

Student Maths Score Prediction

Internship report submitted in fulfilment of the requirements for the Degree of

B.Tech

In

Information Technology

By

Vanshika Jain

Enrollment No.

20104026



Department of Computer Science Engineering and Information Technology

JAYPEE INSTITUTE OF INFORMATION TECHNOLOGY

(Declared Deemed to be University U/S 3 of UGC Act)

A-10, SECTOR-62, NOIDA, INDIA

TABLE OF CONTENTS

Chapter No.	Topics	Page No.
Chapter - 1	Project objective	4
Chapter - 2	Project description	4
Chapter - 3	Contribution	5
Chapter - 4	Skills/ Technologies Used	5
Chapter - 5	Conclusion	6
	References	6

CERTIFICATE

Chapter – 1: PROJECT OBJECTIVE

The objective of this project is to develop a predictive model that can accurately estimate students' math scores based on a set of demographic and educational factors. The dataset includes several categorical variables such as 'gender,' 'race_ethnicity,' 'parental level of education,' 'lunch,' and 'test preparation course,' which provide information about students' background and preparation. The project aims to utilize this information to create a machine learning model capable of predicting math scores.

Chapter – 2: PROJECT DESCRIPTION

- Data Analysis and Preprocessing: The first goal of the project is to thoroughly analyze and preprocess the dataset. This involves handling any missing values, encoding categorical variables, and normalizing or scaling numerical features.
- Feature Selection and Engineering: The categorical variables provided offer valuable insights into students' demographics and educational context. Feature engineering involves selecting the most relevant features and potentially creating new features that could enhance the predictive power of the model.
- Model Development: The primary objective of the project is to build a predictive model that accurately predicts students' math scores. This could involve using regression algorithms, as the target variable (math score) is a continuous variable.
- Model Evaluation: The developed model needs to be evaluated thoroughly to assess its performance. Various evaluation metrics such as mean squared error, mean absolute error, and R-squared could be used to measure the accuracy and generalization ability of the model.
- Interpreting Feature Importance: Since the model uses demographic and educational factors as input, it's important to interpret the impact of different features on the predicted math scores. This could provide insights into which factors play a significant role in determining math scores.
- Deployment: If the model performs well, it can be deployed to provide insights into potential outcomes based on students' backgrounds.

Chapter – 3: CONTRIBUTION

I took on the role of driving the entire project, assuming responsibility for every stage of its development. From inception to completion, my contributions encompassed all aspects of the project's execution, including implementation, and delivery.

Project Vision and Planning: I took the project's purpose and objectives, outline of the roadmap for its execution from the YouTube. This involved defining the problem statement, setting clear goals, and determining the scope of the project.

Architecture and Design: I crafted the project's architecture and design, selecting appropriate technologies, frameworks, and methodologies to ensure optimal implementation. This encompassed designing the data processing pipelines, model development approach, and deployment strategy.

Development and Implementation: My hands-on involvement extended to coding and implementation. I wrote the codebase for data preprocessing, model training, testing scripts, and any other necessary components.

Chapter – 4: SKILLS/TECHNOLOGIES USED

Skills:

- EDA (Exploratory Data Analysis)
- ML (Machine Learning)
- MLOPS (Machine Learning Operations)

Technologies:

- Python
- Python Libraries (Scikit-Learn, Pandas, Numpy, Matplotlib, Seaborn, Catboost, Xgboost, Dill)
- Flask
- GitHub
- VScode

Chapter – 5: CONCLUSION

The project's overarching goal of predicting students' math scores based on demographic and educational factors has been met through a comprehensive approach. The dataset's exploration, preprocessing, feature engineering, and model development stages have collectively contributed to the creation of a robust predictive model. By leveraging machine learning algorithms and best practices, this model demonstrates the potential to provide valuable

As a testament to the significance of MLOps principles, the project also showcased the integration of automation, version control, and deployment strategies.

In conclusion, this project not only achieved its immediate objective of predicting math scores but also serves as a valuable case study for the application of machine learning techniques in real-world scenarios. The insights gained, lessons learned, and skills honed throughout this endeavour lay a foundation for future endeavours that seek to harness the power of data and technology to drive informed decisions and meaningful outcomes. The success of this project is a testament to the dedication, expertise, and collaborative spirit of the team, reaffirming the potential of data science to contribute positively to the academic landscape and beyond.

GitHub Link: <https://github.com/Vanshikajain2721/MLOPS/tree/main>

Deployed Link: <https://maths-score-prediction.onrender.com>

REFERENCES

- <https://youtu.be/f8AtpzmRzFc>
- <https://youtu.be/MrurgA-IkjA>
- <https://www.databricks.com/glossary/mlops#:~:text=MLOps%20stands%20for%20Machine%20Learning,then%20maintaining%20and%20monitoring%20them.>
- <https://www.geeksforgeeks.org/what-is-exploratory-data-analysis/>