

Predicting Student Performance using Advanced Learning Analytics

Ali Daud ^{a,d}

^a Faculty of Computing and
Information Technology,
King Abdulaziz University,
Jeddah, Saudi Arabia
adfmohamad@kau.edu.sa

Naif Radi Aljohani ^a

^a Faculty of Computing and
Information Technology,
King Abdulaziz University,
Jeddah, Saudi Arabia
nraljohani@kau.edu.sa

Rabeeh Ayaz Abbasi ^{a,b}

^b Department of Computer Sciences,
Quaid-i-Azam University,
Islamabad, Pakistan
rabbasi@kau.edu.sa

Miltiadis D. Lytras ^c

^c Computer Information Systems
Department, The American College
of Greece, Greece
mlytras@acg.edu

Farhat Abbas ^d

^d Department of Computer Science
and Software Engineering,
International Islamic University,
Islamabad, Pakistan
farhatabbas421@gmail.com

Jalal S. Alowibdi ^e

^e Faculty of Computing and
Information Technology, University
of Jeddah, Saudi Arabia
jalowibdi@uj.edu.sa

ABSTRACT

Educational Data Mining (EDM) and Learning Analytics (LA) research have emerged as interesting areas of research, which are unfolding useful knowledge from educational databases for many purposes such as predicting students' success. The ability to predict a student's performance can be beneficial for actions in modern educational systems. Existing methods have used features which are mostly related to academic performance, family income and family assets; while features belonging to family expenditures and students' personal information are usually ignored. In this paper, an effort is made to investigate aforementioned feature sets by collecting the scholarship holding students' data from different universities of Pakistan. Learning analytics, discriminative and generative classification models are applied to predict whether a student will be able to complete his degree or not. Experimental results show that proposed method significantly outperforms existing methods due to exploitation of family expenditures and students' personal information feature sets. Outcomes of this EDM/LA research can serve as policy improvement method in higher education.

CCS CONCEPTS

• Computing methodologies~Supervised learning by classification • Applied computing~Education

KEYWORDS

Learning Analytics (LA); Educational Data Mining (EDM); Student Performance Prediction; Family Expenditures; Students Personal Information

1. INTRODUCTION

Students are the main stakeholders of institutions/universities and their performance plays a significant role in a country's social and economic growth by producing creative graduates, innovators and entrepreneurs [26]. Educational Data mining has emerged as very important area of research to reveal presentable and applicable knowledge from large educational data repositories. Data mining algorithms are used to obtain the hidden information and desired benefits from these large data repositories [17]. There is a critical demand for academic institutions to maintain and to integrate large datasets of learners for multipurpose decision making. The use of web technology has also become an integral part of the current era of education in many universities, increasing the actual amount of data about students, teachers and their interactions with learning and educational systems [7,15].

Higher education plays an important role in the development of a society. It is a field which provides a large amount of data about participants such as students, teachers, facilities and curricula [21]. The performance of students is a main concern of various stakeholders including educators, administrators and corporations. For recruiting fresh graduates, academic achievement is the main factor considered by the recruiting agencies. Therefore, graduates have to work hard for excellent grades, so that they may rise up to the expectations of recruiting agencies [26].

The sources of educational data may be broadly divided into two categories. The first category comprises centralized educational systems such as LMS; 'centralized' here means that the educational data in the analytics come from one source [4]. The second comprises the de-centralized educational data retrieved from different systems' resources, such as WWW data, massive open online courses that make use of many systems to deliver the learning materials [5,18].

Recently, analysis of educational data, for instance learning analytics, academic analytics, educational data mining, predictive analytics and learners' analytics has emerged as an innovative area of research [22]. The commonality between all of these terms is the use of educational data for multiple purposes. Recently, a new term has been introduced: 'educational data science' that clarifies how different disciplines and researchers with different research interests and backgrounds can work in this area [15].

© 2017 International World Wide Web Conference Committee (IW3C2),
published under Creative Commons CC BY 4.0 License.
WWW 2017 Companion, April 3-7, 2017, Perth, Australia.
ACM 978-1-4503-4914-7/17/04.

DOI: <http://dx.doi.org/10.1145/3041021.3054164>



Educators, parents and institutions would like to know the answer of the question “Is it possible to predict the performance of a student that is enrolled in an educational institution?” For example, will he complete his degree or not? However, currently the process of learning has been declared as an individual’s effort. Therefore, developing models for evaluating learning efforts of a student is not an easy task [16]. Recently, data mining techniques have been used to provide new insights for this problem. There are several diversified influential factors to evaluate the students’ performance. These can be identified by using data mining approaches in educational sector [17]. Data mining for an educational system is an iterative process for hypotheses development and testing. Fig. 1 shows the applications of data mining in educational systems. Student performance evaluation system can help in decision making for awarding scholarships or in other words targeting the right students.

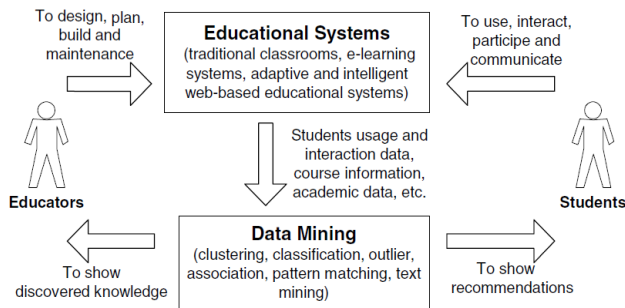


Figure 1 The cycle of applying Data Mining in educational systems [17]

In the past, mostly student performance is predicted by using different types of feature sets, such as, academic record, family income and family assets [16,25,12]. Family income and expenditure feature sets play an important role in student performance prediction. Intuitively family expenditures and personal information feature sets seem equally effective for the said task.

This paper investigates several family expenditure and personal information related feature sets based on learning analytics for improved student performance prediction. Extensive experimentations are conducted to evaluate the impact of existing, proposed and hybrid feature sets. Effectiveness of proposed features sets on real data of scholarship holding students from different Pakistani universities is provided by using both discriminative and generative classification models. Using proposed features outperforms existing methods and 86% accuracy is achieved in predicting that the student will complete the degree or not. According to the best of our knowledge the proposed features are not exploited before and they have a significant impact on the students’ performance in studies. Our method can be used also as a benchmark for similar studies in the future.

The main contributions of this paper are as follows:

1. In this paper, two new feature sets family expenditures and student personal information are investigated.
2. An effective feature set of twenty-three features is constructed by combining proposed features along with existing features. The feature subset selection process is adapted by using information gain and gain ratio metrics.

The impact of these features is determined as per their efficiency and most influential features are shortlisted.

3. The performance of standard models (both discriminative and generative) is analyzed by comprehensive experiments with baseline and proposed features.

The overall finding is that Learning Analytics based on personalized features can improve prediction of students’ performance. For sure the generalization of these findings requires additional research for the incorporation of additional features like talent, skills and personal competencies from different online web sources.

Rest of the paper is arranged as follows: Section 2 describes the related work. The problem definition is presented in Section 3. Section 4 provides applied models, performance measures, data collection and construction of feature space. Section 5 presents experiments conducted by using five classifiers with baseline and proposed feature sets. Finally, section 6 concludes this work.

2. RELATED WORK

The research problem of students’ performance prediction can be analyzed through diverse angles. In the current literature, a number of complimentary approaches provide a baseline for such an analysis. In an ideal scenario, a rich dataset with student identity along with numerous characteristics could be the basis for advanced learning analytics. The problem is that in most of the cases, not all the data are available for the dynamic construction of the student identity, further limited by lack of access to various sources. In Fig. 2, we briefly present some of the most representative methods of Applied Educational Data Mining and Learning Analytics based on a comprehensive literature review.

Student performance prediction has got a lot of attention from the educational data mining researchers. Typical data mining methods have been employed to deal with different tasks related to the students. A survey of data mining techniques for traditional educational systems such as adaptive web-based and content management systems is presented in [16].

An association rule based mining method is applied for selection of weak students in a school and is found effective [8]. Genetic Algorithm is used to assign the weights for the modeling of students’ grade for three levels (binary, 3-level and 9-level) [9]. It shows that the combination of multiple classifiers leads to a significant improvement in classification. A model is proposed for predicting student performance using six machine learning techniques for distance learning education, which is quite different from the traditional educational system [6]. The experimental results show that demographic and performance features are better predictors for predicting student performance. A regression model is applied to predict the test score of subject for school students [14]. It concludes that mixed-effect models present best performance as compared to Bayesian network.

A prediction model (CHAID) is developed to predict the performance of higher secondary school students, which is critical before getting admission into universities [16].

The grades of graduate students are predicted using Naïve Bayesian and Rule Induction classifiers [25]. Clusters are made from students’ data and the outliers are successfully identified. A model is presented to estimate the abilities of students and competence of teachers in order to predict the future student outcomes [19]. It shows that demographic profiles and personality traits features are correlated and have high impact on student performance. Similarly student performance evaluation and engineering students’ abilities are analyzed for improved recruitment process by using data mining methods [12].

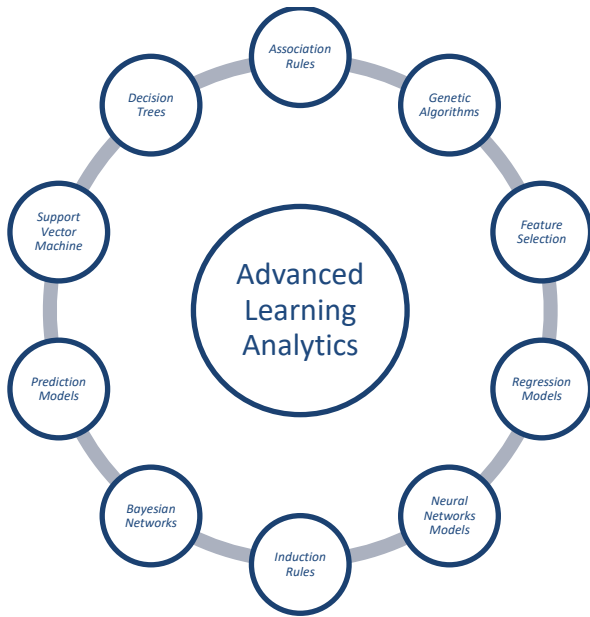


Figure 2 Overview of Data mining methods for Advanced Learning Analytics

A comparison of self-regulation and self-discipline measures of students is provided using hierarchical regression analysis and it shows that SR composite is more effective than SD composite [27]. A prediction model is presented to forecast the Student Academic Performance (SAP) of undergraduate engineering students [20]. An investigation of student performance is made through a longitudinal study [3].

A novel approach [13] predicts the number of times a student repeats a course. It uses neural networks to find relationship between some attributes of students for course assessment. The prediction of grade and prediction of approval/failure problems of students is addressed by [24]. The employability of master level students is predicted by [10] and draws the conclusion that empathy, drive and stress management abilities are major emotional parameters for employability. All the previous studies exploit a number of features related to students. We decide to use two additional characteristics in order to provide more meaningful insights for the association of students' performance and advanced educational decision making.

This paper addresses the problem of students' performance prediction by presenting new features, mostly related to family expenditure and student personal information. The researchers have used some basic characteristics related to the student personal information like family income and family assets information. Therefore, it is required to introduce some influential and effective features of students for performance evaluations in their studies. In this paper, family expenditures (electricity, telephone, gas bills, accommodation and medical) and student personal information features (e.g. self-employed and marital status) are explored. One of the greatest challenges for future Digital Learning Research in WWW is to investigate flexible and reliable methods for the extraction and integration of learners' data from diverse sources in order to support advanced educational decision making.

3. PROBLEM DEFINITION

The formal definition of student performance prediction problem is described as:

Given n training samples $(X_1, z_1), (X_2, z_2) \dots (X_n, z_n)$, where X_i is the feature vector for student a_i and A is the set of n students where $A = \{a_1, a_2, a_3 \dots a_n\}$. The $X_i \in R^m$ and m is the total number of features and z_i is the student performance status (degree completed or dropped) where $z_i \in \{-1, +1\}$. To predict the performance of a student, the following prediction function is proposed:

$$z = F(A/X) \quad (1)$$

Where,

$$F(A/X) = \begin{cases} \geq 0 & \text{if } z = +1, \text{ completed} \\ < 0 & \text{if } z = -1, \text{ dropped} \end{cases} \quad (2)$$

Learning Task: Goal is to learn a predictive function $\hat{F}(\cdot)$ or alternatively to predict whether a student will complete his/her degree or not. It is written as:

$$\hat{z} = \hat{F}(A/X) \quad (3)$$

4 METHODOLOGY

4.1 Models

Two types of classification models (discriminative and generative) are used to learning the desired predictive function $\hat{F}(\cdot)$. Two generative and three discriminative models are used for experimental analysis. They are selected on the basis of their frequent usage in the existing literature. The list of methods are as follows:

1. Support Vector Machine (SVM) [discriminative]
2. C4.5 [discriminative]
3. Classification and Regression Tree (CART) [discriminative]
4. Bayes Network (BN) [generative]
5. Naive Bayes (NB) [generative]

4.2 Performance Evaluation

For performance evaluation, three standard evaluation metrics (precision, recall and F1-score) are used. 5-fold cross validation is used for comparison with baseline methods. These performance evaluation parameters are defined as:

$$\text{Precision} = \frac{TP}{TP + FP} \quad (4)$$

$$\text{Recall} = \text{sensitivity} = \frac{TP}{(TP + FN)} \quad (5)$$

$$\text{F1 score} = 2 \cdot \frac{\text{precision} \cdot \text{recall}}{\text{precision} + \text{recall}} \quad (6)$$

4.3 Data Collection

For experimental purpose, the data of graduate and undergraduate students has been collected from different universities of Pakistan during the period (2004 to 2011). Initially, about 3000 student records were collected. Pre-processing is applied to obtain the most relevant characteristics of students. After removing inconsistencies and duplications in the dataset, we considered 776 student instances for experiments. The main goal of this research is to predict the student's performance i.e. "will he/she complete his/her degree or will he/she drop". The dataset consists of 690 instances of students, who have completed their academic degrees

(true values) and 86 instances that are dropped in the midway or at the end (false values). In the first step, 20 students' instances have been used (10 completed, 10 dropped), and then 40, 60, 80 and 100 instances of dataset are used. However, after 100 records/tuples, the performance remains unchanged as with the increase in instances in the data set. So, 100 instances of students (50 completed, 50 dropped) for experimental setup are selected. The distribution of the dataset is presented pictorially in Fig. 3.

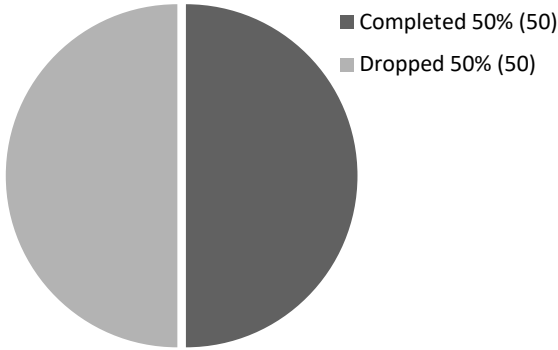


Figure 3 Characteristics of dataset.

4.4 Construction of Feature Space

Feature set is constructed by considering four categories of characteristics related to student and his family. Initially, a pool of 33 features is constructed by combining some existing (baseline) and proposed features and then feature subset selection process is applied to remove/reduce the number of redundant features. Information Gain and Gain ratio are used to select the best feature subset.

Overall, four categories of features are collected (some from literature and some are proposed in this research work). Out of twenty-three, thirteen are our proposed features. Table 1 presents the description of each feature, its category and status (proposed or existing/old). Then feature subset selection process is adapted in the following manner: First of all, for comprehensive analysis of features' comparison and best features selection, two measures, information gain and gain ratio are used. A threshold of 0.01 is used to identify the best feature subset. Finally, we get 23 features in which 13 are new (proposed) and 10 are old as shown in Table 1. We found larger information gain values of selected features as compared to gain ratio values. Because our dataset does not contain equal number of samples for both classes and information gain is biased towards maximum attribute values.

5. EXPERIMENTAL RESULTS

In this section, comprehensive experiments are presented using data set that is designed based on student's information acquired from different universities of Pakistan as described in Section 4.3. The dataset consists of 100 student records (tuples) and 23 features. Therefore, we get a 100×23 feature matrix. Default parameters are used for all classifiers using Weka 3.7.

Five-fold cross validation method is used to evaluate the accuracy of all the classifiers. Experiments are conducted in two ways: Firstly, influence of individual feature for the student's performance is analyzed. Secondly results of classifier on baseline methods and proposed feature sets are evaluated.

5.1 Individual Feature Analysis

This section evaluates the impact of each feature for the prediction of student's performance. Twenty-three features (selected by the feature extraction process) are selected for experiments. In experiments five classifiers are used (BN, NB, SVM, C4.5 and CART) to analyze the influence of each feature for predicting the performance of students as shown in Fig. 4.

We find the "natural gas" expenditure is the best predictor for the desired student's performance using C4.5 classification method as shown in Fig. 4. BN and NB methods show second and third highest F1- scores using same features. Other family expenditure features also play important roles. The "Stock Value" feature has the lowest performance for prediction and all classifiers present same F1-score (0.333). By analyzing performance of best and worst features, that conclude the proposed proposition based on the family expenditure feature set improves classification accuracy.

"Self Employed" is found to be the second-best feature that also belongs to proposed feature set of student's personal information and all classifiers show 0.77 F1-score which represents better performance by using proposed feature. The third best feature is "Location" which belongs to the baseline feature set. If we critically analyze the impacts of other proposed features in comparison with old features, better accuracy is obtained by using our proposed feature space as compared to the old feature space. Hence, it can be concluded that students' "natural gas" expenditure, "electricity" expenditure, "self-employed" and "location" characteristics are most influential for prediction of his/her performance in academics.

5.2 Comparisons

Performance of classifiers is analyzed using four baseline methods and our proposed feature sets based method and results are critically analyzed. The performance of experiments is evaluated by F1 score. The purpose of this experiment is to analyze the influence of proposed and existing features based methods for the student performance prediction task.

The feature sets proposed by [16,12,25,23] are considered as baseline for comparison as shown in Fig 5. Proposed method significantly outperforms baseline methods as shown in Fig. 5. SVM performs best for our proposed feature sets with F1 score of 0.867, which is 13% more as compared to second best method for SVM model. BN and NB classifiers overall perform better in case of most methods as compared to C4.5 and CART. For C4.5 model the performance of most methods is very low and unstable.

5.3 Discussion

This research work presents the student academic prediction methods that use four different types of features namely: family expenditure, family income, student personal information and family assets. It also adapts the process of feature subset selection in order to identify the most effective determinants for student academic performance prediction. It is evident from the comparative analysis that our proposed features are important predictors and achieved F1-score of 86% (Fig. 5) on real life undergraduate students' data.

Table 1 Features Distribution.

Category	Name	Description	Status	Info. Gain	Gain Ratio	Features Used
Family Expenditure	Electricity Bill	Average of Electricity bills for last six months	New	0.38	0.05	✓
	Natural Gas Bill	Average of Gas bills for last six months	New	0.26	0.06	✓
	Telephone Bill	Average of Telephone bills for last six months	New	0.10	0.04	✓
	Water Bill	Average of Water bills for last six months	New	0.10	0.06	✓
	Food Expenses	Average of food expenses for last six months	New	0.09	0.03	✓
	Miscellaneous Expenditure	Average of Miscellaneous Expenditures for last six months	New	0.11	0.02	✓
	Medical	Average of Medical Expenditures for last six months	New	0.06	0.01	✓
	Family Expenditure on Education	Average of Family Expenditure on education for last six months	New	0.35	0.04	✓
	Accommodation Expenses	Average of Accommodation Expenses for last six months	New	0.27	0.25	✓
	Studying Family Members	Total number of studying family members of student	Old	0.008	0.003	
	Dependent Family Member	Total number of dependent family members of student	Old	0.02	0.007	
Family Income	Father Income	Per month income of father/guardian of student	Old	0.29	0.04	✓
	Mother Income	Per month income of mother of student	Old	0.02	0.03	✓
	Land Income	Per month income from land of family of student	Old	0.02	0.05	✓
	Miscellaneous Income	Per month miscellaneous income of family of student	Old	0.08	0.03	✓
	Earning Hands	Total number of Earning hands of student's family	Old	0.007	0.005	
	Father Status	Status of father of student: alive or deceased	New	.0008	0.001	
	Father Retired	Father retired or in service	New	0.002	0.003	
	Guardian Alive	Is student's guardian alive	New	0.003	0.004	
Student Personal Information	Gender	The gender of the student (male or female)	Old	0.004	0.005	
	Marital Status	Marital status of student (married or unmarried)	New	0.003	0.01	✓
	House Owner Ship	Student have his/her own house	New	0.08	0.10	✓
	Previous Program Scholarship	Scholarship received or not in previous academic program	New	.0002	.0003	
	Previous Institution Type	Type of student previous institution	Old	0.001	0.002	
	Self Employed	Is student is self employed	New	0.06	0.04	✓
Family Assets	Land Value	Current value of lands belongs to student's family	Old	0.04	0.02	✓
	Bank Balance	Bank balance of student's family	Old	0.05	0.07	✓
	Stock Value	Value of Shares/Bonds belong to student's family	Old	0.01	0.08	✓
	House Value	Value of house belong to student's family	Old	0.14	0.03	✓
	House Condition	Structure of house belong to student's family	New	0.06	0.04	✓
	Miscellaneous Asset Value	Any other assets related to student	Old	0.04	0.02	✓
	Location	Type of Location where student resides; urban or rural	Old	0.03	0.04	✓
	No of Vehicles at home	How many vehicles belong to family of a student	Old	0.005	0.008	

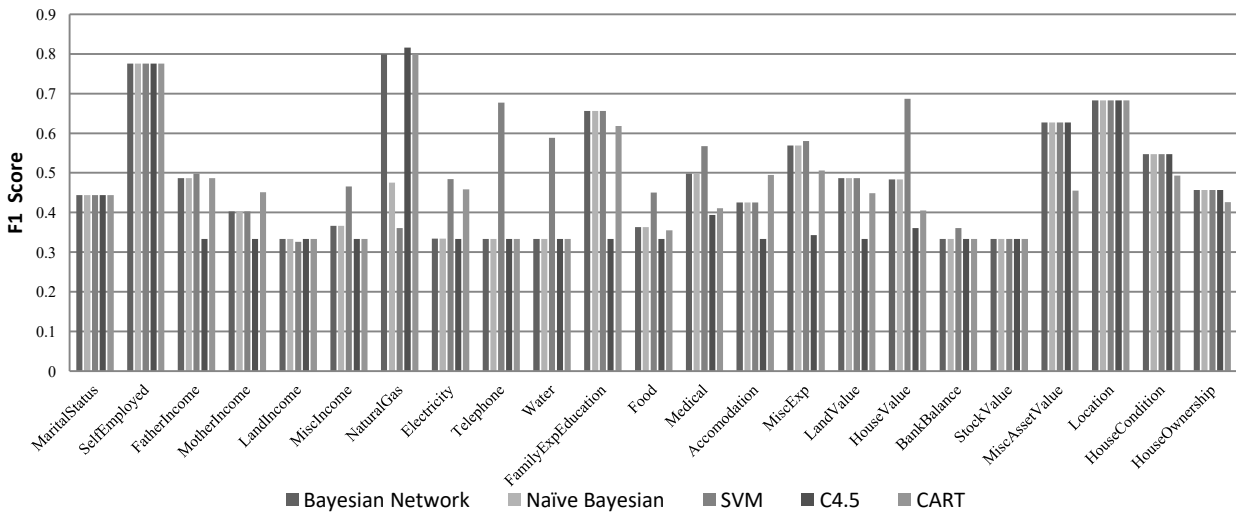


Figure 4 Impact of selected individual features on classification accuracy.

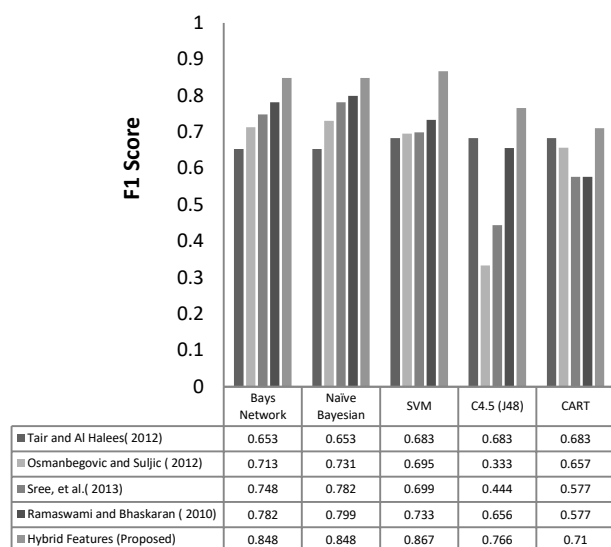


Figure 5 Comparison with baseline and proposed features.

The features related to family expenditure such as natural gas, electricity, telephone, water, accommodation, miscellaneous expenditures, and most importantly family expenditure on education are found to be most effective in predicting academic performance. Most of these features are ignored by the baseline methods and prior studies [12,25]. The best predictive performance is obtained when family expenditure based features are combined with other features (hybrid features). It has been observed that family expenditures affect the students' performance and reduce their concentration and interest in studies. The claims made on the basis of experimental outcomes are verified by 25 students studying on scholarships. Most of these students agree to the following discussions.

An increase in the expenditures of family reduces the opportunities for a student to grow up and excel in their studies because time and money are important factors in life and are directly related with the family expenditure. The increase in expenditure, especially on medical treatments and accommodations dominantly affects the performance of students. More expenditure on medical relates to health issues and more expenditure on accommodation may affect the budget (for education) of a middle class family.

On the other hand, some personal characteristics of students are also important predictors for their performance evaluation, e.g., married students better concentrate on their studies as compared to bachelor students perhaps because of emotional stability in their personal lives. The same is the case with students who themselves or their parents have their own property. Families having their own house definitely saves money by not paying house rent and can spend these savings for the education of their children. They also don't need to keep changing the rented houses which may waste time and energy of a student. Similarly, the self-employment status of a student enables him/her to better schedule time for studies in an efficient way because less worries about finances result in comfort and satisfaction. In addition to this, self-employment develops hard working attitude in the personalities of the students, both these factors are very helpful for students in achieving better performance in their studies. Last but not the least, the house condition of a student is also an influential factor because having a comfortable living accommodation enables

him/her to better utilize his/her abilities in studies. On the other hand, if the house condition is not good, the student's time and energy may be wasted in repairs or in helping his/her parents to get the repairs done. Previous studies [16,23] have also explored the gender and institution type characteristics, which we do not consider.

6. CONCLUSIONS

In this research, an effort is made to find the impact of our proposed features on student performance prediction with the help of generative and discriminative classification models. A feature space is constructed by considering characteristics of family expenditure, family income, personal information and family assets of students. The potential/dominant features selection is unavoidable as it provides us with subset of features. SVM classifier is found effective for our proposed features of family expenditure and student personal information categories. It can be concluded from the results that family expenditure and personal information features have significant impact on the performance of the student due to intuitive reasons provided in discussions.

In future, WWW research on Digital Learning and Learning Analytics should be focused on the following directions:

- Which kinds of methods and flexible applications permit the construction of critical learners' data from the WWW, e.g. mining of social media content can be a basis for personal expenditure?
- Which are the possibilities to proceed to dynamic profiling of personal characteristics of students from the Deep Web?
- Which are the standards for codifying critical students' information in the WWW, and how this can envision future WWW based learning services?

Learning analytics of mobile and ubiquitous learning environments from the perspective of human computer interaction [1,2] also require detailed exploration in addition to aforementioned traditional and web based features.

REFERENCES

- [1] N. R. Aljohani and H. C. Davis, "Learning analytics in mobile and ubiquitous learning environments," in *11th World Conference on Mobile and Contextual Learning*, 2012.
- [2] N. R. Aljohani, H. C. Davis, and S. W. Loke, "A comparison between mobile and ubiquitous learning from the perspective of human-computer interaction," *International Journal of Mobile Learning and Organization*, vol. 6, no. 3/4, pp. 218-231, 2012.
- [3] R. Asif, A. Merceron, and M. K. Pathan, "Investigating performance of students: a longitudinal study," in *Fifth International Conference on Learning Analytics And Knowledge (LAK '15)*, New York, USA, 2015, pp. 108-112.
- [4] M. A. Chatti, A. L. Dyckhoff, U. Schroeder, and H. Thüs, "A reference model for learning analytics," *International Journal of Technology Enhanced Learning (IJTEL)*, vol. 4, no. 5/6, pp. 318-331, 2012.
- [5] N. Fournier, R. Kop, and H. Sitlia, "The value of learning analytics to networked learning on a personal learning environment," in *1st International Conference on Learning Analytics and Knowledge*, 2011, pp. 104-109.
- [6] S. Kotsiantis, C. Pierrakeas, and P. Pintelas, "Predicting

- students' performance in distance learning using machine learning techniques," *Applied Artificial Intelligence*, vol. 18, no. 5, pp. 411-426, 2004.
- [7] E. Lotsari, V. Verykios, C. Panagiotakopoulos, and D. Kalles, "A Learning Analytics Methodology for Student Profiling," in *Artificial Intelligence: Methods and Applications*, 2014, pp. 300-312.
- [8] Y. Ma, B. Liu, C. K. Wong, P. S. Yu, and S. M. Lee, "Targeting the right students using data mining," in *6th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD '00)*, New York, USA, 2000, pp. 457-464.
- [9] B. Minaei-Bidgoli, D. A. Kashy, G. Kortemeyer, and W. F. Punch, "Predicting student performance: an application of data mining methods with an educational Web-based system," in *33rd Annual Frontiers in Education (FIE 2003)*, Westminster, CO, 2003.
- [10] T. Mishra, D. Kumar, and Sangeeta Gupta, "Students' Employability Prediction Model through Data Mining," *International Journal of Applied Engineering Research*, vol. 11, no. 4, pp. 2275-2282, 2016.
- [12] E. Osmanbegović and M. Suljić., "Data mining approach for predicting student performance," *Economic Review*, vol. 10, no. 1, pp. 3-12, 2012.
- [13] O. K. Oyedotun, S. N. Tackie, and Ebenezer O. Olaniyi, "Data Mining of Students' Performance: Turkish Students as a Case Study," *International Journal of Intelligent Systems and Applications*, vol. 7, no. 9, pp. 20-27, 2015.
- [14] Z. A. Pardos, N. T. Heffernan, B. Anderson, C. L. Heffernan, and W. P. Schools, "Using fine-grained skill models to fit student performance with Bayesian networks," in *Handbook of educational data mining.*, 2010, pp. 417-426.
- [15] P. J. Piety, D. T. Hickey, and M. J. Bishop, "Educational data sciences: Framing emergent practices for analytics of learning, organizations, and systems," in *4th International Conference on Learning Analytics and Knowledge*, 2014, p. 193.
- [16] M. Ramaswami and R. Bhaskaran., "A CHAID based performance prediction model in educational data mining," *International Journal of Computer Science*, vol. 7, no. 1, pp. 10-18, 2010.
- [17] C. Romero and S. Ventura, "Educational data mining: A survey from 1995 to 2005," *Expert systems with applications*, vol. 33, no. 1, pp. 135-146, 2007.
- [18] J. L. Santos, K. Verbert, S. Govaerts, and E. Duval, "Addressing learner issues with StepUp!: An evaluation," in *International Conference on Learning Analytics and Knowledge*, 2013, pp. 14-22.
- [19] B. Shalem, Y. Bachrach, J. Guiver, and C. M. Bishop, "Students, teachers, exams and MOOCs: Predicting and optimizing attainment in web-based education using a probabilistic graphical model," in *Joint European Conference on Machine Learning and Knowledge Discovery in Databases*, 2014, pp. 82-97.
- [20] A. Sharabiani, F. Karim, A. Sharabiani, M. Atanasov, and H. Darabi, "An enhanced bayesian network model for prediction of students' academic performance in engineering programs," in *IEEE Global Engineering Education Conference (EDUCON)*, 2014, pp. 832-837.
- [21] G., Siemens and P Long, "Penetrating the fog: Analytics in learning and education," *EDUCAUSE Review*, vol. 46, no. 5, 2011.
- [22] S. Slater, S. Joksimovic', V. Kovanovic, R. S. Baker, and D. Gasevic, "Tools for Educational Data Mining: A Review," *Journal of Educational and Behavioral Statistics*, 2016.
- [23] G. S. Sree and C. Rupa., "Data Mining: Performance Improvement In Education Sector Using Classification And Clustering Algorithm," *International Journal of Innovative Research and Development*, vol. 2, no. 7, pp. 101-106, 2013.
- [24] P. Strecht, L. Cruz, C. Soares, J. Mendes-Moreira, and R. Abreu, "A Comparative Study of Classification and Regression Algorithms for Modelling Students' Academic Performance," in *International Educational Data Mining Society*, 2015, pp. 392-395.
- [25] M. M. A. Tair and A. M. El-Halees, "Mining educational data to improve students' performance: a case study," *International Journal of Information*, vol. 2, no. 2, pp. 140-146, 2012.
- [26] S. K. Yadav, B. Bharadwaj, and S. Pal, "Data mining applications: A comparative study for predicting student's performance," *International Journal of Innovative Technology and Creative Engineering*, vol. 1, no. 12, pp. 13-19, 2011.
- [27] B. J. Zimmerman and A. Kitsantas., "Comparing student's self-discipline and self-regulation measures and their prediction of academic achievement," *Contemporary Educational Psychology*, vol. 39, no. 2, pp. 145-155, 2014.