APPLYING PREDICTIVE ANALYTICS IN ELECTIVE COURSE RECOMMENDER SYSTEM WHILE PRESERVING STUDENT COURSE PREFERENCES

Thesis submitted in partial fulfillment of the requirements for the award of degree of

Master of Engineering
in
Computer Science and Engineering

Submitted By
Ridima Verma
(Roll No. 801632040)

Under the supervision of:
Ms. Anika
Lecturer, CSE Department

THAPAR INSTITUTE
OF ENGINEERING & TECHNOLOGY
(Deemed to be University)

COMPUTER SCIENCE AND ENGINEERING DEPARTMENT
THAPAR INSTITUTE OF ENGINEERING AND TECHNOLOGY
PATIALA – 147004

June 2018

CERTIFICATE

I hereby certify that the work which is being presented in the thesis entitled, "*Applying Predictive Analytics in Elective Course Recommender System while preserving Student Course Preferences*", in partial fulfillment of the requirements for the award of degree of Master of Engineering in *Computer Science and Engineering* submitted in Computer Science and Engineering Department of Thapar Institute of Engineering and Technology, Patiala, is an authentic record of my own work carried out under the supervision of *Ms. Anika* and refers other researcher's work which are duly listed in the reference section.

The matter presented in the thesis has not been submitted for award of any other degree of this or any other University.

Signature:

Ridhma Verma

This is to certify that the above statement made by the candidate is correct and true to the best of my knowledge.

Ms. Anika

Lecturer, CSED

ACKNOWLEDGEMENT

I would like to express my grateful thanks to my supervisor **Ms. Anika**, this work would not have been possible without her encouragement and guidance. I really thank for all the time, patience, discussion and valuable comments. Her enthusiasm and optimism made this experience both rewarding and enjoyable.

I am equally grateful to **Dr. Maninder Singh**, Professor and Head of Computer Science and Engineering Department, who always encouraged us to keep going with the work.

I will be failing in my duty if I does not express my gratitude to **Prof. S. S. Bhatia**, Senior Professor and Dean of Academic Affairs, Thapar Institute of Engineering and Technology, for making all the provisions of infrastructure such as library facilities, computer labs equipped with net facilities, immensely useful for learners to equip themselves with latest in the field.

I am also thankful to entire faculty and staff members of Computer Science and Engineering Department for their direct-indirect help and cooperation, which made my stay at Thapar Institute of Engineering and Technology memorable.

Date: June, 2018 (Ridima Verma)

Place: Thapar Institute of Engineering and Technology

ABSTRACT

In higher education scenarios, elective courses sought to provide a deeper insight of the trending advancements in the field of specialization for undergraduate students. Choice of elective subjects during the pre-final or final year of the undergraduates play a crucial role as they help in shaping their career or area of specialization for future research. However, there exist numerous gaps and concerns that arise due to mismatch of the elective course pre-requisites and the student's possessed skills-set which result in degraded student academic performance as well as quality of education. This research study focuses on filling in these gaps by efficiently predicting the marks in different elective subjects for the current cohort of students, beforehand, as well as side by side preserving their explicit subject preferences.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ABBREVIATIONS

EDM	Educational Data Mining
DT	Decision Tree
LR	Logistic Regression
NN	Neural Network
CBR	Content-based Recommendation
CF	Collaborative Filtering
KNN	K-nearest Neighbor
NB	Naïve Bayes
TN	True Negatives
FN	False Negatives
TP	True Positives
FP	False Positives

CHAPTER 1
INTRODUCTION

The process of extraction of useful information from huge piles of data is termed as data mining. Its applications provide some useful insights in many areas. A comparatively new field known as Educational Data Mining (EDM) came into existence in year of 2005. This new field is intended to develop data mining techniques to obtain information gathered from educational institutions. The main concern of this field is to facilitate decision making in educational scenarios by analyzing academic data. The facts produced by Educational Data Mining can be used by several associates in academics. For instance, it can facilitate instructors appraise course structure and teaching ways. Additionally, students can get course recommendations supporting their performance. Advisors, on other hand can get useful insights from this new field in order to make predictions for student performance. Predicting weak students at early stages can provide extra support for them.

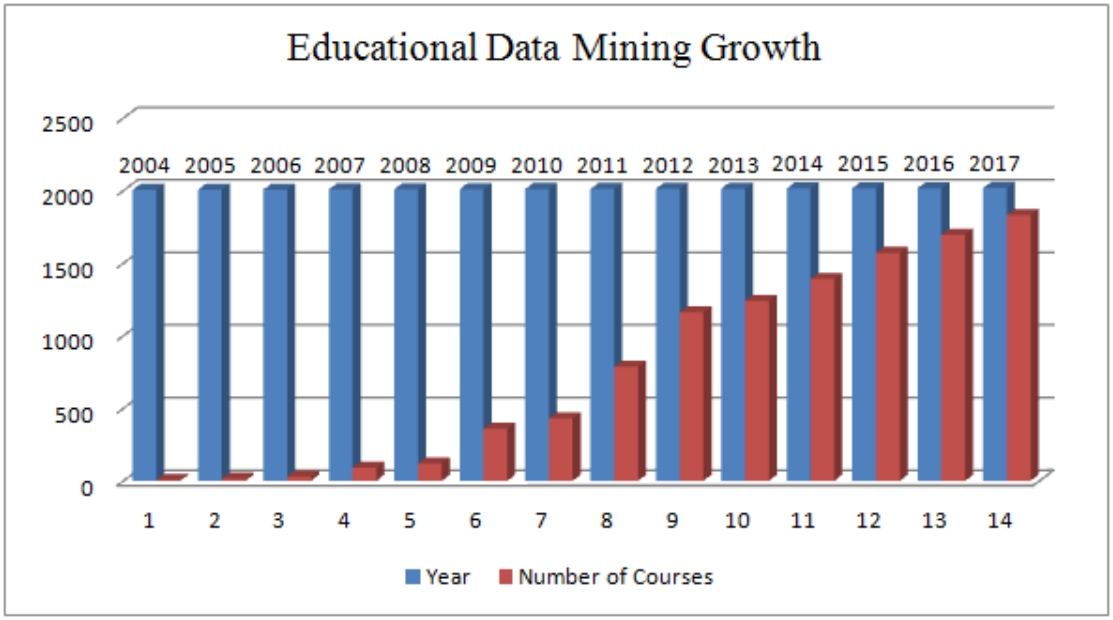

Figure 1.1: Evolution of Educational Data Mining

Relatively hard queries such as whether a student has knowledge or if they are previously occupied can be easily answered through Educational Data Mining. Many new techniques have been practiced by researchers for model building that can predict promising student outcomes.

In case of higher education, as the undergraduates proceed in their academics, they develop a liking for a particular field of study in their area of specialization. The higher academic course curriculum is also arranged in such a way that after imparting the knowledge of core subjects during the initial two and half years of their study, in pre-final or final year they are sought free to choose their own area of specialization in the form of different elective subjects. The elective subjects provide an insight into the trending advancements of their fields and form the basis of their carrier, as these tend to give a glimpse of their future area of research and specialization.

However, as per the current educational scenarios, the undergraduates remain mostly confused on what to choose as they either lack in having the sufficient initial knowledge of the elective subjects or are having knowledge overflow of all subjects and so are unable to decide which one to choose. In such scenarios, they often seek the advice of their instructors or friends and mostly go with the cohort choice. However, going with the flow often creates a gap between their actual skills set and the required skills set for the elective subject that they have preferred as their choice. In later stages, this results in loss of interest of the students in the enrolled elective subject and hence a degraded academic performance is encountered by the institution. On the other hand, from the institutional perspective also elective subjects play a vital role as they need to arrange all the required infrastructural and teaching resources timely for the successful run of the various elective subjects. In worst cases, institution is bound to choose another subject over the students preferred subject due to lack of the resources. Similarly, as a result of this, there can be numerous limitations, gaps or concerns arising either in case of students or institutions in real world educational scenarios.

This research study tries to bridge these concerns by recommending efficient elective course subjects to the institution that indirectly predicts the academic success of different electives beforehand and along with this also preserves the student subject interests. This approach can help to reduce the skill-gap often seen in case of higher education scenarios and timely guide the institutions for the electives that can result in overall increased academic performance and quality education. So, a personalized recommender system is required that can provide recommendations to the students, according to their curriculum, interest and context for choosing most relevant courses. Figure 1.2 shows the proposed framework for generating such elective course recommendations.

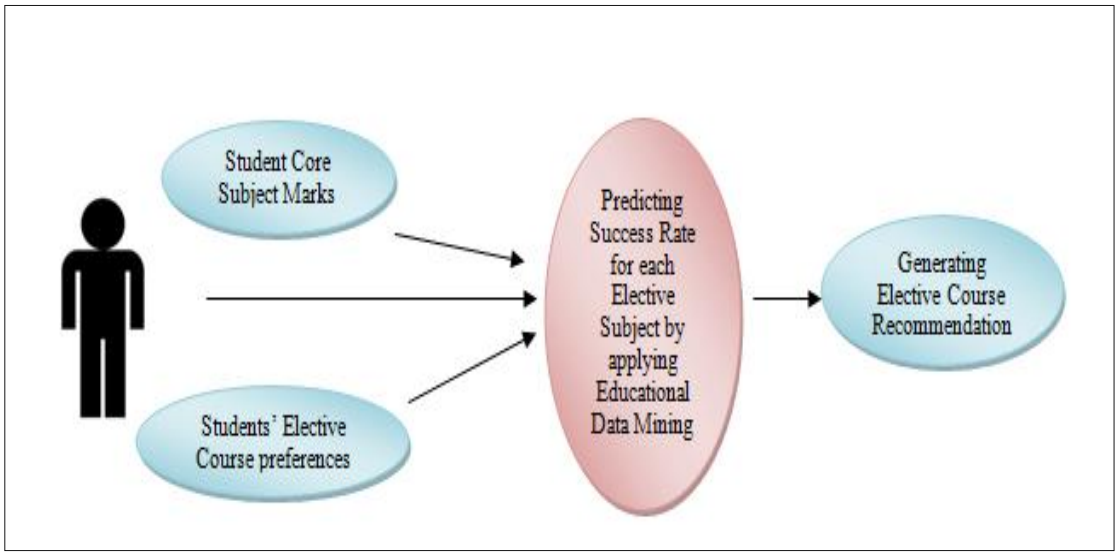

Figure 1.2: Proposed Framework

1.1 E-Learning Recommender Systems

The main task of these systems is to generate learning task recommendations based on learner's previous tasks and performances. The likeliness of the learner can be constructed using the user profiles, which identify the needs of user. Recommendation generation task is broadly divided in two major phases: "learning" and "advising" phase. The former phase identifies learning patterns of previous users and the later phase helps in application of learned model for generation of recommendations. Recommender systems can facilitate a user to find related items according to their interest. The main objective of these systems is to make selected recommendations based on user requirements. They implement the recommendation generation process by asking the user to rate a series of objects, on basis of which new recommendations are generated for the user and other similar users. These evaluations act as input for model generation, which helps in making predictions for the user on basis of either their profiles or the objects evaluated by them in the past.

The recommender systems must be able to produce tailored suggestions based on specific needs, which can help educators to know the strengths and weakness of learner and take measures according to learner's skills. They should be able to facilitate a mechanism to compile the huge amounts of user data altogether at once to produce standard recommendations without any sort of discrepancy.

1.1.1 Recommender System Approaches

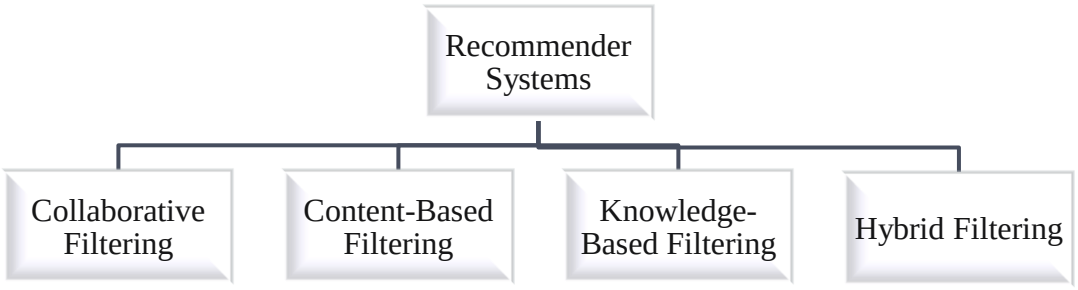

Figure 1.3: Diversification of Recommender System

- Collaborative filtering (CF) is a recommender system approach that is commonly used to generate personalized recommendations. It computes the likeliness either between clients or things. It leverages product transactions to produce recommendations. In this type of approach, for a specific customer, we find similar customers based on transaction history and recommend items that the customer in question hasn't purchased yet and which the similar customers are likely to buy.

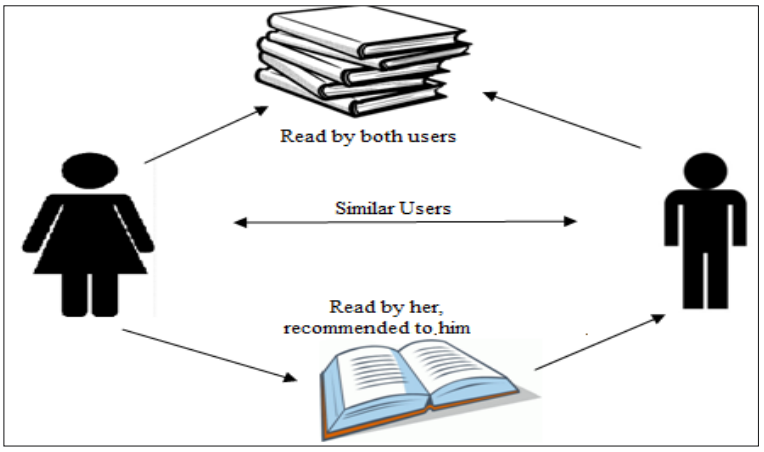

Figure 1.4: Collaborative Filtering Example

- The main task of content-based filtering or cognitive filtering is to generate recommendations on basis of similarity between the list of the items and user profiles. The content of each item is pictured as a set of descriptors or terms, usually the words that occur in a document. Such systems leverage product information for its recommendations. For example, if a person looks at a book on an online bookstore, a content-based system would probably recommend the similar author books, because it would look at the author field.

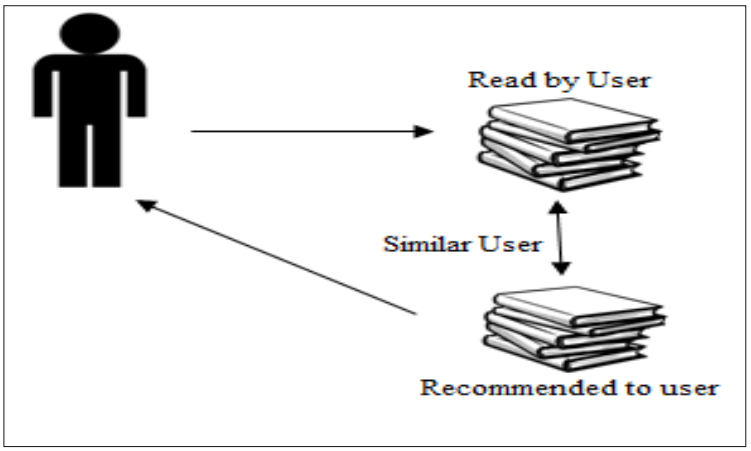

Figure 1.5: Content-based Filtering Example

- The main task of knowledge-based recommender system is to use information about users and items and continue a knowledge based perspective for generation of recommendations, reasoning about what items meet user's perspectives. Such recommender systems are more specific about the task performed by them based on explicit information about item assortment, user preferences and recommendations. These systems work well where approaches like collaborative filtering and content-based filtering cannot be applied. Non-existence of cold-start problem makes knowledge-based recommender system approach better than other existing approaches. For generation of explicit recommendations, there is a need for potential knowledge acquisition bottleneck to be triggered.

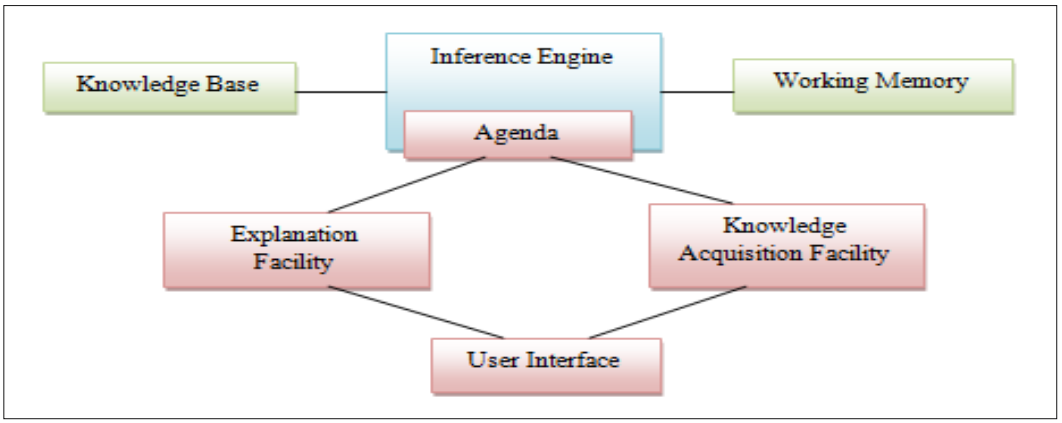

Figure 1.6: Knowledge-based Filtering Approach

- The combination of two or more recommender system approaches for better performance and lesser drawbacks constitute to formation of hybrid recommender systems. Most frequently, collaborative filtering approach is combined with other existing techniques in an attempt to avoid the problem of expansion. Combination

of two or more techniques for building hybrid recommender systems seek to inherit advantages and disadvantages. Establishment of synergy between two or more approaches led to more accurate results for hybrid recommender systems that combine multiple recommendation techniques.

In spite of existence of different recommender system approaches, a number of approaches are practical to merge (i.e. Collaborative, Content-based and Knowledge-based Recommender), work will mainly focus on the combination of collaborative filtering and content based filtering techniques. Depending on the type of data and attributes, different types of combinations might produce dissimilar outputs.

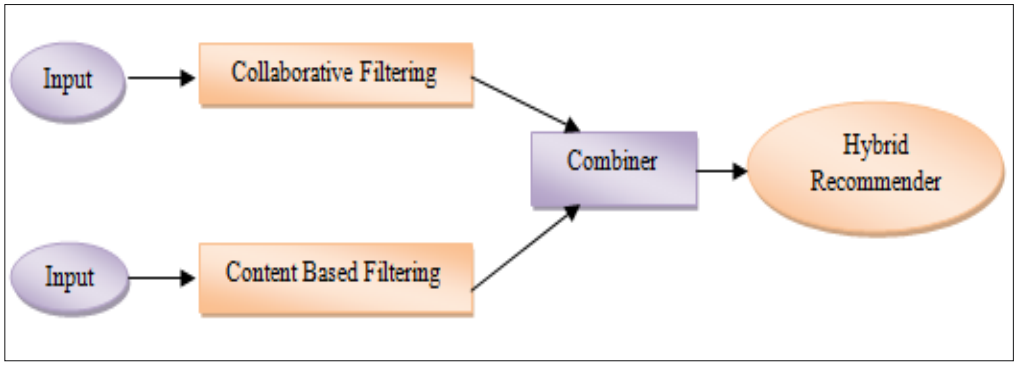

Figure 1.7: Hybrid Recommendations Example

1.2 Predictive Analytics

The process of extraction of useful knowledge from existing datasets in order to obtain insights and to predict future outcomes and trends is termed as predictive analytics. It does not tell what will happen in near future rather it forecasts what might happen in near future with adequate level of authenticity, including what-if scenarios and risk assessments. Predictive models utilize known outcomes to create a model that forecasts that represent to likelihood of objective variable in view of assessed criticalness from arrangement of input variables.

Prerequisites required for implementation of predictive analytics is to collect huge amounts of indistinct data from different authorities. The consolidated data which is improved by internal data is mainly important. Statistical methods such as extrapolation, regression, neural networks and machine learning are used to process data in predictive analytics, which detect data patterns and derive algorithms. These algorithms are optimized based on test data. Also, it works on principle of more the

data, better the developed algorithm. Once the optimization is done, the algorithms and models are applied for dataset whose classification is unknown.

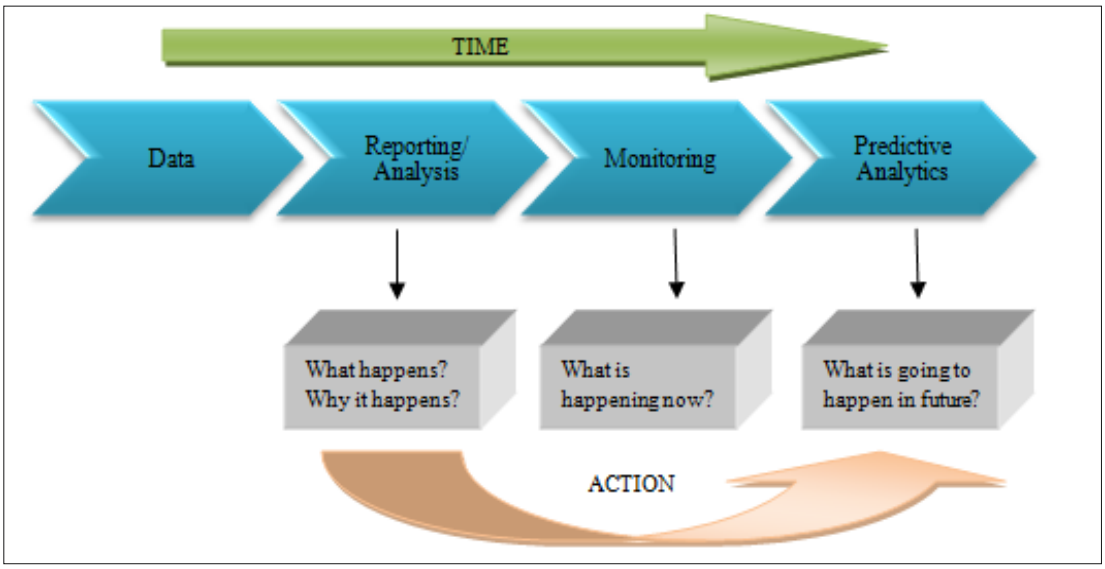

Figure 1.8: Predictive Analytics Approach

1.3 Aim and Scope

For blended learning environments, course recommendations form a bilateral process as institutional resources and student interests both needs to be taken care off. In blended educational scenarios, institutional infrastructural and teaching resources often form a limiting factor. As a result, despite of the trending student subject preferences, institution based pre-defined subject list is imposed on the students considering the institution resource limitations at the last hour. On the student part, as per the student's disinterest in the imposed subjects, the academic performance of the student diminishes that result in quality degradation of the skill set possessed by the students. This indirectly also effects the learning outcomes of the subjects with lower success rates and thus affecting institutional quality education.

On the other hand, a similar case of considering only the student elective subject preferences and allotting them the desired subjects also does not assure complete academic success as there exists a gap between student's desired subject and the possessed skill-set. For practical scenarios both must go hand-in hand as the possessed skill-set or capability must match the course demanding pre-requisite and ongoing skills for better academic success.

Considering the present blended educational scenarios, there is a need of an elective course recommender system that is aimed to provide timely elective course

recommendations for institutions that assures academic success of the students at the backend while preserving student subject interests.

These predictions in turn, can also help to intimate the institution for the necessary arrangements of the resources, beforehand, that will lead to successful academic quality performances.

1.4 Thesis Structure

Rest of this theory is categorized in the following way:

- Chapter 2 discusses about the earlier existing literature in the domain of Educational Data Mining and Course Recommender systems framework.
- Chapter 3 forms the problem statement and presents the objectives of this study.
- Chapter 4 defines the methodology used to achieve the objects for Elective Course Recommendations. It deals with the generation of algorithm along with use of data mining techniques and tools for that task of elective course recommendation.
- Chapter 5 describes the implementation part. It shows the stepwise procedure to implement the proposed system which is mentioned in Chapter 4.
- Chapter 6 analyses the results of the proposed system in course of accuracy, precision, recall and F1 Score and compares the results with other existing systems.
- Chapter 7 concludes and presents the deeper insights into the future that can extend and improve the quality of the work of elective course recommendation generation.

CHAPTER 2
LITERATURE REVIEW

Colossal amounts of research have been done in the field of prediction of student academic performance and provision of course recommendations. Apart from the number of features that can influence student's academic performance, choice of models used for making the predictions also play a crucial role in the prediction procedure. A brief review of the various existing models or techniques is discussed below.

Ray and Sharma [1] proposed a recommender system approach that made use of collaborative filtering approach for generation of elective course recommendations. They made use of both user-based and item based filtering, which is applied on real-time data for prediction of elective courses. Results are based on mean absolute error, calculated for each elective course.

Arsad *et. al.* [2] proposed a neural network student performance prediction model which gave review on what factors can influence the final GPA obtained by students in their last semester. The results calculated were based on correlation and mean squared error between models applied.

Shahiri *et. al.* [3] proposed a recommender system for online course enrolment, which used historical data to show the factors which influences student's choice of elective course selection. They made use of collaborative filtering approach and provided performance based on recall and coverage

Romero *et. al.* [4] proposed data mining techniques for classification of students which made structuring and implementation easier for instructors. They made use of pre-processing techniques such as discretization and rebalancing on collected data in order to obtain better classification results.

Osmanbegović and Suljić [5] presented data mining techniques for making student predictions based on their performance. Different data mining techniques were compared for generation of prediction model, which in return predicted student success. Data was collected by performing survey during the semester. The success of student was evaluated based on grades obtained by students in their final exam.

Bunkar *et. al.* [6] proposed data mining classification techniques. The main approach adopted by them was based on improvement of student performance. They made a

system that facilitates the use of rule generation process. Final grades of graduate students are predicted using decision tree model.

Guo [7] presented neural network approach by making use of statistical methods. These techniques incorporated establishment of dynamic models which helped in predicting student course satisfaction. Out of all the models applied, MLP outperforms in generation of near practical results.

Ramesh *et. al.* [8] presented statistical and data mining approaches. The objective of this study is to make student predictions based on their performance in the curriculum. They also considered influencing factors that can affect student performance and their final grades.

Affendy *et. al.* [9] proposed a model based on data mining approaches. They made predictions for calculating student academic performance based on factors that can affect their overall grades. The main objective of this study is to rank the affecting factors in order to warn the students, so that they are able to maintain their grades.

Chamillard [10] proposed a model based on data mining techniques. The main objective behind this study is to make an analysis in curriculum for generation of approaches that can make predictions based on each student's past performance in their academic curriculum. This approach can help in making student performance predictions in their later courses.

Al-Badarenah and Alsakran [11] presented a collaborative filtering recommender system using clustering techniques for generation of elective courses by making use of association rules to recommend courses based on similarity measure.

Cakmak [12] proposed a recommender system approach that is used for estimation of student course grades. Along with this, they enhanced the proposed approach by implementing automated outlier removal, which improved the quality of result generation.

Tran *et. al.* [13] presented analysis on prediction of student performance in educational scenarios. They made use of different regression and data mining strategies for implementation of proposed approach. They also implemented combination of aforementioned techniques and calculated results based on model accuracies.

Mueen *et. al.* [14] presented an analysis for application of data mining techniques for prediction of student performance. They made use of real-time data collected from undergraduate students.

Bydžovská [15] proposed a prediction model using classification, regression and collaborative filtering approaches for predicting final grades of students based on previous achievements of similar students.

Bienkowski *et. al.* [16] presented enhanced learning and teaching through educational data mining along with learning analytics model to predict student performance based on various influencing factors to support online learning.

Mishra *et. al.* [17] presented analysis on approaches and trends followed in predictive analytics in knowledge discovery domain. They made use of business intelligence data for forecasting and modeling.

Felfernig *et. al.* [18] presented basic approaches in recommender systems.

Gaddam and Shekhar [19] proposed a strategy for detecting anomalies by utilizing a mix of k-mean clustering and ID3 classification tree. They also made comparison between classification performance with the individual ID3 decision tree and K-mean clustering.

Ahre and Lobo [20] demonstrated information mining approach. They made use of clustering and association rule mining and compared their outcome with the open source information mining apparatus Weka. The outcome acquired showed that utilizing joined approach is superior to application of individual methodologies.

Zaiane [21] proposed the web mining systems that prescribed on-line learning exercises in course web pages, in view of learners' history to enhance course route and help the online learning process. They designed a recommender system agent by using association rule mining.

Upendran and Chatterjee [22] proposed a course recommendation system that undertook students as basis of their past performance and learning ability. They constructed model by using previous student data as input. The basic subsequent belief for the technique used is that if a student with certain skill is able to complete the course successfully then a new student with similar skills will be able to complete the course.

Castro and Vellido [23] provided an effective and efficient information about current research and data mining applications in e-learning such as to know about student failure, classification based on students performances, e-learning system recommendation, clustering of students and much more.

Elbadrawy and Karypis [24] showed that how student and the academic feature can regulate the enrollment data. They used this features to specify the student and course

group at various level. They defined how this feature is important in designing the model for making predictions on basis of collaborative filtering and matrix factorization techniques.

The Degree Compass [25] system is implemented to make course recommendations for students studying in tertiary programs. The main motive behind development of this system is to pick out the course selection process that will guide the students and will help them to perform better in their examinations. This system predicts the courses to overcome the problem, which students face while selecting the courses.

The Course Agent [26] is another course recommendation system which contains student data containing course preferences opted by students during their curriculum. It is a general system that is able to manipulate its results on basis of student preferences.

RARE [27] is a recommender system that made use of association rules which generates recommendations for elective courses to be undertaken by students during their curriculum. This system makes use of data mining approaches along with generation of user ratings for prediction of elective courses. It uses historical data to mine the course rules. The rules are used for generating recommendations and also allow user to rate the courses for making future improvements.

Engin and Aksoyer [28] presented a survey in which they discussed two educational systems. The first recommender system is able to generates course recommendations for undergraduate students where as second recommender system suggests a student about their scholarship based on their eligibility.

Garcia and Romero [29] described a methodology to maintain and develop the web-based course system. They utilized information mining strategy named association rule mining to discover the course connection as IF-THEN principles. Afterward, they utilized collaborative filtering method to score and offer the principles acquired by instructive specialists.

This chapter presented a brief introduction to various researches implemented in the field of Educational Data Mining and Course Recommender Systems for efficiently predicting the student academic performances for a better success rate in different courses, in the future.

CHAPTER 3
PROBLEM STATEMENT

Despite of vast literature existing on the efficient models and methodologies being used for making efficient student academic course prediction, elective course recommendations area seems to be unexplored by the researchers. Not enough literature have been found that provides elective course recommendations to the institutions based on the student preferences as well as also merits over the estimated academic performance of the current batch. This study is aimed to bridge this gap and provides a bilateral elective course recommendation that assures student as well as institution success.

This research study tries to bridge gaps by recommending elective course subjects to institutes that indirectly predicts success rate of different elective courses beforehand and also preserves the student subject explicit preferences. These predictions in turn can also help to intimate the institution for the necessary arrangements of resources, beforehand, that will lead to successful academic quality performance. Thus, the aim of this study is to propose a methodology that can be implemented for generating efficient elective course recommendations for assuring two-sided success.

The study targets the following research objectives:

- Identifying the efficient data mining techniques for predicting marks in proposed elective subjects.
- Propose an algorithm that also considers the contextual information of student varying preferences.
- Generate efficient list of elective course recommendations that assures bilateral academic success of students as well as the institutions.

CHAPTER 4
METHODOLOGY

4.1 Data Set Description

Real time university anonymous dataset, for 2 years, consisting of undergraduate student's core subject marks, subject preferences, student allocated elective subject and marks in the respective elective subject of computer science and engineering department during their final year, were used in this study [32]. The dataset consisted of approximately 658 student entries with 13 attributes, namely marks obtained by students in their ten core subjects of computer science and engineering curriculum, actual allotted elective subject, actual marks in the elective allotted subject, along with student preferred interest subjects.

4.2 Feature Selection

Within the anonymous datasets, out of 15, 13 relevant features for the respective domain were selected. These selected features were further used for model construction and to train and test the models for obtaining efficient elective course recommendations. The 13 considered features consisted of 11 continuous attributes and 2 discrete attributes. Table 4.1 and 4.2 represents the considered elective course subjects list and a brief description about the actual features considered for this study, respectively.

TABLE 4.1: ELECTIVE COURSE SUBJECT LIST

Code Used	Attributes
SVV	Software Validation and Verification
CC	Cloud Computing
CS	Cyber Security
IoT	Internet of Things
ML	Machine Learning
NLP	Natural Language Processing
AR_VR	Augmented Reality and Virtual Reality

TABLE 4.2: SELECTED ATTRIBUTES LIST FROM THE DATASET

No.	Attributes	Description	Selected ?
1	Student Gender	Male, Female	No
2	Student Family Status	Lower, Middle, Upper	No
3	Student Marks in 10 Core Computer Science and Engineering Subjects (Computer programming, Data Structures and Algorithms, Database Management Systems, Microprocessor, Computer Architecture, Software Engineering, Computer Networks, Operating System, Web Development, Automata)	Marks ranging from 0 to 100	Yes
4	Student Preference (Interest) for Elective Subject	Only first subject preference was considered	Yes
5	Student Allotted Elective Subject	Assigned elective Subject	Yes
6	Student Marks in Allotted Subject	Marks ranging from 0 to 100	Yes

4.3 Data Preprocessing and Categorization

As discussed, in the dataset out of thirteen, eleven were continuous marks attributes. Although, in small number, but the missing values in case of marks were replaced by the average value obtained by students for that particular subject. Missing values for allotted and preferred subjects attributes were not encountered as these were mandatory fields for every student.

Data categorization was also done within the dataset by converting the obtained marks in the allotted subject from continuous variable to a discrete variable through the process of binning of data. The allotted subject marks were binned into different classes or bins with a fixed difference of 5 marks each. So, the actual binning of data for allocated marks started from [0 to 5], [6 to 10], [11 to 15] and so on till [96 to 100]. The motive behind this categorization was to fit the data for various data mining classification techniques used further in this study. A glimpse of final pre-processed and categorized marks dataset is shown in Table 4.3.

TABLE 4.3: DATASET AFTER PREPROCESSING AND CATEGORIZATION

S1	S2	S3	S4	S5	S6	S7	S8	S9	S10	Allocated Elective Marks	Categorized Elective Marks
72	75	92	67	55	70	79	75	68	81	86	[86 to 90]
41	38	53	68	64	56	71	46	81	21	64	[60 to 65]

The complete methodology adopted for generating elective course recommendations is shown in Figure 4.1.

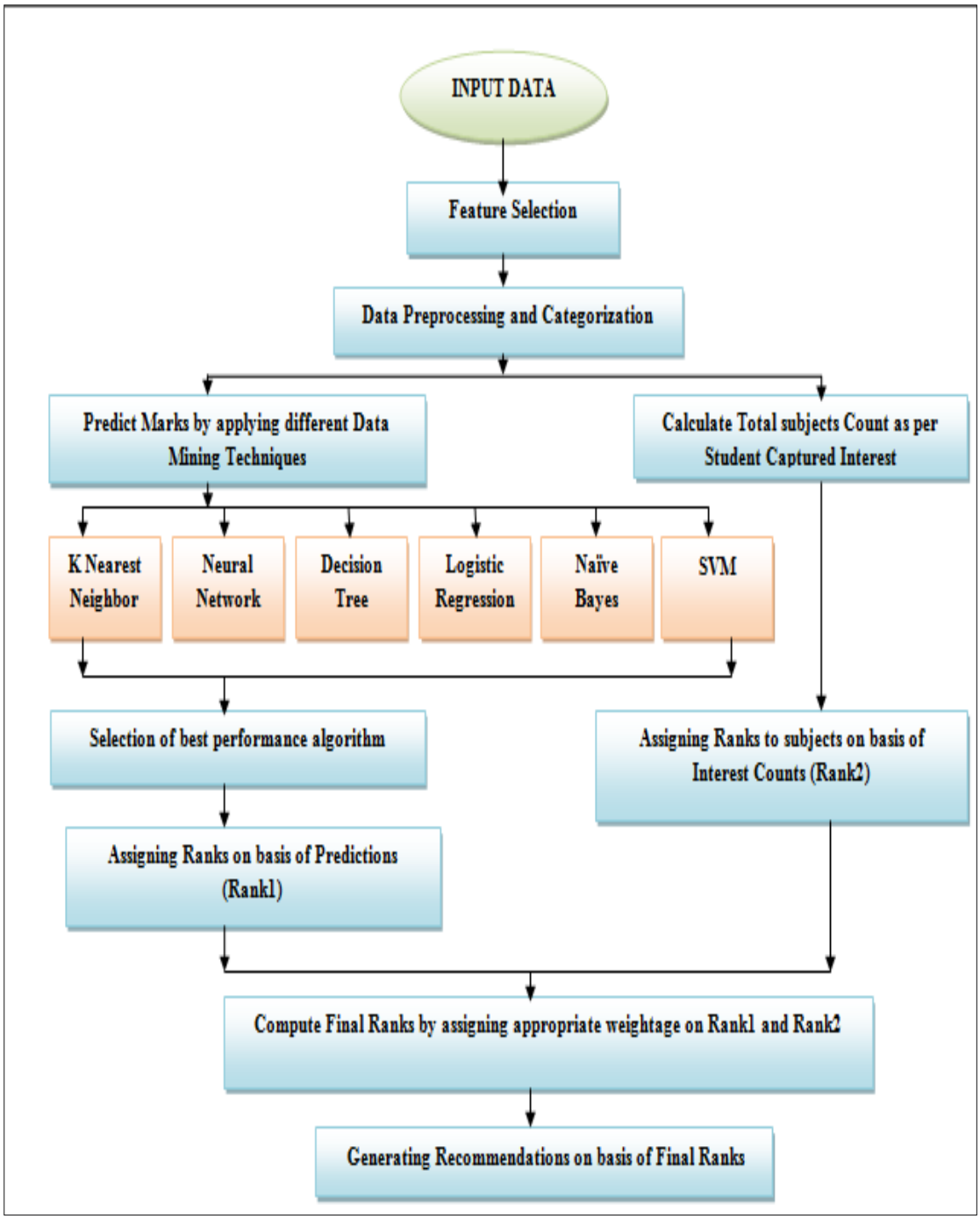

Figure 4.1: Complete Flowchart of Proposed System for generating Elective Course Recommendations

4.4 Applying Educational Data Mining

After data preprocessing and data categorization, various supervised learning classification models were imposed on the final dataset. The selection of the various classification techniques used for generating predictions were done on the basis of previous studies. According to the previous studies, six common classification models were identified that are frequently used by researchers and outperformed the marks based predictions on their respective datasets. The identified six techniques were: K-Nearest Neighbor used within collaborative filtering approach, Naïve Bayes, Decision Tree, Neural Network, Logistic or Linear Regression (for continuous data) that is mostly preferred for marks prediction and Support Vector Machines.

So, further a brief description of the final six models (K- Nearest Neighbor, Neural Network, Decision Tree, Naïve Bayes, Support Vector Machines and Logistic Regression) applied on the final datasets is given below.

4.4.1 k-Nearest Neighbor

k-Nearest Neighbor (kNN) is lazy learning as well as a non-parametric algorithm which is used for Classification and Regression. In kNN distance measure is used to determine closeness between instances. kNN acts as a collaborative filtering technique, which stores all possible events and classifies new events based on the similarity with other similar events. Euclidean distance is used to find the separation amongst students and classified them into the similar neighbor groups. kNN forms the basis of the collaborative filtering technique used in recommender systems. kNN is a simple algorithm which classifies new cases based on similarity measure. It is based on feature similarity. This is better choice for classifications where accuracy is important. For predicting the various categories of allocated subject marks various values of k (1, 2, 5, 10, 20… 100) were experimented. Initially the accuracy varied on larger scales but with the k-size between 50-60, the results yielded highest accuracy and after this the accuracy was stabilized. So, the experimentation proceeded by assigning a k-size of 60.

4.4.2 Decision Tree

The Decision tree (DT) is a conventional, tree shape structure which is used to determine every possible result and statistical probability, used in supervised learning as decision support tools. They contains conditional control statements and promise

an output in either case *i.e.*, whether the condition is met or not. This model breaks down the dataset into smaller subsets on the basis of conditions and also incrementally develops the tree along side. The result of this is a decision tree with decision nodes as conditions and leaf nodes as outputs. For the datasets used, decision tree's ID3 algorithm was used for classification and allotted subject marks category as output. Figure 4.2 shows a small portion of the decision tree generated for the dataset.

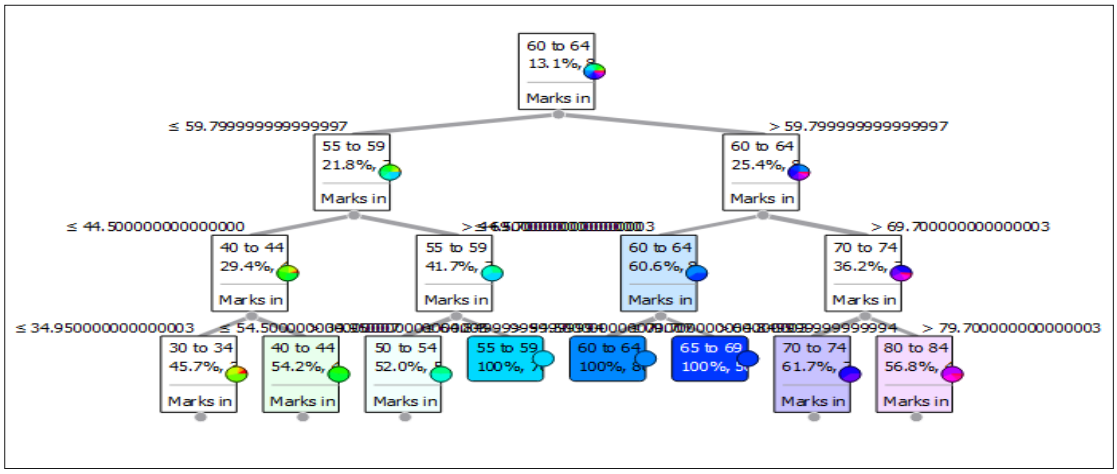

Figure 4.2: Pictorial Representation of generated Decision Tree

4.4.3 Neural Network

Neural Network (NN) simulates the principle of human brain that is composed of highly inter-connected network of neurons. The data nodes in this model act as input nodes of the neurons. The model is trained by iteratively adjusting the weights of the hidden layers in order to obtain the desired output. Neural Network thus modifies itself as it progresses further and learns from the training sets. However, neural network is said to work faster on large datasets. Here, the model was trained by supplying of ten core subject marks as input data nodes and actual allotted subject marks category as desired output.

4.4.4 Naïve Bayes

Naïve Bayes (NB) classifiers are probabilistic classifiers based on Bayes theorem and assume that the features involved are of independent nature. It uses maximum likelihood for parameter estimation. Because of the excessive use of Naïve Bayes classifier and its variants in educational scenarios for student knowledge estimation, this was also selected and used for making the prediction for the allocated subject marks category of students. Figure 4.3 represents the nomogram for Naïve Bayes.

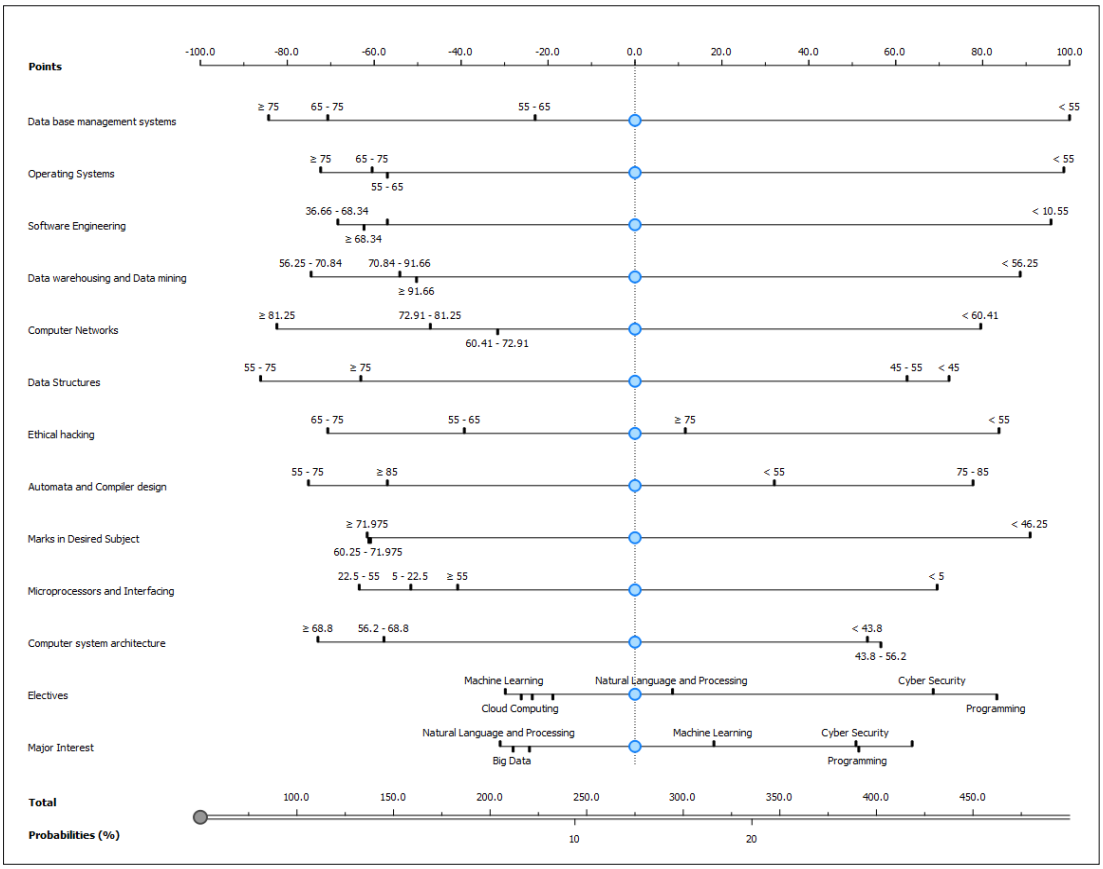

Figure 4.3: Nomogram representing Naïve Bayes

4.4.5 Support Vector Machine

In supervised learning, Support Vector Machine (SVM) is used as a discriminative classifier which classifies on the basis of separating hyper-plane. When it is supplied with supervised learning training dataset, it outputs an optimal hyper-plane that further helps in classifying or categorizing the testing data. This forms the first choice when the dataset size is small. Inputting the labeled dataset, the classifier was trained and further used to test and predict the allocated subject marks category as the output. Figure 4.4 represents the output of the Support Vector Machines.

4.4.6 Logistic Regression

Linear Regression is most commonly used for marks prediction. But due to the categorical nature of the target variable here, Logistic Regression was used for predictive analytics. Logistic Regression analysis is similar to linear regression except that the outcome in case of Logistic Regression is dichotomous. However, because of its flexibility and adaptive nature it can be used for categorical data too. The categorical variable considered for prediction was the allocated marks subject. Figure 4.5 shows the graphical representation of the Logistic Regression.

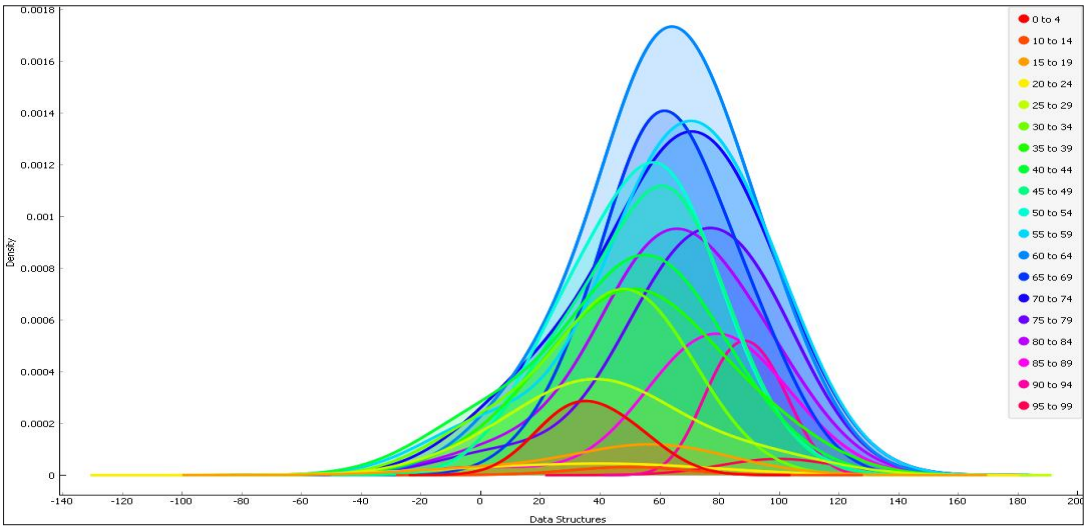

Figure 4.4: Output of Support Vector Machine according to the Core Subjects based on Categorized Elective Subjects respectively

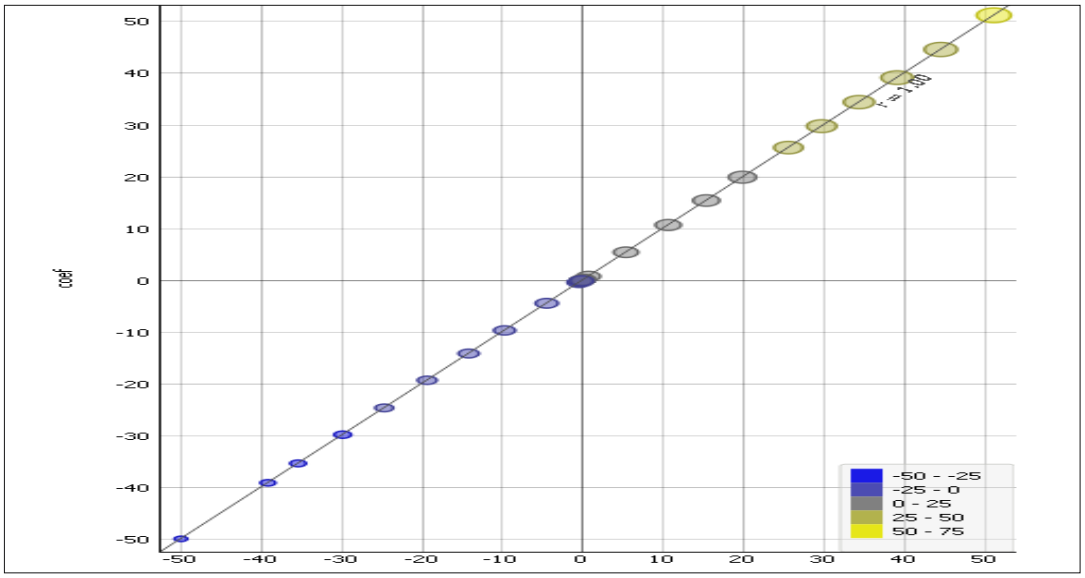

Figure 4.5: Graphical Representation for Logistic Regression

4.5 Rank-I Calculation

After the application and evaluation of the different data mining techniques used in this study, the model producing the highest accuracy was selected. In this case Support Vector Machine's were used as it was found to be the best in terms of accuracy when comparing different data mining techniques used in this study. After this the marks predictions were obtained for each elective separately. Once the predicted marks categories for each elective subject for each student were obtained, efforts were put for finding average marks category slots for that entire elective subject using student allocated subject column. Cumulative frequency concept was used for determining the average marks category slot for particular subject. Once the

average marks category slot was calculated for each elective subject, ranks were assigned to the subjects, where highest average marks category was awarded with rank 1 and lowest average marks category was awarded with rank n, n being the total number of elective subjects (7 in this case).

4.6 Rank-II Calculation

In parallel, the student subject preferences were also considered and actual count of students preferring a particular subject was calculated. This procedure was repeated for each existing elective course. Once all the counts were calculated, again ranks were provided based on the most preferred and least preferred elective subject. Highest preferred subject was assigned a rank of 1 and least preferred a rank of n (n is 7 in this case).

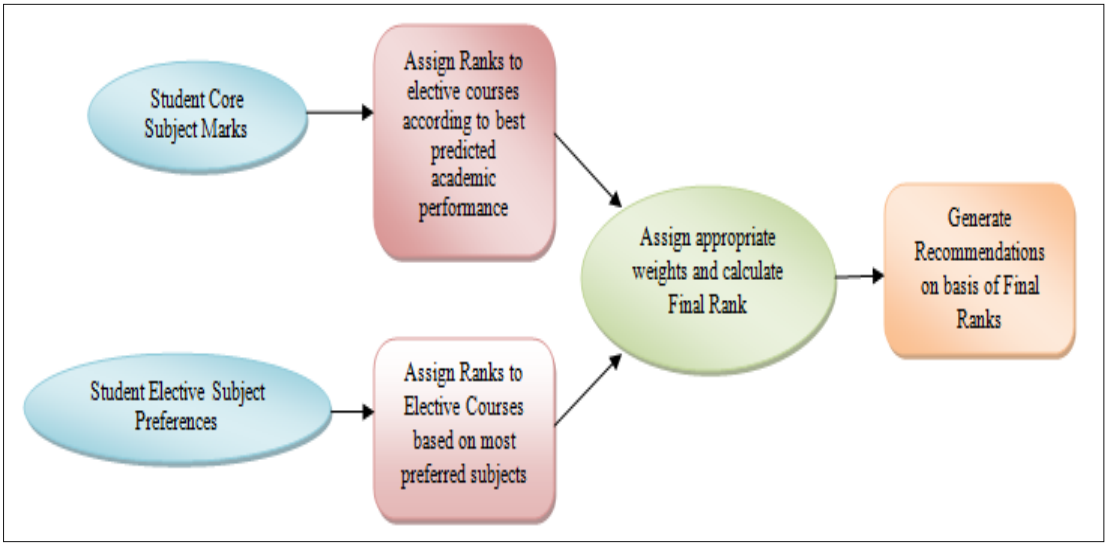

Figure 4.6: Steps for generation of Final Elective Course Recommendations

4.7 Weighted Rank Calculation

Since efforts are put in to preserve the student preferences along with assuring quality academic performance of the students collectively, the weighted rank concept was proposed. After the calculation of Rank1 and Rank2 from the above steps, final weighted rank can be calculated by assigning w1 to Rank1 and w2 to Rank2 as given in the following equation:

$$Weighted\ Rank_i = w1 * Rank1_i + w2 * Rank2_i$$

Here,

$Rank1_i$ = Rank1 for i^{th} elective subject preserving academic performance for i^{th} elective subject.

$Rank2_i$ = Rank2 for i^{th} elective subject preserving student subject preference.

w1 = 0.5 and w2 =0.5

In this study, equal weightage was given to both, so w1 and w2 both were assigned equal weights of 0.5. However, these can further vary depending upon the institutional requirements and preferences.

4.8 Final Rank Generation

On the basis of weighted ranks obtained in earlier step, final ranks were calculated again, by assigning most preferred (having lowest weight) a rank of 1 and least preferred elective subject with a rank of n (n is 7 in this case).

The subjects are then arranged in increasing order of their Final Ranks and recommended to the concerned authorities for further considerations. Figure 4.6 and Table 4.4 briefly describes the complete procedure of generating the elective course recommendations.

TABLE 4.4: CALCULATION STEPS USED FOR GENERATING FINAL RANKS

Elective Subjects	Average Marks Category Slot	Rank1	Preferred Subjects	Subject Preference Count	Rank2	Weighted Ranks	Final Rank
Cloud Computing	[60-64]	4	Cloud Computing	50	7	(4*.5+7*.5) = 5.5	7
Internet of Things	[55-59]	5	Internet of Things	68	2	(5*.5+2*.5) = 3.8	4
Natural Language and Processing	[70-74]	2	Natural Language and Processing	55	6	(2*.5+6*.5) = 3.6	3
Cyber Security	[45-49]	7	Cyber Security	65	3	(7*.5+3*.5) = 5.4	6
Software Verification and Validation	[65-69]	3	Software Verification and Validation	62	4	(3*.5+4*.5) = 3.4	2
Machine Learning	[75-79]	1	Machine Learning	79	1	(1*.5+1*.5) = 1	1
Augmented and Virtual Reality	[50-54]	6	Augmented and Virtual Reality	62	4	(6*.5+5*.5) = 5	5

CHAPTER 5
IMPLEMENTATION

For experimentation, educational data mining, analysis and predictions, python's Orange [30] library was used. The final results of the predictions were stored in the form of excel sheet which was further processed as per the proposed algorithm. The analysis was performed on a dataset consisting of 658 rows. Each data mining model used their own learning algorithm and generated a model that fitted the input data and generated predictions. Predictions that are generated via models are expected to have good generalization capability and can efficiently produce a correct class label or categorization for the previously unknown data.

5.1 Tools and Technology

Python programming dialect was utilized for execution work. This is a scripting language that gives an inbuilt library for machine learning, information mining, recommendations, mining and substantially more. It supports all data mining operations with the help orange library which is used for data mining task. Because of all these components, Python was utilized as an execution language. Orange python library was also used for data preprocessing and data categorization task. It was used for implementing the data mining algorithms which are useful for generating the recommendation for Elective Courses.

5.2 Data Mining using Orange

Orange3 is used for implementation purpose and it provides data visualization in an interactive manner and it is also available in the form of python library. Orange3 is visual programming software that provides services such as data visualization, data mining, and machine learning. It provides simplicity to the researcher to use its graphical interface for their research work. As it is a tool that contains widgets in its interface that takes the user input, process that data and simply provide output to the user without any hurdle as well as provides graphical results.

Figure 5.1 and Figure 5.2 demonstrates the trial setup and work flow used for the generation of Elective Courses Predictions.

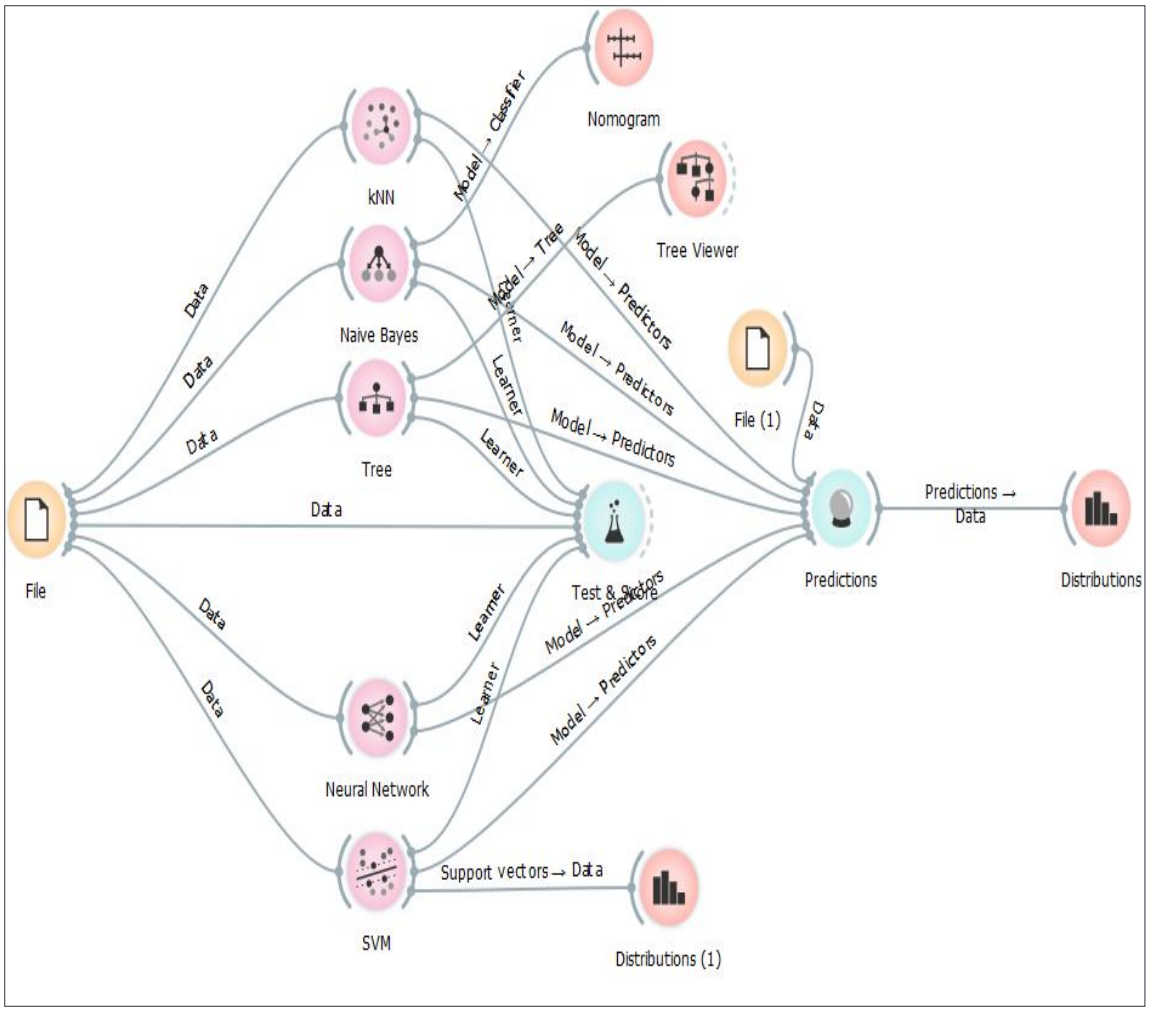

Figure 5.1: Experimental setup for Elective Course Predictions using Classification Algorithms

Figure 5.1 workflow comprises of the accompanying steps:

- Dataset file is uploaded into File Widget and target variable is fixed.
- Various Classification algorithms are being applied on the dataset by using their respective widgets.
- Results of Classification algorithms can be viewed in widgets attached to its Classification widget.
- At that point classifiers are validated and tested using Test Widget of Orange.
- Predictions are made using Prediction Widget of Orange, which can be viewed in Distributions Widget.
- Final results can be viewed in Distributions widget.

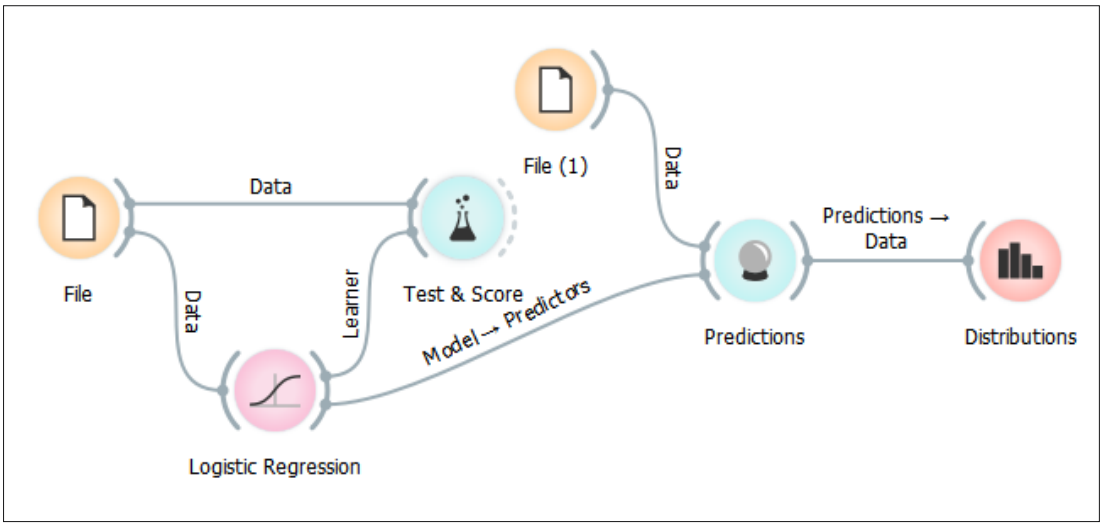

Figure 5.2: Experimental setup for Elective Course Predictions using Logistic Regression

Figure 5.2 workflow comprises of the accompanying steps:

- Dataset file is uploaded into File Widget and target variable is fixed.
- Logistic Regression algorithm is being applied on the dataset by using their respective widgets.
- At that point classifiers are validated and tested using Test Widget of Orange.
- Predictions are made using Prediction Widget of Orange, which can be viewed in Distributions Widget.
- Final results can be viewed in Distributions widget.

CHAPTER 6
RESULTS AND DISCUSSIONS

The final results of the predictions that were obtained with the help of different models, were stored in the form of excel sheet which was further processed as per the proposed algorithm. Predictions that were generated via models are expected to have good generalization capability and can efficiently produce a correct class label or categorization for the previously unknown data. Classification model's performance is evaluated on the basis of how many correct and incorrect predictions are made by the model on the testing dataset. The confusion matrix for a two-class classification problem is shown in Table 6.1.

TABLE 6.1: CONFUSION MATRIX FOR BINARY CLASSES

		Predicted Class	
		Class = 1	**Class = 0**
Actual Class	**Class = 1**	True Positive (TP)	False Negative (FN)
	Class = 0	False Positive (FP)	True Negative (TN)

For efficient comparison of these different data mining models used within this study, four different evaluation metrics were used for judging the quality of the predictions made, namely: accuracy, precision, recall and F1 score. Also, for training and testing purpose of each data mining model, 10-fold cross validation approach was used for statistical analysis of the datasets.

6.1 Accuracy

Accuracy is the measure of degree of closeness of the quantity's measurement to that quantity's true value. It can be calculated from the confusion matrix using the following equation:

$$Accuracy = (TP + TN) / (TP + TN + FP + FN)$$

where TP, FP, TN, FN being number of True Positive, False Positive, True Negative and False Negative, respectively.

The accuracy (in percentage) for different educational data mining models used, in case of different elective subjects is shown in Table 6.2 and Figure 6.1.

TABLE 6.2: ACCURACY IN PERCENTAGE FOR DIFFERENT MODELS

Accuracy	SVV	CC	CS	IoT	ML	NLP	AR_VR
Logistic Regression	77.2	76.3	75.0	77.1	77.4	76.9	76.5
Naïve Bayes	80.4	80.0	80.3	80.0	81.0	80.1	80.2
Neural Networks	87.0	87.1	87.0	87.3	87.6	86.9	87.3
SVM	88.7	88.6	87.7	87.4	89.1	88.2	88.3
Decision Tree	84.3	84.2	83.8	84.1	84.6	83.9	84
kNN	79.0	79.3	78.6	78.5	80.5	78.1	78.2

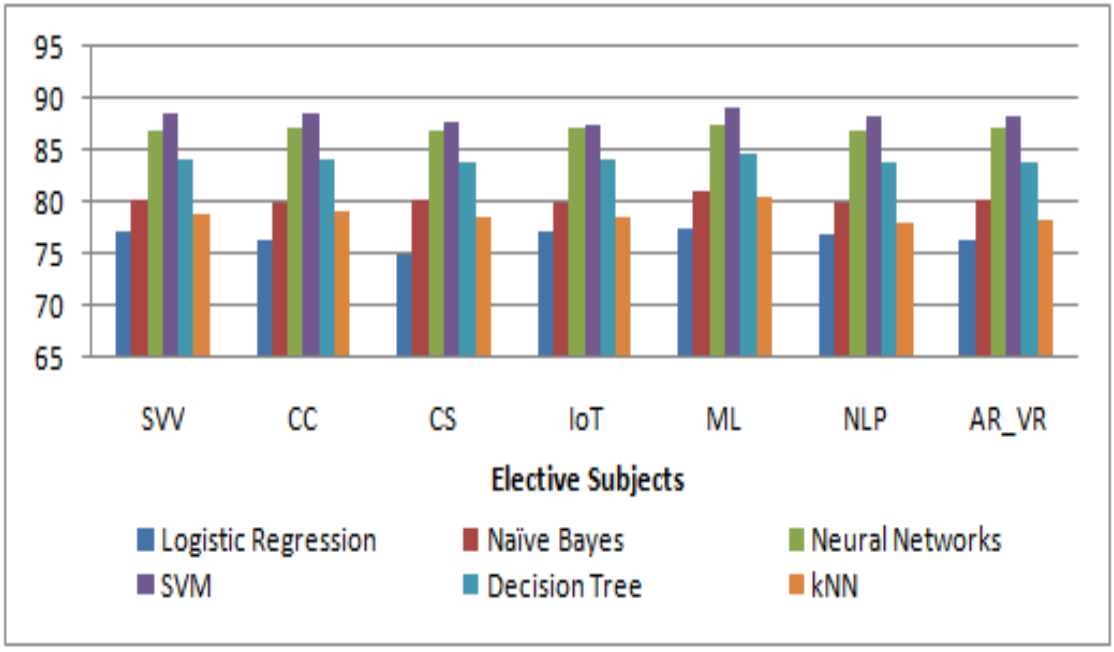

Figure 6.1: Comparison of Accuracy in Percentage for Different Models

6.2 Precision

It is the ratio of correctly predicted positive observations to the total predicted positive observations. It can be calculated from the confusion matrix using the following equation:

$$Precision = TP / (TP+FP)$$

where TP and FP stands for True Positive and False Positive respectively. The precision value ranging from a scale of 0 to 1 for different data models used, for different elective subjects is shown in Table 6.3 and Figure 6.2.

TABLE 6.3: PRECISION ON 0 TO 1 SCALE FOR DIFFERENT MODELS

Precision	SVV	CC	CS	IoT	ML	NLP	AR_VR
Logistic Regression	0.265	0.259	0.261	0.263	0.268	0.262	0.263
Naïve Bayes	0.006	0.006	0.006	0.005	0.007	0.006	0.006
Neural Networks	0.405	0.403	0.405	0.403	0.407	0.400	0.402
SVM	0.405	0.402	0.402	0.402	0.408	0.400	0.401
Decision Tree	0.397	0.395	0.391	0.395	0.400	0.392	0.392
kNN	0.358	0.352	0.352	0.352	0.36	0.352	0.355

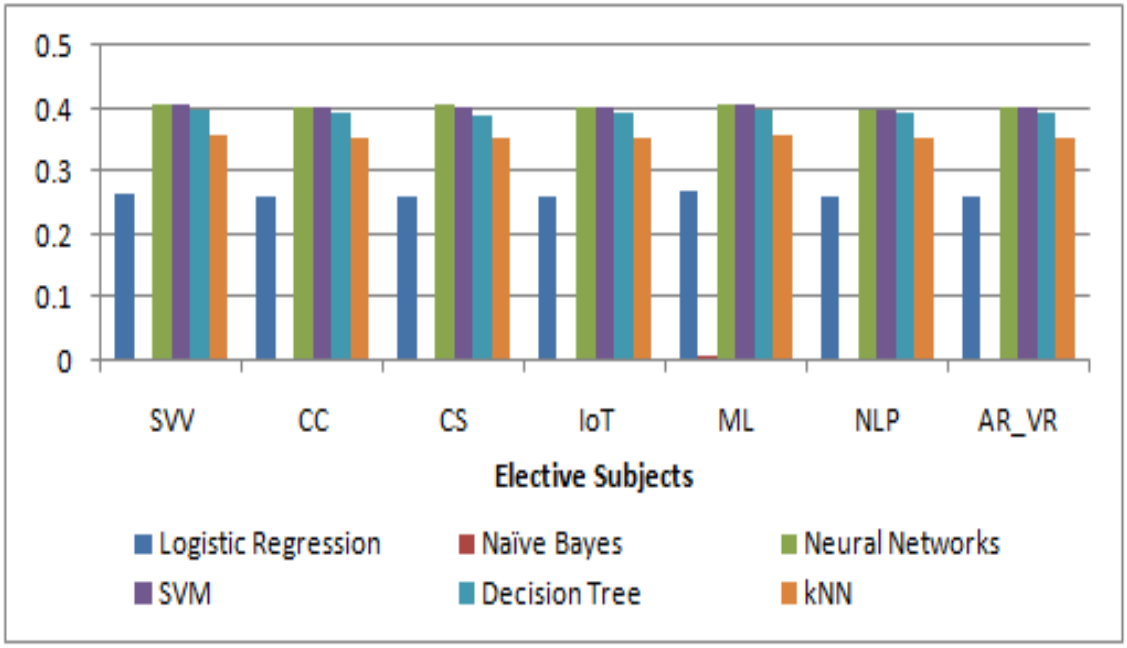

Figure 6.2 Comparison of Precision on 0 to 1 scale for Different Models

6.3 Recall

Recall is defined as the number of true positives over the number of true positives plus the number of false negatives. It can be calculated from the confusion matrix using the following equation:

$$Recall = TP / (TP+FN)$$

With TP and FN stands for True Positive and False Negative respectively. The recall value ranging from a scale of 0 to 1 for different data models used, for different elective subjects is shown in Table 6.4 and Figure 6.3.

TABLE 6.4: RECALL ON 0 TO 1 SCALE FOR DIFFERENT MODELS

Recall	SVV	CC	CS	IoT	ML	NLP	AR_VR
Logistic Regression	0.274	0.271	0.272	0.271	0.276	0.269	0.272
Naïve Bayes	0.016	0.016	0.016	0.016	0.017	0.016	0.016
Neural Networks	0.424	0.42	0.421	0.419	0.427	0.423	0.419
SVM	0.408	0.406	0.407	0.404	0.410	0.406	0.407
Decision Tree	0.390	0.389	0.386	0.383	0.400	0.387	0.384
kNN	0.342	0.340	0.339	0.341	0.345	0.339	0.341

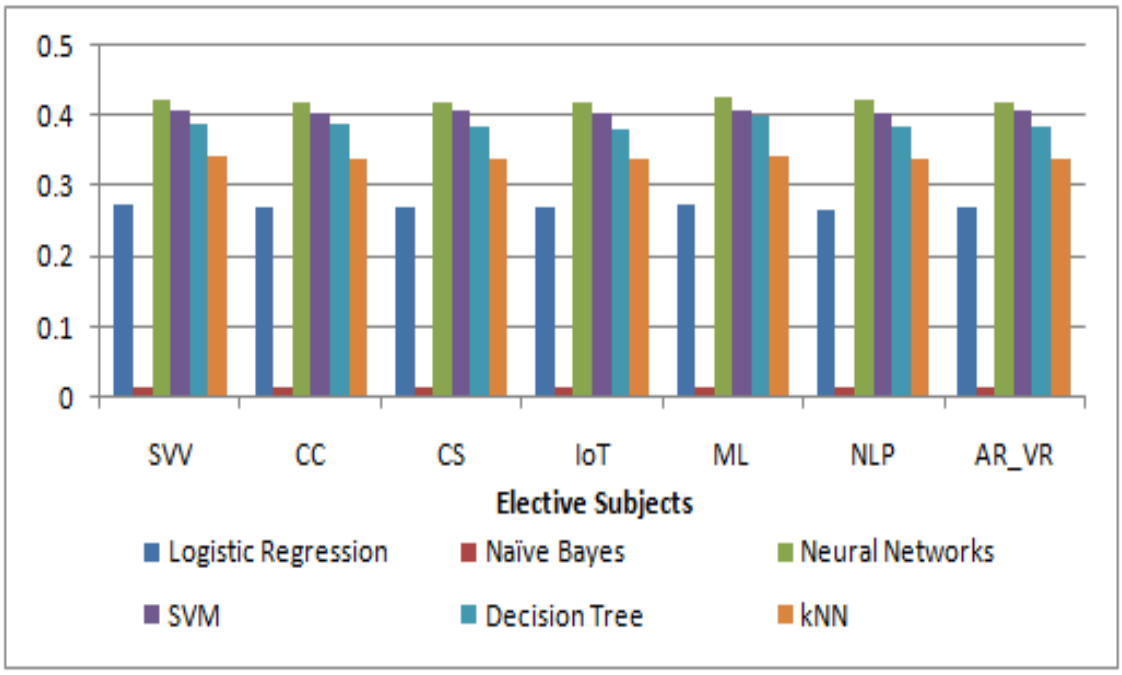

Figure 6.3 Comparison of Recall on 0 to 1 scale for Different Models

6.4 F1 Score

For binary classification, F1 score measures the test accuracy. It is the weighted harmonic mean of precision and recall. It is calculated by considering both recall and precision as given in the following equation:

$$F1\ Score = 2 / ((1/Recall)+(1/Precision))$$

The F1 score value ranging from a scale of 0 to 1 for different data models used, for different elective subjects is shown in Table 6.5 and Figure 6.4.

TABLE 6.5: F1 SCORE ON 0 TO 1 SCALE FOR DIFFERENT MODELS

F1 Score	SVV	CC	CS	IoT	ML	NLP	AR_VR
Logistic Regression	0.243	0.265	0.266	0.267	0.272	0.265	0.267
Naïve Bayes	0.008	0.009	0.009	0.008	0.010	0.009	0.009
Neural Networks	0.403	0.411	0.413	0.411	0.417	0.411	0.410
SVM	0.373	0.404	0.404	0.403	0.409	0.403	0.404
Decision Tree	0.394	0.392	0.388	0.389	0.400	0.389	0.388
kNN	0.350	0.346	0.345	0.346	0.352	0.345	0.348

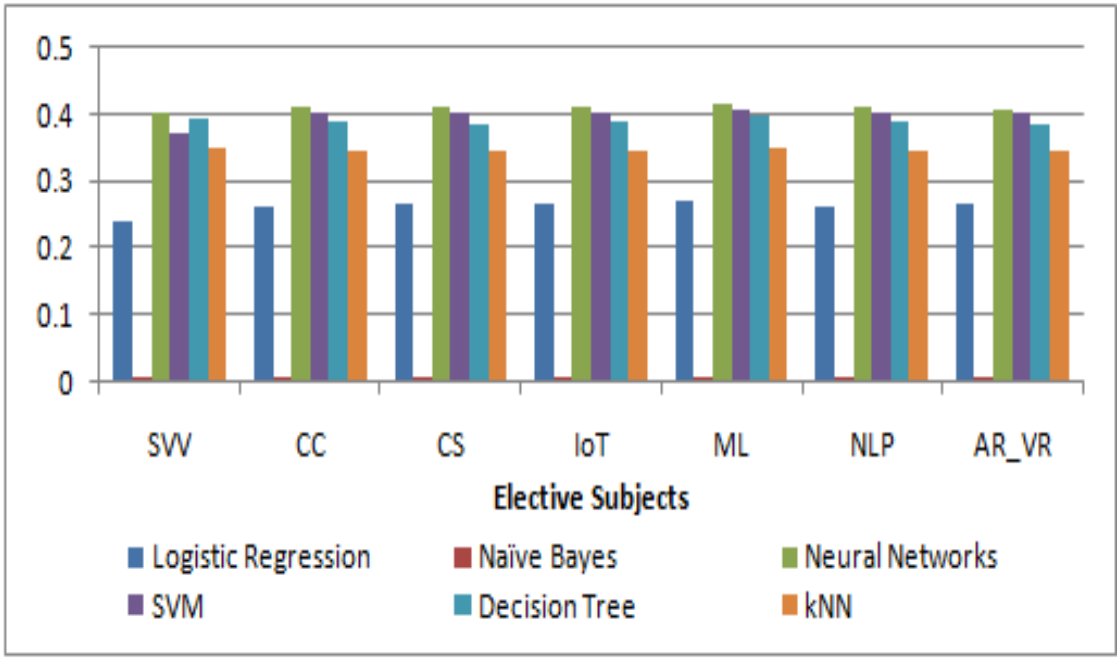

Figure 6.4 Comparison of F1 Score on 0 to 1 scale for Different Models

6.5 Discussion

After analyzing the different models used in this study on the performance evaluation parameters Support Vector Machines outperformed in terms of accuracy with an average accuracy of 88 percent across all the elective subjects. Support Vector Machines were ahead in terms of Accuracy whereas Neural Networks were ahead by fractions, in all other evaluation parameters, namely Precision, Recall and F1 Score. Except Support Vector Machines and Neural Networks that yielded somewhat comparable results, other models were far behind in terms of accuracy, precision, recall and F1 score. An exceptional behavior of Naïve Bayes which have exceptionally low values in precision, recall and F1 score parameters is also observed.

Naïve Bayes specifically focuses on the independence of the features and hence, this property of Naïve Bayes can self- explains the reason behind the lower values of precision, recall and F1 score.

Out of Neural Network and Support Vector Machines, Neural Network is said to have higher accuracy, but as already discussed, Neural Network requires large amount of data for the efficient training of its hidden layer nodes. However, here as the dataset considered is small, so this can be the reason why Support Vector Machines produced higher accuracy as compared to the Neural Networks, because Support Vector Machines act as extremely good classifiers on smaller datasets. Support Vector Machines with the help of its hyper-planes, yielded an accuracy of 88.7 percent in one of the elective subject datasets. However, considering other evaluation parameters of precision, recall and F1 score, Neural Network model can be said to have better adapted for the considered dataset and in case of large educational datasets it may in future, even yield better and promising results. A comparison of the proposed system with other similar or partially similar system and approaches are shown in Table 6.6.

TABLE 6.6: COMPARISON OF PROPOSED SYSTEM WITH OTHER APPROACHES

REFERENCE	DATASET	TECHNIQUES USED	RESULTS	LIMITATIONS
[1] Ray & Sharma	Real time data (255 students)	User-based and Item-based Collaborative Filtering	Predicted grades of students. (MAE=0.38)	Smaller Dataset Single Model Approach
[2] Arsad *et. al.*	Real time data (391 matriculation and 505 diploma students)	Neural Network	Predicting students' academic performance. matriculation students and MSE=0.0488 for diploma students)	Smaller Dataset Single Model Approach
[4] Romero *et. al.*	Real time data (438 students)	Decision tree, Rule induction algorithm, Neural Network, Statistical Classifier, Fuzzy Rule	Predicted students' final marks. (Decision Tree outperformed with Accuracy= 67.02%)	Smaller Dataset Less Attributes
[5] Osmanbegović & Suljić	Real time data (257 students)	Naïve Bayes, Multi Layer Perceptron (MLP), C4.5	Predicted success in a course. (Naïve Bayes outperformed with Accuracy= 76.65%)	Smaller Dataset Less distinction between attributes
[7] Guo *et. al.*	Real time data (43 course data)	Neural Networks	Identified most relevant factors in student courses (MSE=0.0016 and Correlation=0.943)	Very Small Dataset Single Model Approach

REFERENCE	DATASET	TECHNIQUES USED	RESULTS	LIMITATIONS
[8] Ramesh *et. al.*	Real time data (900 students)	Naïve Bayes, Multilayer Perception, SMO, J48, REPTree	Predicted grades of students. (MLP outperformed with Accuracy=72.4%)	Smaller Dataset Results based on psychometric factors
[10] Chamillard	Real time data (285 students)	Regression Analysis	Predicted student performance. (Correlation=0.579)	Smaller Dataset Single Model Approach
[11] Al-Badarenah *et. al.*	Real time data (2000 students)	Collaborative filtering, K-Means, Similarity distance measure.	Recommended elective courses. (Precision=0.95, Recall=0.5)	More Techniques can be used and experimented
[12] Cakmak	Real time data. (55475 course data)	User–based collaborative filtering, Item–based collaborative filtering, Similarity distance measure.	Predicted students' grades for recommending new elective courses. (MAE=0.32)	Single Model Approach
[14] Mueen *et. al.*	Real time data. (60 students)	Naïve Bayes, Neural Network, Decision tree	Predicted student academic performance. (Naïve Bayes outperformed with Accuracy=86%)	Very Small Dataset
Proposed System	Real time data (658 students)	Naïve Bayes, Neural Network, Decision tree, Logistic Regression, Support Vector Machines, k- Nearest Neighbor	Elective Course Recommendations by preserving student interests (Support Vector Machines outperformed with Accuracy=88.7%)	Smaller Dataset. Subject marks treated a independent observations

CHAPTER 7
CONCLUSION AND FUTURE SCOPE

7.1 Conclusion

In this study efforts were put to propose an efficient algorithm that assures academic success in the elective courses via its predictions as well as preserves the student subject preferences for greater achievement of bilateral academic quality learning outcomes. The distinction of this approach lies in the fact that it assures two-way success and takes care of institutional and student preferences at the same time. While considering the academic predictions, support vector machines were found to be the best predictor classifier model for predicting the marks in the respective elective subjects based on past academic score obtained in core subjects. Once the student subject preferences and subjects having highest academic success rate are identified, with the help of weighted ranks, the proposed algorithm helps to generate efficient elective course recommendations that can be used for assuring bilateral academic success of students as well as the institutions.

This research study has helped to fulfill the set objectives in the following ways:

- Support Vector Machines, were found to be the best predictor classifier model in the present education scenario that helped to make efficient course predictions.
- An algorithm is proposed that side-by-side considered the contextual information of individual student varying preferences.
- With the help of the predictions and individual student varying preferences, the proposed system was able to generate efficient elective course recommendations.

7.2 Future Scope

The proposed elective course recommender system can be extended by:

- Consideration of cross-domain large datasets from different educational institutions for better generalized results.
- Other contextual attributes like gender, age etc. can also be researched upon for incorporating more efficiency within the proposed algorithm for generating more efficient elective course recommendations.

REFERENCES

[1] S. Ray and A. Sharma, "A collaborative filtering based approach for recommending elective courses," in *International Conference on Information Intelligence Systems, Technology and Management*, Springer, pp. 330-339, 2011.

[2] P.M. Arsad and N. Buniyamin, "A neural network students' performance prediction model (NNSPPM)," in *IEEE International Conference on Smart Instrumentation, Measurement and Applications*, IEEE, pp. 1-5, 2013.

[3] A.M. Shahiri and W. Husain, "A review on predicting student's performance using data mining techniques," *Procedia Computer Science*, vol. 72, pp. 414-422, 2015.

[4] C. Romero, S. Ventura, P.G. Espejo and C. Hervás, "Data mining algorithms to classify students," in *Educational Data Mining 2008*, pp. 1-10, 2008.

[5] E. Osmanbegović and M. Suljić, "Data mining approach for predicting student performance," *Economic Review*, vol. 10, no. 1, pp. 3-12, 2012.

[6] K. Bunkar, U.K. Singh, B. Pandya and R. Bunkar, "Data mining: prediction for performance improvement of graduate students using classification," in *IEEE Ninth International Conference on Wireless and Optical Communications Networks*, IEEE, pp. 1-5, 2012.

[7] W.W. Guo, "Incorporating statistical and neural network approaches for student course satisfaction analysis and prediction," *Expert Systems with Applications*, vol. 37, no. 4, pp. 3358-3365, 2010.

[8] V. Ramesh, P. Parkavi and K. Ramar, "Predicting student performance: a statistical and data mining approach," *International Journal of Computer Applications*, vol. 63, no. 8, pp. 35-39, 2013.

[9] L.S. Affendey, I.H.M. Paris, N. Mustapha, M.N. Sulaiman and Z. Muda, "Ranking of influencing factors in predicting students' academic performance," *Information Technology Journal*, vol. 9, no. 4, pp. 832-837, 2010.

[10] A.T. Chamillard, "Using student performance predictions in a computer science curriculum," *ACM SIGCSE Bulletin*, vol. 38, no. 3, pp. 260-264, 2006.

[11] A. Al-Badarenah and A. Jamal, "An automated recommender system for course selection," *International Journal of Advanced Computer Science and Applications*, vol. 7, no. 3, pp.1166-175, 2016.

[12] A. Cakmak, "Predicting student success in courses via collaborative filtering," *International Journal of Intelligent Systems and Applications in Engineering*, vol. 5, no. 1, pp.10-17, 2017.

[13] T. Tran, H. Dang, V. Dinh, T. Truong, T. Vuong and X. Phan, "Performance prediction for students: a multi-strategy approach," *Cybernetics and Information Technologies*, vol. 17, no. 2, pp. 164-182, 2017.

[14] A. Mueen, B. Zafar and U. Manzoor, "Modeling and predicting students' academic performance using data mining techniques," *International Journal of Modern Education and Computer Science*, vol. 8, no. 11, pp. 36-42, 2016.

[15] H. Bydžovská, "A comparative analysis of techniques for predicting student performance," in *Proceedings of the 9th International Conference on Educational Data Mining*, pp. 306-311, 2016.

[16] M. Bienkowski , M. Feng and B. Means, "Enhancing teaching and learning through educational data mining and learning analytics: an issue brief", *U.S. Department of Education*, Office of Education Technology, vol. 1, pp. 1-57, 2012.

[17] N. Mishra and S. Silakari, "Predictive analytics: a survey, trends, applications, opportunities & challenges," *International Journal of Computer Science and Information Technologies*, vol. 3, no. 3, pp. 4434-4438, 2012.

[18] A. Felfernig, M. Jeran, G. Ninaus, F. Reinfrank, S. Reiterer and M. Stettinger, "Basic approaches in recommendation systems," in *Recommendation Systems in Software Engineering*, Springer, pp. 15-37, 2014.

[19] S. R. Gaddam, V. V. Phoha and K. S. Balagani, "K-Means + ID3: A novel method for supervised anomaly detection by cascading K-Means clustering and ID3 decision tree learning methods," *IEEE Transactions on Knowledge and Data Engineering*, vol. 19, no. 3, pp. 345-354, 2007.

[20] S.B. Aher and L.M.R.J. Lobo, "Combination of machine learning algorithms for recommendation of courses in E-Learning system based on historical data," *Knowledge-Based Systems*, vol. 51, pp. 1-14, 2013.

[21] O.R. Zaíane, "Building a recommender agent for e-learning systems," in *International Conference on Computers in Education*, pp. 55-59, 2002.

[22] D. Upendran, S. Chatterjee, S. Sindhumol and K. Bijlani, "Application of predictive analytics in intelligent course recommendation," *Procedia Computer Science*, vol. 93, pp. 917-923, 2016.

[23] F. Castro, A. Vellido, À. Nebot and F. Mugica, "Applying data mining techniques to e-learning problems," *Evolution of teaching and learning paradigms in intelligent environment*, Springer, pp. 183-221, 2007.

[24] A. Elbadrawy and G. Karypis, "Domain-aware grade prediction and top-n course recommendation," in *Proceedings of the 10th ACM Conference on Recommender Systems*, ACM, pp. 183-190, 2016.

[25] T. Denley, "How predictive analytics and choice architecture can improve student success," *Research & Practice in Assessment*, vol. 9, pp. 1-9, 2014.

[26] R. Farzan and P. Brusilovsky, "Social navigation support in a course recommendation system," in *International Conference on Adaptive Hypermedia and Adaptive Web-Based Systems*, pp. 91-100, 2006.

[27] N. Bendakir and E. Aïmeur, "Using association rules for course recommendation," in *Proceedings of the AAAI Workshop on Educational Data Mining*, vol. 3, pp. 1-10, 2006.

[28] G. Engin, B. Aksoyer, M. Avdagic, D .Bozanlı, U. Hanay, D. Maden and G. Ertek, "Rule-based expert systems for supporting university students," *Procedia Computer Science*, vol. 31, pp. 22-31, 2014.

[29] E. García, C. Romero, S. Ventura and C. De Castro, "An architecture for making recommendations to courseware authors using association rule mining and collaborative filtering," *User Modeling and User-Adapted Interaction*, vol. 19, no. 1-2, pp. 99-132, 2009.

[30] J. Demsar, T. Curk , A. Erjavec , M. Milutinovicˇ, and M. Mozina , "Orange: Data Mining Toolbox in Python," *The Journal of Machine Learning Research*, vol. 14, pp. 2349–2353, 2013.

[31] A. Gupta, "Applying Data Mining Techniques in Job Recommender System for Considering Candidate Job Preferences," M.E. thesis, Department of Computer Science, Thapar University, Patiala, India, 2014.

[32] D. Dua and E.K. Taniskidou, "UCI machine learning repository", *Irvine, CA: University of California, School of Information and Computer Science*. 2017. [Available Online: http://archive.ics.uci.edu/ml].

PUBLICATIONS

Communicated Paper

R. Verma, Anika, "Applying Predictive Analytics in Elective Course Recommender System while preserving Student Course Preferences," communicated in *6th IEEE International Conference on MOOCS, Innovation and Technology in Education*, MITE 2018.

PLAGIARISM REPORT

Turnitin Originality Report

Processed on: 09-Jul-2018 12:23 +0530
ID: 794817249
Word Count: 8946
Submitted: 2

Ridima_Thesis By Anika Gupta

Similarity Index	Similarity by Source	
7%	Internet Sources:	5%
	Publications:	6%
	Student Papers:	0%

Ridima_Thesis

ORIGINALITY REPORT

7%	5%	6%	0%
SIMILARITY INDEX	INTERNET SOURCES	PUBLICATIONS	STUDENT PAPERS

PRIMARY SOURCES

1	"Smart and Innovative Trends in Next Generation Computing Technologies", Springer Nature, 2018 Publication	<1%
2	Meena Jha, Sanjay Jha, Liam O'Brien. "chapter 15 Social Media and Big Data", IGI Global, 2017 Publication	<1%
3	epress.lib.uts.edu.au Internet Source	<1%
4	Alashwal, Hany Deris, Safaai Othmanb, Ra. "Comparison of domain and hydrophobicity features for the prediction of protein-protein interactions ", International Journal of Information Tec, Jan 2006 Issue Publication	<1%