Student's Performance Prediction Using Data Mining

BY

SANJIDA AHMED
ID: 151-15-5095
AND

ISRAT JAHAN FATEHA
ID: 151-15-5036

This Report Presented in Partial Fulfillment of the Requirements for the Degree of Bachelor of Science in Computer Science and Engineering

Supervised By

Dr. Syed Akhter Hossain

Professor and Head
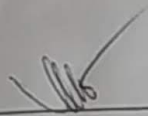
Department of CSE

Daffodil International University

DAFFODIL INTERNATIONAL UNIVERSITY

December, 2018

APPROVAL

This Project/internship titled **"Student's Performance Prediction Using Data Mining"**, submitted by SANJIDA AHMED, ID No: 151-15-5095 and ISRAT JAHAN FATEHA, ID No: 151-15-5036 to the Department of Computer Science and Engineering, Daffodil International University has been accepted as satisfactory for the partial fulfillment of the requirements for the degree of B.Sc. in Computer Science and Engineering and approved as to its style and contents. The presentation has been held on 11th December 2018.
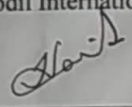
BOARD OF EXAMINERS

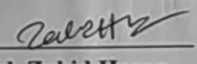**Dr. Syed Akhter Hossain**   Chairman
Professor and Head
Department of Computer Science and Engineering
Faculty of Science & Information Technology
Daffodil International University

Dr. Sheak Rashed Haider Noori   Internal Examiner
Associate Professor & Associate Head
Department of Computer Science and Engineering
Faculty of Science & Information Technology
Daffodil International University

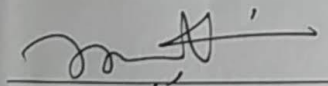**Md. Zahid Hasan**   Internal Examiner
Assistant Professor
Department of Computer Science and Engineering
Faculty of Science & Information Technology
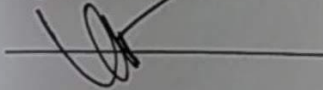Daffodil International University

Dr. Mohammad Shorif Uddin   External Examiner
Professor
Department of Computer Science and Engineering
Jahangirnagar University

Declaration

We hereby declare that, this project has been done by us under the supervision of **Syed Akhter Hossain, Professor and Head, Department of CSE** Daffodil International University. We also declare that neither this project nor any part of this project has been submitted elsewhere for award of any degree or diploma.
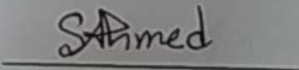
Supervised by:

Dr. Syed Akhter Hossain

Professor and Head

Department of Computer Science and Engineering

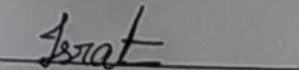Daffodil International University

Submitted by:

Sanjida Ahmed
ID: 151-15-5095
Department of CSE
Daffodil International University

Israt Jahan Fateha
ID: 151-15-5036
Department of CSE
Daffodil International University

Acknowledgement

First we express our heartiest thanks and gratefulness to almighty Allah for His divine blessing makes it possible to complete the final year thesis successfully.

We are really grateful and wish our profound indebtedness to **Syed Akhter Hossain**, **Professor and Head**, **Department of CSE** Daffodil International University, Dhaka. Deep Knowledge & keen interest of our supervisor in the field of *Data Mining* helped us a lot to carry out this research. His endless patience, scholarly guidance, continual encouragement, constant and energetic supervision, constructive criticism, valuable advice, reading many inferior drafts and correcting them at all stages have made it possible to complete this project.

We would like to express our heartiest gratitude to Almighty Allah and Head**,** Department of CSE, for his kind help to finish our project and also to other faculty members and the staffs of CSE department of Daffodil International University.

We would like to thank our entire course mates in Daffodil International University, who took part in this discussion while completing the course work.

Finally, we must acknowledge with due respect the constant support and patients of our parents.

Abstract

Amelioration of the education system is one of the most intensive issue in Bangladesh. Likely most of the developing countries, in Bangladesh this actually plays an important role. Higher education in Bangladesh, faced a phenomenal growth in quantity of third ranked student from 1.12 million to 2.61 million within few years. But in terms of quality of education could not improve corresponding with this growth. In the necessity of improvement of education system and student s quality we were trying to propose a system to predict performance of a student, so that the necessary steps can be taken by the authority and the lacking's can be filled up. In this is era of Data, datasets are producing from everywhere as in every institution. We also get a huge dataset from educational institutions and thus have so much of information. Using **Data Mining** techniques, we extracted some valuable information from the educational dataset and tried to build up a suitable model in the purpose of predicting student performance.

Table of Contents

Contents	Page

Chapter

List of Figures

Chapter 1
Introduction

1.1 Introduction

Establishment of quality education in now-a-days is a global concern. The demand for quality education is increasingly growing at national and global levels. The possible reason for this is the increasing of students in tertiary education. And key to the success and sustainability of quality education is basically quality culture. Thus developing n quality culture requires incessant process and attention to the educational area which refers to ongoing process. With this motives we just worked with a portion of education system which is student's grade and we tried to build a model which will help to predict the upcoming result of student grade by processing the data of previous performance data of related subjects to the predicted one.

1.2 Motivation

The world is in a need of quality education. Quality education can change the structure of a nation. As considering in Bangladesh, the rate of tertiary education is increasingly growing with is population. So, it has become a major topic in Bangladesh. In the necessity of quality education we approached to work with educational or institutional data. As data mining technique could be very useful to this area as it extracts hidden patterns and generate knowledge as the system or the model requires, we wanted to work with this technique.

1.3 Rationale of the Study

It is obvious that in research area, working with educational data is not something new to explore hidden patterns and generating knowledge. Throughout the dataset, an operation of a systematic inquiry to pursue hypotheses and predicting object point and throw some questions for future work is done. In necessity of the improvement of student quality we

tried to focus on some basic points which might help the education system in terms of improvement.

The research objectives are given below:

✓ To identify the weak students who need extra care.

✓ To identify stronger students who can also help the weak students.

✓ To know the student's capabilities and help to find suitable career guideline according to his best performance.

✓ To help a student for knowing his or her inner strength.

✓ To decrease the percent of dropout students.

1.4 Research Questions

1. Why different student acquire different grades while having the same course teacher
2. Why a student gets poor grades?
3. How to identify the students who are lagging behind?
4. How to have a self-assessment on individual grades?

1.5 Expected Output

In this project we wanted to find the pattern of students who are getting's bad grades in their exams. So, we have collected the grade sheet of the courses of a particular semester of a student. By using those data we wanted to find out the relation between interrelations of various courses on the basis of the grades achieved by the students. Throughout this research we tried get this done –

- Predicting student performance to a specific subject.
- Building a relation according to the performance with related subjects.
- Understand the lacking from the patterns.
- Specify the area to improve.

1.6 Report Layout

This report is structured in a way that all the issues that needs to be understand can be understand throughout the whole report.

Chapter 1 which contains the introductory part such as objective of this project, the reason that motivated us to work with that research, some logics to work with and our expected outcome from this project.

Chapter 2 is about the related works and attached the brief of background study to further carry on with this project.

Chapter 3 contains the research methodology. The procedure of data collection and statistical analysis.

Chapter 4 contains the details of the model we built and the way we generated our knowledge base and extracted the hidden patterns. And how we tested our dataset for a perfect model creation.

Chapter 5 has the summary of this project and deals with the actual result or outcome we gained so far. The future scopes of this project, recommendations and conclusion.

Chapter 2
Background

2.1 Introduction

As we wanted to work with educational data which is not a new field to work on. Data mining is a technique which can generate knowledge from dataset as the system requires and extract hidden patterns, many of research work tried to build some system using data mining technique in necessity of improvement of educational field.

2.2 Related Works

In field there are so many research works although data mining technique was not very much used to work on.

Pauziah Mohd Arsad, Norlida Buniyamin, Jamalul-Lail Ab Manan, Noraliza Hamzah proposed **academic students' performance prediction model: A Malaysian case study** in 2011 [1] describes the performance of Engineering Electrical Degree students at the Faculty of Electrical Engineering, University Teknologi MARA, and Malaysia. The study was based on longitudinal progress of CGPA for every semester until graduation and for three consecutive intakes of matriculation students namely July 2005, July 2006 and July 2007. Then, the methodology which was possible to work on that was applied to consecutive intakes of students that are Diploma holders for July 2006, July 2007 and July 2008. The outcomes of the research indicates that the there is a correlation between student's abilities in semester one with the final overall academic performance irrespective of gender.

Olugbenga Wilson Adejo & Thomas Connolly proposed **Predicting student academic performance using multi-model heterogeneous ensemble approach** [4] and the purpose of this paper is to investigate and compare the use of dataset, different methods and ensembles of classifiers technique in predicting student academic performance. This study will compare the performance and the efficiency of ensemble techniques that make use of different combination of data sources with that of base classifiers with single data source.

George Kondraske proposed **Performance theory: implications for performance measurement, task analysis, and performance prediction** [2] targeting improved quantification of human performance identified conceptual shortcomings, motivating development of General Systems Performance Theory and lastly the Elemental Resource Model for human performance.

Asraful Alam Pathan, Mehedi Hasan, Md. Ferdous Ahmed, and Dewan Md. Farid, Department of Computer Science and Engineering, United International University, Bangladesh proposed **Educational Data Mining: A Mining Model for Developing Students' Programming Skills** [7].

Dewan Md. Farid, and Hasan Sarwar, Department of Computer Science and Engineering United International University proposed **Knowledge Mining for Effective Teaching and Enhancing Engineering Education** [10].

Pooja Thakar, Anil Mehta, Manisha proposed **Performance Analysis and Prediction in Educational Data Mining: A Research Travelogue** [16] in 2015 presents a comprehensive survey, a travelogue (2002- 2014) towards educational data mining and its scope in future.With the help of their works we could actually carry on with this project.

2.3 Research Summary

We are generating a lot of data in educational databases, which is not properly being utilized. To get benefited from such a huge set of educational data, we can use some data mining techniques which is powerful tool for investigation, generation of knowledge and prediction. This data mining techniques is used in so many field but there is not much related works in the field of education. Though there are some works is done in this field, but still there are many more areas to work with. Thus in this thesis we build a model that can help to evaluate and predict the result of a particular course with the help of the knowledge of previous similar course results which will also help to identify similar courses taken by student and their result evaluation. This model will help both the students and the teachers even also support the University for Quality Assurance in education as for quality assurance in education it is very much important to know about the strength and weakness of all students.

2.4 Scope of the Problem

Now-a-days one of the biggest challenge that higher education faces is prepare students skilled. Many universities and institutes are not in a shape to guide their students because of lack of information and assistance from their teaching-learning systems. To better administration and serve the students, the universities or institutions need better assessment, analysis, and prediction tools. Although considerable amount of work is done in analyzing and predicting academic performance, but all of these works are segregated. There is a clear need for unified approach. Unlike academic works, there are large numbers of primary areas that plays a significant role in prediction. Sustainable data mining techniques are required to compute, overlook and attention to factors for prediction. Finally the input vector with qualitative values can be increased the accuracy rate of prediction as well. Those models are required for all the stakeholders of an Institution. Therefore ensuring sustainable growth for all. If a system as such can be precise the field of artificial intelligence will introduce to another level in HCI.

2.5 Challenges

To discuss about the challenges we must include challenges of the suitable data collection problem and get a high accuracy from the generated model.

Chapter 3
Research Methodology

3.1 Introduction

Application of Data Mining concepts and techniques are really vast. Market basket analysis, Future Healthcare, Manufacturing Engineering, CRM(Customer Relationship Management), Real Estate, Fraud Detection, Financial Banking, Criminal Investigation, Bio Informatics, Web Mining, Education – all of these are some of the most usable applications of Data Mining. Educational Data Mining is the new arising technique of Data Mining that can be applied to the data related to the field of education. Many techniques are used in educational Data Mining such as – Decision Trees, Neural Networks, Naïve Bayes, K-nearest neighbors, Regression and many more.

These techniques lead us to many different kinds of knowledge that can be discovered like Classification, Association Rule, Regression and Clustering. The process of this discovering knowledge known as KDD can be used for predicting the grade of a particular course of an individual student.

So basically we are going to use here the KNN (K-Nearest Neighbors) algorithm for better forecasting. This algorithm is also robust to noisy training data and that is so much helpful for us. Though it's a very simple algorithm, this algorithm can work on huge dataset and also time efficient. It works based on minimum distance from the query instances to the training samples for determining the K-nearest neighbors. After gathering K-nearest neighbors, we take simple majority of these K-nearest neighbors to be the prediction of those instances.

3.2 Research Subject and Instrumentation

In this section, we discussed about the tools that we used for the implementation. First of all, the data was collected through DIU Student Portal and VUS of Academic Result. In this purpose, python Web scrapping method was used. The data was received and recorded as a CSV file (CSV=Comma Separated Values). After that, we processed the responses and prepared the dataset using Microsoft Excel 2010. Next, we implemented the data mining models on the dataset using Scikit-Learn library (a package of machine learning algorithms) of Python. We also used 'Pandas', 'Numpy', 'Seaborn' and 'Matplotlib' library of python during implementation phase. We assessed the models by calculating accuracy using sklern preprocessing. Finally we created different box-plot for visualizing our data and evaluating the models using Microsoft Excel 2010. Finally, all of the implementation and procedure was done in a 64 bit, Windows 10 machine.

3.3 Data Collection Procedure

3.3.1 Data Collection

For doing this thesis we need to have all course result data of at least 100 of the students. So we collected 150 student's ID from Google Classroom and another resources. Then we collected the result through the VUS (Academic Result) of DIU. We collected the result of 51 courses of around 150 students and that includes almost all of their semester. We recorded the data in a Microsoft Excel document. The data was collected using Python Web scrapping and also manually.

No data was generated. All of the data was real.

	ENG113	MAT111	CSE112	PHY113	PHY124	PHY123	CSE122	CSE123	ENG123	MAT121	...	CSE331	CSE333	CSE334	ACT301	CSE332	CS
0	3.25	2.50	3.00	3.50	3.50	3.25	2.75	2.75	3.75	3.00	...	3.75	3.75	3.75	3.75	3.50	3.7
1	2.75	2.25	4.00	2.25	3.50	2.50	3.50	3.75	3.25	3.50	...	3.25	3.75	3.75	3.25	3.50	3.7
2	3.50	4.00	4.00	4.00	3.75	3.75	4.00	4.00	3.50	4.00	...	4.00	4.00	3.50	3.50	3.75	3.7
3	3.75	3.00	3.75	3.25	3.75	3.00	2.75	4.00	3.75	3.25	...	3.75	4.00	3.00	3.00	4.00	3.5
4	3.25	3.25	3.50	3.50	4.00	3.50	3.00	3.25	3.75	2.25	...	4.00	3.75	3.00	3.75	3.75	4.0

5 rows × 52 columns

Fig 3.1: Sample Dataset

Here we also have an extra column that contained the overall CGPA achieved by the student. Columns contained the data of the courses and row contained single student data. It took almost two and a half weeks for gathering and preparing the data in a proper format.

3.3.2 Data Preparation

The toughest part of our thesis was data cleaning, preprocessing and preparation. Because there are a lot of students who don't have the result of the some courses. There was also some student whose result was showing blocked for some semester. Some of the student dropped their course or semester; some didn't pass and took that course again. So while cleaning process we could properly collect around the data of 92 students. Then we preprocessed the data. The student who failed was not showing a Float value instead of that it was showing NaN. We fixed the problem by filling the NULL value with grade 0.0. We ensured the data type was all in float.

```
Out[8]:  CSE225      float64
         CSE224      float64
         CSE213      float64
         CSE212      float64
         CSE132      float64
         CSE133      float64
         CSE123      float64
         CSE122      float64
         CSE423      float64
         dtype: object
```

Fig 3.2: Data type of the courses

3.4 Statistical Analysis

We need to know some information like Mean, Median, Mode, Standard Deviation etc. before going into deep. So we used the Python library for knowing those necessary things.

```
          ENG113       MAT111      CSE112      PHY113      PHY124      PHY123   \
count  85.000000   85.000000   85.000000   85.000000   85.000000   85.000000
mean    3.370588    3.050000    3.520588    3.276471    3.544118    3.138235
std     0.435773    0.715267    0.488273    0.612401    0.549860    0.762306
min     2.250000    0.000000    2.000000    2.000000    0.000000    0.000000
25%     3.000000    2.500000    3.250000    2.750000    3.250000    2.750000
50%     3.500000    3.250000    3.750000    3.250000    3.750000    3.250000
75%     3.750000    3.500000    4.000000    3.750000    4.000000    3.750000
max     4.000000    4.000000    4.000000    4.000000    4.000000    4.000000
```

Fig 3.3: Statistical Description

We also plotted histogram of each of the courses for viewing frequency measurements.

```
In [10]: df.CSE450.plot('hist')

Out[10]: <matplotlib.axes._subplots.AxesSubplot at 0x8a18390>
```
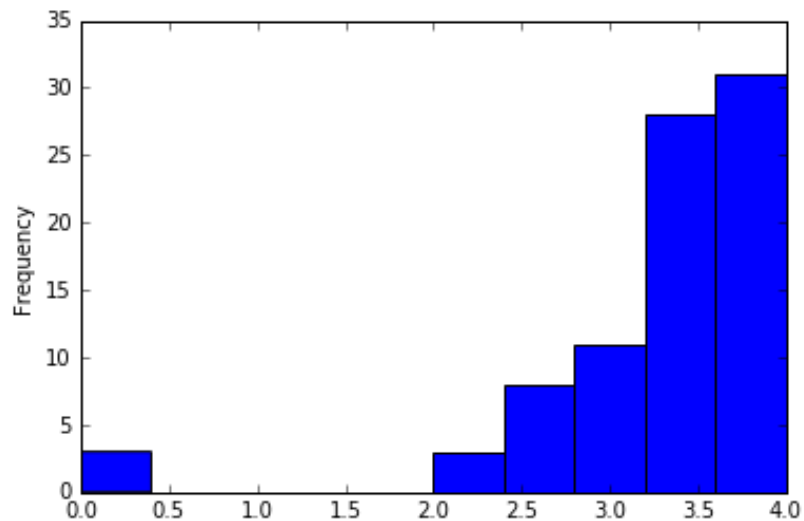
Fig 3.4: Frequency Histogram for CSE450

3.5 Implementation Requirements

For implementation, we needed to process all of the data in accurate form and shape with same data type. There were various kind of data type that was needed to be converted into one fixed data type.

```
Out[101]: 3          13              Out[13]: 3.75    21
          2.25       12                       3.50    16
          2.5        12                       3.25    12
          3.5        10                       3.00    11
          3.25        8                       4.00    10
          2           8                       2.75     5
          3.75        7                       2.25     3
          4           7                       0.00     3
          2.75        6                       2.50     3
          0           2                       NaN      1
          R block     1              Name: CSE450, dtype: int64
          Name: MAT121, dtype: int64
```

Fig 3.11: Before Fixing Data Type

```
Out[20]: 3.00    13              Out[21]: 3.75    20
         2.50    12                       3.50    16
         2.25    12                       3.00    12
         3.50    10                       3.25    12
         2.00     8                       4.00    10
         3.75     7                       2.75     5
         4.00     7                       2.25     3
         3.25     7                       0.00     3
         2.75     6                       2.50     3
         0.00     2              Name: CSE450, dtype: int64
         Name: MAT121, dtype: int64
```

Fig 3.12: After Fixing Data Type

We dropped and cleaned the data as needed. Cleaning is mandatory for applying algorithm.

There are lots of courses and also some groups of similar courses. At first, we selected a course and then we selected all of the courses which was similar to that course or whose knowledge was useful for getting a good grade on that specific course. Here we selected Embedded System and its course code is CSE423. Then we selected some course which

was anyhow related with CSE423 by our prior knowledge. Then we visualized all of that course result one by one with CSE423 and carefully noticed which of the grade of the courses was directly related to that course and finally selected only those courses result for our KNN training. For visualizing we created box-plot as it can help to visualize the median and frequency of the grades.
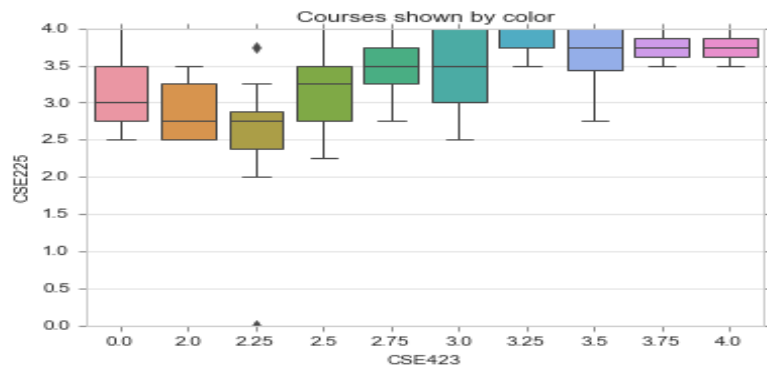

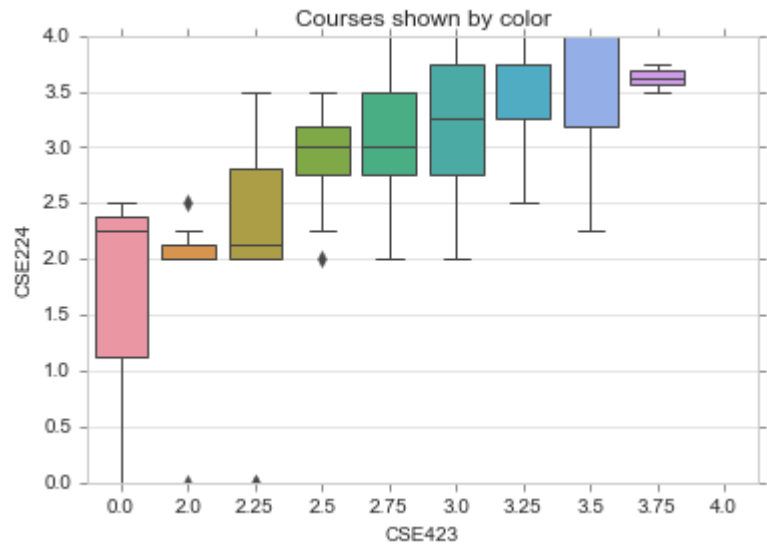Fig 3.3: Visualizing Relation between CSE423 and CSE225

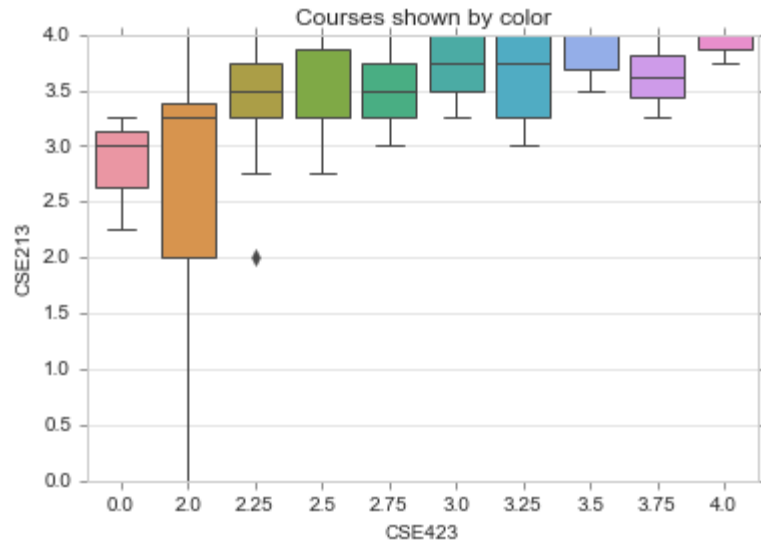Fig 3.4: Visualizing Relation between CSE423 and CSE224

Fig 3.5: Visualizing Relation between CSE423 and CSE213

Fig 3.6: Visualizing Relation between CSE423 and CSE221

Fig 3.7: Visualizing Relation between CSE423 and CSE132
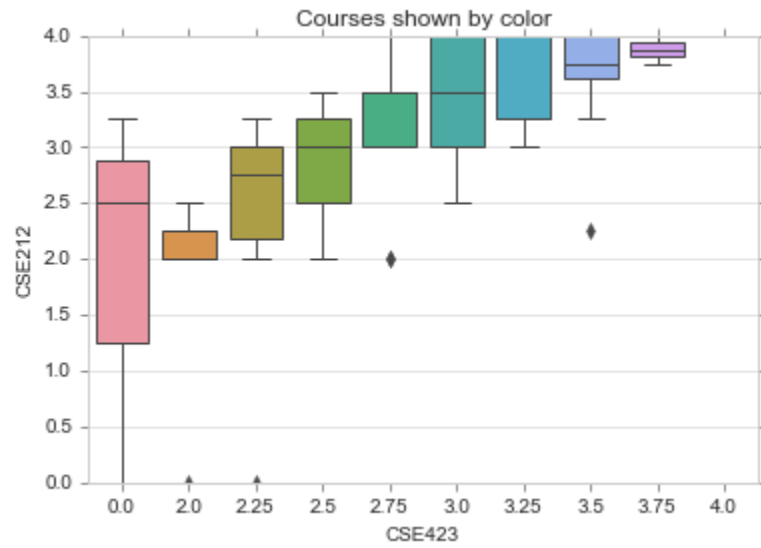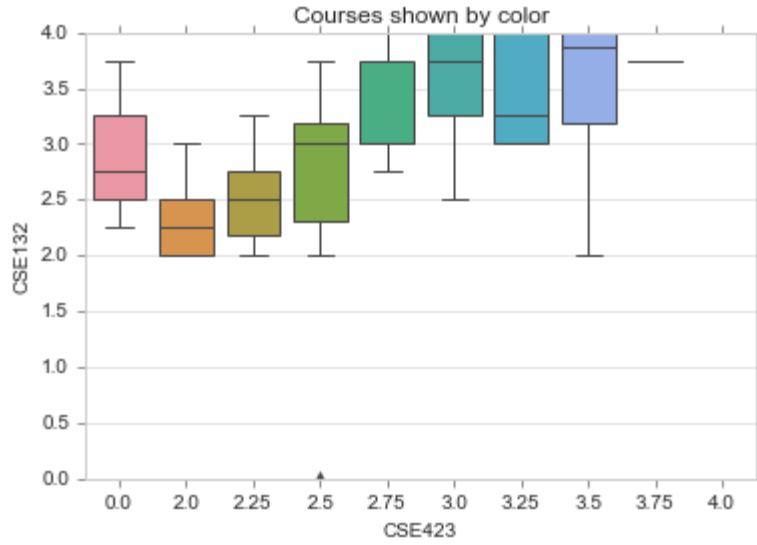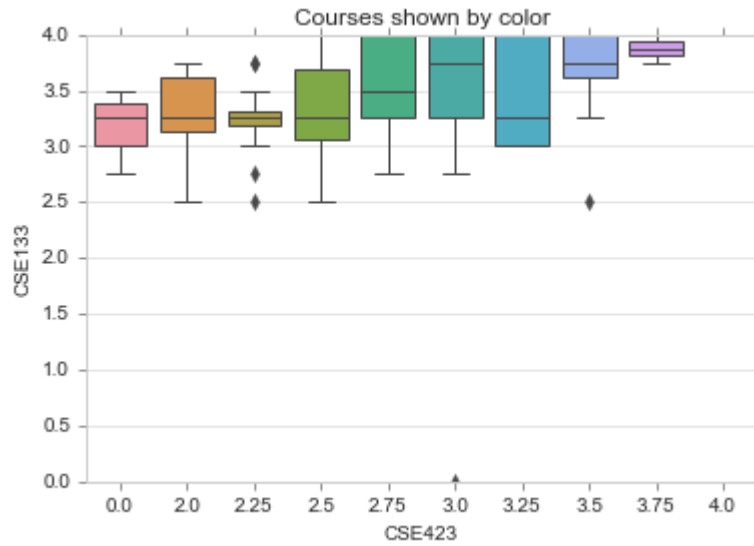

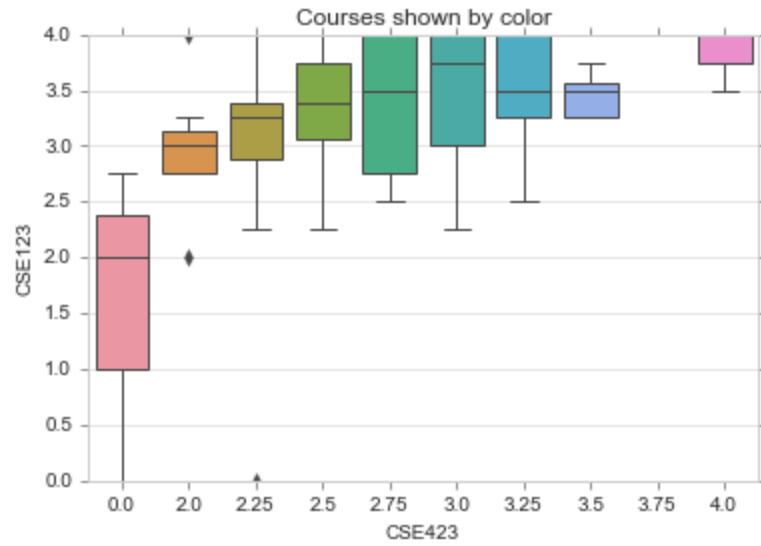Fig 3.8: Visualizing Relation between CSE423 and CSE133

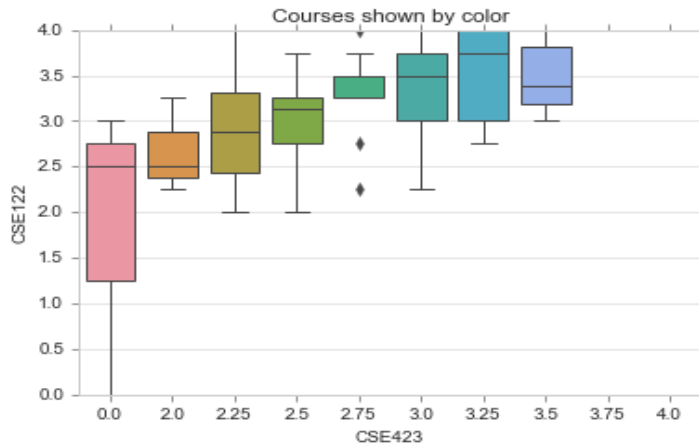Fig 3.9: Visualizing Relation between CSE423 and CSE123

Fig 3.10: Visualizing Relation between CSE423 and CSE122

We found these 8 courses (CSE225, CSE224, CSE213, CSE221, CSE132, CSE133, CSE123, CSE122) was directly correlated with CSE423 by visualizing the relationship. So we took all of these courses for training.

Chapter 4

Experimental Results and Discussion

4.1 Introduction

As far we know there are similarities in many courses that we complete through our 4 years BSc in CSE journey. There are also some similarities in the grades of a student in all of those similar courses. This is the main concept of this thesis. We used this concept for predicting a range of the grade of a specific course by the help of similar course results of that student.

The process used to collect information and data for understanding the purpose of correlation between different courses taken by students.

4.2 Experimental Results

We split the data set into test dataset and train dataset. The size of the train data set was 70 and the size of the test data set was 14. In both of that the number of the column was 9. We labeled the range for prediction as (0.0-0.0: 0), (2-2.25: 1), (2.5-2.75: 2), (3-3.25: 3), (3.5-4: 4). After that we scaled the desired data using Pandas library and scikit-learn preprocessing.

	CSE225	CSE224	CSE213	CSE212	CSE132	CSE133	CSE123	CSE122	CSE423
count	84.000	84.000	84.000	84.000	84.000	84.000	84.000	84.000	84.000
mean	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
std	1.006	1.006	1.006	1.006	1.006	1.006	1.006	1.006	1.006
min	-4.338	-3.142	-5.613	-3.488	-4.081	-5.884	-4.235	-4.740	-3.738
25%	-0.681	-0.500	-0.370	-0.632	-0.810	-0.302	-0.425	-0.651	-0.633
50%	0.150	0.110	0.034	0.224	-0.156	0.128	-0.019	0.093	0.058
75%	0.982	0.652	0.840	0.795	0.826	0.665	0.954	0.836	0.403
max	0.982	1.193	0.840	1.081	1.153	0.987	0.954	1.208	1.783

Fig 4.1: Scaled Information

Then we applied K-nearest neighbor algorithm and our model predicted 78.5% of those accurately.

```
          precision   recall  f1-score   support

   0.0       1.00       0.33      0.50       3
   1.0       0.71       0.45      0.56      11
   2.0       0.00       0.00      0.00       0
   3.0       0.00       0.00      0.00       0

avg / total   0.78       0.43      0.54      14
```

Fig 4.2: Classification Report

```
In [15]: new_df_test = pd.DataFrame(new_df_test_features, columns = num_cols)
         new_df_test['predicted'] = KNN_mod.predict(new_df_test_features)
         new_df_test['correct'] = [1 if x == z else 0 for x, z in zip(new_df_test['predicted'], new_df_test_labels)]
         accuracy = 100.0 * float(sum(new_df_test['correct'])) / float(new_df_test.shape[0])
         print(accuracy)

         78.5714285714
```

Fig 4.3: Accuracy of the predictive model

4.3 Summary

In this thesis we build a model that can help to evaluate and predict the result of a particular course with the help of the knowledge of previous similar course results. This model will also help to identify similar courses taken by student and their result evaluation. This model will help both the students and the teachers even also support the University for Quality Assurance in education as for quality assurance in education it is very much important to know about the strength and weakness of all students.

Chapter 5

Summary, Conclusion, Recommendation and Implication for Future Research

5.1 Summary of the Study

We are generating a lot of data in educational databases, which is not properly being utilized. To get benefited from such a huge set of educational data, we can use some data mining techniques which are powerful tool for investigation, generation of knowledge and prediction. This data mining techniques is used in so many field but there is not much related works in the field of education. Though there are some works is done in this field, but still there are many more areas to work with. Thus in this thesis we build a model that can help to evaluate and predict the result of a particular course with the help of the knowledge of previous similar course results which will also help to identify similar courses taken by student and their result evaluation. This model will help both the students and the teachers even also support the University for Quality Assurance in education as for quality assurance in education it is very much important to know about the strength and weakness of all students.

5.2 Conclusions

Applying data mining on educational data, reveals some significant areas in education field, where prediction with data mining has reaped benefits, such as finding set of weak students, determining student's satisfaction for a particular course, comprehensive student evaluation, predicting students' dropout course registration planning, predicting the enrollment headcount, evaluation of collaborative activities etc. The research mainly investigated and compared the performance accuracy and efficiency of student grade and ensemble of classifiers that make use of single and multiple data sources. The study has developed a significant model that can be used for predicting student performance that is high in accuracy and efficient in performance. And this research study advances the understanding of the application of ensemble techniques to predicting student

performance using learner data actually. This research presents the performance in specific courses of students in the Bachelor Degree program.

5.3 Recommendation

The major limitations and possible future works are described below -

- Concluding the strong ability of students at semester one or high CGPA students will finally pulled through with good performance upon graduation. Some Engineering fundamentals and foundation will help students to maintain their CGPA as shown by the CGPA follow through until graduation. Future work might focus on model of NN based on the performance of students spanning of three generations of graduates.

- Considerable amount of work is done in analyzing and predicting academic performance in specific and related courses, but all of these works are relative. There is a need for specific approach. Other than academic attributes, there are large numbers of factors that play significant role in prediction, which includes no cognitive factors Suitable data mining techniques are required to measure, overlook and infer the factors for prediction. Thus enriching the input may increase the accuracy of prediction and model as well.

The mentioned are the next steps of improving performance and after implementation of these the system will become a more suitable of open world applications.

5.4 Implication for Further Study

This research can be a source of future projects related educational areas. Through the implementations of future scopes we mentioned above can be more suitable and specific as we have done some quality work precisely.

References

1. Proposed academic students' performance prediction model: A Malaysian case study by Pauziah Mohd Arsad ; Norlida Buniyamin ; Jamalul-Lail Ab Manan Noraliza Hamzah 2011 International Conference on. IEEE, 2011.

2. *Performance theory: implications for performance measurement, task analysis, and performance prediction.* By George Kondraske 2000, International Forum on IEEE, 2000.

3. *Study of comprehensive evaluation method of undergraduates based on data mining* by Wu, X., Zhang, H., & Zhang, H. (2010, October). In Intelligent Computing and Integrated Systems, 2010 International Conference on. IEEE.

4. *Predicting student academic performance using multi-model heterogeneous ensemble approach* by Olugbenga Wilson Adejo and Thomas Connolly 2017 Emerald Publishing Limited.

5. *Prediction and Analysis for Students' Marks Based on Decision Tree Algorithm* by Liu, Zhiwu, and Xiuzhi Zhang. Intelligent Networks and Intelligent Systems, 2010 3rd International Conference on. IEEE, 2010.

6. *Significance of Classification Techniques in Prediction Of Learning Disabilities* by Balakrishnan, Julie M. David. International Journal Of Artificial Intelligence & Applications.

7. *Educational Data Mining: A Mining Model for Developing Students' Programming Skills* by Asraful Alam Pathan, Mehedi Hasan, Md. Ferdous Ahmed, and Dewan Md. Farid, Department of Computer Science and Engineering, United International University.

8. *Educational data mining: a review of the state of the art* by Romero, Cristóbal, and Sebastián Ventura Systems, Man, and Cybernetics, Part C: Applications and Reviews, IEEE Transactions on 40.

9. *The state of educational data mining in 2009: A review and future visions* by Baker, Ryan SJD, and Kalina Yacef. JEDM-Journal of Educational Data Mining.

10. *Knowledge Mining for Effective Teaching and Enhancing Engineering Education* by Dewan Md. Farid, and Hasan Sarwar, Department of Computer Science and Engineering United International University International Forum on IEEE, 2012.

11. *The Application of Data Mining to Build Classification Model for Predicting Graduate Employment* by Jantawan, Bangsuk, and Cheng-Fa Tsai. International Journal of CSIS (2013).

12. *Clustering Analysis for Empowering Skills in Graduate Employability Model* by Bakar, Noor Aieda Abu, Aida Mustapha, and Kamariah Md Nasir. Australian Journal of Basic and Applied Sciences (2013).

13. *Performance Analysis of Engineering Students for Recruitment Using Classification Data Mining Techniques* by Singh, Samrat, and Vikesh Kumar. International Journal of CSIT (2013).

14. *An exploratory study of factors affecting undergraduate employability* by Finch, David J, Leah K. Hamilton, Riley Baldwin, and Mark Zehner. Education (2013).

15. *Non-technical skill gaps in Australian business graduates* by Jackson, Denise, and Elaine Chapman. Education Training (2012).

16. *Performance Analysis and Prediction in Educational Data Mining: A Research Travelogue* by Pooja Thakar, Anil Mehta, Manisha proposed International Journal of Computer Applications 110(15):60-68, January 2015.

17. *The survey of data mining applications and feature scope* by Padhy, Neelamadhab, Dr Mishra, and Rasmita Panigrahi. Asian Journal of CSIT (2012).

18. *Data mining approach for predicting student performance* by Osmanbegović, Edin, and Mirza Suljić. Economic Review (2012).

19. *Predicting and analyzing secondary education placement-test scores: A data mining approach* by Şen, Baha, Emine Uçar, and Dursun Delen. Expert Systems with Applications (2012).