Grade Prediction Using Supervised Machine Learning Techniques

Conference Paper · March 2016

2 authors:

Khurum Nazir Junejo
IBEX

Eman Eman
Technische Universität Kaiserslautern

GRADE PREDICTION USING SUPERVISED MACHINE LEARNING TECHNIQUES

Emaan Abdul Majeed
College of Computing and Information Sciences
PAF-Karachi Institute of Economics and Technology
emaan.amajeed@gmail.com

Khurum Nazir Junejo
Singapore University of Technology and Design, Singapore
Karachi Institute of Economics and Technology, Pakistan
Khurum_junejo@sudt.edu.sg

ABSTRACT

Data Mining and Machine Learning can provide invaluable insights to Educational Psychology and Learning Sciences. Predicting student's performance beforehand can help management, faculty and as well students to make timely decisions. Using machine learning approaches we predict student marks and grades before the final examination for 2500 student course records of an institute of higher education. This assessment helps the department to plan for the courses of the upcoming semester beforehand, and helps the students to make more informed decisions. Our findings suggest that the grade of more than 96% students can be accurately predicted even before the conduct of final exam.

Keywords: *Educational data mining, Machine Learning, Student performance predication, Grade prediction, Marks prediction.*

---

1. Introduction

There is an increasing trend towards using data mining and machine learning to aid educational learning, management, and assessment. This new emerging field, called Educational Data Mining, is concerned with developing methods that discovers knowledge from data originating from educational environments. One of the biggest challenges for modern higher education institutes is to analyze their performance in this competitive environment, and build a strategy for further development and actions. Most of the planning in any university is centered around the performance of the students. If student performance can be predicted beforehand, it could help all the stake holders to make better and timely decisions. We therefore highlight three major problems and challenges faced by management, faculty and students at the end of every semester.

Before the end of the semester, the management has to decide which courses to offer. Depending on this question, the management has to make time table, and engage visiting / part time faculty accordingly. Courses are then assigned to the faculty, so that they can prepare the course material. This process is unfortunately delayed because of the delay in the submission of the result of final exam by the faculty. This leads to last minute offering and dropping of courses, based on how many students passed or failed the course. This creates an atmosphere of uncertainty and frustration for management, faculty, and students alike.

On the other hand, a poor performing student is faced by a dilemma of his own; "would I clear the course or not". He wants to decide whether to drop the course before the final exam, or, take the exam and risk an "F" on his transcript. While the teacher is concerned about the poor performing students of his course and wants to counsel them. Timely counseling can help student to make an extra effort or make him drop a course to avoid going on probation.

A solution for these above mentioned three problems needs to be specialized for each student, but even more importantly should also consider which teacher is teaching the course. Some teachers are lenient and thus are prone to giving higher grades, while some teachers are quite strict and hence give poor grades. This behavior can be seen from Figure 1.

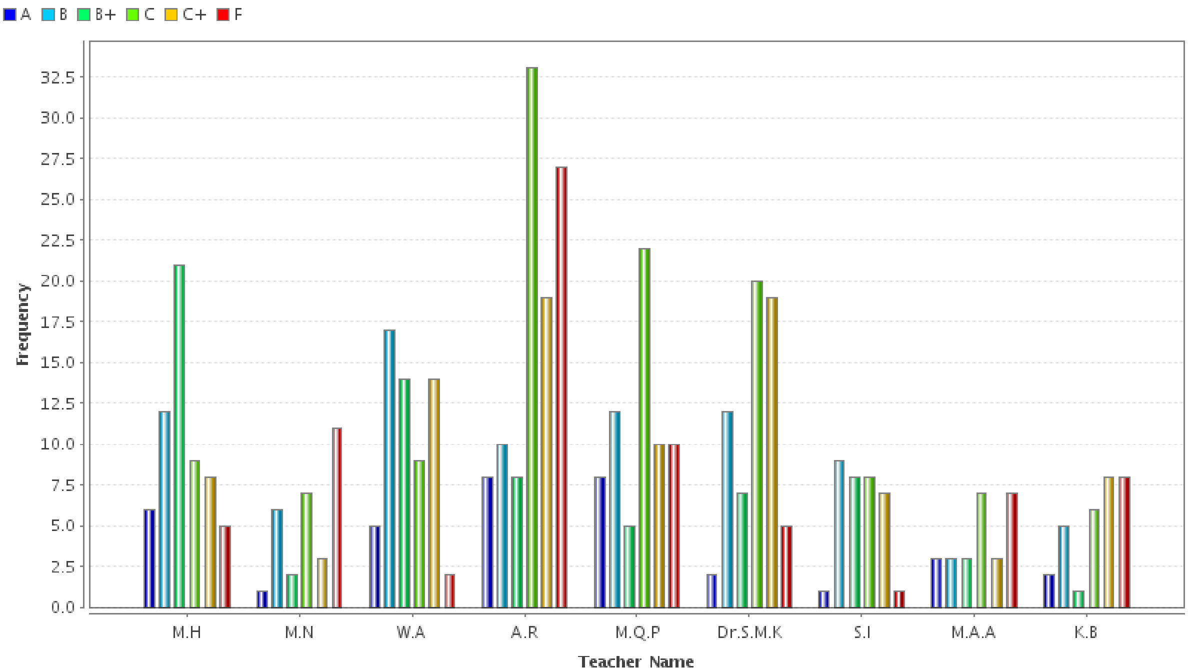

Figure 1: *Trend of grades by different teachers. Horizontal axis lists the abbreviation for faculty names, while the vertical axis represent the count of individual grades.*

Similarly some courses are easier to score in than others. If the students do not have a strong base in mathematics than they tend to perform badly in mathematics courses, for example, the grades in stochastic systems course tend to be significantly lower at most of the university even with different teachers. Figure 2 depicts this behavior e.g. about one third students fail the DLD (digital logic design) course while the failure ratio for DBMS (database management systems) is less than 3%. Similarly Figure 3 shows that the grades vary from department to departments as well.

We address the above mentioned problems by building a machine learning tool that automatically predicts the marks and grades of the students with very high accuracy even before the final exams have been conducted. It takes into consideration the teacher teaching the course and the type of the course as well. Hence it provides the management an estimate about how many students will not pass the course, hence decide on what courses to offer and what not. The student can see their expected grade and take a timely decision of taking the final exam or not. The faculty on the other hand can counsel students who they deem to be week.

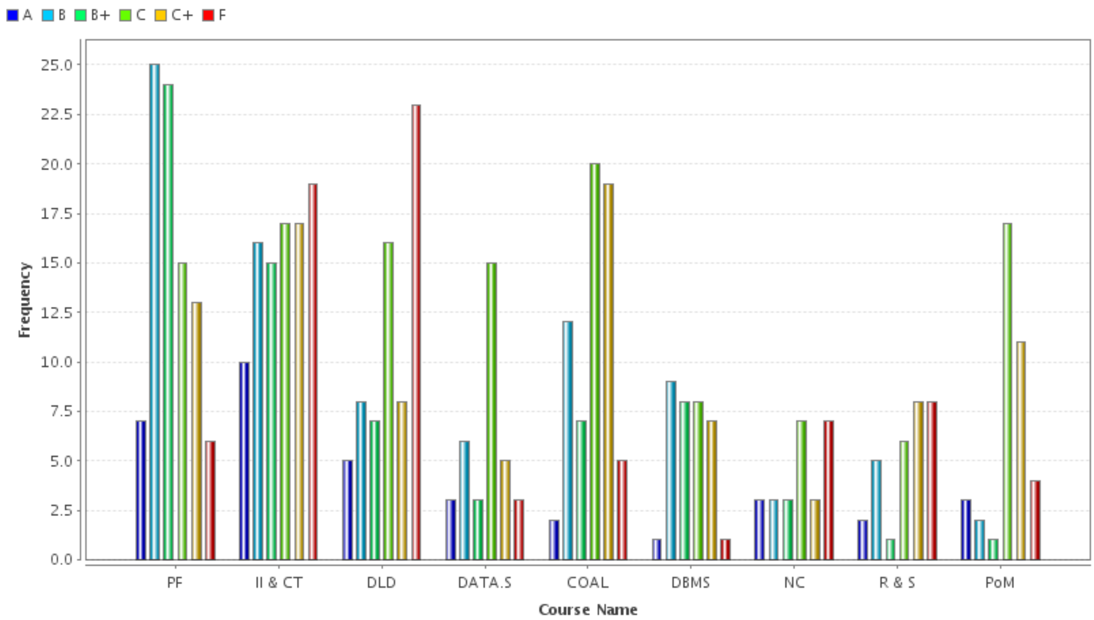

Figure 2: *Trend of grades for different courses. Horizontal axis lists the abbreviation for course names, while the vertical axis represent the count of individual grades*

We accomplish this task by first collecting student course records from College of Computing and Information Sciences of a degree awarding institution. After much preprocessing, the remaining 2500 records are used to train different machine learning algorithms such as naïve Bayes, ID3 and K-nearest neighbor classifier. We have been able to achieve an accuracy of more than 96% which is very promising.

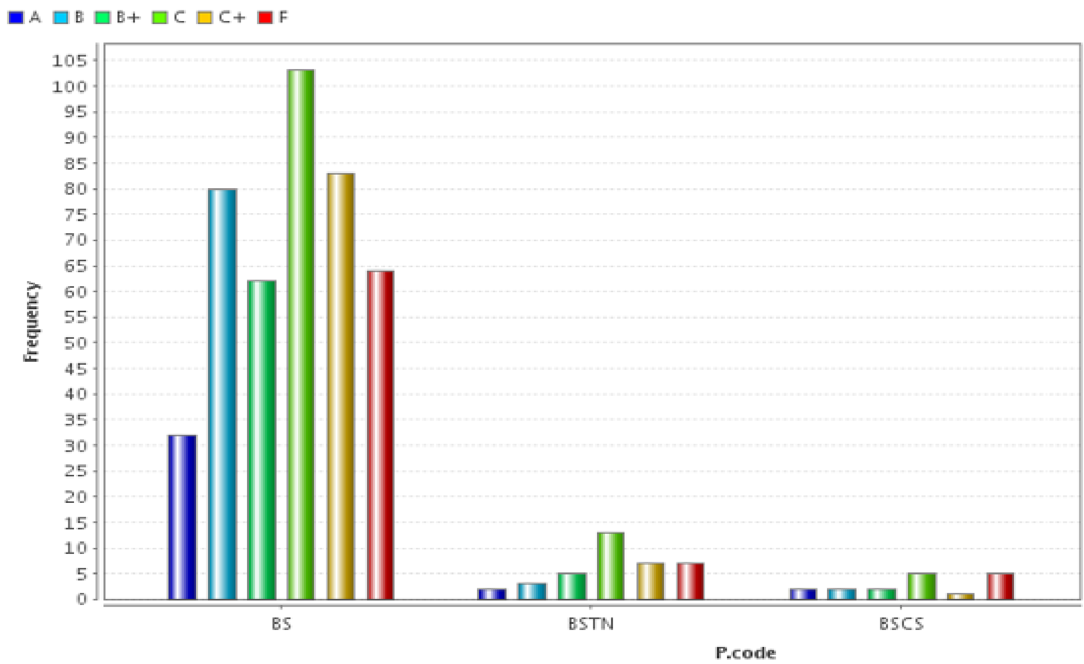

Figure 3: Trend of grades in different departments. Horizontal axis lists the abbreviation for different departments, while the vertical axis represent the count of individual grades.

The remaining section of this paper is organized as follows, review of the related research is provided in Section 2, the research methodology is described in Section 3, the obtained results and the comparative analysis are given in Section 4. We finally conclude in section 5.

2. Related Research

This section discusses the prior research on educational data mining techniques and student performance analysis where different researchers worked on similar or related issues, detail of some of the related work is follows:

(Miranda et al, 2013) conducted a study to identify the attributes impacting the student performance the most. They used only genetic algorithm on a sample of 120 students. Their analysis model considers quantitative factors such as theoretical, mathematical, practical, and other departmental marks. Their approach can be considered as a feature selection problem. We also perform this feature selection as a preprocessing step into our model. A similar study was done by (Brijesh and Pal, 2011). They applied only ID3 algorithm on a very small set of 50 students to identify the attributes that affected the performance of the students the most. They classify student's performance into poor, good, and very good but do not report any accuracy. Their results are biased as they used the training data for testing as well. Attributes like teachers performance were also left out.

(Khan, 2005) perform a study to identify the discriminating measures of cognition, personality and demographic variables for success at higher secondary level in science stream. They used a sample of 400 students (200 boys and 200 girls) of a secondary school. They applied clustering to identify these important measures. It was found that "girls with high socio-economic status had relatively higher academic achievement in science stream and boys with low socio-economic status had relatively higher academic achievement in general". (Kristjánsson, 2010) studied the relationship between academic performance and health behaviors, body mass index, and self-esteem of 6346 adolescents using machine learning approaches and found that "Lower BMI, physical activity, and good dietary habits were all associated with higher academic achievement; however, health behaviour was positively and robustly associated with greater self-esteem."

(Hijazi& Naqvi, 2006) used linear regression on a sample of 300 students (225 males, 75 females) from a group of colleges affiliated with University of Punjab. They used social and demographic variables alongside the grades of the students and found that mother's education and student's family income were highly correlated with the student academic performance. A similar finding was reported by (Brijesh& Pal, 2011) using naïve Bayes classification on a similar sample of 300 students.

(Alaa el-Halees, 2012) tried to predict the performance of more than 3000 students before the time of admission into self-defined labels of excellent, very good, good, fair and poor performance based on their gender, city, high school marks etc. They mainly used unsupervised machine learning approaches such as association rule mining, clustering, and outlier detection to identify low grade students. (Pandey& Pal, 2011) carried a similar investigation on a sample of 600 students by means of Bayes classification on category, language and background qualification. They tried to predict whether newly enrolled students will be able to perform well or not in the university.

(Ramaswami et al., 2012 & 2010) applied Bayesian belief networks and CHAID algorithms on 5650 and 772 students respectively. He tried to predict the performance of higher secondary school students based on 35 personal, social, and demographics attribute such as body mass index, eye's visual acuity, family size, family status, internet, health, etc. They divided the class attributes into pass, and fail; very good, good, and poor; and so on. Using a similar set of attributes, (Cortez& Silva, 2008) conducted a study on 778 Portuguese students using four different machine learning algorithm and found decision tree to be the best. They tried to predict whether a student would pass or fail Mathematics and Portuguese Language course or not.

(Pardos et al., 2007) predict 300 8th grade students within their intelligent tutoring system named ASSISTment. They attempt to predict the skill of students based on ASSISTment using Bayes nets. (Baker et al.,2010) perform a similar study but extend it further to on paper post-tests outside of the System. (Bekele et al., 2005, Beaumont& Soyibo., 2001) performed a similar study on Ethiopian and Jamaican students.

3. Data Preparation and Pre-processing

We have used the data collected from the College of Computing and Information Sciences (COCIS) of a well-known university. Student course records of eight semesters is obtained for the undergrad programs from Fall 2009 to Spring 2013 semester. For each semester, we have a record of 25 to 30 courses with an aggregate enrolment of 800 to 1000 students. We removed courses that were offered with very small enrollments, and by faculty members that only had one or two course offerings in the aforementioned period. In the first two weeks (i.e. the add/drop period), students tend to enroll in more courses than they intend to take. At the end of this period many students withdraw from their extra enrolled courses, some courses get dropped also. Some students withdraw from courses after the add drop period as well. We remove all of these withdrawals and course drop records from our data. We also removed courses with missing values e.g. some instructors did not conduct hourly (midterm) exams, or quizzes. After further preprocessing we are left with a big total of 2500 records.

Table 1 lists the different attributes and their data types that we used in our machine learning models. Grade is taken as the categorical class (or target) attribute. It is constructed from total marks of the students based on the university's grading policy. There are a total of six grades; "A", "B+", "B", "C+" ,"C" and "F". Their ranges are given in table 1. A sample view of the data is given in Figure 4.

Sno	Name	Pcode	HT M	AT M	CP M	FT M	Total M	Ceil	Grade
5691	Sherjeel Ha:	BS	27	19	9	33	88	88	A
5702	Maria Musht:	BSCS	28.500	17	9	36	90.500	91	A
5703	Hassan-Ud-	BSCS	28.500	18	9	34	89.500	90	A
5741	Syed Danial	BS	27.500	18	9	32	86.500	87	A
5759	Waqar Muha	BS	25	19	9	34	87	87	A
5763	Hafiz Muham	BS	27	19	8	34	88	88	A
5810	Mirza Danial	BSCS	29	19	10	34	92	92	A
5811	Hafiz Muham	BS	24	20	10	31	85	85	A
5821	Muhammad	BS	23	19	10	35	87	87	A
5864	Muhammad	BS	27	19	9	38	93	93	A
5865	Rida Kaynat	BE	27	19	9	35	90	90	A
5498	Muhamamd	BS	20	19	8	27	74	74	B
5707	Muhammad	BS	23	20	9	20	72	72	B

Figure 4. Sample view of the training data.

Attributes	Description	Type	Range
HT	Hourly Total	Numeric	0-30
QT	Quiz Total	Numeric	0-10
AT	Assignment Total	Numeric	0-5
CP	Class Participation	Numeric	0-5
LW	Lab Work	Numeric	0-10
C.ID	Course ID	Nominal	Nominal semesterwise
S.ID	Student ID	Nominal	Categorical
Grade	Class Variable used for predict students grade	Polynominal	{A > 85% , B+ > 78% & < 85% , B > 71 % & 77% , C+ > 65% & < 71% , C > 60 & <66 % , Fail < 60% }
P.Code	Program Code	Nominal	BS , BS(CS) ,BS (TN)
P.J	Project Marks	Numeric	0-20
T.N	Teacher Name	Polynominal	Text

Table 1: *Details of Data Set. Essentially, we try to learn a machine learning model that predicts the value of the Grade attribute based on the rest of the attributes given in table 1*

4. Result & Experiments

The main objective of the formulation and implementation of this model is to predict the target variable (also known as the class or output) variable (i.e. Grade) using rest of the attributes as input variables which are retained in the model. Before we delve in the techniques and results, it needs to be mentioned that it is unrealistic to get a hundred percent accuracy on this data set because the weightage of the final exam varies from 25% to 50% of the total marks, which is quite a high weightage. Some students perform significantly different on the final exam as compared to their performance before the final exam. Some get ill before the exam or had some emergency because of which they were not able to prepare for the exam, while some students put an extra effort and perform better than their past record. Since we cannot model these type of variables, therefore it is unrealistic to predict the final grade with 100% accuracy.

We apply four different machine learning algorithms, namely, ID3, K-Nearest Neighbor, Naïve Bayes, and Rule Induction method. The choice of these four classifier is pretty interesting as all of them use a different strategy for learning. We used the Rapid Miner Tool's (Rapid Miner) implementation of the above algorithms. Each classifier is evaluated using holdout approach i.e. we train the model on

80% of the data while we evaluate it on the remaining 20% which is referred to as the testing data or the unseen data. This is done to avoid over fitting, a phenomenon that arises when we evaluate the model on the same set of records that we used to learn it from.

4.1. Performance Measures

We used three performance measures to compare the performance of the classifiers, namely; accuracy, precision and recall. Accuracy of a system can be defined as the degree of closeness of the predicted value of a quantity to that quantity's actual value. In other words, it's the percentage of decisions that were correctly made by the classification system i.e. the percentage of times our classifier's predicted grade matched the actual grade of that student. A single accuracy value was calculated for all the grades, whereas precision and recall were calculated separately for each grade. For example consider Figure 5, showing a sample of the prediction made by one of the classifiers used. All items from row 1 till 12 are correct predictions except item number 11. The actual label was C+ but the classifier predicted it as C.

Row No.	S.N	GRADE	prediction(GRADE)	confidence(A)	confidence(B)	confidence(...	confidence(...	confidence(...	confide...	Name	P.code
1	6464	B	B	0	0.987	0.013	0	0	0	Kanwer Sha	BS
2	6465	C+	C+	0	0	0	0.012	0.976	0.012	Asad Iqbal	BS
3	6287	B	B	0	0.987	0.013	0	0	0	Abdul Sattar	BS
4	6356	B	B	0	0.987	0.013	0	0	0	Anas Mubas	BS
5	6271	B+	B+	0.061	0	0.924	0.015	0	0	Muhammad	BS
6	6406	B+	B+	0.061	0	0.924	0.015	0	0	Muhammad	BS
7	6421	C+	C+	0	0	0	0.012	0.976	0.012	Muhammad	BS
8	6405	B	B	0	0.987	0.013	0	0	0	Arif Shehzad	BSCS
9	6658	B	B	0	0.987	0.013	0	0	0	Hafiza Kulso	BS
10	6406	A	A	0.962	0.038	0	0	0	0	Muhammad	BS
11	6427	C+	C	0	0	0	0.981	0.019	0	Ayesha Jave	BS
12	6411	B	B	0	0.987	0.013	0	0	0	Rafaiy Sethi	BS

Figure 5. A sample of predictions made by the rule induction classifier

Precision for a class is defined as the number of items correctly labeled as belonging to a grade (true positives) divided by the total number of items predicted as belonging to that particular grade. Whereas recall is defined as the true positives (TP) divided by the total number of elements that actually belong to that particular grade. They can be more formally defined as:

$$Accuracy = \frac{TP + TN}{TP + FP + TN + FN}$$

$$Precision = \frac{TP}{TP + FP}$$

$$Recall = \frac{TP}{TP + FN}$$

Where FN are the number of false negatives i.e. the number of items that were predicted as not belonging to grade G and they actually did not belong to grade G. FP are the number of false positives i.e. the number of items that were predicted as grade G but actually do not belong to grade G, and FN are the number of false negatives i.e. the number of items that were predicted as not belonging to grade G when actually they did belong to it.

Accuracy, precision and recall vary between 0 and 100%. The maximum value of 100% accuracy means that the machine learning model made no mistakes in prediction. It is the opposite of error i.e. error is 100% - Accuracy. The maximum value of 100% precision for a grade G means that every item labeled as G does indeed have grade G. The maximum value of 100% recall means that every item from grade G was labeled as grade G.

4.2. Decision Tree Classifier (ID3)

Decision tree is one of the well known tree based algorithms for classification (Al-Radaideh et al., 2006, Cios et al., 1992). Decision tree can be visualized as a tree structure that consist of nodes from root to leaf, checks on attributes are set in the internal nodes and the class variable are shown in the leaf nodes. The selection of attributes is based on an information theoretic measure termed as Entropy. A record is traversed through the root node till the leaf, the class of the leaf node is the predicted label for that record. All decision trees can be translated to simple IF-ELSE, OR and AND rules which are intuitive and can be easily understood by users. The results for ID3 are presented in Table 2.

Accuracy: 94.98 %							
Predicted	**Actual**						
	A	**B+**	**B**	**C+**	**C**	**F**	**Precision**
A	36	4	7	0	0	1	75 %
B+	0	62	0	0	2	0	97 %
B	0	3	79	0	0	0	96 %
C+	0	0	0	87	1	1	98 %
C	0	0	0	3	117	0	96 %
F	0	0	0	1	1	74	97 %
Recall %	100 %	90 %	92 %	96 %	97 %	97 %	8+16

Table 2: Results of ID3 classifier.

4.3. k-NN Classifier

The K-Nearest Neighbor algorithm (k-NN) is the most widely used method for classifying records based on closest training records in the feature space (Zhang et al., 2007). It is a type of instance based learning or lazy learning approach, where the function is only approximated locally and all computation is deferred until classification. In k-NN, an object is classified by a majority vote of its neighbors, with the object being assigned to the class most common amongst its *k* nearest neighbors (where *k* is a positive integer, typically an odd number to avoid a tie in the voting). There is no universal best choice of *k,* it varies from data to data and is usually determined through hit and trial. Generally, larger values of *k* improve the accuracy by reducing the effect of noise on the classification, but it make boundaries between classes less distinct. We experimented with different values of *k* including, 1, 3, 5, 7 and 9. The best results are for *k* = 1 and 3 we present them in Table 3.

Accuracy: 87.89 %	
Predicted	**Actual**

	A	B+	B	C+	C	F	Precision
A	31	4	2	0	0	0	84%
B+	4	56	5	0	1	0	85%
B	1	7	73	4	1	0	85%
C+	0	0	5	78	7	1	86%
C	0	2	1	9	111	3	88%
F	0	0	0	0	1	72	99%
Recall %	100%	97%	97%	96%	98%	99%	30+28

Table 3: Results of *k*-NN classifier.

4.4. Bayesian Classifier

The most popular classifiers in pattern recognition and machine learning are Bayesian classifiers (Lewis & D.D, 1998). They are probabilistic classifiers that predict class by means of probabilities, such as the probability that a given sample belongs to a particular class or not. Bayesian networks and naive Bayes variants are the two most fundamental methods that belong to these classes of classifiers. We here are using naive Bayes classifier that assumes conditional independence given the value of the class. i.e. the attributes do not influence each other given the value of the class attribute. Bayesian classifiers are common classification algorithms due to their simplicity, computational efficiency and good performance for real-world problems. The achieved results are presented in Table 4.

Accuracy: 73.48 %							
Predicted	**Actual**						
	A	**B+**	**B**	**C+**	**C**	**F**	**Precision**
A	26	3	1	0	0	0	87%
B+	10	45	14	0	2	0	63%
B	0	21	56	12	0	0	62%
C+	0	0	11	53	6	0	76%
C	0	0	2	22	104	8	76%
F	0	0	2	4	9	68	82%
Recall %	72%	65%	65%	58%	86%	89%	**81+46**

Table 4: Results of Naïve Bayes Classifier

4.5. Rule Induction

Another learning algorithm used in our experiment is Rule induction. The Rule Induction classifier (Cohen& William, 1995) generates a one-level decision tree that expresses a set of rules that all test one particular attribute. It is similar to ID3 classifier discussed above except that ID3 algorithm can grow to n levels, where n is the number of attributes in the data. It is a simple, efficient method that often produces good rules with high accuracy. Its results are presented in Table 5.

Accuracy: 96.25 %							
Predicted	**Actual**						
	A	**B+**	**B**	**C+**	**C**	**F**	**Precision**
A	31	0	2	0	0	0	94%
B+	5	66	0	0	2	0	90%
B	0	3	84	0	0	0	96%
C+	0	0	0	88	1	1	98%
C	0	0	0	3	117	0	97%
F	0	0	0	0	1	75	99%
Recall %	86%	96%	98%	98%	97%	99%	**12+6**

Table 5: Results of Rule Induction classifier.

4.6. Discussion

Both the decision tree based classifier performed significantly better than the rest of the classifiers with Rule Induction taking the lead (refer to Figure 6). Not far behind is the ID3 classifier. The performance of NB has been disappointing even though it is one of the best classifiers for text classification and brain image classification. The classifier with the highest precision is the rule induction classifier, whereas *k*-NN has the highest recall. This is a bit surprising because *k*-NN accuracy is about 9% less than rule induction's accuracy. The reason for this is that *k*-NN is making one mistake more often than the other i.e. it is classifying more examples as FP as opposed to FN.

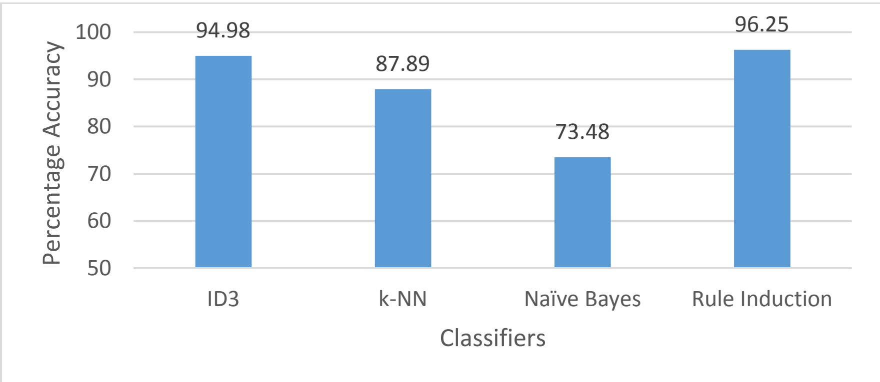

Figure 6. Classifier comparison based on accuracy.

An interesting point to note is that all the mistakes done by the classifier are not far away from the actual labels e.g. classification of a B as a F is a more severe mistake than classifying it as a A because F is more further than B as compared to A. E.g. in Table 5, out of the 86 students who actually had a B grade, two were mistakenly classified as A. Similarly B+ was predicted for 5 students whereas their actual grade was A. This is quite understandable as the student might have performed poorly in the final exam so he got one or two letter grades lower than his projected grade. This means that based on his/her performance before the final exam he/she was set for an A in the course but because of his poor performance on the final he actually ended up with a B.

Another interesting point is that the F grade is the most easiest to predict with the highest precision and recall average over the four classifiers, 94% and 96%, respectively. F grade is the most important grade to predict for the stake holders because this grade determines the enrollment size of the course offerings for the upcoming semesters. At second place is grade C with precision and recall of 89% and 95% respectively. B+ is the most difficult grade to predict at 84% and 87%, precision and recall respectively.

Lastly, total number of misclassifications made by the four classifiers combined are 227 out of which 131 (57%) are underestimates i.e. the values below the diagonal in Tables, 2, 3, 4, and 5. This implies that the number of students whose grade slip because of their performance on the final exam is more than the number of students who make an extra effort and improve their grade. Another explanation could be that the classifiers are biased towards giving a lower grade.

5. CONCLUSION

We have made a first attempt at predicting the final grade of students in undergraduate courses based on their performance in the same course prior to the final exam, type of the course and the teacher teaching that course. After applying various data preprocessing and machine learning algorithm on 2500 course records, we have achieved quite a remarkable prediction performance (varying between 73-96 % accuracy). The following useful conclusions have been drawn by this study:

Grade of more than 96% students can be accurately predicted even without the result of the final exam. This implies that the performance of students in various instruments before the final exam is highly correlated to their performance in the final exam. Another interesting point is that the F grade is the most easiest to predict followed by C grade, while B+ is the most difficult grade to predict. Furthermore, the number of students whose grade slip because of their performance on the final exam is more than those who improve it because of final exam. Lastly, rule induction classifier outperforms ID3 classifier on student performance prediction.

This prediction will help the management to plan in advance the courses that they are to offer in the upcoming semester more effectively thus making course allocation among faculty members less uncertain. This would help students in deciding whether to drop a course or not, and lastly, counselors can timely warn students if they are failing. In future we would like to extend our model to recommend the courses that would maximize his GPA taking into consideration the type of the course, teacher of the course and his past performance in such courses.

Acknowledgments

The first author would like to thank Dr. Khurram Junejo for his efforts and encouragement and College of Computing and Information Sciences of her alma mater for providing the data for this study.

References

T.Miranda Lakshmi1, A. Martin, V.Prasanna Venkatesan (2013). An Analysis of Students Performance Using Genetic Algorithm. Journal of Computer Sciences and Applications, Vol. 1, No. 4, 75-79.

Brijesh Kumar Broadway, Saurabh Pal (2011). Mining Educational Data to Analyze Students' Performance. International Journal of Advanced Computer Science and Applications. Vol. 2, No. 6.

Khan, Z. N. (2005). Scholastic achievement of higher secondary students in science stream. Journal of Social Sciences, 1(2), 84.

Kristjánsson, Á. L., Sigfúsdóttir, I. D., & Allegrante, J. P. (2010). Health behavior and academic achievement among adolescents: the relative contribution of dietary habits, physical activity, body mass index, and self-esteem. Health Education & Behavior, 37(1), 51-64.

Hijazi, S. T., and Naqvi, R.S.M.M. (2006). Factors Affecting Student's Performance: A Case of Private Colleges. Bangladesh e-Journal of Sociology. Vol. 3, No. 1.

Brijesh Kumar Baradwaj, Saurabh Pal (2011). Data Mining: A prediction for performance Improvement using classification. International Journal of Advanced Computer Science and Applications (IJACSA). Vol. 9, No.4.

Mohammed M. Abu Tair and Alaa M. El-Halees (2012). Mining Educational Data to Improve Students' Performance: A Case Study. International Journal of Information and Communication Technology Research, Vol.2 No. 2.

U.K. Pandey, and S. Pal (2011). Data Mining: A prediction of performer or underperformer using classification. International Journal of Computer Science and Information Technology. Vol. 2(2), pp.686-690.

Ramaswami, M. and Rathinasabapathy, R. (2012). Student Performance Prediction Modelling: A Bayesian Network Approach. International Journal of Computational Intelligence and Informatics, Vol. 1 : No. 4.

Ramaswami, M., & Bhaskaran, R. (2010). A CHAID Based Performance Prediction Model in Educational Data Mining. International Journal of Computer Science Issues (IJCSI), 7(4).

Cortez, P., & Silva, A. M. G. (2008). Using data mining to predict secondary school student performance.

Pardos, Z. A., Heffernan, N. T., Anderson, B., & Heffernan, C. L. (2007). The effect of model granularity on student performance prediction using Bayesian networks. In User Modeling 2007 (pp. 435-439). Springer Berlin Heidelberg.

Baker, R. S. D., Corbett, A. T., Gowda, S. M., Wagner, A. Z., MacLaren, B. A., Kauffman, L. R & Giguere, S. (2010). Contextual slip and prediction of student performance after use of an intelligent tutor. In User Modeling, Adaptation, and Personalization (pp. 52-63).

Bekele, R., & Menzel, W. (2005). A Bayesian approach to predict performance of a student (BAPPS): A Case with Ethiopian Students. Algorithms, 22(23), 24.

Beaumont-Walters, Y., & Soyibo, K. (2001). An analysis of high school students' performance on five integrated science process skills. Research in Science & Technological Education, 19(2), 133-145.

Rapid Miner tool at https://rapidminer.com/

Al-Radaideh, Q. A., Al-Shawakfa, E.M., and Al-Najjar, M. I. (2006). Mining Student Data using Decision Trees". International Arab Conference on Information Technology.

Cios, K. J., & Liu, N. (1992). A machine learning method for generation of a neural network architecture: A continuous ID3 algorithm. Neural Networks, IEEE Transactions on, 3(2), 280-291.

Zhang, M. L., & Zhou, Z. H. (2007). ML-KNN: A lazy learning approach to multi-label learning. Pattern recognition, 40(7), 2038-2048.

Lewis, D. D. (1998). Naive (Bayes) at forty: The independence assumption in information retrieval. In European Conference on Machine learning: ECML.

Cohen, William W. (1995). Fast Effective Rule Induction. Proceedings of the Twelfth International Conference on Machine Learning (ICML).