Predicting Students That Are At Risk Of Not Graduating In Record Time

Lenny Bokgoshi
School of Computer Science and Applied Mathematics
The University of the Witwatersrand,
Johannesburg, South Africa
lenny.bokgoshi1@students.wits.ac.za

Ashwini Jadhav
Science Teaching and Learning Unit
Faculty of Science
The University of the Witwatersrand,
Johannesburg, South Africa
ashwini.jadhav@wits.ac.za

Ritesh Ajoodha
School of Computer Science and Applied Mathematics
The University of the Witwatersrand,
Johannesburg, South Africa
ritesh.ajoodha@wits.ac.za

***Abstract*—In this study, we apply machine learning techniques to understand the cause of low throughput of graduates in South African institutions. When students enroll in universities for a degree of their choice, they are selected based on the marks they have achieved from secondary school or other universities. Over the course of the years on enrollment until qualifying, these students face challenges that could put their academic path at risk. This research therefore aims to predict the students that are at risk of not completing their year of study until the end of the degree. Predictive models are applied to a synthetic dataset, generated by a Bayesian network, to classify the students into four risk profiles. This work will benefit staff who provide support to students who are at academic risk of not completing their academic programmes and also to accommodate student who transfer between universities.**

***Index Terms*—Machine Learning, Student Attrition, At-risk,,Classification, Academic Performance,**

I. Introduction

In recent years, the throughput of graduates is getting less and less such that it raises a concern in communities and/or in universities. The rate of enrollment in the first year of study is not proportional to the rate of graduates produced in universities. It turns out that these students do not complete their chosen programmes in the allocated time, and some dropout in the middle of their academic years [15]. This can be caused by various factors such as , and transitioning into a new system used by that particular institution, or adapting to a new environment. In the first year of study, especially those from high school with no university experience, have it much harder to transition since they move to a whole new space. Students who already have a bit of experience in university begin to relax as most university systems do not keep track with regular attendance of lectures, and sometimes tutorials. These students therefore miss out and eventually begin to perform poorly, if not worse, in their studies.

This research study objective is to help in the prediction of a students' risk profile when enrolled for a particular degree based on the streamlines they wish to follow, by using background, individual and prior results achieved from high school and/or academic years. The aim is to predict a students' performance that puts them at risk of not completing a degree using a students biographical information and enrollment features. We will determine the machine learning models that can accurately predict a students' risk profile.

Using information collected on the students features, we classified which of the features significantly contributes to the students' being at risk of dropping out, or not completing a degree. We classify the students' risk profiles into four categorical classes: Low Risk, Medium Risk, High Risk and Highest Risk. We will then represent the results from the selected six machine learning models using accuracy, f1-score and the area under the curve.

The structure of the paper is as follows, the related work section has an overview of previously published work, allowing us to find relevant theories, methods, and research gaps; the methodology section highlights the data processing, features, and classification models used; the Results section explains the results obtained from the models; and the conclusion section outlines the study, highlights our contributions, and makes suggestions for further research.

II. Related Work

This section will reviews prior published work done associated with exploring different features various authors found that affect a students' performance after applying machine learning models.

With an increase to access in higher learning, it has resulted in a large number of students being admitted into institutions. Even though students are enrolled in their choice of study, in the years, 2008 and 2015, about 50% of students who were able to get admitted, they failed to complete their degrees in the allocated time [5].

A. Conceptual Framework

For this research we adopt the conceptual framework of [19] as reasoning to predict student attrition using biographical and enrollment observations. The product of the following characteristics, Background, Pre-College and Individual, play a role in a student's performance in universities, and further says that these three factors determine the individual's ability to set goals, and adds that the importance of social life and involvement in universities play as a factor as well, in students' performance, stated [19]. Figure 1 shows the interface that will be adapted to predict the students' performance which will be used in the next section to develop a list of the features associated with affecting students' performance.

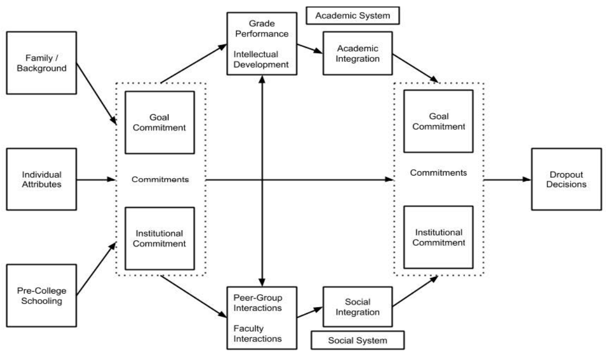

Fig. 1: A model by [19] that shows factors affecting students' performance in a university.

B. Features to predict student success

1) Background and Individual : For this section we will explore variables that are related to the student, as an individual, and the students' background. The variables/features used for this characteristic are commonly, home province, home country, whether the student is from rural or urban, race, etc. The home province feature is found to be one of the most contributing factors to student performance in studies [1], [15]. The features return a reasonable prediction to predict the performance in universities.

A study conducted found that the background and individual factors, as determinants of students' success, they are significant predictors of students' academic success [8]. These factors can improve prediction accuracy, however, using these measures as a tool to research, as measures of student quality they are not straightforward, and hence make performance prediction a more complex process [8].

2) Pre-College : When talking about the Pre-College features, we refer to scores the students obtained before enrolling into a university, and in some cases, refer to the scores obtained for a particular module during the academic year(intra-college). For pre-college, in most universities, enrollment of matriculants is based on the criteria of obtaining a particular APS score. Additionally, for some degrees, the percentages obtained for specific subjects are used to supplement admission protocols. An example to the above statement, should a student want to enroll computer sciences, the mathematical ability is looked at as a primary predictor for success and therefore the mathematics results are used to determine the success of a student in the streamline. This means the student has a fair chance in university mathematics, though there are other significant factors [1].

3) Psycho-Social Factors : In this subsection, we review variables related to the students' social and psychological factors. A study using psychological factors to predict students' academic performance and found that cognitive factors such as attention, memory, and reasoning and non-cognitive factors such as time management, self-regulation, academic mind-sets and learning strategies predict adjustment and academic performance [7].

There is a constructed model of university success using psycho-social factors in study [12]. Based on the students satisfaction and goals, the features are, namely, class communication, time management, satisfaction with academic work, efficacy on the student, extramural activities within the university and lastly, stress pressure.

Table I shows the various features associated with students' performance in higher education and are grouped into their respective categories.

C. Methods for predicting academic performance

Using predictive modelling, various authors have made use of historical data to predict academic performance in institutions around the world.

Table II provides an insight on the combination of features authors used to evaluate the performance of the machine learning models they applied.

III. Methodology

We predict student academic performance in institutions by using the conceptual framework in Figure 1. We achieve this goal through the use of machine learning techniques to classify these students into 4 risk profiles, namely, Low Risk - which means a student will finish in the minimum time required(record time), Medium Risk - a student will exceed the minimum allocated time, High Risk - it will take a student the maximum number to complete a chosen degree and finally, Highest Risk - the student does not complete a degree and eventually dropout of the institution. We apply several machine learning models to deduce the students who would fall into the risk profiles.

A. Data Processing and Collection

The data used in this study is synthetically generated by a Bayesian Network and consists of biographical and enrolment observations that act as variable for the conceptual model in [19]. The data is adopted from a recent study that aimed to identify first year students at risk of not finishing a degree in the allocated time at a South African university [15]. In this data, conditional independence was assumed to convey the relationship between features such as home country, age in each year of study, and a variety of others in Table I. Additionally, this data is of the students who want to pursue studies in the Faculty of Science, in which it consists of science streamlines. Few of these streamlines are Mathematical Science, Physical Science, and other known science streams.

Background Factors (BF)	Pre & Intra-College Scores(PICS)	Individual Features(IF)	Psycho-Social Factors (PSF)
Rural/Urban School	School Quintile	Home language	Class communication
Home Country	APS	Age in year of study	Time management
Home Province	English	Plan Description	Student Absent Days
Financial support	Accounting Science	Interest in sports	Visited Resources
International	Mathematics		Cognitive Difficulties

TABLE I: List of features associated with student performance in institutions.

Author(s)	Factors used	Models	Prediction Accuracy
[17]	BF, PICS, IF and PSF	Random Forest	95.45%
[17]	BF, PICS, IF and PSF	Logistic Model Trees	93.15%
[11]	BF and PICS	Naive Bayes	90.91%
[15]	BF, PICS, IF and PSF	Random Forest	85%
[13]	BF, PICS and IF	Neural Networks	84.80%
[2]	PICS	ANN	84.60%
[18]	BF, PICS, IF and PSF	Decision Tree	79.60%
[1]	BF, PICS and IF	Naive Bayes	69.18%
[6]	PICS	Decision Tree	66.67%
[9]	PSF	Support Vector Machine	63.10%

TABLE II: A table comparing the models various authors applied along with the combinations of categories from Table I to predicts the students' performance, and accuracies achieved in each case.

B. Feature Selection

The dataset contains a total of 41 features. However, not all of the features are meaningful in terms of classifying. For this study a total of 15 features were selected, for their respective year of study. Table III gives a summary of the color coded features according to the developed scheme in Table I.

In the preprocessing phase, the features are evaluated using Information Gain(IG), also called entropy. Using the IG along with the experiments, a subset of 15 features were selected for for each of our 3 cases, to train our machine learning models. The information gain obtained for the features was vast, thus the features are not arranged according to the entropy due to random sampling to determine these features.

C. Classification Models

This subsection will briefly discuss the six machine learning predictive models are used to predict students at risk for each year of study. The models used are : Random Forests, Logistic Regression, Multi-Layer Perceptron(MLP), Naïve Bayes, Decision trees and K-Nearest Neighbour(KNN).

***Decision trees**:* This model is a decision tool that constructs a tree based on the if-then conditions to make decisions on the possible outcomes. It splits the nodes on all available features and selects the split which results in most homogeneous sub-nodes. Though there are different algorithms to construct a tree, for this study, J48(sometimes called the C4.5) and Random Forests are the selected models.

#	Features	1st Year	2nd Year	3rd Year
1	PlanDescription	■	■	■
2	PlanCode	■	■	■
3	NumberOfYearsforDegree	■	■	■
4	YearStarted	■	■	■
5	Language	■	■	■
6	ProgressoutcomeYOS1		■	■
7	Secondyearoutcome			■
8	Firstyearoutcome		■	■
9	AgeatFirstYear	■	■	■
10	AggregateYOS1		■	■
11	AggregateYOS2			■
12	RaceDescription	■	■	■
13	ProbofBioStreamline	■	■	
14	LifeOrientation	■	■	■
15	ProbOfPhysicsStreamline	■	■	
16	Homeprovince	■	■	■
17	ProbOfMatheStreamline	■	■	
18	ProbofEarthStreamline	■	■	
19	EnglishFL	■	■	■
20	ProbofUnknownStreamline	■		
21	AgeatSecondYear		■	■
22	EnglishFirstLang	■		

TABLE III: Features used for classification for each year of study. The features are sorted such that they predict students at risk in $1^{st}, 2^{nd}$ and 3^{rd} year of study.

***Random Forests**:* This model generates a variety of decision trees using a subset of the available features and returns the mode predicted class of each tree(for classification), or the average(for regression). The implementation on this current study is based on [22].

***K Nearest Neighbour(KNN)**:* KNN tries to predict the class by identifying the centroid sample for the testing data by calculating the distance between the centroid and all other training data samples [23]. The implementation is adapted from [27].

***Naïve Bayes(NB)**:* This is a classification models based on probability theory, particularly on Bayes' Theorem. Learning is based on the assumption of independence among predictors. The probability of the outcome is mathematically expressed as:

$$P(A|B) = \frac{P(B|A) * P(A)}{P(B)}$$

In this study, the implementation is based on [24].

***Logistic Regression(LR)**:* The LR predicts the probability of the dependable outcome as a function of the independent variables, for which in this study, it will construct a function of features to predict the category risk a student falls in. The architecture of the model used in this study follows that of [25].

Multi-Layer Perceptron(MLP): The model is a type of neural network that uses back propagation to learn parameters. The architecture of this model has three hidden layers, each with nodes. The first layer is the input value for the features, the middle layer with one or more levels that specify the mathematical function, and, the third layer as the output layer for the predicted outcome. The implementation is based on [26].

D. Prediction And Evaluation

This study is a classification problem since the 4 risk profiles we are predicting are provided in the synthetic data. In this subsection, we look at the scoring metrics that will tell us how accurate the models are by using the following:

1) Confusion matrix : it tells us the number of how many students we are able to classify in their respective classes. Table IV presents the description of the confusion matrix, where TP are the true positives-number correct of positive predictions, FP the false positives- number of incorrect positive predictions, FN are false negatives-number of incorrect negative predictions, and TN are the true negatives-number of correct negative predictions.

	Predicted +ve	Predicted -ve
Actual +ve	TP	FP
Actual -ve	FN	TN

TABLE IV: A confusion matrix structure for a 2-class classification.

For the confusion matrices obtained from this study, please see the tables on the last page of this document, Figures 2, 3 and 4, for first year of study, second year of study and third year of study, respectively.

2) Evaluation and Validation: The effectiveness of the selected models has to be evaluated as well, for our experiments, when applying the 10-Fold cross validation procedure. This procedure involves the partitioning the data into 10 subdata partitions(folds). This technique uses one partition as training data, which is further split into 10 folds consisting of training and testing data, and the remaining 9 unseen partitions for validation of the models. This occurs interchangeably for all 9 folds.

3) Accuracy : In this metric, the evaluation metrics recall and precision obtained from the confusion matrix are used to determin the performance of the machine learning models applied in the study. The equations that follow will show how the two evaluation metrics are calculated and explained:

$$Recall = \frac{TP}{TP + FN}, \tag{1}$$

where it is defined as the measure in which the models correctly identify the True positives, and,

$$Precision = \frac{TP}{TP + FP}, \tag{2}$$

defined the ratio of the true positives and all the positvies.

It therefore follows that the accuracy is calculated as :

$$Accuracy = \frac{TP + TN}{TP + FP + TN + FN} \tag{3}$$

Various literature utilized a similar representation of the accuracy [1], [17]. Other measures include the F1-score, which is also called the F-Score or the F-Measure, defined as the harmonic average of recall and precision, with an optimal value at 1 and worst value of 0. Another measure metric that is used is the Receiver Operating Characteristic curve, also called the ROC curve, which tells us how well can the model can distinguish between the classes. The ROC ranges from 0(indicating perfectly inaccurate test) to 1(reflecting perfectly accurate test). To ensure that the study was not guess work we want our ROC area to be above 0.5.

IV. RESULTS AND DISCUSSION

In this section we present the results of the six prediction models using the scoring metrics in the previous subsection.

A. Prediction Outcomes

This subsection presents the accuracies obtained from equation (3), for each model. The accuracies are summarized in Table V. We note that in each year of study, decision trees obtain the highest accuracy in classifying students while the KNN obtain the lowest accuracy.

	Prediction Accuracy		
Models	1st Year	2nd Year	3rd Year
Decision tree	95.94%	96.66%	97.33%
Random Forests	97.33%	96.94%	97.94%
Naïve Bayes	88.38%	89.50%	89.11%
Logistic Regression	84.88%	85.05%	85.27%
K Nearest Neighbour	70.55%	70.83%	70.88%
Multi-Layer Perceptron	91.61%	92.83%	92.11%

TABLE V: The predictive accuracies obtained by the predictive models after 10-fold cross-validation.

B. Model Performance Evaluation

This subsection presents the results of the scoring metric obtained in the first, second, and third, years of study of classification, along with tables that present performance values of the F1 Score and ROC curve, and analysis on the results obtained.

In Table VI, we note that for each year of study, KNN has the lowest F-Measure, and Random Forests obtains the highest F1-score.

Table VII presents the ROC curve that summarizes the overall diagnostic accuracy of the models used in the study.

	F1 Score		
Models	1^{st} Year	2^{nd} Year	3^{rd} Year
Decision tree	0.957	0.966	0.974
Random Forests	0.973	0.970	0.978
Naïve Bayes	0.882	0.894	0.891
Logistic Regression	0.849	0.850	0.852
K Nearest Neighbour	0.704	0.708	0.708
Multi-Layer Perceptron	0.923	0.911	0.931

TABLE VI: The F1 Scores of the six trained models.

	Area Under the Receiver Operating Characteristic Curve(ROC AUC)		
Models	1^{st} Year	2^{nd} Year	3^{rd} Year
Decision tree	0.971	0.976	0.983
Random Forests	0.997	0.998	0.999
Naïve Bayes	0.971	0.977	0.980
Logistic Regression	0.959	0.960	0.964
K Nearest Neighbour	0.804	0.879	0.806
Multi-Layer Perceptron	0.986	0.988	0.990

TABLE VII: The ROC AUC values of the six trained models.

V. CONCLUSION

This paper contributes to the current body of knowledge that by using an approach to predict students' performance in each year of study, until graduating, we can impact student success in universities. If we accurately predict a students' risk profile, we could argue that, an early warning system for those that might be not be able to graduate, be provided. This way a student is notified early on in the year of the various consequences ahead associated with graduating in record time.

Six machine learning models are applied to predict the students at risk in each year of study using a synthetic dataset. After 10-fold cross validation, the models performed fairly well. The F1 Score and the ROC AUC curve analysis show that we can accurately apply machine learning models to predict the students at risk in each year of study, supporting our argument. For future work of this study, we hope to implement the models on real data to develop a more enhanced warning system for institutions worldwide.

The significance for this paper is to improve the throughput of graduating, not only in record time but should the identified at risk student be warned, the student can still graduate in the maximum years allowed to complete a degree or alternatively, enrol for a course they might be able perform well in and eventually be able to secure employment.

REFERENCES

[1] Tasneem Abed, Ritesh Ajoodha, and Ashwini Jadhav. A prediction model to improve student placement at a south african higher education institution. In 2020 International SAUPEC/RobMech/PRASA Conference, pages 1–6. IEEE, 2020.

[2] Samy S Abu-Naser, Ihab S Zaqout, Mahmoud Abu Ghosh, Rasha R Atallah, and Eman Alajrami. Predicting student performance using artificial neural network: In the faculty of engineering and information technology. 2015.

[3] Silvia Acid, Luis M de Campos, and Javier G Castellano. Learning bayesian network classifiers: Searching in a space of partially directed acyclic graphs. Machine learning, 59(3):213–235, 2005.

[4] Olugbenga Wilson Adejo and Thomas Connolly. Predicting student academic performance using multi-model heterogeneous ensemble approach. Journal of Applied Research in Higher Education, 2018.

[5] Ritesh Ajoodha. Predicting learner attrition for the sciences using background individual attributes and schooling at a south african higher educational institute. Private Communication, 2019.

[6] Raheela Asif, Saman Hina, and Saba Izhar Haque. Predicting student academic performance using data mining methods. International Journal of computer science and network security, 17(5):187–191, 2017.

[7] Riya Bhattacharya and Bani Bhattacharya. Psychological factors affecting students academic performance in higher education among students. International Journal for Research and Development in Technology, 4(1):63–71, 2015.

[8] KG Bokana and DD Tewari. Determinants of student success at a south african university: An econometric analysis. The Anthropologist, 17(1):259–277, 2014.

[9] Iti Burman and Subhranil Som. Predicting students academic performance using support vector machine. In 2019 Amity International Conference on Artificial Intelligence (AICAI), pages 756–759. IEEE, 2019.

[10] Nir Friedman and Moises Goldszmidt. Building classifiers using bayesian networks. In Proceedings of the national conference on artificial intelligence, pages1277– 1284, 1996.

[11] Raza Hasan, Sellappan Palaniappan, Abdul Rafiez Abdul Raziff, Salman Mahmood, and Kamal Uddin Sarker. Student academic performance prediction by using decision tree algorithm. In 2018 4th international conference on computer and information sciences (ICCOINS), pages 1–5. IEEE, 2018.

[12] Elizabeth J Krumrei-Mancuso, Fred B Newton, Eunhee Kim, and Dan Wilcox. Psycho-social factors predicting first-year college student success. Journal of College Student Development, 54(3):247–266, 2013.

[13] ET Lau, L Sun, and Q Yang. Modelling, prediction and classification of student academic performance using artificial neural networks. SN Applied Sciences, 1(9):1–10, 2019.

[14] Yaoran Li, Jeff Allen, and Alex Casillas. Relating psychological and social factors to academic performance: A longitudinal investigation of high-poverty middle school students. Journal of Adolescence, 56:179–189, 2017.

[15] Noluthando Mngadi, Ritesh Ajoodha, and Ashwini Jadhav. A conceptual model to identify vulnerable undergraduate learners at higher-education institutions. In 2020 2nd International Multidisciplinary Information Technology and Engineering Conference (IMITEC), pages 1–8. IEEE, 2020.

[16] Irfan Mushtaq and Shabana Nawaz Khan. Factors affecting studentsâC™ academic performance. Global journal of management and business research, 12(9), 2012.

[17] Ndiatenda Ndou, Ritesh Ajoodha, and Ashwini Jadhav. Educational data-mining to determine student success at higher education institutions. In 2020 2nd International Multidisciplinary Information Technology and Engineering Conference(IMITEC), pages 1–8. IEEE, 2020.

[18] Edin Osmanbegovic and Mirza Suljic. Data mining approach for predicting student performance. Economic Review: Journal of Economics and Business, 10(1):3– 12, 2012.

[19] Vincent Tinto. Dropout from higher education: A theoretical synthesis of recent research. Review of educational research, 45(1):89–125, 1975.

[20] D. M. Powers, "Evaluation: from precision, recall and f-measure to roc, informedness, markedness and correlation," 2011.

[21] Y. EL-Manzalawy, "WLSVM," 2005, you don't need to include the WLSVM package in the CLASSPATH.

[22] L. Breiman, "Random forests," Machine Learning, vol. 45, no. 1, pp.5–32, 2001.

[23] A. Dutt and M. A. Ismail, "Can we predict student learning performance from lms data? a classification approach," in 3rd International Conference on Current Issues in Education (ICCIE 2018). Atlantis Press, 2019, pp. 24–29.

[24] G. H. John and P. Langley, "Estimating continuous distributions in bayesian classifiers," in Eleventh Conference on Uncertainty in Artificial Intelligence. San Mateo: Morgan Kaufmann, 1995, pp. 338–345.

[25] S. Sperandei, "Understanding logistic regression analysis," Biochemia medica: Biochemia medica, vol. 24, no. 1, pp. 12–18, 2014.

[26] T. Gedeon and H. Turner, "Explaining student grades predicted by a neural network", in Proceedings of 1993 International Conference on Neural Networks (IJCNN- 93-Nagoya, Japan), IEEE, vol. 1, 1993, pp. 609–612.

[27] S. Taruna and M. Pandey, "An empirical analysis of classification techniques for predicting academic performance," 2014 IEEE International Advance Computing Conference (IACC), 2014, pp. 523-528, doi: 10.1109/IAdCC.2014.6779379.

		Predicted			
		High	Medium	Lowest	Highest
Actual	High	433	0	17	0
	Medium	0	431	0	19
	Lowest	19	0	431	0
	Highest	0	16	0	434

(a) Decision Trees accuracy of 95.94%

		Predicted			
		High	Medium	Lowest	Highest
Actual	High	432	0	18	0
	Medium	0	425	0	25
	Lowest	4	0	446	0
	Highest	0	3	0	447

(b) Random Forests, accuracy of 97.33%

		Predicted			
		High	Medium	Lowest	Highest
Actual	High	329	52	42	27
	Medium	4	404	0	42
	Lowest	4	3	443	0
	Highest	3	32	0	415

(c) NB, accuracy of 88.38%

		Predicted			
		High	Medium	Lowest	Highest
Actual	High	362	18	68	2
	Medium	12	377	19	42
	Lowest	38	7	388	17
	Highest	0	10	41	399

(d) LR, accuracy of 84.88%

		Predicted			
		High	Medium	Lowest	Highest
Actual	High	312	53	63	22
	Medium	34	394	8	14
	Lowest	44	10	279	117
	Highest	23	10	132	285

(e) KNN, accuracy of 70.55%

		Predicted			
		High	Medium	Lowest	Highest
Actual	High	412	6	32	0
	Medium	4	391	10	45
	Lowest	33	2	414	1
	Highest	0	9	2	439

(f) Multi-Layer Perceptron, accuracy of 91.61%

Fig. 2: First Year Confusion matrices for the six models.

Actual \ Predicted	High	Medium	Lowest	Highest
High	435	0	15	0
Medium	0	431	0	19
Lowest	10	0	440	0
Highest	0	16	0	434

(a) Decision Trees accuracy of 96.66%

Actual \ Predicted	High	Medium	Lowest	Highest
High	439	0	11	0
Medium	0	422	0	28
Lowest	8	0	442	0
Highest	0	6	0	444

(b) Random Forests, accuracy of 96.94%

Actual \ Predicted	High	Medium	Lowest	Highest
High	362	41	26	21
Medium	8	403	0	39
Lowest	8	1	441	0
Highest	1	44	0	405

(c) NB, accuracy of 89.50%

Actual \ Predicted	High	Medium	Lowest	Highest
High	373	18	57	2
Medium	19	379	5	47
Lowest	38	8	394	10
Highest	3	22	39	386

(d) LR, accuracy of 85.05%

Actual \ Predicted	High	Medium	Lowest	Highest
High	306	58	54	32
Medium	30	402	10	8
Lowest	24	2	317	107
Highest	40	3	157	250

(e) KNN, accuracy of 70.83%

Actual \ Predicted	High	Medium	Lowest	Highest
High	416	11	23	0
Medium	1	404	7	38
Lowest	24	3	421	2
Highest	0	16	3	431

(f) Multi-Layer Perceptron, accuracy of 92.83%

Fig. 3: Second Year Confusion matrices for the six models.

Actual \ Predicted	High	Medium	Lowest	Highest
High	438	0	12	0
Medium	0	437	0	13
Lowest	12	0	438	0
Highest	0	12	0	438

(a) Decision Trees accuracy of 97.33%

Actual \ Predicted	High	Medium	Lowest	Highest
High	444	0	6	0
Medium	0	435	0	15
Lowest	8	0	442	0
Highest	0	4	0	446

(b) Random Forests, accuracy of 97.94%

Actual \ Predicted	High	Medium	Lowest	Highest
High	353	49	18	30
Medium	10	409	0	31
Lowest	10	1	430	9
Highest	2	36	0	412

(c) NB, accuracy of 89.11%

Actual \ Predicted	High	Medium	Lowest	Highest
High	379	23	45	3
Medium	33	362	6	49
Lowest	27	10	408	5
Highest	2	18	42	387

(d) LR, accuracy of 85.27%

Actual \ Predicted	High	Medium	Lowest	Highest
High	300	76	46	28
Medium	58	376	1	15
Lowest	17	4	313	116
Highest	22	17	124	287

(e) KNN, accuracy of 70.88%

Actual \ Predicted	High	Medium	Lowest	Highest
High	414	8	25	3
Medium	1	400	7	42
Lowest	10	1	436	3
Highest	0	15	20	415

(f) Multi-Layer Perceptron, accuracy of 92.11%

Fig. 4: Third Year Confusion matrices for the six models.