Towards a Multi Classifier Machine learning Based Approach for Course Cancelation Problem Avoidance

Obada Alhabaasneh*

O.alhabashneh@mutah.edu.jo

Abstract

The course cancelation problem (CCP) is one of the main problems of the university timetabling process. The problem usually occurs when the academic registry at the university cancels a course section for violating the "minimum number" hard constraint which refers to the minimum number of students enrolled in the course section. CCP is common within the universities in which the timetable is created before the student registraion process strats.

This paper discusses the development of a multi-classifier machine learning-based approach for course section cancellation risk estimation. The approach analyzes the enrollment historical data of the university to identify the common features of the canceled sections. These features include the course, section time-slot, the number of students who are eligible to take it, and the lecturer. These features are then associated with section cancellation status. The resulted dataset is fed into a multi-classifier component to predict the risk level of the section cancellation. The proposed approach aims to advice the academic departments in preparing the timetable of the upcoming academic term to avoid including the high risk courses in the time table which ,in turn, is expected to inance the stabilty of the time table by minimzing the number of canclection cases.

Results have shown that the proposed approach has achieved 85% classifying accuracy in identifying the cancelation risk level of sections before including them in the timetable. The classifying accuracy is expected to improve with the growth of the data volume. Also, using different gives the approach the dynamicity to use the most accurate classier to achieve the highest accuracy based on the provided case.

Keywords: Machine Learning, University Timetabling, Course Cancelation, Multi-Classifier.

---

* College of Business, Mutah University.
Received: 22/11/2020. Accepted: 19/1/2021.

نحو طريقة معتمدة على تعلم الآلة ومتعددة المصنفات لتجنب مشكلة شطب الشعب

عباده الحباشنة*

ملخص

تعد مشكلة شطب الشعب من المشاكل الأساسية التي تنطوي عليها عملية التسجيل الفصلي حيث أنه يتم شطب العديد من الشعب التي لم يبلغ عدد الطلبة المسجلين فيها العتبة المحددة لها حسب سياسة الجامعة وتعد هذه المشكل من المشاكل الشائعة في الجامعات التي تقوم بطرح الجدول لدراسي وتجهيزه قبل بدء عملية التسجيل.

تناقش هذ الورقة تطوير طريقة معتمدة على تعلم الآلة ومتعددة المصنفات لحل مشكلة شطب الشعب عن طريق توقع مستوى خطورة شطب الشعبة قبل طرحها وبذلك يتم تجنب طرح الشعب التي من المرجح شطبها أو إلغاؤها. تقوم الطريقة بتحليل البيانات التاريخية للتسجيل في الجامعة لتحدي السمات المشتركة للشعب التي تم إلغاؤها سابقا مثل: المادة، وقت الشعبة، عدد الطلبة المستحقين للمادة، والمدرس. وبعد ذلك يتم إدخال البيانات الناتجة إلى المكون متعدد المصنفات حتى يقوم بتحديد مستوى الخطور المترافق مع طرح الشعبة على الفصل الأكاديمي المقبل حيث يتكون من عدة مصنفات ويم أخذ نتيجة المصنف الأكثردقة. تهدف هذه الطريقة لتوفير آلية إنذار مبكر للأشخاص القائمين على إعداد الجدول الدراسي لتخفيض عدد الشعب الملغاة كل فصل وذلك بتجنب طرح الشعب ذات مستوى الخطورة العالي.

أظهرت النتائج أن الطريقة المقترحة حققت دقة تصنيف قدرها ٨٥% في توقع احتمالية إلغاء الشعب وأن هذه النسبة مرشحة للزيادة مع تزايد حجم البيانات المستخدمة في تدريب المصنفات وباستخدام تعلم الآلة.

الكلمات الدالة: التعلم الآلي، الجدول الزمني للجامعة، إلغاء الدورة، المصنف المتعدد.

---

* كلية الأعمال، جامعة مؤتة.

تاريخ تقديم البحث: 2020/11/22م. تاريخ قبول البحث: 2021/1/19 م .

1- Introduction

Timetabling is one of the key tasks in the enrollment process in the universities as the course offering must be ready for the students every academic term brfore they start their study. Usually, the academic departments prepare the course timimtable in collaboration with the academic registry to ensure it is convenient for the students, complying with the academic plan and the university policy. However, the credit hours system adopted by different universities over the world requires that the student have the right to select the courses offered by the university based on their preferences of the class time and lecture. This requires, usually, to offer different sections of the same course with different timeslots and/or different lecturers to give students enough flexibility on their choices. Moreover, the timetable must also be compatible with the lectures' specialties and availability. Preparing a suitable timetable that meets all these requirements might sound simple at the theoretical level, but it is more complicated in practice as such it must satisfy the students, lecturers, and university constraints including the academic plans and other academic and financial constraints.

Due to the prolonged global financial hardship, financial modeling has become a critical factor at all levels in university planning, in general, and in course offering in particular. To minimize cost, the course offering process often applies a strict section cancelation policy that identifies a minimum number of student enrollments, referred to as the threshold, required to survive section cancelation. The threshold is identified based on different factors such as the nature of the course (lecture-based or lab-based), the program level (BSc, Masters, Ph.D.), the course role in the academic plan (mandatory or elective), and the course's level (university, faculty or program).

Such policy helps the university minimizing the teaching cost, but it causes problems for the students and the academic departments due to regular cancelation of significant sections. The cancellation usually happens in the early weeks of the academic term. This causes a huge disturbance since the students need to look for an alternative section, or even a course, to replace the canceled section. Even though, this might go smoothly in some cases, finding the alternative often involves complexities related to resolving conflicts in sections timeslot. In some other cases, the situation becomes more complicated especially with the students who are graduating on that term or the following one as they need to apply for permission to take a course that is not part of their academic plan and it is the responsibility of the academic department to identify which course to take. This leads to a shift of the actual start of the learning and teaching process as the timetable takes two to three weeks to stabilize.

University Course Timetabling Problem (UCTTP) is an optimization problem that occurs every semester .(Asmunim,2015) (Babaei, Karimpour & Hadidi,2015). The goal of the university course timetabling problem (UCTTP) is to create a timetable in which all events are assigned to a set of predefined timeslots and rooms in such a way all constraints (hard and soft) are satisfied .(Babaei, Karimpour & Hadidi,2015). A set of events include the subsets: courses, students, and lectures. Constraints have two categories; hard and soft. Hard constraints refer to those conditions that must be met such as minimum and maximum number of students in the class and there is no tolerance for violation. Soft constraints refer to those constraints that are preferred to satisfy to make the timetable more efficient such as minimizing the time gaps between students' classes or minimizing the distance between the rooms for students .(Asmunim,2015) (Babaei, Karimpour & Hadidi,2015).

Course/section cancelation problem (CCP) can be defined as the situation in which the academic registry at the university cancels a course or a course section because of the violation of one of the hard constraints which is usually the minimum number of students enrolled in the course. This problem is quite common in the university in which the timetable is created before the enrollment process. In this case, course sections are offered and assigned to timeslots and lectures before students enrollment, Consequently, the students have the choice to select the courses based on their preferences under the condition of satisfying the academic plan constraints such as the prerequisites and the maximum number of courses taken of each requirement type in the academic plan of the student's program. This margin of selection freedom often leads to a lack of enrollment in some sections due to the tendency of students to avoid certain time slots (early morning and evening), lectures, or courses. These sections are usually canceled especially if didn't achieve the minimum number of students which causes a significant distribution for the teaching process as the students in the canceled courses start to look for other courses to enroll again. Course cancellation usually takes place after the first week of the semester. This leads to a delay in the actual start of the teaching process and long queues of students asking for the advice of the academic departments. Another complication of this problem is that the academic departments in many cases must reassign lecturers to courses to meet the academic load of each lecturer in the department.

In the light of the above, providing an effective approach to help the academic departments in preparing an optimal course offering that meets the student's requirements and minimizes the number of the canceled sections has become imperative.

This paper discusses the development of multi-classifier machine learning- based approach for course's section cancellation risk estimation. The approach analyzes the enrollment historical data of the university to identify the common features of the canceled sections. Such features may include the course itself, section-time slot, the number of students who are eligible to take it, the lecturer. These features are then with section cancellation status. The resulted data set then is red into a multi classifier to predict the likelihood of the section cancellation. The proposed approach aims to assist the academic departments in preparing the timetable of the upcoming academic term by predicting the likelihood of the section cancellation before including it in the timetable. This in turn is expected to minimize the number of the canceled section.

2- Related Work

The ACTTP has been studied extensively and several researche have attempted to provide efficient solutions. However, due to the high dimensionality of the problem together with a variance of the rules applied in different countries and universities, providing a comprehensive approach that can overcome all subproblems sounds nonrealistic which makes this research area still lacking.

One of the earliest attempts to solve this problem used graph coloring problem in 1967(Welsh,1967). The graph coloring-based approach was further developed to solve the timetabling problem using a non-directional graph to prevent the conflict in the timetable (Werra,1985). To reduce the number of resulted graphs a separated graph was used on separate students' groups (Selim,1988). This method has been modified to direct the vertices and edges of the graph to common courses to create a uniform distribution of the courses (Amal & Mayez,2010).

Another graph-based approach has been proposed using genetic coloring to find the least colors of the graph (Asham, Soliman& Ramadan,2011). A genetic algorithm has been proposed for maximizing the number of enrolled students in the course using bipartite graphs (Long,2019). A modified grouping-based algorithm has been proposed to solve the course timetabling problem (Kralev, Kraleva& Kumar,2019).

The integer programming(IP) method is used to split the timetabling into two subproblems: timetabling and grouping. An IP-based heuristic approach was applied to the two groups until finding the optimal timetable (Aubin,Ferland,1989). Another IP-based method was introduced to solve the timetabling problem. The method aimed to produce the best fit timetable by allocating a set of courses to a lecturer, groups of students, and timeslots (Daskalaki,Birbas & Housos,2004). An IP-based relaxation method was proposed to create an optimal timetable on two steps (Daskalaki,Birbas,2005). The courses with one timeslot were timetabled in Step.1 while the courses with more than one timeslot were timetabled in Step.2. The IP has also been used to improve the rule finite and constraint satisfaction in timetabling (Bakir & Aksop 2008). Lecturer university timetabling was looked at as an NP-hard problem (Luisa& Meurant,2020)and both algorithms were evaluated on this base. Genetic Algorithm was shown to outperform Hybrid Genetic Algorithms-Hill Climbing in finding the optimal solution. (Yusoff,Roslan,2019).A binary integer programming formulation has been used to solve the university timetabling problem(Erim Chung & Kim,2019).

Constraint satisfaction programming (CSP)has also been used to solve the time baling problem. A constraint-based approach has been integrated with a genetic algorithm to find the optimal timetable (Deris, Omatu & Ohta,1999). In another research, the object-oriented approach and constraint-based reasoning have been combined to find an optimal timetable (Deris,Omatu,&Ohta,2000).In more recent research, constraint satisfaction programming has been used to find the objective function to allocate resources required for an optimal timetable[19](Lixi, SimKim,2005). A Constraint satisfaction (CS) model was introduced in (Lixi, SimKim,2005) to automatically construct the university timetable. The model uses the university constraints such as the academic plan, time slots lecture availability to produce the timetable. A list of hard and soft problems relating to university timetabling was provided in (Herres, Schmitz,2019).

The list included different problems such as the time overlapping and the lecturer's double booking.

Other researchers adopted a heuristic Approach. For example, in the Greedy Randomized Adaptive Search Procedure (GRASP), a maximum flow partial solution and simulated Annealing were integrated to find the optimal university timetable (Lixi&SimKim,2005). The introduced componential model was developed using the second International Timetabling Competition (ITC-2007). A meta-heuristic algorithm was integrated with a non-dominated sorting genetic algorithm to find the best solution for the timetabling problem (Al-ghamdi,2012). The education quality measure was used as a base for the solution selection. The education quality included different factors such as the teaching capabilities of lecturers, student counseling, and tutorials time during the day. The Simulated Annealing (Kirkpatrick,Gelatt&Vecchi,1983) was used to optimize the Reheating (SAR) (Goh, Kendall& Sabar,2017) algorithm to find the optimal solution for the post-enrollment course problem. The introduced approach started with finding a feasible solution first then optimizing that solution (Goh, Kendall& Sabar,2019). Ant colony optimization (ACO) algorithm has been used to optimize the university timetabling. The algorithm was applied on Faculty of Informatics and Computing (FIC) datasets and achieved an acceptable accuracy level in assigning the offered courses to the available time slots (Mazlan, Makhtar& Ahmad Khairi,2019). A Genetic Algorithm was introduced in (Mazlan, Makhtar& Ahmad Khairi,2019) for the timetabling problem. The algorithm was designed to increase the convergence rate of the new schedule. An extensive computational experiment was conducted on the International Timetabling Competition ITC-2007 data set (Gülcü& Akkan,2020). In their experiment, the researchers applied a multi-objective simulated annealing (MOSA) algorithm to achieve the optimal solution for the soft constraint problem. Tabu search based approaches for timetabling were proposed in (Gaspero & Schaerf 2001) (Abdalla,Obit&Alfred,2019). The proposed approaches used the diversification of the neighborhood to find the optimal solution. The approach focused on reducing the number of potential solutions to speed up the optimization process.An agent-based framework for the university course timetabling was proposed in (Abdalla,Obit&Alfred,2019). The framework used distributed multi-agent system based on integer programming. The approach consisted of a central agent that is coordinating between the other agents to find the optimal solution. A model-driven approach has been proposed to solve the time tabling problem (Eke.,et al 2019). The approach used a model-driven UI to generate an initial solution then give the user the ability to adjust the solution.

Recently, machine learning has been used for university timetabling problem. A naïve Bayesian approach was proposed to identify preferred timeslots by lecturers to resolve the soft constraint (Tharwat,2018). Support Victor Machine and Leaner Regression were both used for feature extraction and selection to identify the hard and soft constraints. Neural Network was integrated with the heuristic approach for better performance in the class/teacher timetabling problem (CTTP) (la Rosa-Rivera.,et al 2020). Fuzzy logic has been used to solve the problem of course timetabling. A fuzzy c-mean based approach has been introduced to handle the uncertainty in clustering students' courses in a weekly program. The approach reduced conflict in the student's preferences and class time (Asmuni.,et al 2020). Another research (Rachmawati& Srinivasan,2005) proposed a hybrid fuzzy approach to students' projects. The approach used an objective satisfaction method for assigning students to projects. A fuzzy genetic heuristic has been applied to the University course timetabling problem (Chaudhuri&De,2010). The genetic heuristic has been used to match the events(classes) with the features while the fuzzy sets have been used to handle the soft constraints' violation. In another research, A fuzzy genetic algorithm has been integrated with local search to solve UCTP (Kohshori & Abadeh,2012). The fuzzy genetic algorithm has been used to produce the optimal timetable and the local search for improving the efficiency of the algorithm. Fuzzy logic has been integrated with a local search algorithm to address the ambiguity in soft constraints of the timetable and the local search was used to find the optimal solution (Babaei,Karimpour&Hadidi,2019). In another research (June .,et al 2020), the fuzzy logic has been integrated with a sequential constructive algorithm to optimize the course timetable. Fuzzy logic has been used expand produce multiple sequences that are normalized to find the best sequence. Fuzzy sets have been used in modeling the violations of the soft constraints (Perzina&Ramík,2013). The sets were incorporated in a self-learning genetic algorithm with indirect representation to produce a low-cost optimal course timetable.

3- The Proposed Approach and its Results.

As shown in Figure(1), this paper proposes a multi classifier machine learning based framework to support the decision of the academic departments in building the timetable. It uses the classifiers to identify the cancelation risk level of the suggested course sections during the timetable construction process. The aim of using different classifiers is to maximize classifying accuracy. This can be achieved by evaluating the accuracy performance of different classifiers first. Consequently, the classifier with the highest accuracy performance is selected.

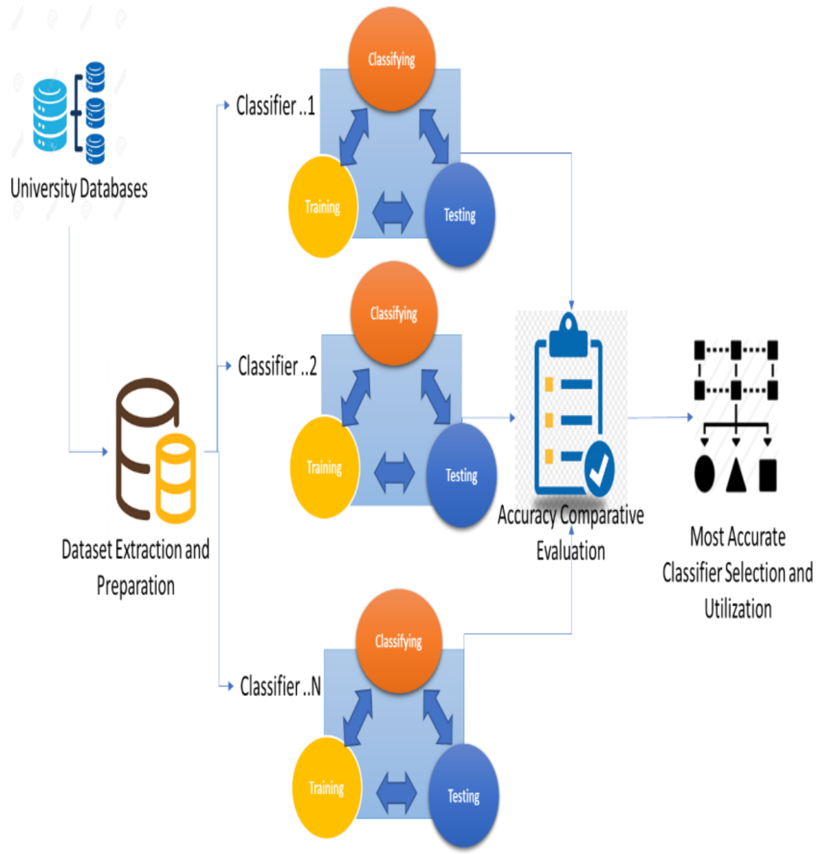


Figure 1: The Proposed Approach

1.1 Dataset Extraction and Preparation.

1.1.1 Data Collection

In this step, a dataset is collected from the university enrollment and registration databases. The collected data may include data related to the number of the students who are eligible to take each course, the minimum number of students allowed for each course section, the academic plan for each academic program, the average number of sections canceled for each lecturer in the past, the average number of sections that have been previously canceled for each course and the status of each course section in the time table history.

The proposed framework has been developed, trained, and evaluated based on real data collected from MU academic registry system databases. The dataset used in this study has been extracted from the Mutah University (MU) enrollment records focusing on the data of the canceled course' sections over the last 5 years. The dataset consisted of 5000 records and included data on the timetable history, lectures, courses, and academic plans of the academic programs.

MU is one of the largest public universities in Jordan. It has been founded in 1981 to be the first university in the southern part of Jordan. MU encompasses two wings; Military and Civilian. The Civilian wing contains 14 colleges: Arts, Business, Law, Social Sciences, Social Sciences, Educational Sciences, Sport Sciences, Medicine, Pharmacy, Nursing, Engineering, Science, Information Technology and Agriculture. The university offers 56 academic programs on Bachelor, Master, and Doctorate levels. The university has around 18,000 students currently enrolled in all programs.

Like the other Jordanian universities, MU adopts the credit hour system in which each credit hour equates to three class hours per week throughout the academic semester which spans over 16 weeks. Each program has an academic plan that defines the program requirements and their dependencies. The program completion requirements are released in the form of courses which are of different requirement types such as university-level mandatory, university-level elective, college-level mandatory, college level elective, program level mandatory, and program level elective. Each requirement type has a specific number of credit hours that should be completed successfully by the students to graduate. The academic plan defines course dependencies by enforcing a prerequisite mechanism by which the student can't enroll in a course until he/she completed (either successfully or unsuccessfully) the prerequisite in a previous semester.

1.1.2 Dataset Prepossessing and construction

In this step, the collected data is processed and prepared for training the classifiers. In our case, the dataset consisted of six features (predictors) and one target as shown in Figure (2).

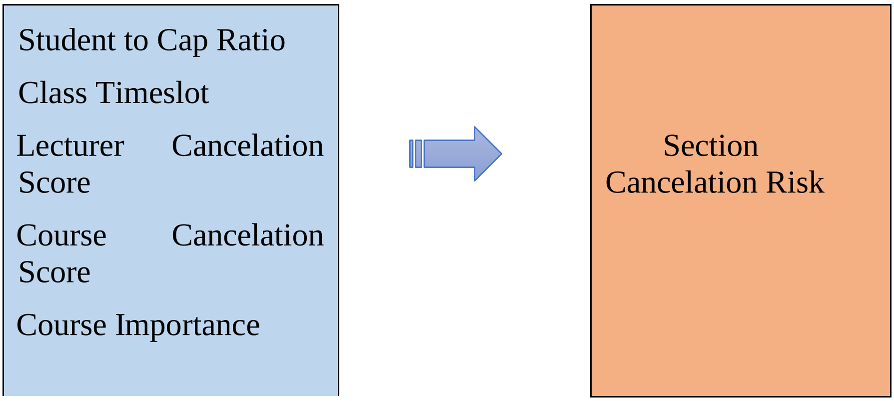


Figure (2) Predictors – Target

The table below describes how these features have been constructed and calculated

Table (1) Features Values Calculation

	FEATURE	DESCRIPTION	VALUE GUIDE
1	The student to Cap Ratio	Represents the scaled ratio between the number of students who are eligible to take the course (according to the program academic plan constraints) and the minimum number of students to keep the section according to the university policy.	1: High (=<02) 2: Medium (01< X <02) 3: Low (01>= X)

2	Class Timeslot	Refers to the time of the day in which the section is scheduled. This feature can take one of three values.: Early, Middy or Late.	1: Early (08:00-10:00) 2: Middy (10:00-14:00) 3: Late (14:00-18:00)
3	Lecturer Cancelation Score	This is a score of the cumulative number of sections that were canceled for the lecturer in the past. It is a scaled ratio between the number of the sections canceled for the lecture and the average number of the section canceled for lectures across the university.	1: High (=<02) 2: Medium (01< X <02) 3: Low (01>= X)
4	Course Cancelation Score	This is a score of the cumulative number of sections that were canceled of a specific course in the past. It is a scaled ratio between the number of the sections canceled for the course and the average number of the section canceled for courses across the university.	1: High (=<02) 2: Medium (01< X <02) 3: Low (01>= X)
5	Course Importance	This is a scaled weight of the importance of the course for the students in the academic plan and takes its value based on the type of the requirement the course belongs to in the academic plan such as mandatory or elective and if the course is a prerequisite of other courses.	1: High 2: Medium 3: Low

1.1.3 Labeling

The data was labeled by associating the five features in the previous table with the scaled cancelation risk which has been calculated based on the cancelation history of the previous timetables and as follows:

1: High: refers to the case in which the course section was canceled from the first round of the timetable review due to not meeting the minimum number of students' constraints. In most cases, the number of students enrolled in these sections is significantly below the threshold. Therefore, sections in this category do not pass to the second round of the timetable review.

2: Medium: refers to the case where the course section passes the first round of the timetable review with several students that is slightly below or above the threshold. Hence, another chance in the second-round timetable review is granted for such sections.

3: Low: refers to the case in which the course section passes the first round with a students' number that is fairly above the threshold.

The table below shows a sample of the labeled dataset.

Table 2: Dataset Sample

Student to CapRatio	Timeslot	Lecture Cancelation Score	Course Cancelation Score	Course Importance	Cancelation Risk
1	1	1	1	3	1
1	2	1	1	3	1
1	1	1	1	3	1
1	1	1	1	3	1
1	3	1	1	3	1
1	1	1	1	3	1
1	1	1	1	3	1
1	2	1	1	3	1
1	1	1	1	3	1
1	1	1	1	3	1
1	1	1	1	3	1

1.2 Multi Classifiers training and testing

A supervised machine learning approach has been used to develop a model to predict the cancelation risk of course sections. As shown in Figure (3), different classifiers have been trained and tested on the dataset to compare their accuracy performance. In turn, this comparison provides an indicator of the optimal classifier for the data set. The used classifiers included Neural Networks, Decision Tree, Naïve Bayes, SMV, Random Forest, and KNN.

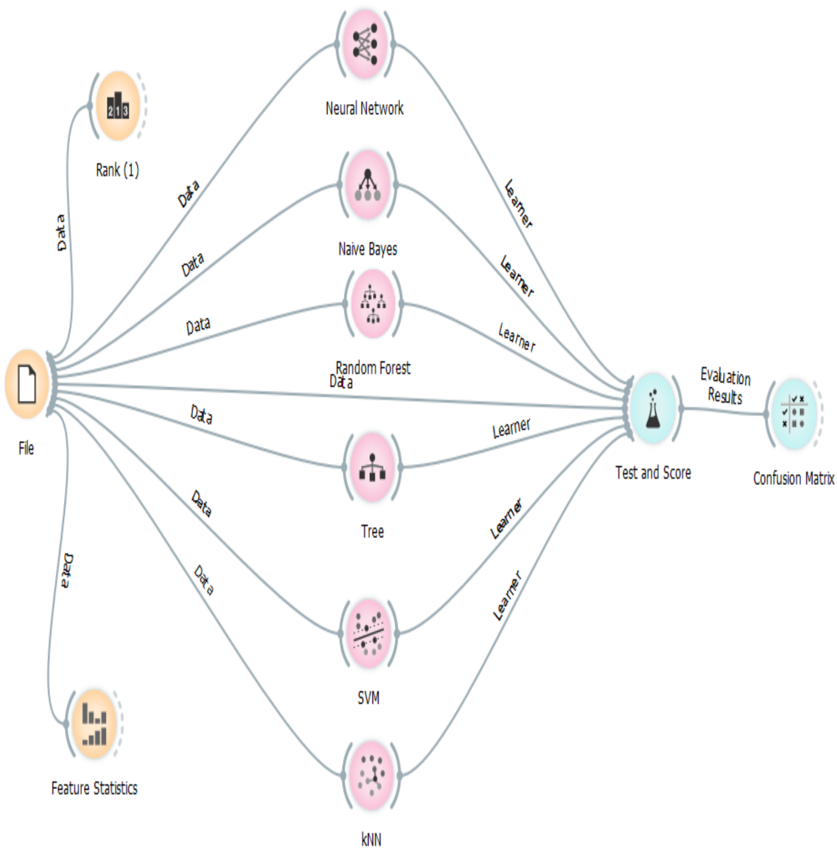


Figure 2: Multi-classifier Model

To train and evaluate the model's accuracy, a 5-fold cross-validation has been used. The dataset is divided into training and testing subsets for each fold where the model is trained and tested using the training and testing sets, respectively.

1.3 Classifying Accuracy Evaluation and Best Classifier Selection.

In this stage, the resulted classifying accuracy of the used classifiers are compared to identify the most accurate classifier to be used for the prediction of section risk of cancelation. The evaluation is performed based on two matrices which are the Classifying Accuracy (CA) (Rakesh Kumar.,et al 2013), which shows the average accuracy for each classifier for all classes, and the confusion matrix (Tharwat,2018), which identifies the classifying accuracy of each classifier for each class.

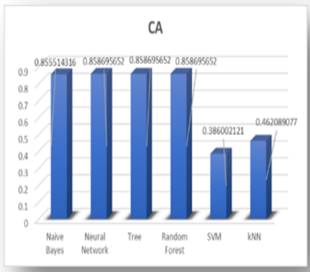
$$CA = (TP + TN)/ (TP + TN + FP + FN) \qquad \text{Equation1:CA}$$

Where: T: True; F: False; P: Positive; N: Negative

Table (3) shows the accuracy performance of each classifier.

Table 3: Accuracy Performance

MODEL	CA
NAIVE BAYES	0.8555
NEURAL NETWORK	0.8586
TREE	0.8586
RANDOM FOREST	0.8586
SVM	0.3860
KNN	0.4620



The table shows that Neural Networks, Decision Tree, Naïve Bayes, and Random Forest have achieved 85% classifying accuracy which is relatively acceptable having the size of the dataset. However, both SMV and KNN have poor classifying accuracy 38 % and 46%.

For more insight on the accuracy performance of well-performing classifiers, a confusion matrix test has been conducted as shown in the table below:

Table (4) Confusion Matrix Test

Neural Networks

Actual \ Predicted	1	2	3	∑
1	62.2%	9.9 %	27.9 %	**1410**
2	0.0 %	100.0%	0.0 %	**737**
3	0.0 %	0.0%	100.0%	**1625**
∑	**877**	**876**	**2019**	**3772**

Naïve Bayes

Actual \ Predicted	1	2	3	∑
1	90.1%	9.9 %	0.0 %	**1410**
2	0.0 %	100.0%	0.0 %	**737**
3	25.0 %	0.0 %	75.0%	**1625**
∑	**1677**	**876**	**1219**	**3772**

Random Forest

Actual \ Predicted	1	2	3	∑
1	62.2%	9.9 %	27.9 %	**1410**
2	0.0 %	100.0%	0.0 %	**737**
3	0.0 %	0.0 %	100.0%	**1625**
∑	**877**	**876**	**2019**	**3772**

Decision Tree

Actual \ Predicted	1	2	3	∑
1	62.2%	9.9 %	27.9 %	**1410**
2	0.0 %	100.0%	0.0 %	**737**
3	0.0 %	0.0 %	100.0%	**1625**
∑	**877**	**876**	**2019**	**3772**

The table shows that Neural Networks, Random Forest, and Decision Tree have the same classifying accuracy for all classes (1, 2 and 3) as they achieved 100% classifying accuracy for classes 2 and 3. However, they performed well in classifying class three correctly as they achieved 62.2 % classifying accuracy. On the other side, Naïve Bayes performed well in classifying class 1 (90.1%) correctly, however it was not as accurate as class 3 (75%). In our case, class 1 might be more important than other classes as it represents the courses' sections that have a high probability of cancelation; Therefore, such classes need to be detected in advance to avoid post cancelation disruption. In light of this fact, Naïve Bayes could be considered the most well-performing classifier for our dataset. However, the accuracy performance may vary based on the nature and the size of the data collected in the future.

2- Conclusion

This paper presents an intelligent approach for early-warning mechanisms that could be used by universities to minimize course cancelation cases. The mechanism is intelligent and adaptive such that it utilized the historical timetable and student enrollment data to predict the risk of course section cancelation. It is based on using the machine learning approach using an optimal classifier that is selected among different classifiers. The selection process is achieved through a comparative analysis based on their accuracy performance. In this paper, the data collected has been scaled and reorganized to consists of five main features that have been used in the classification process. These included Student to Student to Cap Rati, Class Timeslot, Lecturer Cancelation Score, Course Cancelation Score, and Course Importance.

In this paper, different classifiers were trained and tested on the collected dataset. The classifiers included Neural Networks, Decision Tree, Naïve Bayes, SMV, Random Forest, and KNN. Neural Networks, Decision Tree, Naïve Bayes, and Random Forest has performed similarly well (85% Classifying Accuracy) and outperformed SMV and KNN which bother had poor classifying accuracy(38 % and 46%). Although Neural Networks, Decision Tree, Naïve Bayes and RandomForest had a similar average classifying accuracy (85%), Naïve Bayes outperformed the Neural Networks, Random Forest, and Decision Tree in classifying class 1 which is the most important class in our case. However, the classifying accuracy is subject to change for all classifiers based on the nature and the size of the collected data. Based on the results, the paper concludes that the proposed approach is promising, and using different classifiers gives the advantage of selecting the most accurate one to use.

References

Asmuni,H.(2015) Fuzzy Methodologies for Automated University Timetabling Solution Construction and Evaluation, n.d.

Abdalla ,M., Obit ,J., Alfred ,R.& Bolongkikit, J.(2019). Agent based integer programming framework for solving real-life curriculum-based university course timetabling, in: Lect. Notes Electr. Eng., Springer Verlag, 2019: pp. 67–76. https://doi.org/10.1007/978-981-13-2622-6_7.

Al-ghamdi, A. et.(2012). An Expert System for Advising Postgraduate Students, Ijcsit. 3 (2012) 4529–4532.

Algorithms and Complexity - Bozzano G Luisa, Gerard Meurant - Google Books, (n.d.). https://books.google.jo/books/about/Handbook_of_Theoretical_Computer_Science.html?id=DypHwwEACAAJ&redir_esc=y (accessed November 15, 2020).

Asham ,M., Soliman ,M .& Ramadan ,R.(2011). Trans Genetic Coloring Approach for Timetabling Problem, 2011.

Asmuni, H., Garibaldi, E., Burke ,E., Garibaldi ,J.(2020). Fuzzy Multiple Heuristic Ordering for Course Timetabling, Proc. 5 TH UNITED KINGDOM Work. Comput. Intell. (UKCI05. (2005) 302--309. http://citeseerx.ist.psu.edu/viewdoc/summarydoi=10.1.1.59.6678 (accessed November 15, 2020).

Aubin, J.& Ferland ,J.(1989). A large scale timetabling problem, Comput. Oper. Res. 16 (1989) 67–77. https://doi.org/10.1016/0305-0548(89)90053-1.

Babaei, H., Karimpour, J& Hadidi,A.(2015). A survey of approaches for university course timetabling problem, Comput. Ind. Eng. 86 (2015) 43–59. https://doi.org/10.1016/j.cie.2014.11.010.

Babaei, H., Karimpour,J.& Hadidi, A.(2019). Generating an optimal timetabling for multi-departments common lecturers using hybrid fuzzy and clustering algorithms, Soft Comput. 23 (2019) 4735–4747. https://doi.org/10.1007/s00500-018-3126-9.

Bakir ,A & Aksop,C.(2008). A 0-1 integer programming approach to a university timetabling problem.73(1):41-55.

Chaudhuri, A.&De ,K.(2020). Fuzzy Genetic Heuristic for University Course Timetable Problem, 2010. www.i-csrs.org (accessed November 15, 2020).

Dandashi,A. & Al-Mouhamed,M.(2010). Graph coloring for class scheduling, in: 2010 ACS/IEEE Int. Conf. Comput. Syst. Appl. AICCSA 2010, IEEE Computer Society, 2010. https://doi.org/10.1109/AICCSA.2010.5586963.

Daskalaki, S. & Birbas ,T.(2005). Efficient solutions for a university timetabling problem through integer programming, in: Eur. J. Oper. Res., North-Holland, 2005: pp. 106–120. https://doi.org/10.1016/j.ejor.2003.06.023.

Daskalaki, S., Birbas, T.&Housos,E.(2004). An integer programming formulation for a case study in university timetabling, in: Eur. J. Oper. Res., North-Holland, 2004: pp. 117–135. https://doi.org/10.1016/S0377-2217(03)00103-6.

Deris, S., Omatu, S., Ohta, H. & Saad, p. (2000) P., Incorporating constraint propagation in genetic algorithm for university timetable planning, Eng. Appl. Artif. Intell. 12 (1999) 241–253. https://doi.org/10.1016/S0952-1976(99)00007-X.

Deris, S., Omatu, S., Ohta, H., (2000). Timetable planning using the constraint-based reasoning, Comput. Oper. Res. 27 819–840. https://doi.org/10.1016/S0305-0548(99)00051-9.

Eke ,C., Garcia-Dominguez ,A., Van Mourik, J.& Khan, I.(2019). A case study of model-driven engineering for automated timetabling, in: Proc. - 2019 ACM/IEEE 22nd Int. Conf. Model Driven Eng. Lang. Syst. Companion, Model. 2019, Institute of Electrical and Electronics Engineers Inc., 2019: pp. 203–210. https://doi.org/10.1109/MODELS-C.2019.00034.

Erim Chung, Y. & Kim, H. (2019). The University Examination And Course Timetabling Problem With Integer Programming, J. Korea Soc. Comput. Inf. 24 9–20. https://doi.org/10.9708/jksci.2019.24.09.009.

Gaspero, L & Schaerf,A.(2001). Tabu Search Techniques for Examination Timetabling**.** | Selected papers from the Third International Conference on Practice and Theory of Automated Timetabling III, (n.d.). https://dl.acm.org/doi/10.5555/646431.692915 (accessed November 15, 2020).

Goh, S., Kendall ,G& Sabar, N.(2019). Simulated annealing with improved reheating and learning for the post enrolment course timetabling problem, J. Oper. Res. Soc. 70 (2019) 873–888. https://doi.org/10.1080/01605682.2018.1468862.

Goh, S., Kendall ,G.& Sabar, N.(2017), Improved local search approaches to solve the post enrolment course timetabling problem, Eur. J. Oper. Res. 261 (2017) 17–29. https://doi.org/10.1016/j.ejor.2017.01.040.

Gülcü ,A.& Akkan ,C.(2020). Robust university course timetabling problem subject to single and multiple disruptions, Eur. J. Oper. Res. 283 (2020) 630–646. https://doi.org/10.1016/j.ejor.2019.11.024.

Herres, B.&Schmitz, H.(2019). Decomposition of university course timetabling: A systematic study of subproblems and their complexities, Ann. Oper. Res. (2019) 1–19. https://doi.org/10.1007/s10479-019-03382-0.

June ,T., Obit, J., Leau ,Y., Bolongkikit, J.& Alfred, R.(2020). Sequential Constructive Algorithm incorporate with Fuzzy Logic for Solving Real World Course Timetabling Problem, in: Lect. Notes Electr. Eng., Springer Verlag, 2020: pp. 257–267. https://doi.org/10.1007/978-981-15-0058-9_25.

Kirkpatrick ,S., Gelatt ,C.& Vecchi, M.(1983). Optimization by Simulated Annealing, Science (80-.). 220 (1983).

Kohshori ,M & Abadeh,M .(2012). Hybrid Genetic Algorithms for University Course Timetabling. International Journal of Computer Science Issues ,9(2),446-455.

Kralev ,V., Kraleva, R. & Kumar, S.(2019). A modified event grouping based algorithm for the university course timetabling problem, Int. J. Adv. Sci. Eng. Inf. Technol. 9 (2019) 229–235. https://doi.org/10.18517/ijaseit.9.1.6488.

la Rosa-Rivera, F., Nunez-Varela, J., Puente-Montejano, C.& Nava-Muñoz S.(2020) Measuring the complexity of university timetabling instances, J. Sched. (2020). https://doi.org/10.1007/s10951-020-00641-y.

Lixi, Z.& SimKim L.(2005). Constructing university timetable using constraint satisfaction programming approach, in: Proc. - Int. Conf. Comput. Intell. Model. Control Autom. CIMCA 2005 Int. Conf. Intell. Agents, Web Technol. Internet, 2005: pp. 55–60. https://doi.org/10.1109/cimca.2005.1631445.

Long, D.(2019). An enhanced genetic algorithm-based courses timetabling method for maximal enrollments using maximum matching on bipartite graphs, Vietnam J. Sci. Technol. 57 (2019) 734. https://doi.org/10.15625/2525-2518/57/6/13501.

Mazlan, M., Makhtar,M., Khairi, A.&Mohamed ,M.(2019). University course timetabling model using ant colony optimization algorithm approach, Indones. J. Electr. Eng. Comput. Sci. 13 (2019) 72–76.

https://doi.org/10.11591/ijeecs.v13.i1.pp72-76.

Perzina, R.& Ramík, J.(2013). Timetabling Problem with Fuzzy Constraints: A Self-Learning Genetic Algorithm, Undefined. (2013).

Rachmawati ,L.& Srinivasan ,D.(2005). A hybrid fuzzy evolutionary algorithm for a multi-objective resource allocation problem, in: Proc. - HIS 2005 Fifth Int. Conf. Hybrid Intell. Syst., 2005: pp. 55–60. https://doi.org/10.1109/ICHIS.2005.10.

Rakesh Kumar, S., Gayathri,N., Muthuramalingam, S., Balamurugan, B., Ramesh,C.&Nallakaruppan, M.(2019). Medical Big Data Mining and Processing in e-Healthcare, in: Internet Things Biomed. Eng., Elsevier, 2019: pp. 323–339. https://doi.org/10.1016/b978-0-12-817356-5.00016-4.

Selim,S.(1988).SPLIT VERTICES IN VERTEX COLOURING AND THEIR APPLICATION IN DEVELOPING A SOLUTION TO THE FACULTY TIMETABLE PROBLEM., Comput. J. 31 (1988) 76–82. https://doi.org/10.1093/comjnl/31.1.76.

Tharwat,A.(2020). Classification assessment methods, Appl. Comput. Informatics. (2018). https://doi.org/10.1016/j.aci.2018.08.003.

Welsh,D.(1967) An upper bound for the chromatic number of a graph and its application to timetabling problems, Comput. J. 10 (1967) 85–86. https://doi.org/10.1093/comjnl/10.1.85.

Werra, D.(1988). An introduction to timetabling, Eur. J. Oper. Res. 19 (1985) 151–162. https://doi.org/10.1016/0377-2217(85)90167-5.

Yusoff, M. & Roslan ,N.(2019). Evaluation of genetic algorithm and hybrid genetic algorithm-hill climbing with elitist for lecturer university timetabling problem, in: Lect. Notes Comput. Sci. (Including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics), Springer Verlag, 2019: pp. 363–373. https://doi.org/10.1007/978-3-030-26369-0_34.