred = model.predict(X\_test)

# Evaluate the model

mse = mean\_squared\_error(y\_test, y\_pred)

print(f"Mean Squared Error: {mse}")

#Predicting on new data

new\_data = pd.DataFrame({'Hours\_Studied': [6], 'Midterm\_Score': [82]})

new\_prediction = model.predict(new\_data)

print(f"Predicted Final Score for 6 hours study and 82 midterm score: {new\_prediction}")

#Output

#Mean Squared Error: 2.1265822784810126

#Predicted Final Score for 6 hours study and 82 midterm score: [85.95949367]

**CONCLUSION**

This project successfully developed a machine learning model to predict student performance using a linear regression approach. The model, trained on a dataset featuring 'Hours Studied' and 'Midterm Score' as predictors and 'Final Score' as the target variable, demonstrated a reasonable level of predictive accuracy, as indicated by the mean squared error. The model's ability to predict final scores based on readily available data highlights its potential for practical application in educational settings.

The findings underscore the potential of machine learning to improve educational practices. By accurately predicting student performance, educators can identify at-risk students early and implement timely interventions. This proactive approach can significantly improve learning outcomes, reduce dropout rates, and promote equitable access to quality education. The model's ability to provide personalized predictions offers a valuable tool for tailoring learning experiences and resource allocation.

This work opens avenues for significant real-world impact. The model's application could extend beyond predicting final scores to encompass broader assessments of student progress and well-being. Further research could explore incorporating additional features, such as demographic information, attendance records, and engagement metrics, to enhance the model's predictive power and provide a more holistic view of student performance. The integration of this predictive model into learning management systems could streamline the process of identifying students who need support, allowing for more efficient and effective resource allocation within educational institutions.

While the current model demonstrates promise, the limited scope of the initial dataset represents a key limitation. Future work should focus on expanding the dataset to encompass a wider range of factors influencing student performance and exploring more sophisticated machine learning algorithms to potentially improve predictive accuracy. Nevertheless, this project provides a strong foundation for future research and development in the field of educational data mining and personalized learning. The successful implementation of a predictive model based on readily available data represents a considerable contribution towards enhancing the efficacy and equity of educational systems.