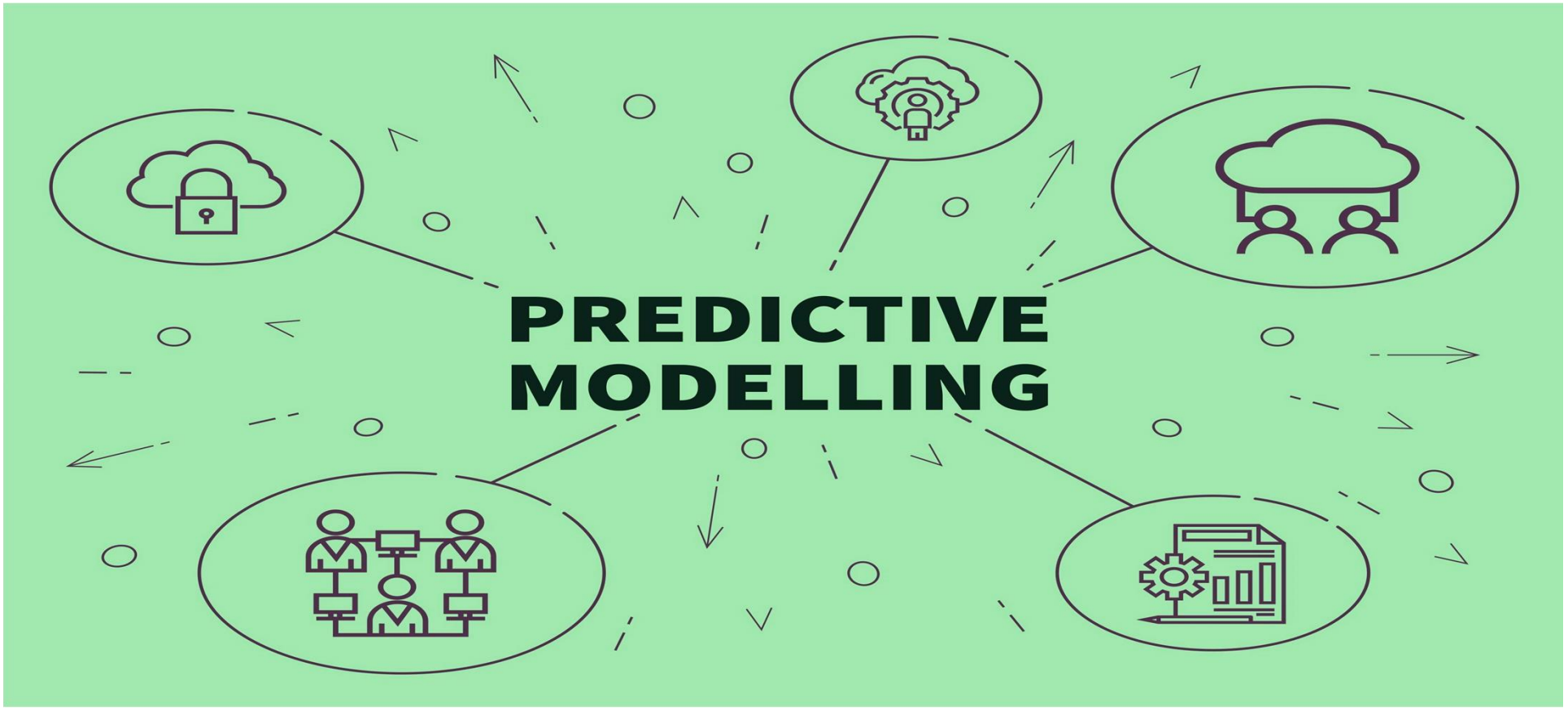
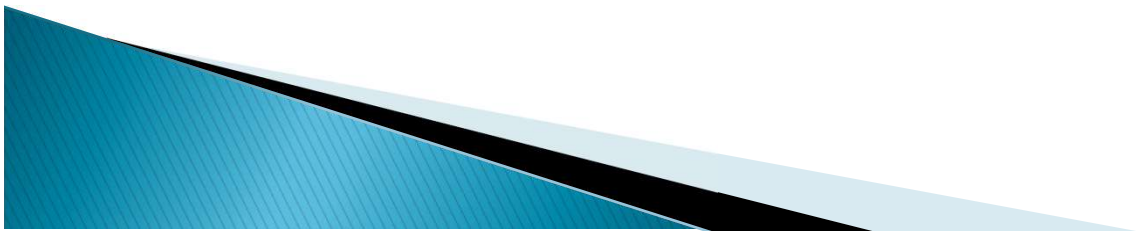


Md Firoz Alam
AWS Machine learning Intern

Table Of contents

- Introduction
- Dataset Details
- Work flow
- Feature Selection
- Implementation
- Conclusion
- Q&A Session
- References

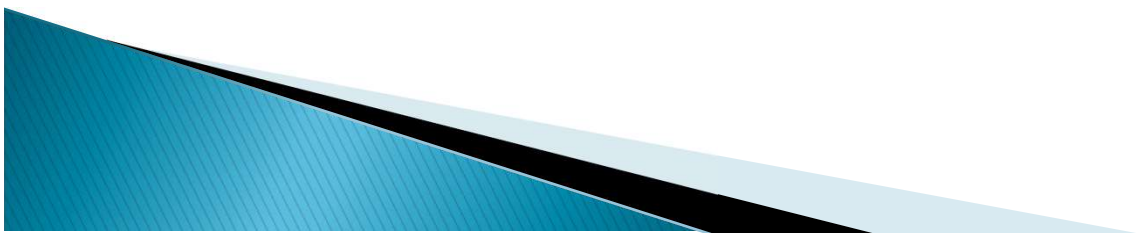


STORY HOOKS

- Imagine you are a teacher,tasked with predicting your student's performance in Mathematics.
- School Board
- Data Science Team
- Factors like Gender,race,parents education, lunch type ,test preparation

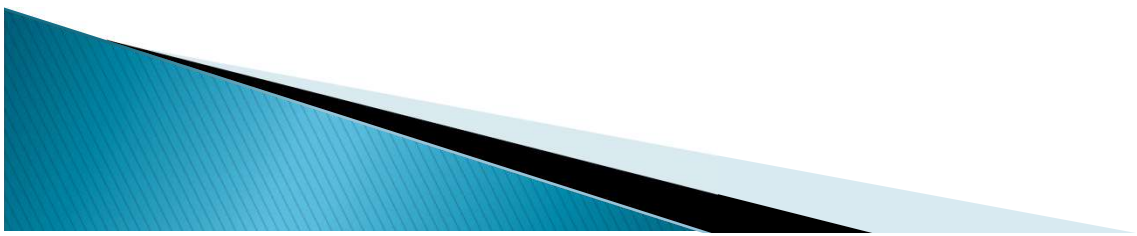
Introduction

- Predictive modeling is a form of data mining that analyzes historical data with the goal of identifying trends or patterns and then using those insights to predict future outcomes.
- It is the powerful technique used to make predictions about Future outcome based on historical data
- predicting students math score using another features.

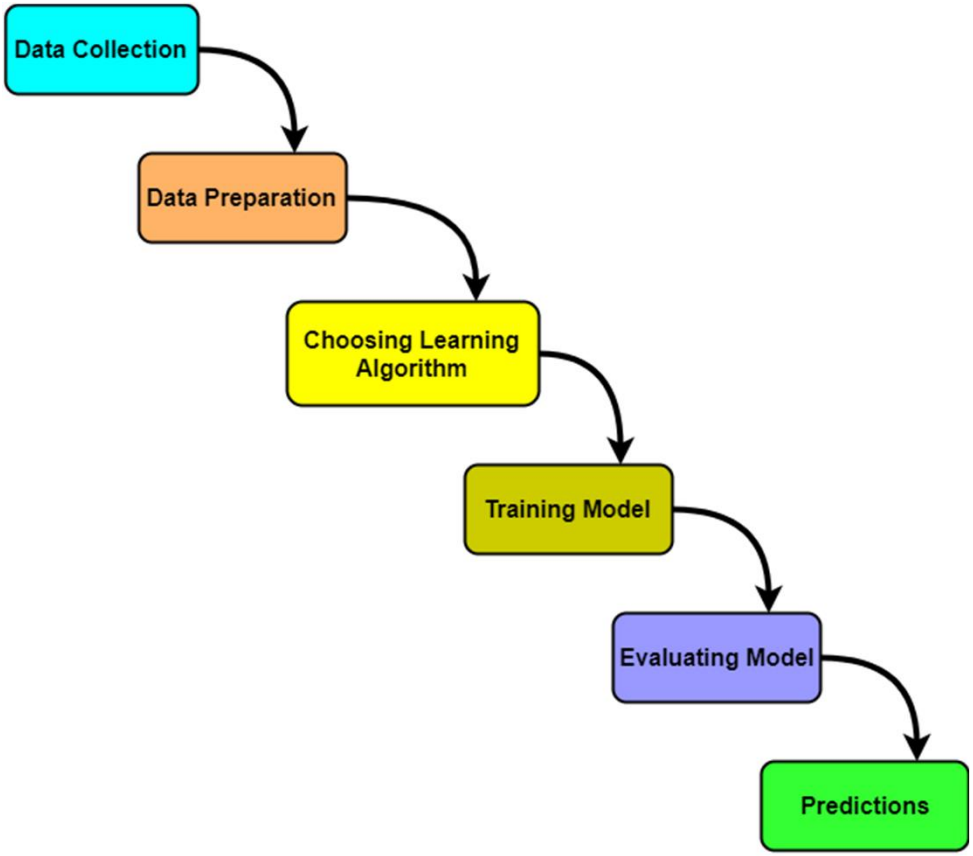


Dataset details

- Students performance dataset have 1000 rows and 8 columns.
- dataset include 8 features.
- Gender
- race/ethnicity
- parental level of education
- Lunch
- test preparation course
- math score
- reading score
- writing score



Work flow:-



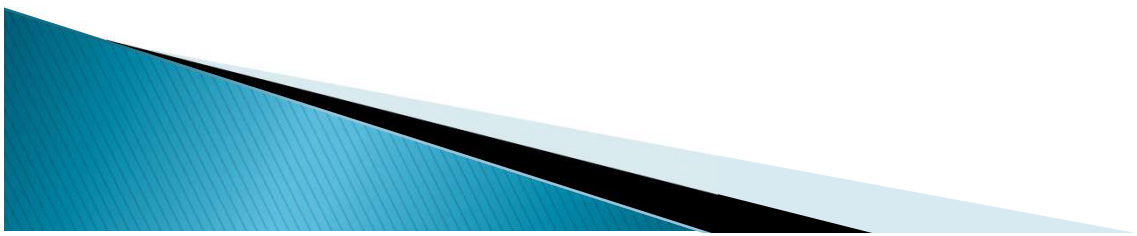
Machine Learning Workflow

Feature selection

- Feature selection the method of reducing the input variable to your model by using only relevant data.
- Here I have used chi-square test and correlation matrix for finding feature.
- In Chi-square test it calculate chi-square score between feature and target variables.
- Features have high chi-square score are more relevant in finding target variables.
- Correlation matrix is table that display correlation between feature and target variable.

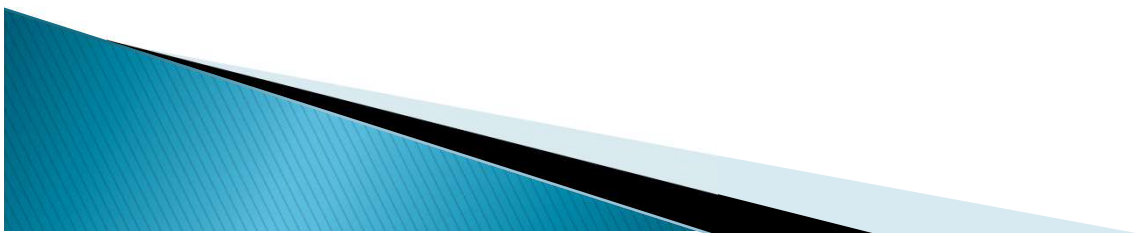
Implementation

- Live Demonstration of Predictive modeling
- After Demonstration we analyze our model
- Our model gives the 94.91% accuracy



Conclusion

- Predictive modeling is a powerful tool that can provide significant insights and benefits across various domains. By understanding its applications, techniques, benefits, and challenges, organizations can leverage predictive modeling to enhance their decision-making processes and gain a competitive advantage.



References

- Breiman, L. (2001). Random forests. Machine Learning, 45(1), 5-32.Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., ... & Duchesnay, E. (2011). Scikit-learn: Machine learning in Python. Journal of machine learning research, 12(Oct), 2825-2830.scikit-learn. (2023). train_test_split function. Retrieved from https://scikit-learn.org/stable/modules/generated/sklearn.model_selection.train_test_split.html.
- YouTube
- HackVeda Reference Videos


Q&A

firozalam1160006@gmail.com