ACKNOWLEDGEMENT

We would like to place on record our deep sense of gratitude to **Dr. Hariom Gupta**, Director, Jaypee Institute of Information Technology, India for his generous guidance, help and useful suggestions.

We express our sincere gratitude to **Mrs. Akanksha Bhardwaj**, Department of CSE, JIIT, for her invigorating guidance, continual encouragement and supervision throughout the course of present work.

We also wish to extend our thanks to our batch mates for their insightful comments and constructive suggestions to improve the quality of this project work.

Signature(s) of Students

RupanshVij (9915103098)

Arpit Gupta (9915103113)

Ashwyn Singh (9915103114)

Disha Nagpal (9915103117)

ABSTRACT

Educational Data Mining refers to techniques, tools, and research designed for automatically extracting meaning from large repositories of data generated by or related to people' learning activities in educational settings. This project highlights the importance of using student data to drive improvement in education planning. The need for prediction of a student's performance is to enable the university to intervene and provide assistance to low achievers as early as possible Working on the same principle we are analysing the result (grades) of students from a particular university by first collecting the data via Google form. Pre-processing the collected data, selecting the attributes and then the dimensionality of data set is reduced by a technique known as PCA. Moreover Decision Tree is applied on test set to predict their grades at the end of the first year.

Keywords—Data Mining; Education; Students; Performance

TABLE OF CONTENTS

ACKNOWLEDGEMENT i

ABSTRACT ii

LIST OF FIGURES v

LIST OF TABLES v

ABBREVIATIONS v

1. INTRODUCTION **1**

1.1 Educational Data Mining 1

1.2 Purpose 1

1.3 Approach 1

2. BACKGROUND STUDY **2**

2.1 Dimensionality Reduction 2

2.2 Decision Tree 3

2.3 Literature Survey 4

3. REQUIREMENT ANALYSIS **6**

3.1 Software requirements 6

3.2 Hardware requirements 6

3.3 Functional Requirements 6

3.4 Non Functional Requirements 6

3.5 User Requirements 6

3.6 UML Diagrams 7

3.6.1 Use Case 7

4. DETAILED DESIGN **8**

4.1 Process Model 8

5. IMPLEMENTATION **9**

6. TESTING REPORT **16**

7. CONCLUSION AND FUTURE SCOPE **17**

8. REFERENCES **18**

LIST OF FIGURES

Figure	Title	Page No.
1.	Transformation of a high dimensional data to low dimensional data using PCA.	2
2.	A decision tree layout	3
3.	Classification accuracy after pruning a test domain.	4
4.	UML Diagram	7
5.	Flow Chart	8
6.	Complete decision tree after applying PCA.	13
7.	Complete decision tree before applying PCA.	13
8.	Accuracy of test set applied on train set after PCA.	14
9.	Accuracy of test set applied on train set before PCA.	14

LIST OF TABLES

1. Google form of Surveyed Dataset…….9
2. Table describing result after each set…… 15
3. Test Table…….. 16
4. Testing Report……... 16

ABBREVATIONS

1. DM: Data Mining
2. RDM: Relational Data Mining
3. KDD: Knowledge Discovery in Database
4. EDM: Educational Data Mining
5. PCA: Principle Component Analysis