PAPER • **OPEN ACCESS**

Student Performance Prediction Model based on Supervised Machine Learning Algorithms

To cite this article: Ali Salah Hashim *et al* 2020 *IOP Conf. Ser.: Mater. Sci. Eng.* **928** 032019

Student Performance Prediction Model based on Supervised Machine Learning Algorithms

Ali Salah Hashim[1]
[1]*College of Computer Science and Information Technology / University of Basrah, Basrah, Iraq.*
alishashim2009@gmail.com

Wid Akeel Awadh[2]
[2]*College of Computer Science and Information Technology / University of Basrah, Basrah, Iraq.*
umzainali@gmail.com

Alaa Khalaf Hamoud[3]
[3]*College of Computer Science and Information Technology / University of Basrah, Basrah Iraq*
Alaak7alaf@gmail.com

Abstract

Higher education institutions aim to forecast student success which is an important research subject. Forecasting student success can enable teachers to prevent students from dropping out before final examinations, identify those who need additional help and boost institution ranking and prestige. Machine learning techniques in educational data mining aim to develop a model for discovering meaningful hidden patterns and exploring useful information from educational settings. The key traditional characteristics of students (demographic, academic background and behavioural features) are the main essential factors that can represent the training dataset for supervised machine learning algorithms. In this study, we compared the performances of several supervised machine learning algorithms, such as Decision Tree, Naļve Bayes, Logistic Regression, Support Vector Machine, K-Nearest Neighbour, Sequential Minimal Optimisation and Neural Network. We trained a model by using datasets provided by courses in the bachelor study programmes of the College of Computer Science and Information Technology, University of Basra, for academic years 2017–2018 and 2018–2019 to predict student performance on final examinations. Results indicated that logistic regression classifier is the most accurate in predicting the exact final grades of students (68.7% for passed and 88.8% for failed).

Keywords: Supervised Machine Learning, Educational Data Mining, Decision Tree, Naive Bayes, Logistic Regression, K-Nearest Neighbour, Multi-layer Perceptron, Neural Network

1. Introduction

The rapid development of information technology (IT) has greatly increased the amount of data in different institutions. Huge warehouses contain a wealth of data and constitute a valuable information goldmine. This dramatic inflation in the amount of data in institutions has not kept pace with the efficient ways of investing these data. Thus, a new challenge has recently emerged, that is, transitioning from traditional databases that store and search for information only through questions asked by a researcher to techniques used in extracting knowledge by exploring prevailing patterns of data for decision making, planning and future vision. One of these techniques is data mining (DM) technology [1].

DM uncovers useful correlations amongst attributes, hidden trends and patterns by analysing large amounts of datasets stored in warehouses. It is also used as a pattern recognition technique and mathematical and statistical method to reduce costs and increase revenue. In addition, DM is a field of knowledge discovery in databases [2].

The management process of educational institutions is one of the difficulties faced by managers due to the complexity of data structure, multiple sources and the huge size of data. Educational institutions face many other administrative, financial and educational problems whilst managing educational procedures. All these problems must be analysed to generate recommendations and conclusions that support decision makers in making their decisions for coordinating and managing the educational process [3, 4].

Educational data mining (EDM) is a DM used in educational and academic institutions. It is theory-oriented and aims to develop computational approaches that combine theory and data to assist with and enhance the quality of academic performance of students and graduates and faculty information of these institutions [5, 6]. EDM uses different techniques, such as Decision Trees (DTs), K-Nearest Neighbour (KNN), Naḭve Bayes (NB), Association Rule Mining and Neural Networks. Many types of knowledge, such as predication, association rules, classifications and clustering, can be discovered using these techniques.

EDM is a useful tool for academic institutions. Universities can use EDM to predict which students will pass or fail and have poor educational performance, to know who will pass the examinations in particular subjects and to obtain the ratio of graduates. EDM is also used for other strategic information. These universities can then develop and enhance their educational policies to help failed students raise their education level or guide them to specialisations that suit their preparations, preferences and abilities. The enhanced measures and policies resulting from EDM can be used for enhancing the academic performance of institutions [7, 8].

In addition, EDM is a common research area to explore data from educational fields by using DM techniques and machine learning approaches [9]. Research on machine learning aims to learn how to automatically recognise complex hidden patterns and create smart data for decision making [10, 11]. The predictive path of DM is a special DM process that performs prediction on

current data [12]. Obtaining a machine for adapting the action is the main interest in machine learning approach, and this interest can enhance the accuracy of certain actions or experiences. The expectation in classification approach is that computers must learn to classify techniques of observation examples, whereas in regression approach, the output should continue to have numeric quantity instead of discrete quantity [13]. Supervised machine learning is used for solving classification and regression problems [14].

This paper discussed the effectiveness of supervised machine learning algorithms in students' success prediction and academic performance in higher education. Specifically, the supervised machine learning algorithms measured student performance on the basis of actual grade or status (passed or failed). Different supervised machine learning algorithms were applied, and the performance criteria were evaluated. The experiments showed that the Logistic Regression classifier algorithm performed the best. Waikato environment for knowledge analysis (Weka) 3.8.0 (an open-source DM software environment) was used to implement the supervised machine learning algorithms.

2. Related Works

Several works that used machine learning algorithms, such as DT, NB, Logistic Regression, Support Vector Machine (SVM), KNN, Sequential Minimal Optimisation (SMO) and Neural Network, to predict student outcomes were reviewed. The details are shown in Table 1.

Table 1. Review of Literature

No.	*Algorithm*	*Reference*
1	DT	[12], [13], [15], [16]
2	NB	[14], [15], [16], [17]
3	Logistic Regression	[15], [16]
4	SVM	[15], [16]
5	KNN	[14]
6	SMO	[17]
7	Neural Network	[14], [15], [16], [17]

S. Natek and M. Zwilling [15] conducted a study on DM with small-sized datasets of students by comparing two different DM methods. Their conclusions were positive and revealed that integrating DM tools is an important part of information management systems in higher education institutions (HEIs). The dataset contained three years of data: 2010–2011 (42 students), 2011–2012 (32 students) and 2012–2013 (32 students). The data collected covered various aspects of the histories of students, including past academic records, family background and demographics. Three classifiers, namely, Rep Tree, J48 and M5P models, were applied to obtain students' academic performance. The experiments showed that J48 was less accurate but more sensitive than Rep Tree. However, the number of classifiers used to compare student performance was less than the number of algorithms in supervised and unsupervised machine learning approaches.

Alaa Khalaf et. al. [16] used Weka to evaluate university students' performance and obtain the factors that affect student success/failure. A total of 161 questionnaires were written on Google forms, and an open source application (LimeSurvey) was used to conduct a student survey at the College of Computer Science and Information Technology, University of Basrah. The authors used classification techniques (J48, Random Tree and Rep Tree) on the questionnaires filled out by students. In terms of accuracy, J48 outperformed the other two. The DTs used in this paper produced outstanding and accurate results, but many other fields in machine learning can achieve more accurate prediction results. Moreover, the model only predicted student status as 'passed or failed' and did not predict their actual grades.

Erman Yukselturk et. al [17] focused on identifying dropout students by using DM approaches in an online application. They applied four DM approaches, namely, KNN, DT, NB and Neural Network. KNN performed the best amongst all classifiers, with 87% accuracy. However, the model only examined four algorithms to predict the dropouts and not the actual grades of the students.

Previous studies analysed the performance of five popular machine learning algorithms that classify students at risk in advance and predict the difficulties they face in higher education at a distance [18, 19]. These algorithms were Artificial Neural Networks (ANNs), SVM, Logistic Regression, NB classifiers and DTs. ANNs and SVM are more accurate (57%) when only using demographic data than other algorithms [18], whereas NB has adequate accuracy but not as promising as the other models [19].

Acharya and Sinha [20] used machine learning to predict student performance. The input features in their study included gender, revenue, board marks and attendance. The techniques applied were C4.5, SMO, NB, 1-Nearest Neighbourhood and Multi-layer Perceptron (MLP). The researchers revealed that SMO is ideal for improving the model performance for all students in a course, with higher average test accuracy (66%) than other approaches. However, the accuracy is less outstanding than other model performances.

3. EDM

This term spread during the first workshop on the concept of EDM in 2005, and this workshop has become an international conference in Montreal since 2008 [21]. Periodical societies have shown interest in publishing the latest research on EDM. The most popular societies created in 2011 and 2012 are the International EDM Society (http://www.educationaldatamining.org/) and the IEEE Task Force of EDM (http://datamining.it.uts.edu.au/edd/), respectively. EDM uses DM methods to study the extracted data from educational systems (students and instructors) and analyse student learning processes in educational institutions.

EDM methods often have multiple levels of meaningful hierarchies, which must often be decided on the basis of data properties, rather than in advance, whether taken from university

administrative data, collaborative learning data based on computers or students' use of interactive learning environments [22]. Similar with the traditional methods of DM, EDM should identify the main goal of the study and the required data, extract data from educational environment, pre-process the data (clean and arrange a selection of techniques that can be applied), interpret the results and verify the applied techniques. The objectives and techniques used in EDM are derived from the specificity of the instructional environment and purpose of exploration [23]. The applications that adopt EDM follow several steps, as shown in Figure 1.

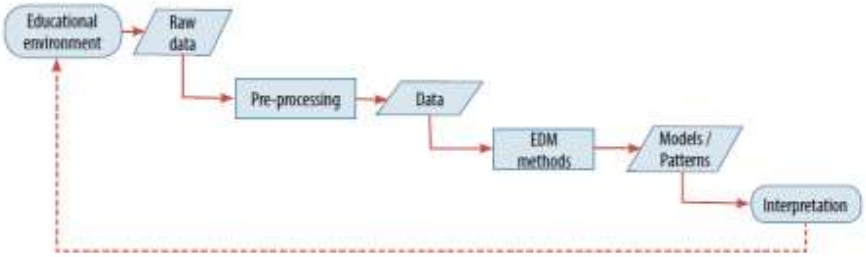

Figure 1. Model Implementation Flowchart [24]

4. Weka Tool

One of the most common machine learning applications is Weka, which is a tool written and developed in Java language at the University of Waikato, New Zealand. Weka is a free open-source software under the GNU general public license. The Weka workbench provides a collection of algorithms and tools for analysing data and implementing predictive modelling. Weka provides a graphical user interface for easy access and performs all DM algorithms [25]. The non-java Weka version is developed using TCL/TK, with modelling algorithms and data pre-processing using C language with Makefile-based system for running machine learning experiments. Weka is originally used for analysing the data of agricultural domains, whereas the Weka Java version released in 1997 is used in many fields, such as educational and research fields [26].

Weka supports different standard DM tasks, such as data pre-processing, data visualisation, clustering, classification, regression and feature selection. All Weka techniques are based on the assumption that the data are accessible as a single flat file or relationship, where each data point is represented by a fixed number of attributes (numerical, nominal or some other types of attributes). Weka can access many dataset files through connections or gateways by using Java database connectivity, SQL database, and comma separated values and many other dataset types. Multi-relational DM is impossible in Weka, where a separate software should be used to convert a collection of linked database tables into a single table suitable for Weka processing. Sequence modelling is another important area not currently covered by Weka distribution algorithms [27, 28].

5. Supervised Machine Learning Techniques

The field of machine learning has gained the attention of computer science and IT researchers. The data analysis field has become more essential than before, owing to the increasing amounts of huge data processed every day. The three basic types of machine learning are supervised, unsupervised and semi-supervised learnings [29]. In supervised learning, the training dataset only consists of labelled data. A supervised function is trained during the learning process, with the aim of predicting the future labels of unseen data. The two basic supervised problems are regression and classification, especially for discrete function classification and continuous regression [30]. Unsupervised learning aims to find meaningful, regular patterns without human intervention on unlabelled data. Its training set is made up of unlabelled data, and no instructor is present to help identify these patterns. Some popular supervised methods include clustering, novelty identification and dimensionality reduction [4, 31]. Semi-supervised learning is a combination of supervised and unsupervised learning processes. It is used to achieve enhanced results with few labelled examples. Its training dataset consists of labelled and unlabelled data. DT, NB, Logistic Regression, SVMs, KNN, SMO and Neural Network are well-known supervised techniques with accurate results in different scientific fields [8, 28, 32, 33].

5.1 DT is a supervised machine learning algorithm that uses branching methodology to show all possible outcomes of a decision in accordance with certain parameters. The tree structure consists of sets of rules hierarchically organised, starting with root attributes and ending with leaf nodes; each tree branch represents one or more outcomes from the original dataset [32, 33]. Root node is the top node in the tree without incoming branches, and all outgoing branches represent all the rows on the basis of the dataset. The internal node in the tree is the node with incoming and outgoing branches and can be used to test the attribute. Terminal node or leaf is the down node with only incoming branch. This node represents the final node in the tree, which may have many leaf nodes representing the final calculations [34].

5.2 NB is an algorithm built on the basis of the theorem of Bayes. This hypothesis is formulated by Thomas Bayes. This model is easy to build and is mainly used for very large datasets [35]. NB aims to calculate the process of conditional probability distribution of each feature. The conditional probability of a vector being classified into class C is equal to the probability product of each of the characteristics of the vector in class C. This algorithm is called ‘naive’ because of its core assumptions of conditional independence. All input features are believed to be independent from one another. If the conditional assumption of independence actually holds, then an NB classifier can converge faster than other models, such as Logistic Regression[36, 37].

5.3 Logistic Regression is mostly used for analysing and explaining the relationship between a binary variable (e.g. ‘pass’ or ‘failed’) and a series of predicted variables [38]. It aims to find the best model that fits to explain the relationship between the dependent and independent variable sets. Logistic Regression is developed together with Linear Regression, but they differ in binary variable and continuous variable response [39].

5.4 SVM is based on Vapnik's principle of theoretical learning. SVM embodies the concept of systemic risk minimisation [40]. SVMs have been applied to many regression, classification and outlier detection fields. The original input space in an SVM is mapped through a kernel into a product space of high-dimensional dot. The new space is called the feature space, where an optimal hyper plane is defined to optimise the ability to generalise. A few data points called support vectors can decide the optimal hyper plane. An SVM can provide strong generalisation output for classification problems, although it does not implement problem-domain knowledge [41].

5.5 KNN is a simple machine learning algorithm where an object is graded by its neighbours' majority vote. The object is being assigned to the most common class amongst its closest neighbours. K represents a positive number and is usually small. If k equals to 1, then the object is assigned to its closest neighbour's class. Choosing k as an odd number in binary (two class) classification problems is good for eliminating tied votes. Selecting parameter k in this algorithm may be important [34, 42, 43].

5.6 SMO is a new training algorithm for SVMs. In 1998, John Platt proposed an easy and fast method called SMO algorithm to train an SVM. The key idea is solving the problem of dual quadratic optimisation by optimising the minimum subset at each iteration, including two components. SMO separates the huge problem of quadratic programming into a collection of smallest problems solved analytically. When SMO is managed with little amount of training sets, the amount of memory needed is linear. Given that matrix computation is avoided, SMO scales somewhere between linear and quadratic in the size of the training set, whereas the regular chunking SVM algorithm scales somewhere between linear and cubic. Thus, SMO is the fastest amongst linear SVMs [43, 44].

5.7 Neural Network is another common technique used in EDM. A multi-layer neural network is composed of several units (neurons) linked together in a pattern. The units in a net are divided into three classes: input, output and hidden units [8, 45, 46]. The benefit of the neural network is its capability to detect all possible interactions amongst variables. It can also perform a full detection without any doubt, even in the nonlinear relationship between dependent and independent variables [47].

6. Methodology

This study used several supervised machine learning algorithms to predict the academic performance of students in their final examinations, and the results were compared. The proposed methodology consisted of two stages. The first stage involved data pre-processing, in which the data are prepared, consolidated and cleaned to prepare for the second stage. The second stage involved the classification performance of the most commonly and frequently used algorithm per mentioned machine learning technique.

6.1 Data pre-processing

Data acquisition focused on the courses for bachelor study programmes at the College of Computer Science and Information Technology (CSIT), University of Basra for the academic years 2017–2018 and 2018–2019. The data were imported from the Examination Committee System to Microsoft (MS) Excel table tools with DM add-in, which was installed on a laptop. The first step in data pre-processing is preparing the data by removing records with empty values and converting the data for processing. A total of 50 records with empty values were under one or more columns. After removing these records, a total 499 records were obtained. The record values were then converted for data processing in Weka 3.8 with its built-in classifiers. The research sample (499 students) was an acceptable sample CSIT population with an error margin of 10% [48]. The second step involves measuring the consistency of the dataset by finding the Cronbach's alpha [49, 50], as show in table 2. The formula is calculated as follows:

$$\alpha = \frac{k}{k-1}\left(1 - \frac{\sum S_i^2}{S_T^2}\right) \ldots\ldots. (1)$$

Where k represents the item number, S_i^2 is the variance of the *i*th item and S_T^2 represents the total scored variance of all items of the dataset.

Table 2. Cronbach's Alpha of the Dataset

Number of Features	*Sum of Features' Variances*	*Sum of All the Records' Variances*	*Cronbach's Alpha*
8	38	155.3452	0.863295

The calculated Cronbach's alpha (0.86) showed a very satisfying consistency and a high internal reliability amongst the dataset items. The datasets of the model are the key steps for creating the DM model with the following columns of student datasets listed in Table 3. Table 3 shows students' dataset, where the attributes taken are the number of students (1–499), study year (2017–2018, 2018–2019), gender (female or male), students' birth year (e.g. 1997), registration (first or repeat), employment (yes or no), activity points (0–50), examination points (0–50) and final points (0–50). The last column was set as the predictable attribute which is the grade ('F,' 'P,' 'M,' 'G,' 'V' and 'E').

Table 3. Examples of Students' Dataset

Student Number	*Study Year*	*Gender*	*Birth Year*	*Registration*	*Course*	*Employment*	*Activity Point (40)*	*Examination Point (60)*	*Final Point (100)*	*Grade*
1	2016–2017	Female	1997	1	P1	Yes	32	31	63	M
2	2016–2017	Female	1997	1	P1	Yes	20	19	39	F
3	2016–2017	Female	1997	1	P1	No	20	8	28	F
4	2016–2017	Female	1997	1	P1	Yes	15	6	21	F
5	2016–2017	Female	1997	1	P1	No	16	6	22	F

6	2016–2017	Female	1997	1	P1	No	34	26	60	M
7	2016–2017	Female	1997	1	P1	No	26	13	39	F
8	2016–2017	Male	1997	1	P1	No	34	14	48	F
9	2016–2017	Male	1997	1	P1	No	43	49	92	E
.										
.										
499	2017–2018	Female	1998	1	P2	No	25	13	38	

The DM technology was selected for the next step. MS provides three analytical options for the DM level. The basic, intermediate and expert levels include MS Excel table tools, MS Excel DM add-in features and MS SQL Server DM capabilities, respectively. The basic level was selected in this research.

6.3 Attribute Evaluation

Four criteria were used to measure the efficiency of the seven algorithms and the performance of the supervised machine learning algorithms of the model. These criteria were true positive (TP) rate, false positive (FP) rate, precision and recall attributes. Equations (2)–(5) display these attributes. TP rate (sometimes called sensitivity) indicates the proportion of the number of true predictions in the positive prediction:

$$TP\ Rate = \frac{TP}{P} = \frac{TP}{TP+FN}, \quad \text{..... (2)}$$

Where TP and FN are the numbers of true detected and undetected errors, respectively. FP rate (sometimes called specificity) is the proportion of the number of expected negatives:

$$FP\ Rate = \frac{FP}{P} = \frac{FP}{FP+TN} \text{..... (3)}$$

Precision is the percentage of complete TP matches out of all the TP matches:

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

If precision is close to one, then the expectations slowly become precise. Recall is the percentage of the TP matches out of all the possible positive matches:

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

7. Result

This research explored the possibility of predicting students' exact grade, success and failure on the basis of different input variables obtained in HEIs. The model was developed using several supervised machine learning algorithms, and the results were compared. Weka was installed and loaded on these algorithms. Classifiers were used for testing option–cross

validation, and the data size was 499, which was divided into 70% training data (349 instances) and 30% test data (150 instances) for all algorithms to be implemented. Table 4 lists the performance criteria of different supervised machine learning algorithms after implementing the model to predict the exact final grade of students. The Logistic Regression Classifier was the most accurate (66%) amongst other algorithms.

Table 4. Performance Criteria of the Actual Grade Prediction Model

No.	*Category*	*Algorithm*	*TP Rate*	*FP Rate*	*Precision*	*Recall*
1	DT	Decision Stump	0.612	0.132	0.483	0.612
		Hoeffding Tree	0.586	0.283	0.490	0.586
		J48	0.673	0.133	0.632	0.673
		LMT	0.681	0.129	0.646	0.681
		Random Forest	0.659	0.110	0.646	0.659
		Random Tree	0.624	0.120	0.602	0.624
		Rep Tree	0.667	0.128	0.646	0.667
2	NB	Bayes Net	0.683	0.122	0.661	0.683
		NaiveBayes	0.675	0.139	0.642	0.675
		NaiveMulti	0.520	0.520	0.270	0.520
		Naiveupdate	0.675	0.139	0.642	0.675
3	MLP	NeuralN	0.663	0.135	0.615	0.663
4	SMO	SMO	0.631	0.214	0.538	0.631
5	Logistic Regression	Logistic	0.687	0.119	0.658	0.687
		SimpleLogistic	0.681	0.129	0.646	0.681
6	KNN	IBK (K Nearest)	0.633	0.119	0.611	0.633
		KStar	0.665	0.143	0.612	0.665
		LWL	0.618	0.130	0.510	0.618
7	Others	JRip	0.629	0.257	0.532	0.626
		OneR	0.661	0.130	0.617	0.661
		PART	0.649	0.135	0.610	0.649

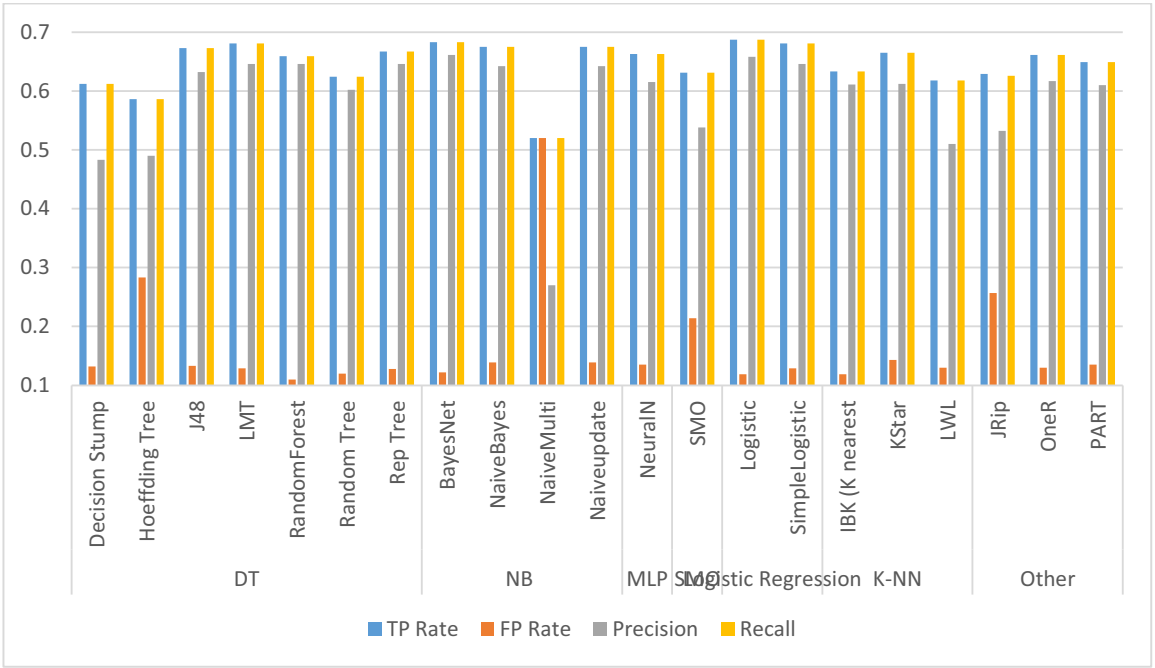

Figure 2. Performance Criteria of the Actual Grade Prediction Model

Figure 2 illustrates the performance criteria of the seven supervised machine learning algorithms. In the DT section, LMT presented the highest score in the TP rate (68.1%), followed by J48 (67.3%), Rep Tree (66.7%), Random Forest (65.9%), Random Tree (62.4%), Decision Stump (61.2%) and Hoeffding Tree (58.6%). In terms of BN, the highest TP rate score (68.3%) was demonstrated by BayesNet, followed by NaļveBayes and Naļveupdate (67.5%) and NaļveMulti (52%). NeuralN scored 66.3% in the MLP field, and SMO scored 63.1%. In Logistic Regression field, Logistic scored the highest score amongst all the algorithms with 68.7%, whereas SimpleLogistic scored 68.1%. In the KNN field, KStar, K Nearest and LWL scored 66.5%, 63.3% and 61.8%, respectively. Other algorithms, such as OneR, PART and JRip, scored 66.1%, 64.9% and 62.9%, respectively.

Low FP rate lowers the false prediction rate. Random Forest scored the lowest value (11%), followed by Random Tree (12%), Rep Tree (12.8%), LMT (12.9%), Decision Stump (13.2%), J48 (13.3%) and Hoeffding Tree (28.3%). In terms of NB, BayesNet scored 12.2%, followed by NaļveBayes and Naļveupdate (13.9%) and NaļveMulti (52%). In the MLP field, NeuralN scored 13.5%, followed by SMO with 21.4%. For Logistic Regression, Logistic and SimpleLogistic scored 11.9% and 12.9%, respectively. In the KNN field, K Nearest, LWL and KStar scored 11.9%, 13% and 14.3%, respectively. OneR, PART and JRip, scored 13%, 13.5% and 25.7%, respectively.

For the precision criterion in the DT section, LMT, Rep Tree and Random Forest showed the highest score of 64.6%, followed by J48 (63.2%), Random Tree (60.2%), Decision Stump a (61.2%), Hoeffding Tree (49%). In terms of BN, the highest precision score for all algorithms was obtained by BayesNet (66.1%), followed by NaļveBayes and Naļveupdate (64.2%) and NaļveMulti (27%). NeuralN scored 61.5% in the MLP field, whereas SOM scored 53.8%. In terms of Logistic Regression, Logistic and SimpleLogistic scored 65.8% and 64.6%, respectively. In the KNN field, KStar, K Nearest and LWL scored 61.2%, 61.1% and 51%, respectively. Other algorithms, such as OneR, PART and JRip, scored 61.7%, 61% and 53.2%, respectively. The field of recall criterion scored the same values as that of the TP rate.

Table 5. Performance Criteria of Students' Status Prediction Model

No.	*Category*	*Algorithm*	*TP Rate*	*FP Rate*	*Precision*	*Recall*
1	DT	Decision Stump	0.884	0.120	0.886	0.884
		Hoeffding Tree	0.845	0.156	0.845	0.845
		J48	0.876	0.127	0.877	0.876
		LMT	0.880	0.122	0.880	0.880
		Random Forest	0.876	0.127	0.876	0.876
		Random Tree	0.851	0.149	0.851	0.851
		Rep Tree	0.878	0.125	0.879	0.878
2	NB	BayesNet	0.884	0.120	0.886	0.884
		NaiveBayes	0.876	0.127	0.877	0.876
		NaiveMulti	0.520	0.520	0.270	0.520

		Naiveupdate	0.876	0.127	0.877	0.876
3	MLP	NeuralN	0.873	0.129	0.874	0.873
4	SMO	SMO	0.876	0.128	0.878	0.876
5	Logistic Regression	Logistic	0.888	0.114	0.888	0.888
		SimpleLogistic	0.880	0.122	0.880	0.880
6	KNN	IBK (K Nearest)	0.855	0.145	0.855	0.855
		KStar	0.884	0.120	0.886	0.884
		LWL	0.882	0.122	0.884	0.882
7	Others	JRip	0.871	0.132	0.873	0.871
		OneR	0.878	0.125	0.878	0.878
		PART	0.871	0.130	0.872	0.871

Table 5 shows the performance criteria of different machine learning algorithms after implementing the model to predict if the students passed or failed; the accuracy of logistic regression was 89%. Figure 3 displays the chart of performance criteria for all categories and algorithms used for predicting student status (passed or failed).

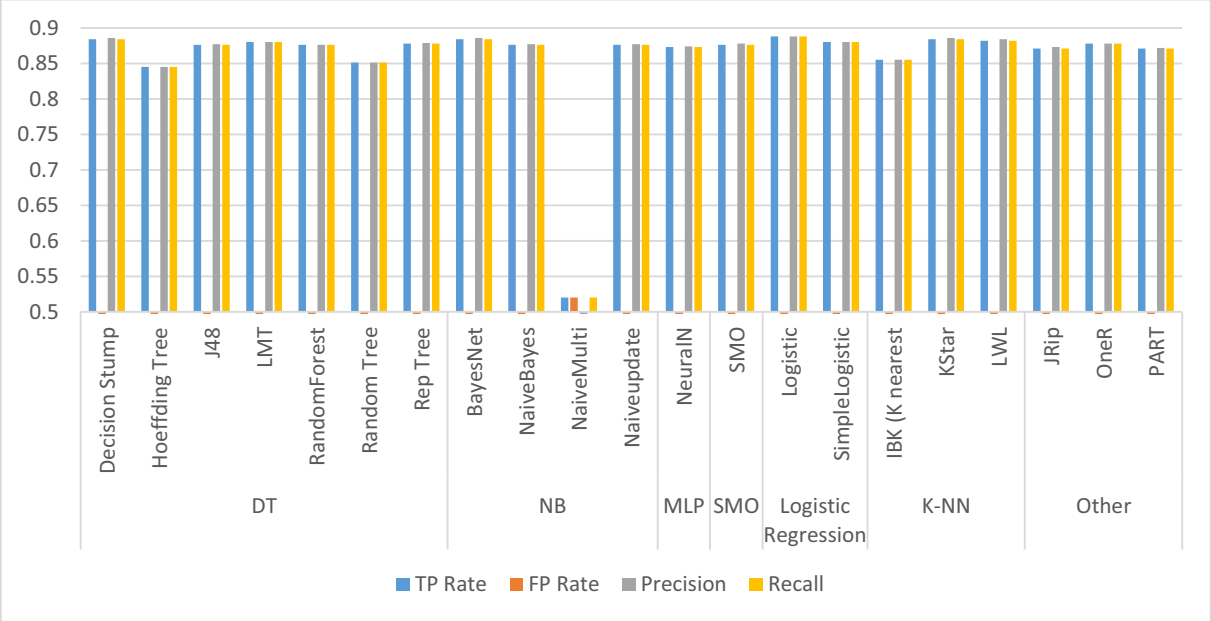

Figure 3. Performance Criteria of Students' Status Prediction Model

Figure 3 shows the performance criteria of the seven supervised machine learning algorithms. In the DT section, Decision Stump presented the highest score in the TP rate (88.4%), followed by LMT (88%), Random Forest and J48 (87.6%), Random Tree (85.1%) and Hoeffding Tree (84.5%). For BN, the highest TP rate score (88.4%) was achieved by BayesNet, whereas NaļveBayes and Naļveupdate scored 87.6%, and NaļveMulti obtained 52%. SMO scored 87.6%, whereas NeuralN scored 87.3% in the MLP field. For Logistic Regression, Logistic scored the highest score amongst all the algorithms with 88.8%, whereas SimpleLogistic

scored 88%. In the KNN field, KStar, LWL and K Nearest scored 88.4%, 88.2% and 85.5%, respectively. OneR scored 87.8%, and PART and JRip scored 87.1%.

In the field of the FP rate with DT algorithms, Decision Stump scored the lowest value (12%), followed by LMT (12.2%), Rep Tree (12.5%), J48 and Random Forest (12.7%), Random Tree (14.9%) and Hoeffding Tree (15.6%). In the NB field, BayesNet scored 12%, NaļveBayes and Naļveupdate scored 12.7% and NaļveMulti scored 52%. NeuralN scored 12.9% in the MLP field, whereas SOM scored 12.8%. The Logistic Regression field scored the lowest score with 11.4% for Logistic and 12.2% for SimpleLogistic. In the KNN field, KStar, LWL and k nearest scored 12%, 12.2% and 14.5%, respectively. Other algorithms, such as OneR, PART and JRip scored 12.5%, 13% and 13.2%, respectively.

For the precision criterion in the DT section, Decision Stump scored the highest score amongst all DT algorithms with 88.6%, followed by LMT (88%), Rep Tree (87.9%), J48 (87.7%), Random Forest (87.6%), Random Tree (85.1%) and Hoeffding Tree (84.5%). In terms of BN, BayesNet scored the highest precision amongst all algorithms with 88.6%, whereas NaļveBayes and Naļveupdate scored 87.7%, and NaļveMulti scored 27%. In the MLP field, NeuralN scored 87.4%, whereas SMO scored 87.8%. In the Logistic Regression field, Logistic scored the highest score amongst all algorithms with 88.8%, whereas SimpleLogistic scored 88%. For KNN, KStar, LWL and K Nearest scored 88.6%, 88.4% and 85.5%, respectively. Other algorithms, such as OneR, JRIP and PART scored 87.8%, 87.3% and 87.2%, respectively. The field of recall criterion scored the same values as that of the TP rate.

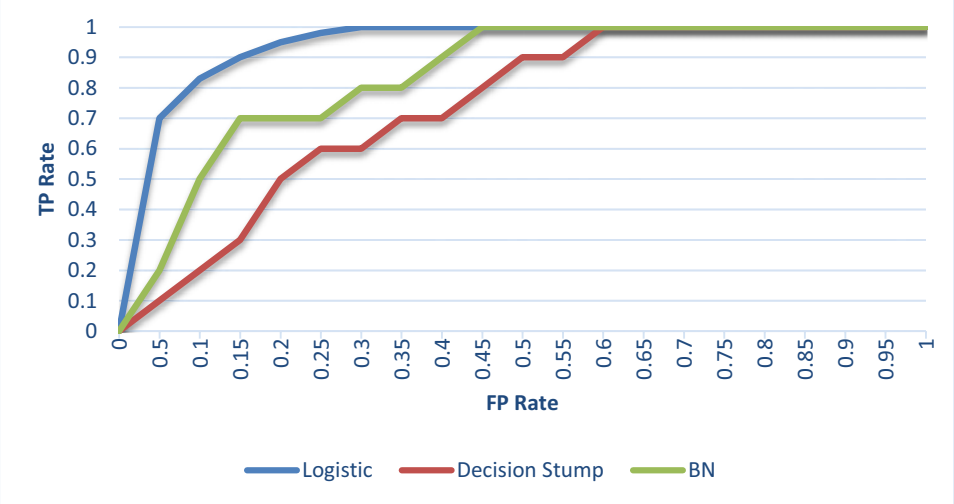

Figure 4. Receiver Operating Characteristic (ROC) Curve of Students' Status Prediction

ROC is a useful tool for evaluating the performance of classifiers. ROC is used for signal detection theory and radar image analysis. ROC basically presents the trade-off between TP and FP rates as a plot diagram. ROC can help analyse and recognise how the model can accurately

perform or mistakenly identify negative cases as positive. When the ROC curve reaches the value of 1, the model is accurate in predicting the positive cases. When the ROC curve reaches the diagonal line or 0.5, the model has low accuracy in predicting the values [51, 52]. Figure 4 shows the ROC of the best three accurate supervised algorithms (Logistic, Decision Stump DT and BN) in predicting student status as 'pass.' Given that the logistic curve reached the value of 1 in most cases, the ROC curve of the first model showed that the Logistic algorithm was the best in predicting student status, followed by Decision Stump DT and BN. The ROC curve of the Logistic algorithm started from the value above 0.5 and moved on to reach the value of 1. The area under curve (AUC) measures the entire area under the ROC curve. The AUC value can reflect the model performance; if the value is close to 1, then the model performs well. The AUC values of Logistic, BN and Decision Stump algorithms were 0.9541, 0.9346 and 0.8627, respectively.

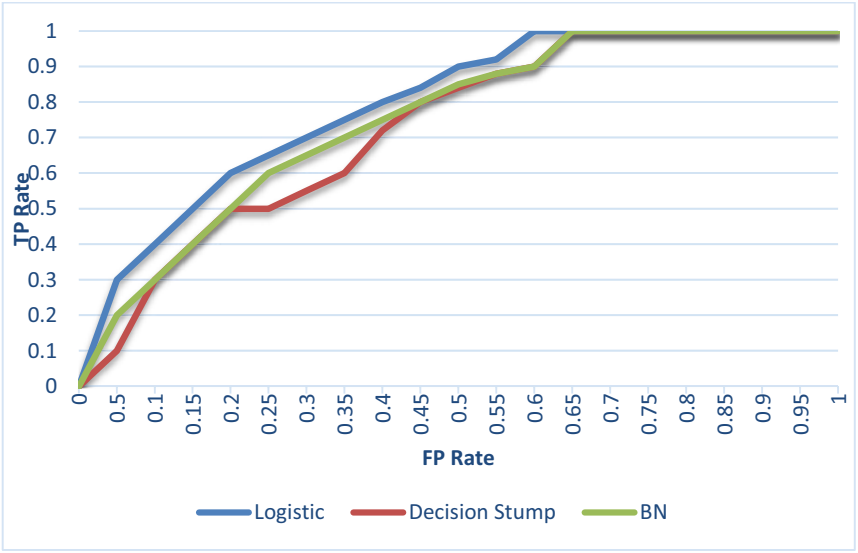

Figure 5. ROC Curve of the Actual Grade Prediction

Figure 5 shows the ROC curves of the three most accurate supervised algorithms (Logistic, Decision Stump DT and BN) in the second model for predicting students' actual grade 'M.' The ROC curve of the second model showed that the best model for predicting student status was the Logistic algorithm because its curve reached the value of 1 in most cases, followed by BN and Decision Stump DT. The AUC values of Logistic, BN and Decision Stump algorithms were 0.9012, 0.8893 and 07984, respectively. The AUC value of the Logistic algorithm was the best amongst the three, indicating the accuracy of model prediction. The ROC of the first model was better than that of the second one because the number of items in the final class was two,

whereas that of the predicted items in the second model was six, suggesting that the second model was more accurate than the first model.

8. Conclusion

Predicting student performance is important in the educational domain because student status analysis helps improve the performance of institutions. Different sources of information, such as traditional (demographic, academic background and behavioural features) and multimedia databases, are often accessible in educational institutions. These sources help administrators find information (e.g. admission requirements), predict the timetable scale of the class enrolment and help students decide how to choose courses depending on how well they will do in the chosen courses. The proposed model predicted student performance. This model was trained, and data were tested with students' data over two semesters by using several supervised machine learning algorithms, such as Decision Tree, NB, Logistic Regression, KNN, MLP, SMO, Neural Network, PART, JRip and OneR. The performance criteria of all algorithms were examined to predict two groups of results (the actual grade and the final status of students). The most perfect and precise results were obtained. The exact final grade and predicted student status (passed or failed) were shown by Logistic Regression, with accuracies of 68.7% and 88.8%, respectively. Many factors affected the accuracy of the results obtained after implementing the algorithms. These factors included the cleaned data, the domain of the features, the number of the features, the dataset size and the domain of the final class. The accuracy increased when the number of the predicted values was decreased. When the number of the values in the final class was six, the accuracy did not exceed 68.7%, but when this number became two, the accuracy exceeded 88.8%. Dataset size also affected the accuracy, that is, the accuracy increased when the size increased. The ROC and AUC can help determine the model accuracy by observing the curve values. After observing the ROC of the first model in predicting the student status as 'passed,' Logistic Regression was found to be the best algorithm, with the AUC of 0.9541, which is considered a good score. The observation of ROC for the second model in predicting the student actual grade 'M' also showed that Logistic Regression was the best algorithm for prediction, with the AUC of 0.9012. The AUC and ROC can help evaluate model performance and reflect the real accuracy in predicting the determined case.

REFERENCES

[1] J. Luan, "PhD Chief Planning and Research Officer," *Cabrillo College Founder, Knowledge Discovery Laboratories "Data Mining Applications in Higher Education*.

[2] R. S. Baker, "Educational data mining: An advance for intelligent systems in education," *IEEE Intelligent systems,* vol. 29, pp. 78-82, 2014.

[3] A. Hamoud, A. Humadi, W. A. Awadh, and A. S. Hashim, "Students' success prediction based on Bayes algorithms," *International Journal of Computer Applications,* vol. 178, pp. 6-12, 2017.

[4] A. K. HAMOUD, "CLASSIFYING STUDENTS'ANSWERS USING CLUSTERING ALGORITHMS BASED ON PRINCIPLE COMPONENT ANALYSIS," *Journal of Theoretical & Applied Information Technology,* vol. 96, 2018.

[5] S. K. Mohamad and Z. Tasir, "Educational data mining: A review," *Procedia-Social and Behavioral Sciences,* vol. 97, pp. 320-324, 2013.
[6] M. Berland, R. S. Baker, and P. Blikstein, "Educational data mining and learning analytics: Applications to constructionist research," *Technology, Knowledge and Learning,* vol. 19, pp. 205-220, 2014.
[7] Hamoud, Alaa, Ali Salah Hashim, and Wid Akeel Awadh. "Clinical Data Warehouse: A Review." *Iraqi Journal for Computers and Informatics* 44.2 (2018).
[8] A. K. Hamoud and A. M. Humadi, "Student's Success Prediction Model Based on Artificial Neural Networks (ANN) and A Combination of Feature Selection Methods," *Journal of Southwest Jiaotong University,* vol. 54, 2019.
[9] B. Guo, R. Zhang, G. Xu, C. Shi, and L. Yang, "Predicting students performance in educational data mining," in *2015 International Symposium on Educational Technology (ISET)*, 2015, pp. 125-128.
[10] I. A. Najm, A. K. Hamoud, J. Lloret, and I. Bosch, "Machine Learning Prediction Approach to Enhance Congestion Control in 5G IoT Environment," *Electronics,* vol. 8, p. 607, 2019.
[11] G. MeeraGandhi, "Machine learning approach for attack prediction and classification using supervised learning algorithms," *Int. J. Comput. Sci. Commun,* vol. 1, pp. 11465-11484, 2010.
[12] J. Han, J. Pei, and M. Kamber, *Data mining: concepts and techniques*: Elsevier, 2011.
[13] M. Gopal, *Applied Machine Learning*: McGraw-Hill Education, 2018.
[14] C. Verma, Z. Illés, and V. Stoffová, "Age group predictive models for the real time prediction of the university students using machine learning: Preliminary results," in *2019 IEEE International Conference on Electrical, Computer and Communication Technologies (ICECCT)*, 2019, pp. 1-7.
[15] S. Natek and M. Zwilling, "Student data mining solution–knowledge management system related to higher education institutions," *Expert systems with applications,* vol. 41, pp. 6400-6407, 2014.
[16] A. Hamoud, A. S. Hashim, and W. A. Awadh, "Predicting Student Performance in Higher Education Institutions Using Decision Tree Analysis," *International Journal of Interactive Multimedia and Artificial Intelligence,* vol. 5, pp. 26-31, 2018.
[17] E. Yukselturk, S. Ozekes, and Y. K. Türel, "Predicting dropout student: an application of data mining methods in an online education program," *European Journal of Open, Distance and e-learning,* vol. 17, pp. 118-133, 2014.
[18] M. Hussain, W. Zhu, W. Zhang, S. M. R. Abidi, and S. Ali, "Using machine learning to predict student difficulties from learning session data," *Artificial Intelligence Review,* vol. 52, pp. 381-407, 2019.
[19] F. Marbouti, H. A. Diefes-Dux, and K. Madhavan, "Models for early prediction of at-risk students in a course using standards-based grading," *Computers & Education,* vol. 103, pp. 1-15, 2016.
[20] A. Acharya and D. Sinha, "Early prediction of students performance using machine learning techniques," *International Journal of Computer Applications,* vol. 107, 2014.
[21] R. S. J. de Baker, T. Barnes, and J. E. Beck, "Educational Data Mining 2008," in *The 1st International Conference on Educational Data Mining Montréal, Québec, Canada*, 2008.
[22] A. S. Hashima, A. K. Hamoud, and W. A. Awadh, "Analyzing students' answers using association rule mining based on feature selection," *Journal of Southwest Jiaotong University,* vol. 53, 2018.
[23] E. Costa, R. S. Baker, L. Amorim, J. Magalhães, and T. Marinho, "Mineração de dados educacionais: conceitos, técnicas, ferramentas e aplicações," *Jornada de Atualização em Informática na Educação,* vol. 1, pp. 1-29, 2013.
[24] R. S. Baker and P. S. Inventado, "Educational data mining and learning analytics," in *Learning analytics*, ed: Springer, 2014, pp. 61-75.
[25] S. B. Jagtap, "Census data mining and data analysis using WEKA," *arXiv preprint arXiv:1310.4647,* 2013.

[26] I. H. Witten, "Data mining with weka," *Department of Computer Science University of Waikato New Zealand,* 2013.
[27] F. Akter, M. A. Hossin, G. M. Daiyan, and M. M. Hossain, "Classification of Hematological Data Using Data Mining Technique to Predict Diseases," *Journal of Computer and Communications,* vol. 6, p. 76, 2018.
[28] A. Hamoud, "Selection of best decision tree algorithm for prediction and classification of students' action," *American International Journal of Research in Science, Technology, Engineering & Mathematics,* vol. 16, pp. 26-32, 2016.
[29] G. Kostopoulos, S. Kotsiantis, and P. Pintelas, "Predicting student performance in distance higher education using semi-supervised techniques," in *Model and data engineering*, ed: Springer, 2015, pp. 259-270.
[30] K. P. Murphy, *Machine learning: a probabilistic perspective*: MIT press, 2012.
[31] X. Zhu and A. B. Goldberg, "Introduction to semi-supervised learning," *Synthesis lectures on artificial intelligence and machine learning,* vol. 3, pp. 1-130, 2009.
[32] P. Guleria, N. Thakur, and M. Sood, "Predicting student performance using decision tree classifiers and information gain," in *2014 International Conference on Parallel, Distributed and Grid Computing*, 2014, pp. 126-129.
[33] A. Hamoud, "Applying association rules and decision tree algorithms with tumor diagnosis data," *International Research Journal of Engineering and Technology,* vol. 3, pp. 27-31, 2017.
[34] K. Basu, T. Basu, R. Buckmire, and N. Lal, "Predictive Models of Student College Commitment Decisions Using Machine Learning," *Data,* vol. 4, p. 65, 2019.
[35] E. Osmanbegovic and M. Suljic, "Data mining approach for predicting student performance," *Economic Review: Journal of Economics and Business,* vol. 10, pp. 3-12, 2012.
[36] A. Géron, *Hands-On Machine Learning with Scikit-Learn, Keras, and TensorFlow: Concepts, Tools, and Techniques to Build Intelligent Systems*: O'Reilly Media, 2019.
[37] B. Godsey, *Think Like a Data Scientist: Tackle the data science process step-by-step*: Manning Publications Co., 2017.
[38] R. B. Millar, *Maximum likelihood estimation and inference: with examples in R, SAS and ADMB* vol. 111: John Wiley & Sons, 2011.
[39] G. Fitzmaurice and N. Laird, "Multivariate Analysis: Discrete Variables (Logistic Regression)," 2001.
[40] Y. Liu and H. Huang, "Fuzzy support vector machines for pattern recognition and data mining," *International journal of fuzzy systems,* vol. 4, pp. 826-835, 2002.
[41] J. Grus, *Data science from scratch: first principles with python*: O'Reilly Media, 2019.
[42] T. Hastie, R. Tibshirani, and J. Friedman, *The elements of statistical learning: data mining, inference, and prediction*: Springer Science & Business Media, 2009.
[43] V. Chaurasia and S. Pal, "A novel approach for breast cancer detection using data mining techniques," *International Journal of Innovative Research in Computer and Communication Engineering (An ISO 3297: 2007 Certified Organization) Vol,* vol. 2, 2017.
[44] J. Platt, "Sequential minimal optimization: A fast algorithm for training support vector machines," 1998.
[45] P. M. Arsad and N. Buniyamin, "A neural network students' performance prediction model (NNSPPM)," in *2013 IEEE International Conference on Smart Instrumentation, Measurement and Applications (ICSIMA)*, 2013, pp. 1-5.
[46] M. A. Ulkareem, W. A. Awadh, and A. S. Alasady, "A comparative study to obtain an adequate model in prediction of electricity requirements for a given future period," in *2018 International Conference on Engineering Technology and their Applications (IICETA)*, 2018, pp. 30-35.

[47] Ulkareem, Maysaa Abd, Wid Akeel Awadh, and Ali Salah Alasady. "A comparative study to obtain an adequate model in prediction of electricity requirements for a given future period." *2018 International Conference on Engineering Technology and their Applications (IICETA)*. IEEE, 2018.
[48] G. D. Israel, "Determining sample size," 1992.
[49] U. Sekaran and R. Bougie, *Research methods for business: A skill building approach*: John Wiley & Sons, 2016.
[50] M. Tavakol and R. Dennick, "Making sense of Cronbach's alpha," *International journal of medical education,* vol. 2, p. 53, 2011.
[51] A. P. Bradley, "The use of the area under the ROC curve in the evaluation of machine learning algorithms," *Pattern recognition,* vol. 30, pp. 1145-1159, 1997.
[52] J. Han, M. Kamber, and J. Pei, "Data mining concepts and techniques third edition," *Morgan Kaufmann,* 2011.