ACKNOWLEDGEMENT

My sincere thanks to our internal project guide **Dr. Amit Thakkar, Prof. Purvi Prajapati, Prof. Chadni Shah, Prof. Jay Patel** for giving us valuable inputs and ideas right from the selection of topic for project till its successful completion. Without his constant encouragement we could not have been able to achieve what we have.

I thank **Dr. Parth Shah(HOD IT department)** for his ongoing support and encouragement in every aspect. Last but not the least, entire staff of Department of IT Engineering for guiding me thoughts and vision. The successful completion of my project would not have been possible without the dedicated support from all our mentors, family and friends.

It's my good fortune that we had support and well wishes of many. We are thankful to all and those names which have been forgotten to acknowledge here but contributions have not gone unnoticed.

With Sincere Regards,
Priya Nasit(16IT057)
Dhvani Raval(16IT106)

ABSTRACT

Student performance prediction is an area of concern for educational institutions.Machine Learning plays an important role in the business world and it helps to the educational institution to predict and make decisions related to the students' academic status. The existing system is a system which maintains the student information in the form of numerical values and it just stores and retrieve the information what it contains. So the system has no intelligence to analyze the data. The prediction with high accuracy in student's performance in beneficial as it helps in identifying the student with low academic achievements at the early stage of academics.

In this work, classifier techniques based on six representative learning algorithms, namely Naive Bayes, Logistic Regression, k-Nearest Neighbour, Support Vector Machine, Decision Tree classification, Random Forest classification and ANN. These six learning methods have been compared separately with respect to the training and test sets. Random Forest Classification is found to be the best classifier for predicting the student's result based on the marks obtained in the semester, Permanent City, Gender, Actual Cast Category,12^{th} Percentage, Board. We would also discuss how these machine learning models can help to improve an education system by considering the different factors in terms of accuracy, Specificity, Precision, Prevalence, Recall, F-Measure in results.

Keywords and terms: student performance, machine learning, naïve Bayes classification, Logistic Regression, k-Nearest Neighbour, Support Vector Machine, Decision Tree classification, Random Forest classification and ANN(Artificial Neural Network).

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES