
Palestine Polytechnic University
Deanship of Graduate Studies and Scientific Research
Master of Informatics

# Intelligent Model for Suitable University Specialization Selection in Palestine

Submitted by:

Lubaba Abd El-Mjeed Tamiza

Thesis submitted in partial fulfillment of requirements of the degree "Master of Science in Informatics"
August, 2017

The undersigned hereby certify that they have read, examined and recommended to the Deanship of Graduate Studies and Scientific Research at Palestine Polytechnic University the approval of a thesis entitled: **Intelligent Model for Suitable University Specialization Selection in Palestine**, submitted by **Lubaba A. Tamiza** in partial fulfillment of the requirements for the degree of Master in Informatics.

**Supervisors:**

Dr. Ghassan Shahin, Palestine Polytechnic University.

Signature:_____ Date:_____

Dr. Radwan Tahboub, Palestine Polytechnic University.

Signature:_____ Date:_____

Dr.Mahmoud Al-Ewiwi, Palestine Technical University-Kadoorei.

Signature:_____ Date:_____

Dr. Faisal khamayseh, Palestine Polytechnic University.

Signature:_____ Date:_____

**Thesis Approved**

Dr. Murad Abusubaih
Dean of Graduate Studies and Scientific Research
Palestine Polytechnic University

Signature:_____ Date:_____

# DECLARATION

I declare that the Master Thesis entitled **"Intelligent Model for Suitable University Specialization Selection in Palestine"** is my original work, and herby certify that unless stated, all work contained within this thesis is my own independent research and has not been submitted for the award of any other degree at any institution, except where due acknowledgement is made in the text.

**Lubaba Abedelmjeed Tamiza**
Signature:_____ Date:_____

# STATEMENT OF PERMISSION TO USE

In presenting this thesis in partial fulfillment of the requirements for the master degree in Informatics at Palestine Polytechnic University, I agree that the library shall make it available to borrowers under rules of the library.
Brief quotations from this thesis are allowable without special permission, provided that accurate acknowledgement of the source is made.
Permission for extensive quotation from, reproduction, or publication of this thesis may be granted by my main supervisor, or in his absence, by the Dean of Graduate Studies and Scientific Research when, in the opinion of either, the proposed use of the material is for scholarly purposes.
Any coping or use of the material in this thesis for financial gain shall not be allowed without my written permission.

# DEDICATION

*To my mother and father, who taught me patience and success*

*To my family, my beloved husband Abelhameed Tamiza my kids Ameer, Jana and Dana*

*To my brothers and sisters for their supporting and encouragement*

# ACKNOWLEDGEMENT

After finish my journey in this thesis with fatigue and pleasure,
Thanks to God for his infinite graces.
I would like to thank my family for their continued support, encouragement, and patience in a master's study trip.
I greatly appreciate the my supervisors Dr. Ghassan Shahin and
Dr. Radwan Tahboub for their support and the time they devoted.
Special thank for Dr. Hashim Tamimi, Dr. Sami Taha and Dr. Noureddine Sadawi for their experience.

## الملخص

اختيار التخصص المناسب هو مفتاح النجاح في الحياة الأكاديمية والمهنية. فكلما كان التخصص مناسبا للطالب. فإن الزيادة في أداء الطلبة هي النتيجة الطبيعية. وبالإضافة إلى ذلك، هو موضوع هام في حياة الشخص وهو واحد من الموضوعات الحياتية التي تستحق البحث. العديد من الدراسات بحثت في العوامل المؤثرة باختيار التخصص باستخدام الأساليب الإحصائية، إلا أن أيا منها لم يدرس هذه العوامل ووظفها في أساليب التعلم الآلي و تطوير نموذج التصنيف الذي يمكن أن يساعد على اختيار التخصص المناسب.

في هذا البحث استخرجنا العوامل المحلية المؤثرة في اختيار التخصص في فلسطين باستخدام طريقة الفلترة (Filter approach) واستخدام تقنية الارتباط القائم على اختيار الميزة ( Correlation-based Feature Selection) (CFS) وتحليل المكونات الأساسية (Principle Component Analysis) (PCA). ووفقا للنتائج حددنا خمسة عوامل أساسية تؤثر في اختيار التخصص الجامعي في فلسطين وهي عامل الشخصية, عوامل ديموغرافية، عامل الجامعة، عامل خلفية العائلة وعامل التحقيق الأكاديمي. بعد ذلك قمنا بتطوير نموذج تصنيف والذي يعتبر النموذج المقترح الأول الذي درس العوامل المؤثرة في اختيار التخصص الجامعي ولديه القدرة على تنبؤ التخصص الجامعي المستقبلي لطلبة المرحلة الثانوية باستخدام قواعد (Rules). ولانجاز ذلك تم إجراء مراجعة موسعة للأدب النظري، وحددنا العوامل المؤثرة التي تؤثر على اختيار التخصص استنادا إلى مراجعة الأدبيات، مثل درجات الطلبة وتأثير أولياء الأمور وملائمة مواهب الطلبة والظروف المالية وعوامل أخرى.

وقد تم إعداد استبيان خاص يغطي مختلف المسائل المتعلقة بالعوامل المؤثرة، وتم اختبار الاستبيان وتحكيمه من قبل 10 خبراء. وبالتالي، نموذجنا المقترح يعتمد على استخراج المعرفة السابقة والتجارب الطلابية. البيانات التي تم جمعها تستخدم كمدخلات لبناء نموذج التصنيف لدينا باستخدام PART. PART المصنف هو المصنف على أساس القاعدة التي بنيت على شجرة قرارات جزئية دون الحاجة إلى التحسين العالمي. المنهجية تجمع بين خوارزميتي C4.5 لتوليد شجرة القرار من مجموعات التدريب (Training set ) و RIPPER تستخدم لتوليد قواعد من خلال مرحلتين هي تكرار مرحلة النمو ومرحلة التقليم . بعد تطوير نموذج التصنيف، تم اختباره على مجموعة فحص لم تستخدم في نموذج التدريب تحتوي على 70 حالة طالب لا يعرفون تخصصهم (Class) ، لكنهم يرغبون في معرفة ما التخصص المستقبلي المناسب لهم. بعد اجابتهم على الأسئلة، فإنه يمكن لنموذج التصنيف التنبؤ بالتخصص المناسب لكل طالب.

وفقا للنتائج، توصل الباحث إلى أن دقة النموذج المقترح هي 77.4٪ لمجموعة التدريب. وبالنسبة لمجموعة الاختبار، أظهرت النتائج أن دقة النموذج المقترح كانت 73.7٪. ومن ثم اعتمد النموذج المتطور 49 قاعدة نهائية التي تعتبر خارطة لقيادة خطوات طلاب المدارس الثانوية نحو اختيار التخصص المناسب.

# Abstract

Choosing suitable track is a key success in the academic and professional life. Whenever the specialization is appropriate for the student; an increase in students' performance is the natural result. In addition, it is an important theme in a persons life and it is one of the vital topics which deserve to be searched. Many studies investigated the influential factors affecting specialization selection by using statistical methods, but none of the researches studied these factors and employed machine learning methods to a develop classification model which can help to choose a suitable specialization.

In this research, we extracted the local influential factors in our area (Palestine) by using filter approach Correlation-based Feature Selection (CFS) and factor analysis approach Principle Component Analysis (PCA). According to the results, we identified five basic influential factors affecting specialization selection at the university in Palestine, which is Personal factors, demographic factor, university factor, family background factor and academic achievement factor. Then we developed a classification model which might consider the first proposed model studying the influential factors affecting the specialization selection and has the ability to predict the specialization selection for high school students by identifying the suitable specialization based on rules. To accomplish this; an extensive literature review was conducted. We identified the influential factors affecting the specialization selec-

tion based on the literature review, such as students grades, parents influence, suitability for students talents and financial conditions and other factors as well.

A special questionnaire was developed which covers various questions relating the influential factors. The questionnaire was tested and arbitrated by 10 experts. Hence, our proposed model depends on extracting the previous knowledge and student experiments. The collected data used as inputs to build our classification model using PART. PART classifier is a rule-based classifier that is built on partial decision tree without the need for global optimization. The method combined between C4.5 method to generate a decision tree from the training data and RIPPER approach used to generate rules by two stages which are repeated growing stage and pruning stage. After developing the classification model, it was tested on a set which was not used in the training model. It contains 70 student instance, who do not know their class (major), but they want to know what future specialization fits them. After they answered the questions, the classification model can use to predict the suitable major for each student.

According to the results, the researcher found that the accuracy of the proposed model is 77.4% for the training group. And for the testing group, the results showed that the accuracy of the proposed model was 73.7%. So the developed model adopted final 49 rules which considered as a map to lead high school students steps toward choosing the suitable specialization.

# Table of Contents

# List of Figures

# List of Tables

# List of Abbreviations

| | |
|---|---|
| **AUC** | Area Under Carve |
| **PCA** | Principle Component Analysis |
| **CFS** | Correlation Feature Selection |
| **SL** | Supervised Learning |
| **FS** | Feature Selection |
| **IG** | Information Gain |
| **ML** | Machine Learning |
| **ROC** | Receiver Operating Characteristic |
| **TP** | True Positive |
| **TN** | True Negative |
| **ACC** | Accuracy |

# Chapter 1

# Introduction

Universities specialization selection considered as one of the main decision students can make in their social and career life. The National Career Development Association (NCDA) in the United States pointed out that helping individuals increase self-understanding of their abilities, interests, values, and goals are vital foundations of the career development process [6]. Whenever the specialization is appropriate for the student, performance of the student will be better as a natural result. Choosing a suitable specialization can help students to be creative and more interesting with their study, then the students can complete their high education efficiently. According to 2noted that unemployment rate related to inefficient university graduates. According to [7, 8, 9], there are many factors affecting university specialization selection. These factors may be related to student characteristics as abilities, interests and other factors. Sarwar and Masood found the influential factors affecting in specialization selection which represented with academic factors, social capital factors, future prospect factors, human capital factors, market demand factors and finally job prospect factors [10]. Another study focused on the necessity of compatibility between the career selection and the char-

acteristics as in Hollands theory [11], he supposed that most people are one of six personality types: Realistic, Enterprising, Investigative, Artistic, Social and Conventional. According to a study at Tafila Technical University (TTU), they found that the family and friends, future job factor, career option and occupational prestige considered as important factors in universities specialization selection [7]. As there are influential factors in specialization selection, there are problems which face students in specialization selection.

According to [12], found that 25% of students at 12th grade are unable to decide their suitable major career choices. Crites (1969, in Zunker, 1990) [8] suggested that about 30% of the students in high school and college are undecided about a career. This problem continues with the students during the undergraduate period. Moreover, it will affect their performance during their future careers . When students select specialization which does not fit their interests and/or abilities, this may lead to incompetence and lack of qualifications needed by companies and other institutions [13]. This is because the students abilities and orientations are not suited to the course they have taken. Hence, the students career success can be best achieved when the specification is suitable for their personality, ability, intellect and their interests as well choosing a college major [8]. According to [4], many of high school students have failed when they chose the wrong undergraduate program, due to the inappropriate selection decision. According to [12], the results of 12th grade found that counseling process helped them to understand more about their future specialization. Thus 83% of the students felt that counseling process can help them in selecting academic fields at the university, then led to choose their occupational future choice. [14] concluded that specialization selection problem is the biggest problem that faces high school students. The first steps to solve specialization selection problem is

identifying the influential factors in specialization selection at the university. Some of these factors are social factors, future prospect factors, personal factors, grades and academic achievement, financial factors, university factors and demographic factors [8, 9, 10, 14, 15]. In this work, a new proposed model is presented to Palestinian students to select and predict their university specialization better using different influential factors through machine learning algorithm and tools.

## 1.1 Research Problem

Specialization selection problem is one of the main problems which students suffer. Choosing inappropriate specialization with student skills and intellectual, abilities is leading in the future to existence graduated students unqualified to works in professional life thus leading to increased unemployment among young people.

## 1.2 Research Objectives

The main goal of this thesis is to suggest an intelligent model to help high school students in specialization selection at the university. To achieve this suggestion; the researcher studied the influential factors that impact choosing the specialization. The thesis objectives are to:

1. identify the influential factors in specialization selection at the university.

2. identify the local influential factors affecting specialization selection in Palestine.

3. developing an intelligent model that can predict a suitable specialization for high school students in the future university study.

## 1.3 Research Questions

These questions can be answered through an intensive literature review:

1. What are the most influential factors in specialization selection?
   This question could be answered throw literature review.

2. What are the local influential factors in specialization selection?
   This question will be answered through filter approach which is (CFS) Correlation- based feature selection method and statistical method which is (PCA) Principle Component Analysis both will be combined to extract the local factors in Palestine.

3. What is the suitable university specialization for high school students?
   This question will be answered during developing intelligent model.

4. How could an intelligent model be built?
   This question could be answered using the classification algorithm (PART).

## 1.4 Thesis Contribution

The thesis goal is to help high school students in specialization selection before attending the university. To achieve that, a machine learning model was developed by using classification algorithms (PART). Based on utilizing the previous knowledge from undergraduate university students, The research was built as a predictive model to determine what specialization fits high school student taking aptitude in consideration. The main contribution of

this thesis is none of the works was studying the various influential factors in specialization selection and employed it to develop a classification model by using machine learning algorithm.

## 1.5 Thesis Organization

This thesis is organized as follows, in chapter two we surveyed the literature review and the theoretical material related to the subject. Where chapter three presented the methodology followed. Then chapter four contained our experiments which were conducted and achieved results and their interpretation. The last chapter we presented and discussed the conclusion and the recommended future works.

# Chapter 2

# Literature Review

In this chapter, we reviewed the previous studies related to our research. Then it covered the background to understand the rest of the thesis and the used concepts.

Many studies contribute to understanding what factors are affecting student specialization selection at the university. Most of the studies examined the reasons which make the students to select a particular specialization at the university by using statistical methods. In this thesis, the researcher employed machine learning to find the local influential factors to help high school students in specialization selection at the university. In addition, we aimed to determine the suitable specialization for them.

In [9], identified the factors affecting the students selection of their future career at junior high school students. Based on the results of the study, factors were divided into four elements; which are, personal factors, career exploration factor, school factor and group factor. The factors are investigated and analyzed statistically. The result showed that personal factor was the most factor affecting the student's decision, then family factor, career exploration factor, while school factor was the lowest effected.

Kenneth Eduard etal.[14] determined the factors that affect the high school students in choosing their college courses by using statistical analysis. They concluded that the financial stability of the family is the most factor affected specialization selection. While Pascual [8] determined the influential factors in selecting favorite specialization as seen in students sample in various specializations at the Philippine University. The data were gathered using a standard questionnaire, Brainard Occupational Preference Inventory (BOPI) scale, (BOPI) is a standard questionnaire contains a list of questions to study students interests and educational plans and future career in addition to students general grades at third year. The data were statistically analyzed by SPSS, The research concluded that the availability of work after finishing university studying is the first factor in specialization selection. [10] determined what the factors affected in choosing a business specialization at the university. According to their statistical study, six factors defined to contribute in specialization selection decision making, which are academic factors, social capital factors, future prospect factors, human capital factors as (personal interests, preferences and student tendencies), market demand factors, and finally job prospect factors. The data were gathered by a questionnaire; they concluded the six influential factors using AMOS for confirmatory factors analysis.

Al-Rfou [7] studied the influence of future job factors and personal factors in selecting a business major. It aimed to examine the relationship between students gender, branch in general secondary, general secondary average and the choice of business as a specialization. The results of the study found that there was a significant relationship between the assumption factors and selecting a business major. [15] determined the student's demographic variables to predict their retention study as age, gender, ethnicity, color, and

high school academic achievement. [12] designed self-counseling and career development programs to improve students problem-solving skills. Results, after counseling the students, showed that they were understanding their strength and weakness, also more able to solve problems.

In this research, we picked the most influential factors in specialization selection as conducted in [14, 8, 10, 7, 9, 15] the target of that is to design a questionnaire for data collection to conclude the local influential factors in Palestine and to help high school students to select university specialization. From the previous studies, we found none of them use machine learning algorithm to help high school students at university specialization selection, but it is intersecting with our research in studying the influential factors in university specialization selection. Hollands [11] supposed six model environment that can be interpreted to career orientation - realistic, investigative, artistic, socially enterprising and conventional. As long as there is a relationship between personality model and environmental model, the compatibility between them can lead to career and psychological stability.

**The six environmental model**

1. Realistic environments: concentrate on practical activities that often use machines and tools within the specialization related with realistic environments such as electrical and electronic engineering, mechanical engineering and military studies.

2. Investigative environments: associated with activities that focus on the creation and use knowledge. They prefer investigation and thinking in problem-solving, thus finding the relationships between things. The specialists associated with the investigative environment are mathematics, physics, economy, sociology and civil engineering.

3. Social environments: focus on teaching and social guidance. The specialists associated with the social environment are political science, nursing, philosophy, special education, and history.

4. Enterprising environments: includes persuasion skills, self confidence, leadership tasks.

5. Artist environments: includes with creative activities, they are unconventional, intuitive imaginative and ambiguous. The artistic environment includes specialist of arts, music, speech architecture.

6. Conventional environments: required regular and routine and deal with tangible verbal information, mathematics, numbers, routine tasks, required few social skills as working with office equipment and tools. The examples of the conventional environments are accounting, economy, secretarial, business office data processing.

According to Hollands theory, it helps to find the compatible specialization with student personality.

From the previous studies, we concluded that we should study students characteristics, their abilities, interests as well as their living conditions before choosing their specialization.

In [4], propodes a classification model for selecting an undergraduate program by using data mining tool WEKA, This model aimed to help undergraduate students in selecting a suitable major. The model has been built on a number of variables (attributes) as student gender, old major (The major program that student got), old GPA (GPA that student got from old school), students education type, faculty (current one), they used C4.5 classification algorithm to extract a decision tree rules. [5] used data mining technique to help high school students in choosing their study track at school by analyzing

the experience of previous students in the same academic achievement. Both of [4, 5] intersection of our research in using machine learning algorithm to help students in their future selection but none of them studied the influential factors in specialization selection.

In [16], proposed data mining- educational analysis model call (DMEDU). The model was presented as a guideline for the higher educational system to improve decision-making processes as planning, evaluation and consulting. This study showed how employing data mining tools in education.

Kovacic [17] examine socio-demographic variables as ( gender, education, age, work status, ethnicity, and disability) and course program in what extent that affect on student pass and fail during collage by using data mining methods and statistical analysis; two decision trees were generated using CART and CHAID, the accuracy of CART tree was 60.5% and 59.4% for CHAID but the feature selection results were not used in the classification trees but they depend on the chi-square results and the generated tree to conclude the important factors that affect on students success and fail.

Another study identified the priorities of vocational education courses as well determined the required courses and less required by using the statistical analysis by studying the relationship between the vocational education and undergraduate students background as student gender, caste, course of study, year of study and the living area. Based on the results of study there is no relationship between student gender, living area and caste with the vocational education interests [18].

Cortiz et al. [19] predicted the secondary students grade of two classes by using the previous students school grades, demographic and social students data. To achieve that they used four data mining techniques which are Decision Trees (DT), Random Forests (RF), Neural Networks (NN) and Support

Vector Machines (SVM).

Ma, Yiming et al.[20] identified the gifted students and weak students, where weak students were to set for remedial classes; to achieve that they used data mining technique which is Association rules, Naive Bayes and C4.5. The results showed that the mining using association rules are better than Naive Bayes and C4.5. Sacin, Cesol et al.[1] proposed recommendation system to support students to select the suitable courses and the number of courses. They depended on the real data of students considered as attribute as courses enrolled in, name of course, accumulative GPA, term start, grade. To build this system, C4.5 decision tree was used. Figure 2.1 illustrated the proposed recommendation system.

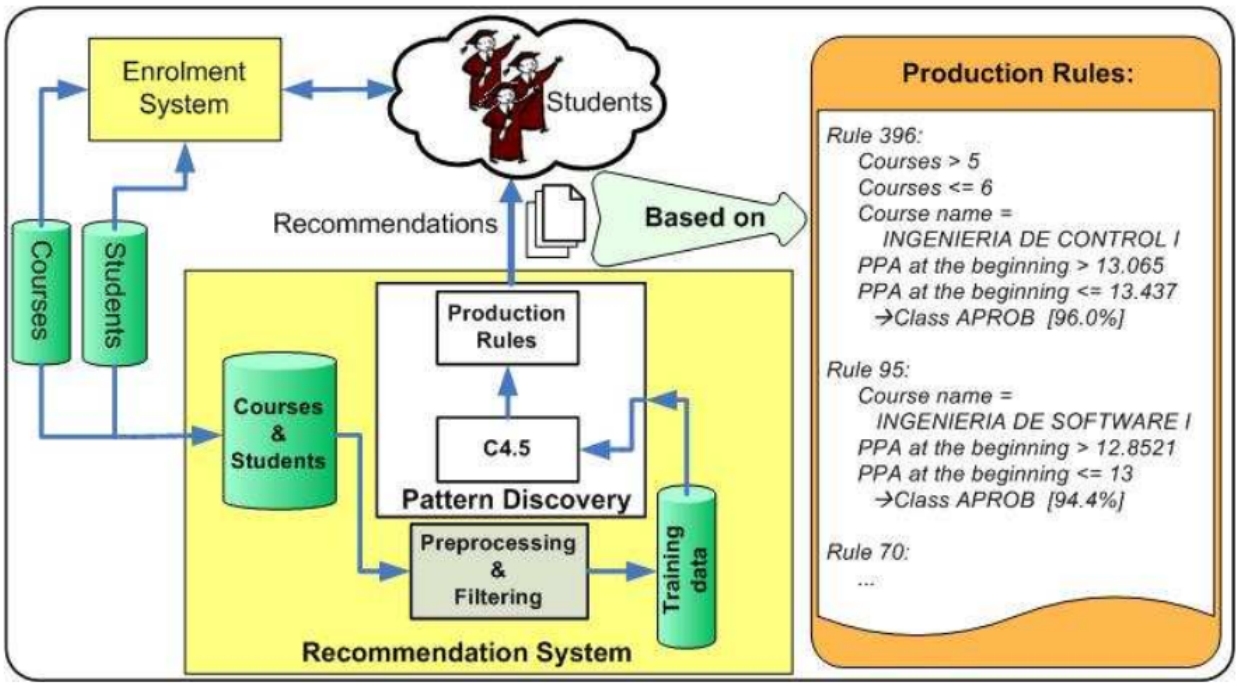


Figure 2.1: Recommendation system architecture [1]

## 2.1 Identifying the Influential Factors in Specialization Selection

Influential is a cognitive factor that tends to have an effect on what you do. Factor: anything that contributes causally to a result [21]. In this thesis, we make a survey to find the most significant factors that influence in specialization selection at the university.

1. Social factors (Family, peers, parents) in some societies, the family is the most effective factor in students specialization selection [8]. Parents may force their children to enroll some specialization, and students may consult their elder sibling to help them choosing the specialization. Parents, elders sibling, and peers are considered as impressive factors in specifying a major [7]. According to the study cited in [8], it found that 26% of sampled students took into account their parents consult and advice. [10] found that social factors as family and peers are the second influential factors in specialization selection. Another study investigated the role of parents (parental involvement) in choosing career and education in Hmong /Mong area [22]. Parents must understand their children mental abilities, interests, and tendencies.

2. Future Prospect Factors (future earning, career option, occupational prestige and type of work ) the future earning, career option, occupational prestige, and type of work considered as influential factors in selecting the specialization [7]. According to the results of [10], it is the third main influential factors in selection specialization. According to [8] results found that the availability and opportunity for work after finishing high school degree is the first influential factors in selection specialization. (Pascarella and Terenzini,1991) [11] have found career

opportunities and rewards significantly affected by academic specialization.

The problem in this factor makes students to choose misfit specialization to find a chance of work [8].

3. Personal factors ( Interests, Talents, Mental abilities, Intellectual skills) According to Holland as cited in [12] argued that personality characteristics are the major variable that effects in choosing a career. And the second variable effects are career interest.[23] defines interest as some thing that concerns, involves, draws the attention of, or arouses the curiosity of a person. Academic achievement, personal aptitudes, Interests, mental abilities, all affect career choice and specialization [8]. Both of [24], [10] they found that students interests in a type of work which fits their abilities are a significant factor in specialization selection. According to [8] results, specialization selection has an important relation with the favorite subject the students choice. Hence, it has been considered as influential factors in specialization selection. Intellectual skills are logical thinking, problem-solving, communication skills, design skills, organizational management and administration skills [25]. Based on literature review can divide personal factors to interests, talents and mental abilities. Students abilities playing a great role in specialization selection [14].

4. Grades and Academic achievement In [4], [5] both of them adopted a Grade Point Average (GPA) as parameters in their classification model for helping high school students in the university specialization selection. [10] found that there is a relationship between Cumulative Grade Point Average (CGPA) and specialization selection.

5. Family Background factors: (Parents educational degree, Parents occupational classification, Parents socioeconomic status [9]) Parents occupational degree, especially father occupational degree considered as an important factor in specialization selection since father considered as a provider with financial needs at the family [8].

6. Financial factors (Income, Economic status) In some societies considered the financial income as a factor in specialization selection at the university. Each of [8], [10] researches whether if the financial factor has an important impact with students specialization selection.

7. University factors (Location, Academic program, Financial aids, Reputation) It is considered as an important factor in our Arab environment. Many of the students involved with study in such university for some reasons as location; nearer to the place of living, academic program, financial aids and university reputation [26, 14].

8. Demographic factor( Gender) According to Osakinle and Adegoroye [6] identified that gender is one of the factors affected in major choice at university. According to (Hagedorn, Nora, Pascarella, 1996; Leslie and Oaxaca, 1998; National Research Council, 1991) as cited in [11] finds that Gender is an important component in choosing college majors. In contrast, in [10] found that there is no relationship between gender and choosing the specialization.

According to Gribbons and Lohnes as cited in [12] they provided some guidelines for career guidance programs in high schools:

1. Awareness of curriculum choices in terms of abilities, interests and values.

2. Awareness of occupational choices in terms of abilities, interests, and values and requirements.

3. Awareness of personal strengths and weaknesses relating to educational and occupational requirements.

4. Accuracy and evidence of self rating in terms of relevant abilities for educational and occupational requirements.

5. Accurate description of occupation

6. A willingness of individuals to take responsibility for their choices.

These guidelines supported our research to inference some of the influential factors can help students in specialization selection at the university.

## 2.2 Machine Learning

Machine learning is one of the artificial intelligence domain aims to design and improve algorithms to make computer to be able to learn. Herbert Simon defined machine learning changes in the system that are adaptive in the sense that they enable the system to do the same task or tasks drawn from the same population more efficiently and more effectively the next time [2]. In another meaning, it is the ability to gain knowledge and improve it. The main machine learning methods are supervised learning and unsupervised learning.

### 2.2.1 Machine Learning Methods

#### Supervised Learning (SL)

SL is defined as "The process of learning a set of rules from instances. In other words, creating a classifier that can be used to generalized from new instances". [27]

Suppose $f$: X $\rightarrow$ Y. Where $f$ is a supervised learning function whereas $f(X)$ predicts the corresponding value of $y$. $X$ is the inputs contain feature vector denoted by $xi = x_1$,$x_2$,..,$x_n$, and $Yi$ is the output values contain the labels denoted by $yi = y_1$,$y_2$,..,$y_n$ where $i$ = 1,2,....n. A pair of $(xi, yi)$ is the training examples and a set of training examples represents the training set [28, 29]. Supervised learning method illustrated in Figure 2.2 [28]. In this research we used the rule based classifier which classified as supervised learning.

#### Unsupervised Learning

Discovers unseen patterns in unlabeled input data sets. There is no training group or class labels as in clustering [2].

## 2.3 Reducing Dimensionality

Reducing feature contains two main methods which are feature selection and feature extraction. In this thesis, we used reduction feature to reduce the number of features and extracting the important.

### 2.3.1 Feature Selection

Feature selection (FS) is a technique for selecting and identifying a subset of a relevant feature in learning model to enhance the generalization and

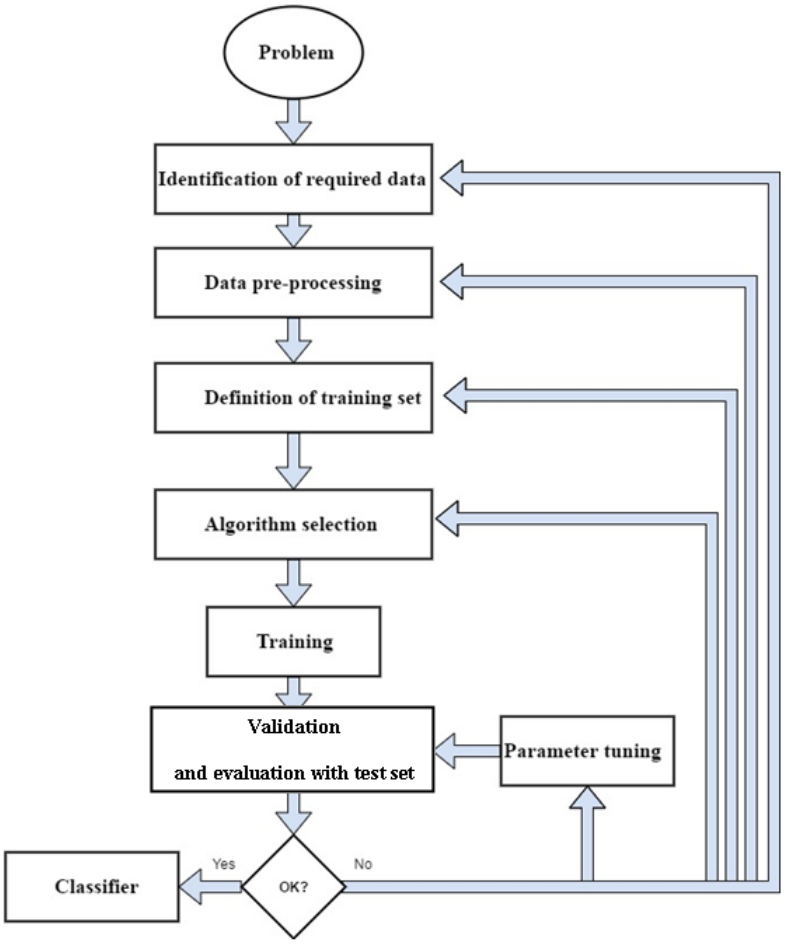


Figure 2.2: Supervised Learning Process [2]

reduce the overfitting. The benefits of this process help to eliminate irrelevant features, increase model learning performance, accelerate it and decrease computational costs, data reduction. Hence, decrease storage requirements. (FS) technique contains three major approaches: wrapper approach, filter approach and embedded approach [30, 31, 32].

**The Steps of Feature Selection Procedure**

1. The starting point is identified to decide the direction of search space. In the case of backward elimination method starting with whole features selection. Respectively remove and decrease features until finding the required. Forward selection method starts without features and adds new required features has the least error [32].

2. Organization of the search: determine the searching strategy to find the subset features in the searching space. The number of possible subsets is ($2^N$). Where N is the number of available features.

3. Evaluation strategy: used to evaluate and estimate the subset feature selection accuracy after feature selection process using the filter, wrapper and embedded approach[32] .

4. Stopping criteria: feature selector stops when the alternative stop improves the classification accuracy, or it continues to review the feature while merit does not break down [32]

### 2.3.2 Feature Selection Methods

Feature selection technique contains three main methods which are wrapper approach, filter approach, and embedded approach. Figure 2.3 illustrates feature selection methods.

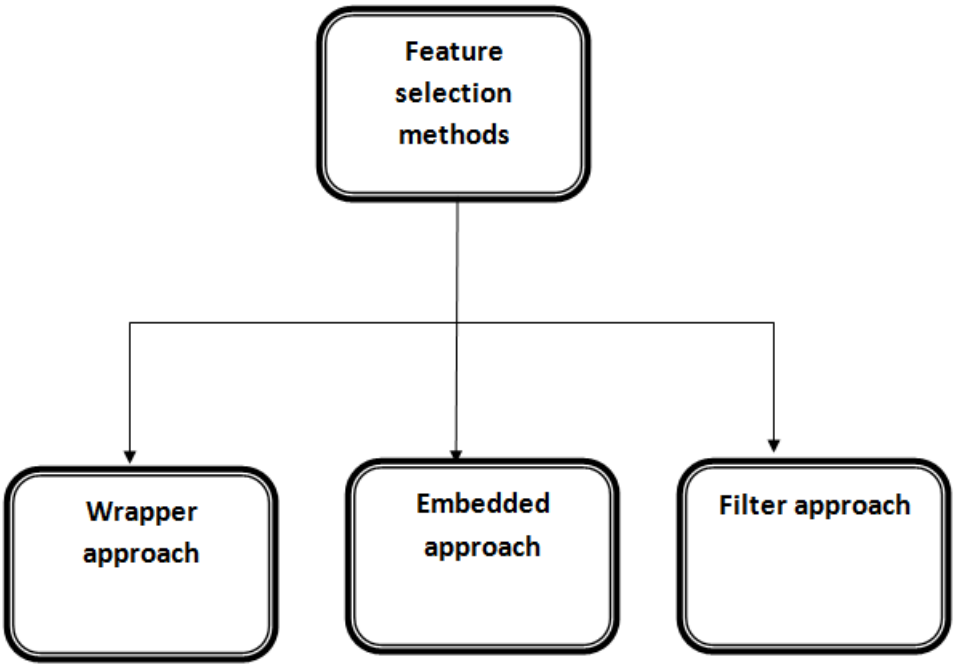


Figure 2.3: Feature selection methods

### Wrapper approach

This approach chooses the best subset features which fit the learning algorithm as well used the learning algorithm to evaluate the features. It depends on a predictive model in subset feature selection. Each new feature subset is used to train a model. As a result of which, wrapper approach is more expensive and more accurate than filter approach. But it is difficult to be used in large data sets that include many features [31, 3, 32].

### Filter approach

Filter approach finds the best set of feature without relying on the predictive model (classifier). Hence it is more speed up than wrapper [23]. As well do not need to rerun when changing to another learning algorithm [30, 3]. In this thesis, we used the filter correlation based feature selection since the accuracy value of the model has the best value by using this filter.

**Correlation-based Feature Selection (CFS)** One of the filter methods measures a correlation based heuristic evaluation function for each feature subset and ranks it based on heuristic value. (CFS) filter basically relies on measure correlation between feature with feature and feature with class. The subset feature has a high correlation with the class label and weak correlation with a feature- to- feature, so it was preferred and considered as useful attributes [3].

#### Embedded approach

This approach joins the advantages of wrapper and filter approaches. It chooses the best feature and evaluates it at the same time. Hence it has the lowest complexity than wrapper approach [31].

### 2.3.3 Feature Extraction

In the feature extraction process, the goal is to reduce the number of dimensions to extract new dimensions. According to [33], these methods may be supervised or unsupervised depending on whether or not they use the output information. Feature extraction has several methods as Factor Analysis (FA), Principle Component Analysis (PCA) and Linear Discriminant Analysis (LDA). In our thesis, we used Principal Component Analysis to extract the influential factors affected in specialization selection.

## 2.4 Rules -Based Classification Algorithms

It is an integrated technique includes two important techniques of data mining, which is association rules mining and classification model. This method basically contains from three stages which are rule generation, rule pruning,

and classification. It is important because it can produce more accurate classification models than traditional classification models. This technique considers the class label as resultant of the rule. The examples of it are (CBA) which is Classification based on Association rules, (CMAR) Classification based on Multiple Association rules, PART which a combination between C4.5 [34] and RIPPER [35], also it combined between both of advantages [36, 37]. In our research, we used another rule-based classifier algorithm to compare our prediction model using PART classifier and the following classification algorithms:

### 2.4.1 OneR

OneR is an abbreviation of "one rule", It generates one rule for each feature. Then chooses the total minimum error feature as one rule [38].

For each predictor (feature) rules value is made as follows:

- Counting how many times the one value of class will appear.
- Determine the most frequented class.
- Assigning the class by the rule value.
- Calculating the rules total errors for each predictor.
- At the end, the predictor will be chosen as the one that has the smallest total errors.

### 2.4.2 RIPPER

It is an abbreviation for Repeated Incremental Pruning to Produce Error Reduction. It generates rules by two stages, repeated growing stage and pruning stage.

During the growing phase the rules are made more restrictive in order to

fit the training data as close as possible. During the pruning phase, the rules are made less restrictive in order to avoid over-fitting, which can cause poor performance on unseen instances [28].

### 2.4.3 PART

It is a rule based classifier algorithm combines C4.5 [34] and RIPPER [35] paradigm. Hence, it is more efficient and more accurate than both. It makes induction rule without global optimization. PART infers rules by creating a partial C4.5 decision tree for each loop instead of fully explored one and the best leaf is made into a rule. The building of decision tree is done by dividing a set of instances into a partial tree frequently. First starting with a testing node, then testing the entropy for the next node. After that expanding the sub-tree by adding the node has the biggest entropy. In this thesis, we used PART rule-induction to generate our rules and to build our prediction model [36], the reason to choose PART is the combination of association rules and classification at the same time. Also, it helps in attaining research goals by extracting rules to predict the future suitable specialization for high school students, in addition; it can predict what factors affecting in specialization selection.

## 2.5 Performance measurement

AUC is used to evaluate the performance of the classification model. These steps are done after building the classifier by training set. The Receiver Operating Characteristics (ROC) curve is a graphical representation used to measure the performance of the classifier. Each junction of ROC curve represents two points (tpr, fpr). It contains two coordinates, true positive

rate, and false positive rate. In the true positive rate, the examples were classified correctly [39]. In our thesis, we find the ROC curve for seven classes (majors) and making the comparison between them. In the case of two classes positive and negative, the classifier figures the values as follows:

$$tn = \frac{NegativesCorrectlyClassified}{TotalNegatives} \tag{2.1}$$

$$tp = \frac{PositivesCorrectlyClassified}{TotalPositives} \tag{2.2}$$

Specificity or the true negative ratio measures the instances incorrectly classified as shown in equation 2.3.

$$Specificity = \frac{tn}{fn + tn} \tag{2.3}$$

Sensitivity measures ratio of instances that are correctly classified or the true positive ratio (tpr) as shown in equation 2.4.

$$Sensitivity = \frac{tp}{tp + tn} \tag{2.4}$$

Confusion Matrix A probability table displays the differences between actual classes and predicted classes. Furthermore, it can help to understand the classification model in terms of where it works well and the place of the errors. Given $n, p$. Where $n$ is negative class and $p$ is the positive class they are actual. Based on a confusion matrix if the actual class label $n$ and the predicted class label is $n$ then it is true negative, and if $n$ classified as $p$ it is a false negative, and if the actual class label is $p$ and the predicted class is $p$ then it is true positive otherwise it is false $p$ [39]. In our prediction model, we calculate the confusion matrix after building it. A confusion matrix is

shown in Figure 2.4 .

| Actual Class | Predicted Class | |
|---|---|---|
| | Positive | Negative |
| Positive | Tp | Fn |
| Negative | Tn | Fp |

Figure 2.4: A confusion matrix

### 2.5.1 Cross Validation (CV)

Is a method for evaluating the performance and estimating the risks of the classification algorithm. In cross-validation method, the data set divided into k-folds. (k-1) fold for the training set and one fold for testing. Each fold contains a subset of the data set. Each data set contains a number of instances have the same ratio of class labels.

For each $K$ time the classifier generate the performance average $f_i$. The average of the results of $f_1$, $f_2$, $f_3 ... f_k$ is figured as follows:

$$\overline{f} = \sum_{t=1}^{k} \frac{f_i}{k} \tag{2.5}$$

cross-validation guarantee better schema in terms of bias and variance, in addition, to avoid over-fitting. Therefore, we used this technique in our thesis and divided data sets for validation and training [40].

## 2.6 Random Sampling Technique

Random sampling is categorized into two types which are the random sample with the replacement, in which each element of the sample can be selected more than once, whereas the element of random sample without replacement, cant be selected more than one. In this research, population samples were identified by using the random sample method [41].

## 2.7 Questionnaire

It is a tool used to acquire information and to collect data about the study population. It contains a group of questions for research and study purposes. The basic characteristics of the questionnaire should be delivery mode, sampling techniques, and data analysis. The questionnaire should be built upon the logical organization of questions; in which each section includes the related questions together [42]. In this thesis, we used the questionnaire as a tool for data collection, taking into consideration the logical series of the questions, in order to make it easier for students to choose. This tool used by many researchers and diverse domains. Each of [7, 8, 9, 10, 26] used the questionnaire as a tool for data collection in their researches.

## 2.8 WEKA environment

It is an abbreviation of Waikato Environment Knowledge Analysis. It is an open source software, provides implementations of a machine learning and data mining algorithm written in Java, developed at the University of Waikato in New Zealand, WEKA contains tools for association rules, regression, classification, clustering, visualization, and data pre-processing. Many

of researches used WEKA environment. In this research, WEKA environment has been employed to develop our proposed model. Few advantages of using WEKA can be mentioned:

1. The possibility of using GUI of WEKA directly and visualizing the results on the output screen.

2. The possibility of calling WEKA libraries by using JAVA language [43, 44].

## 2.9 Discussion

Many of researches covered one or more sides of our study. But there is no research used our methodology to solve the specialization selection problem. The bulk of the studies was exhaustive to find the influential factors affected in specialization selection by using statistical analysis as in [9, 10, 26, 22]. Table **??** illustrates other framework contributed in solving specialization selection. In this research, researcher studied the factors theoretically. Based on that, a special questionnaire was built to investigate these factors in our area, especially in Palestine. Then input data were coded (questionnaire) and analyzed to build the classification model. In contrast, some of the research as in [4, 5] employed machine learning tools and methods to build classification model using decision tree (C4.5 classifier). Both of the previous proposed classification models are good models because they are achieved the desired goals. But in this research may will not achieved our desired goals when using their methodologies because the number of features and the data set are different and the used classification algorithm (C4.5) do not perform our goals which identify the specialization selection for high school students and identify the influential factors in specialization selection. To achieve that

we need rule-based classification algorithm as PART algorithm. Table2.1 illustrates the dataset and the results of the classification models.

Table 2.1: The dataset and the results for other frameworks

| **The classification model** | **Number of students** | **Number of features** | **Number of Classes** | **Accuracy** |
|---|---|---|---|---|
| Classification model (Al-Radaideh Q,2011) | 218 | one feature | 6 class label | 87.9% |
| Classification model (Jirapanthong. W, 2009) | 4000 | seven features | 22 class label | 87.9% |

Table 2.2 illustrates the other methodologies to contribute to solving specialization selection problem.

Table 2.2: Other frameworks to solve specialization selection problem

| **Frame work model** | **Summary** | **Method** |
|---|---|---|
| Classification model (Jirapanthong. W, 2009) | Found classification model for selecting an undergraduate program. | Decision tree- C4.5 algorithm |
| Classification model (Al-Radaideh Q,2011) | Built classification model for helping high school students choosing the suitable branch at Tawjihi. | Decision tree- C4.5 algorithm |
| Career planning program (Ming-Shang Su, et al.,2016) | Identified factors affecting the Students career decision-making of Junior High School Students. | Statistical analyzing |
| Career Planning Program (Pascual. N, 2014) | Identified Factors affecting the Students Career Decision-Making for high school Students. | Statistical analyzing |

# Chapter 3

# Research Methodology

### Introduction

In this chapter, we illustrate the methodology we followed in the research. We develop an intelligent model to predict the fitted specialization for high school students using the classifiers (PART). The data set was gathered by questionnaire. The questionnaire was built according to the literature review about the influential factors affecting in specialization selection process. It was tested and validated by a number of specialists. Then it was improved according to their notes and comments. After building the proposed model, a comparison between other machine learning methods as RIPPER algorithm was made to compare the accuracy results between them. Palestinian local influential factors also extracted and released on reduction features methods which are feature selection and feature extraction. In figure 3.1 illustrated our research methodology.

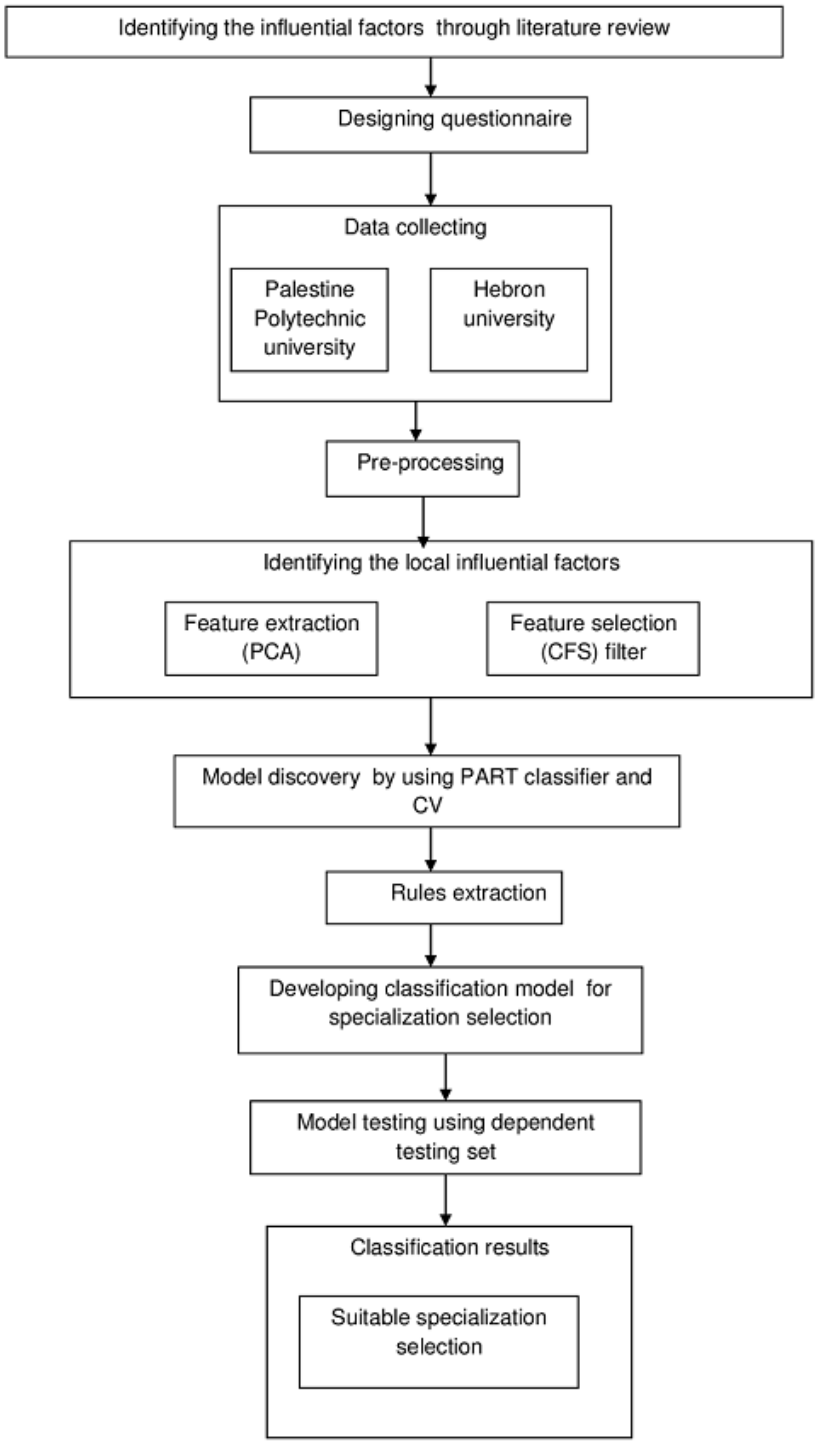


Figure 3.1: The research methodology

## 3.1 Data Collection

A questionnaire was used for data collection, the questionnaire was in two forms, soft form and hard form.

### 3.1.1 Designed Questionnaire

Our questionnaire designed based on previous literature which studied the factors and parameters affecting with students when they have chosen their specialization as mentioned in the previous chapter. Hence, the investigation is compatible with the influential factors that affecting students in their specialization. It was tested and validated by experiments and specialist. To ensure questionnaire distribution among studying sample, an electronic questionnaire was designed using Google forms [45] in addition to the hard copy questionnaire. Each of them contains several parts. Hence, each student answers the questions based on two criteria. First, their studying branch at the high school as (Scientific, Literary, Industrial or Commercial). Secondly; their specialization at the university. The questionnaire basically consisted of three parts, profile part, the second was specialization questions, and the third was general questions. Each student had to go through all parts.

**Profile questionnaire:** consists of questions related to participants with the investigation. These questions availed a background for each participant in the research. As Track, the current specialization, Number of family members, place of living and the gender.

**Particularly questions for specialization:** This part consists of special questions for each specialization and it is different from other questions in the other specialization.

**The general questions:** Consists of common questions related to each

specialization. Figure 3.2 shows questionnaire with reference.

### 3.1.2 Questionnaire validation

Questionnaire validation is considered as one of the important steps of questionnaire building which helps in evaluation and improvement. It depends on the opinions and comments of arbitrators. In our research after building questionnaire, it was arbitrated and tested by specialist and scholars in various fields as management, education and computer science. According to their notes and comments, we screen out the lowest related questions and improved a questionnaire in the final form to be prepared for data collection.

### 3.1.3 Sampling

Random sampling is categorized into two types which are the random sample with the replacement, in which each element of the sample can be selected more than once, whereas the element of a random sample without replacement, cant be selected more than. In this research, population samples were identified by using the random sample without replacement method [41]. In the data collection process, a random hour was chosen for the questionnaires to be distributed during the lecture for all specialization in Palestine Polytechnic University and Hebron University.

**Training sample:** the goal of the training sample is to train and teach our predictive model. The training set was collected from undergraduate university students at Palestine Polytechnic University (PPU) and Hebron University.

**Testing sample:** the goal of the testing sample was to test our predictive model students at Secondary School certification (Tawjihi) to predict the future specialization at the university. The testing sample was collected

| Question | References |
|---|---|
| الجنس؟ | (Lazarus, Kelechi U., and C. Ihuoma,2011)<br>(Porter, Stephen R., and Paul D. Umbach.,2006) |
| مكان السكن الأصلي؟ | (Nora,Amaury,2004)<br>(Alba, Kenneth Eduard Castillo et al.,2010) |
| الفرع؟ | |
| كم هو معدلك في التوجيهي(الثانوية العامة)؟<br>كم هو معدلك التراكمي في الجامعة؟ | (Jirapanthong, Waraporn.,2009)<br>(Sarwar, Aamir, and Rizwana Masood,2010)<br>(Al-Radaideh, Qasem A.,et al.,2011) |
| ما السبب الرئيسي لاختيارك لجامعتك الحالية؟ | |
| هل وجدت تشجيعا من قبل الأهل لاختيارك؟<br>هل وجدت تشجيعا من قبل الأصدقاء لاختيارك لهذا التخصص؟ | (Pascual, Nancy T,2014)<br>(Al-Rfou, Ahmad Nahar.,2013)<br>(Thao, Nealcheng Xeng.,2009)<br>(Sarwar, Aamir, and Rizwana Masood,2010) |
| ما هو مستوى تعليم الأب؟<br>ما هو مستوى تعليم الأم؟<br>ما هي مهنة الأم؟<br>ما هو القطاع الذي يعمل به الأب؟<br>ما هو القطاع الذي تعمل به الأم؟ | (Pascual, Nancy T,2014)<br>(Su, Ming-Shang, et al.,2016) |
| ما الدخل الشهري للأسرة؟ | (Pascual, Nancy T,2014)<br>(Sarwar, Aamir, and Rizwana Masood,2010) |
| ما هي أكثر المهارات الفكرية التي تتمتع بها؟<br>ما هي أكثر القدرات العقلية والذهنية التي تتمتع بها؟<br>كيف تقيم مهارتك باستخدام الحاسوب قبل التحاقك بهذا التخصص؟<br>كيف تقيم مهارتك في التواصل مع الفريق والعمل الجماعي؟<br>ما هي الألعاب المفضلة لديك في طفولتك؟<br>ما هو الدور الذي كنت تلعبه وتتقمصه في طفولتك؟<br>هل لديك مهارة حل المشكلات(problem-solving)؟<br>ما أهم هواياتك؟<br>هل كنت ترغب بتصليح الأجهزة والمعدات قبل التحاقك بهذا التخصص؟<br>هل سبق لك وان قدت سيارة في المرحلة الثانوية وما قبلها؟<br>هل سبق وان أصلحت جهاز الكتروني أو ميكانيكي في طفولتك؟<br>بماذا تتميز شخصيتك؟<br>إي من القضايا تهتم بها؟<br>ما هي التجربة العلمية التي قمت بها في قمت بها في طفولتك بهدف الاستكشاف؟<br>هل سبق وأن قمت بزراعة نبته في طفولتك؟<br>هل سبق وأن قمت بتربية حيوان بطفولتك؟ | (Pascual, Nancy T,2014)<br>(Sarwar, Aamir, and Rizwana Masood,2010)<br>(Kartika, Aniva.,2008)<br>(Alba, Kenneth Eduard Castillo et al.,2010<br>(Kim,David etal., 2002) |



Figure 3.2: The general questions of the questionnaire

from students at Hebron Governorate by the same questionnaire used for data collecting at the appendix.

## 3.2 Data set

In our research, students of Hebron University and Palestine Polytechnic University (PPU) were the study sample, the number of students at Hebron University is around 9000 students [46]. Also, the number of students at Palestine Polytechnic university is around 6000 students [47] till the time of data collection. Our questionnaire was distributed to all college students in both universities randomly [41] for students of 2nd, 3rd, 4th and 5th years. Hence, the sample population size around 15000 samples. In our research, we distributed around 700 questionnaires in hard form. In addition to electronic questionnaire form. But the number of valid instances was 700 samples. In this thesis, we adopted their responds since their Cumulative Grade Point Average (CGPA) over than 65%. Hence, we supposed they are passed in their specialization. Then we develop our prediction model on their responds long with previous students experiments to discover patterns that can help in finding the impact factors in specialization selection. We used Excel workbook to save data set.

According to data collection by questionnaire, one feature was gotten (attribute) to each question, considered as inputs in the classification model. Hence, the whole number of features is 52 features, and there were 7 classes labels (major), each major contains a sample of students as illustrated in Figure 3.3.

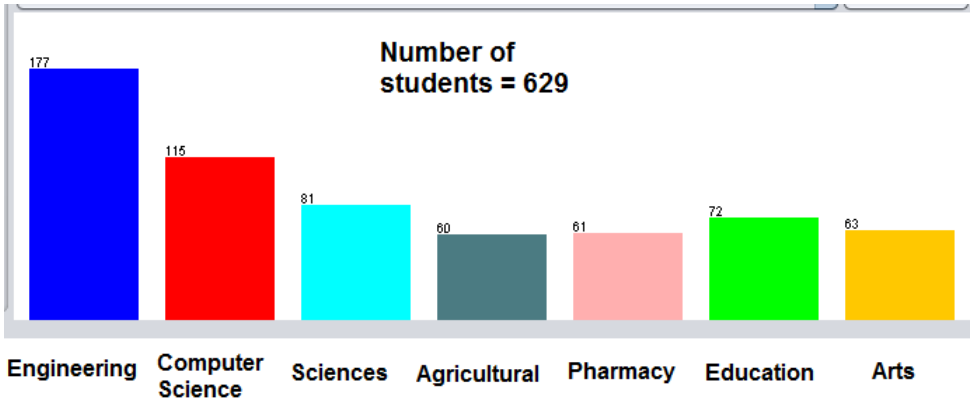


Figure 3.3: Number of students in each major

### 3.2.1 Data Entry

Data entry is considered the first step after data collection. In this thesis, we used Microsoft Office Excel 2007 for data entry and storage. We used a comma-separated values (CSV) files to save our collected data, in addition, to use Attribute-Relation File Format (ARFF) because WEKA environment can handle this type of files. The input data contains students answer which will be used to build the classification model, each column of the work-sheet included one factor.

### 3.2.2 Data Cleaning

The main goal of data cleaning is to identify the missing data, incomplete, inaccurate, and improving the quality of error correction and replacing the missing value [48]. One of the main problems we faced in this thesis was the missing values problem. The missing values are compensated by several ways such as calculating the arithmetic mean or mode of a set of values by sum all of the values and divide it by the number of values. Assume we have the following values $x_1$, $x_2$, $x_3$...$x_n$ the arithmetic mean is calculated as follows.

$$\overline{x} = \sum \frac{x}{n} \tag{3.1}$$

In our approach, we used the arithmetic mean for filling the missing values to improve the accuracy of our proposed model. The reason to filling the missing values because our questionnaire designed for different specialization and different branches as well each specialization contains different questions than others. Hence, will not answered all the questions and will be as missing values.

## 3.3 Identifying the influential factors in specialization selection

Identifying the influential factors affected in specialization selection based on the previous studies is the basis of our research. The main influential factors were displayed in details in the literature review chapter and we shall briefly present it.

Table 3.1: The influential factors

| **The Influential Factor** | Components |
|---|---|
| Social factor | Represented by family, parents, and peers impacts |
| Future prospect factor | Represented By career option, future earning, occupational Prestige type of work |
| Personal factors | Represented by mental abilities, talents, and interests |
| Grade and academic achievement | Represented Cumulative grade, point average in high school |
| Family background factor | Occupational degree, educational degree, and socioeconomic status |
| Financial factor | Monthly income, economic status of the family |
| University factor | Represented by available academic program, University location and reputation, providing students financial aids |
| Demographic factor | Represented With the student gender |

According to these factors, we designed a questionnaire contains a variety of questions used as a tool for data collection. Figure 3.4 illustrates the basic

influential factor in specialization selection.

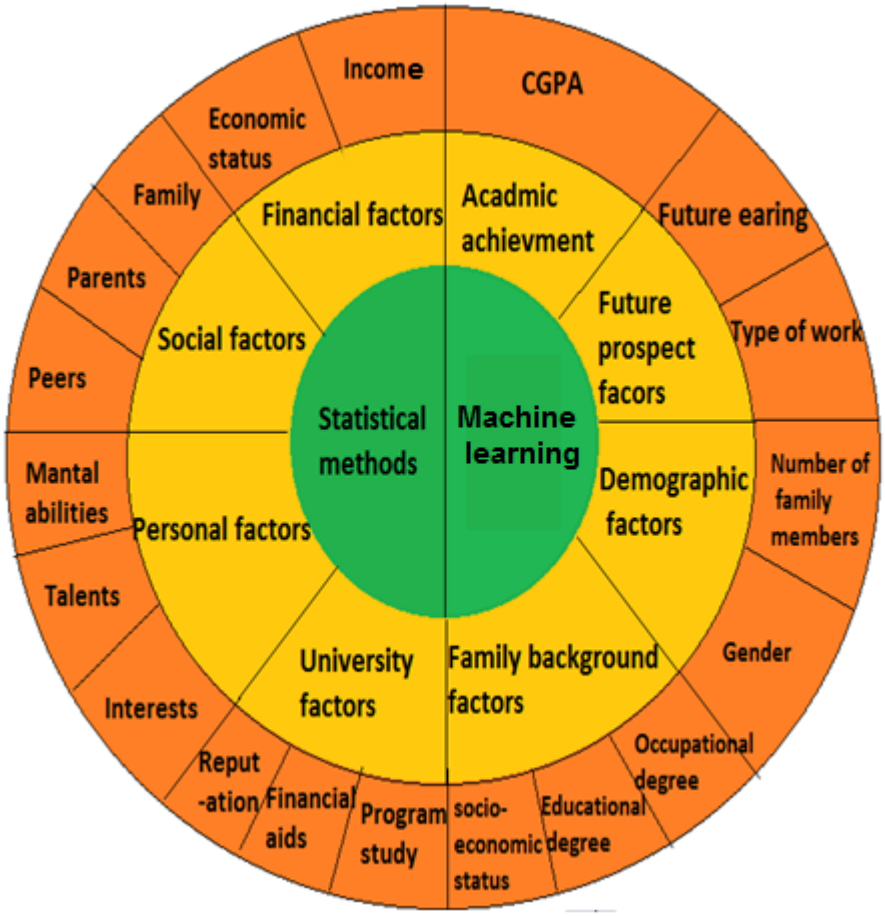


Figure 3.4: The basic influential factors

## 3.4 Identifying the Local Influential Factors in Specialization Selection

In order to find the local influential factors, we used reducing dimensionality methods which are feature selection and feature extraction. To implement that, we used filtering approach and statistical analysis.

Reducing the influential factors can help in understanding the most important factors influencing in specialization selection in Palestine. To reduce these factors we used two approaches which are feature selection and feature

extraction [33]. In feature selection, we use filter method which is Correlated based Feature Selection (CFS) [3]. In addition, and to use the extracting feature, Principle Component Analysis (PCA) was used [33]. (PCA) can find a set of correlated factors among, which lead to more general factors (Components) not correlated with each other. From this procedure, we can extract and select the important factors that contribute to specialization selection process in our local area (Palestine). The reducing feature process is illustrated in figure 3.5

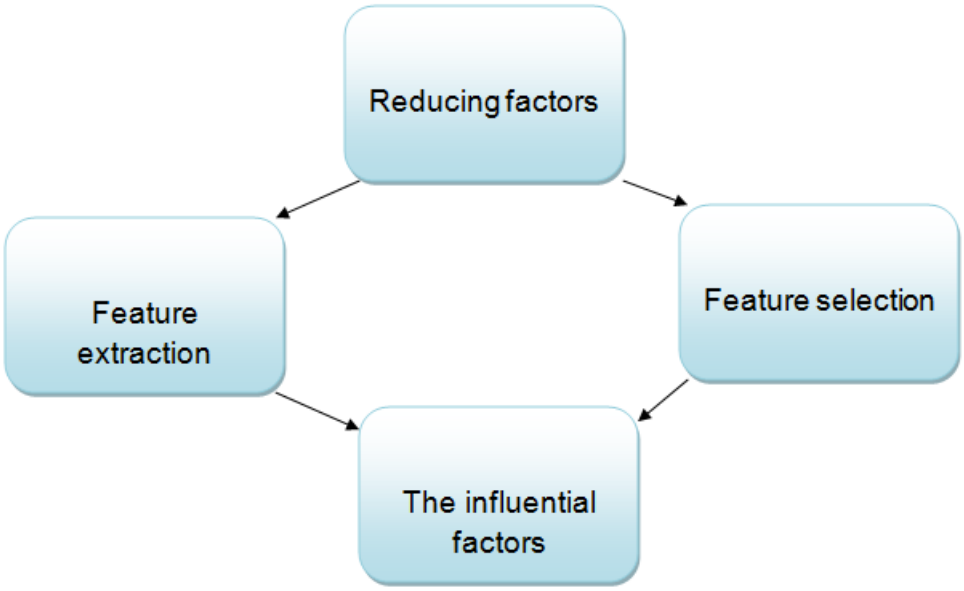


Figure 3.5: Reducing Features

## 3.5 Developing our Proposed Classification Model

First, our data sets split into a training set, validation set, and testing set. PART algorithm [49] has been used to build the predictive model, and looking forward to develop a model that has the ability to keep the old one and to contain the old training data by which, the new supplied data for a new

student can be used by the old model to choose the suitable specialization. PART algorithm combined between RIPPER and C4.5 classifiers. The importance of using PART classifier is the accuracy and simplicity of produce rules sets by incrementally generated partial decision tree without the need for global optimization. In addition build pruned partial decision tree for a set of instances to produce a single rule; which means this algorithm generates and explores the partial tree to induction one rule. Hence, this decreases the complexity time and improves the performance. PART classifier proposed by Ian Wettin and Eibe Franke. Figure 5.1 illustrates our methodology for developing the proposed model.

### 3.5.1 Model Training

This step involves taking the data set which is known by the class label (major) to build and train the classifier and rule extraction. The training set was 629 students of undergraduate students to build the learning classifier.

### 3.5.2 Model Testing

In our approach we tested the building model in two stages: first, using cross-validation technique to validate and test the building model through the building process. In each time, dividing the training set into k folds. (k-1) folds for training the model and the remaining fold for testing. In this research, we split our data set into 5 folds, 4 folds for learning model represented by gray and 1 for testing model represented by blue, then we measured the performance each time and found the performance average. In the second stage after building the model, we tested it by using a testing set which was not used in the learning model. This set contained 70 students with the unknown class label (major) to predict it. The testing set contains

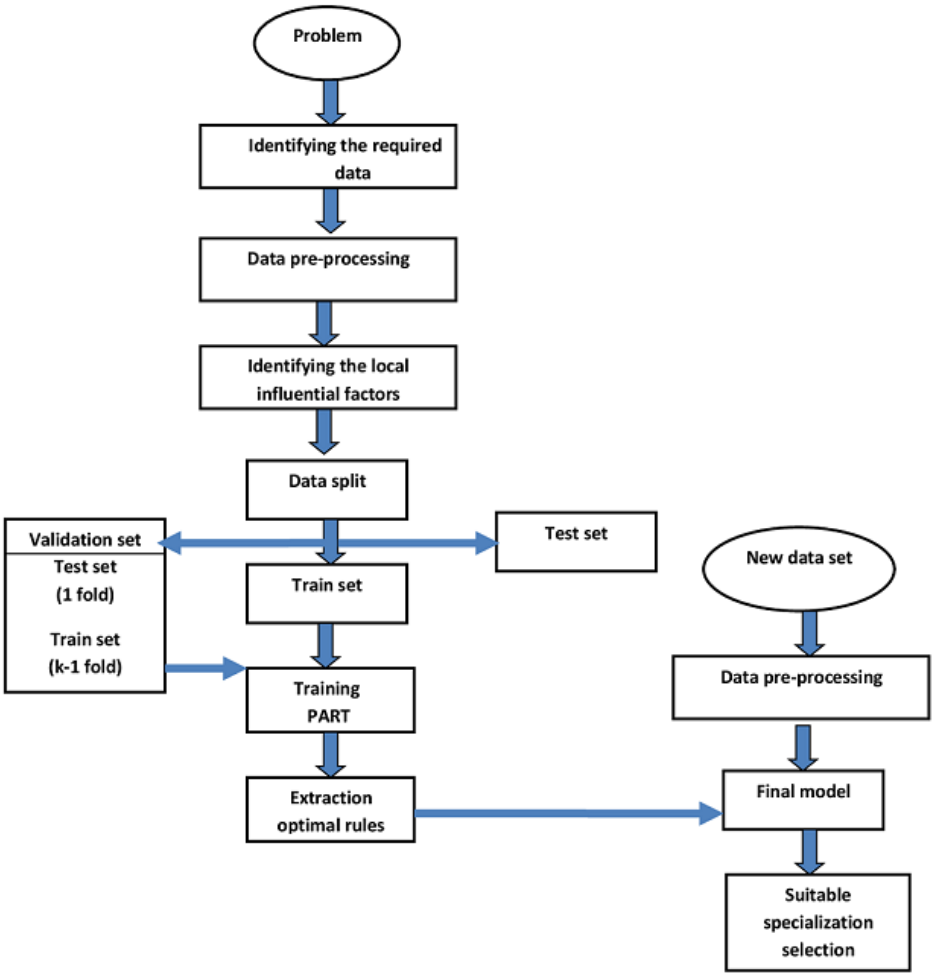


Figure 3.6: The proposed classification model

students in the second secondary grade (Tawjihi stage). Figure 3.7 illustrates the k-fold cross validation, which used in the building model.

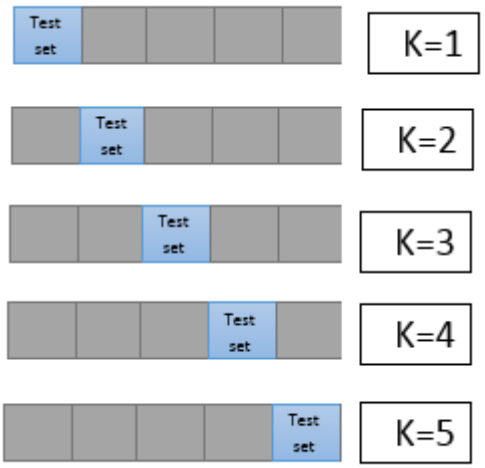


Figure 3.7: Cross-Validation

# Chapter 4

# Results and Finding

**Introduction**

In this chapter, We used rule-based classifier (PART) to build our classification model. Because our proposed model was built on the prediction rules that help to understand the influential factors in specialization selection, then chose the suitable specialization for high school students. The dataset was divided into three parts which are training set, validation set, and testing set. To do the experiments in this research, we used the lab top with processor Intel Core(TM) i5, 4200M @ 2.50GHz, memory 4GB and the operating system Windows 7 with 64 bit. We used net Beans IDE 8.0.2 in coding, in addition, WEKA tool version 8.3.1 was used and SPSS statistic version.18. A series of experiments were held to achieve the objectives of the research as will be illustrated later.

## 4.1 Identifying the Local Influential Factors

Identifying the influential factors in specialization selection in Palestine is one of the main objectives of the thesis. To achieve this goal we used reducing

feature method which contains feature selection and feature extraction [33].

### 4.1.1 Identifying the Local Influential Factors Using Feature Selection

Before adopting this method, as a way to find the local influential factors, we tried to use wrapper approach for feature selection, but did not offer the required results because wrapper approach could not handle with high numbers of features. But using CFS filter we found it is more suitable and compatible than wrapper as well we get the required results.

CFS technique illustrated in the figure 4.1. In our research, we used the filter approach which is Correlation-based Feature Selection (CFS) [3] for selecting the relevant features (factors). According to (CFS) filter and the extraction rules, we find the main influential factors in our Area especially in Palestine as illustrated in the Table 4.1. In this method the selected features(factors) is arranging according to the merit wight. Thus the feature has the highest merit considered as the main effected factor in specialization selection. Another approach was used to identify the influential factors which is the statistical analysis by [8] determined the factors affecting with high school students by calculating the mean value and standard deviation of the factors affecting.

### 4.1.2 Identify the Influential Factors by Using Feature Extraction

Another method for reducing factors which is feature extraction method. In this approach, we used Principle Component Analysis (PCA) to identify the influential factors. To achieve that, we tested the correlation between the in-

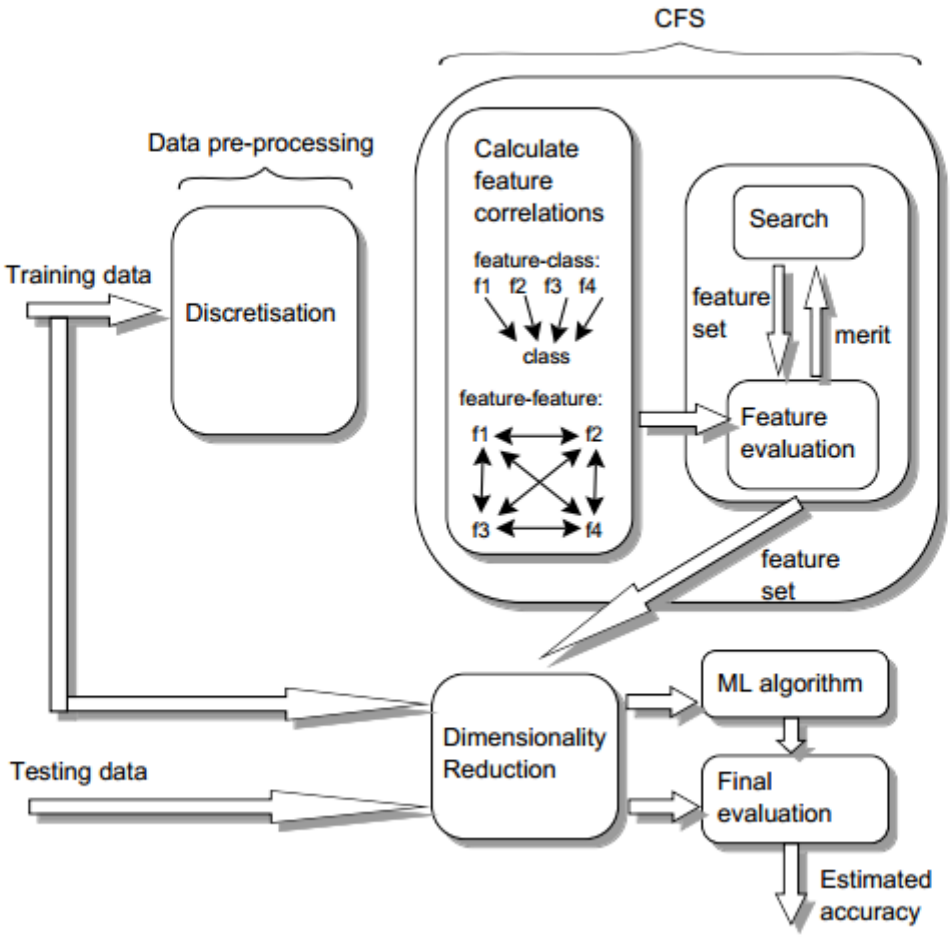


Figure 4.1: Correlation based feature selection technique [3]

Table 4.1: The local influential factors extracted by using (CFS) Filter

| **Factor Number** | **Factor Name** |
|---|---|
| Factor 1 | Track |
| Factor 2 | Gender |
| Factor 3 | Skills |
| Factor 4 | Breeding animal at childhood |
| Factor 5 | Tawjihi Average |
| Factor 6 | The reason to choose university |
| Factor 7 | English language average |
| Factor 8 | Problem solving skill |
| Factor 9 | Fixing devices |
| Factor 10 | Driving cars |
| Factor 11 | The preferred subject |
| Factor 12 | Cultivation of plants in childhood |
| Factor 13 | Making scientific experiment at childhood |

put factors also to test the strength of the relationship between factors thus measure the adequacy of the sample , based on the value of Kaiser Meyer Olkin (KMO). We found the (KMO) value is 0.931, which is an excellent relationship between factors. To find associated factors we based on the Rotated Component Matrix table after tuning the parameters. We considered the accepted loading values >= 0.4. The number of factors were identified according to Eigen -value. The reason to use (PCA) because it is considered as one of the most accurate and popular methods of mathematical analysis.

Table 4.2 shows the factors which had a strong correlation that can be generalized to the academic achievement factor.

Table 4.2: The Extracted Component

| **Component Name (Academic achievement )** | **Loading's values** |
|---|---|
| Track | 0.8 |
| Contemporary Issues subject Average | 2.220 |
| Islamic subject Average | 1.820 |
| History subject | 2.168 |
| Health subject Average | 1.365 |
| Geography subject Average | 1.809 |
| English subject Average | 1.449 |
| The preferred subject | 1.513 |
| Characteristic analyzing factor | 0.614 |
| Favorite documentaries | 0.541 |

In the Table 4.3 illustrated the influential factors affected in specialization selection using Principle Component Analysis (PCA).

Table 4.3: The Extracted Factors by Using (PCA)

| **The Factor Number** | **The Factor Name** |
|---|---|
| Factor 1 | Academic achievement |
| Factor 2 | The preferred subjects |
| Factor 3 | The character you cast at childhood |
| Factor 4 | Father career |
| Factor 5 | The reason to choose your university |
| Factor 6 | Reason to choose your major |

### 4.1.3 Identifying the Final Local Influential Factors by Using (CFS) and (PCA)

To understand the main influential factors in our area, we used Principle Component Analysis (PCA) and the filter Correlation-based Feature Selection (CFS) method. Then we identified from both of results the local influential factors affected specialization selection at Palestine. The tables below illustrate the influential factors. According to our study. The first influential factor is grade and academic achievement, this factor represented by Tawjih average and the track at the secondary stage this factor was identified by [4, 5]

According to the results, we proved their effects in Palestine. The second influential factor is the demographic factor which represented with the student gender, this influential factor was identified by [6, 11, 10] further more our proposed model proved this factor as affecting factors in specialization selection. The third influential factor is the personal factor which represented with the student interests, skills, talents and mental abilities, it was studied and identified by [8, 10, 12, 24], the fourth influential factor is family background factors as in [9, 8], affected by parents educational degree, parents occupational classification and parents socioeconomic status. According to our results, family factor represented with father career, which affected specialization selection in Palestine.

The fifth influential factor is university was identified by [26, 14] which represented with the reasons to choose the university as reputation and place of the university, financial aids, availability of academic programs. According to our results, we proved this factor as an influential factor in specialization selection in Palestine. figure 4.2 illustrated the final local influential factors in Palestine.

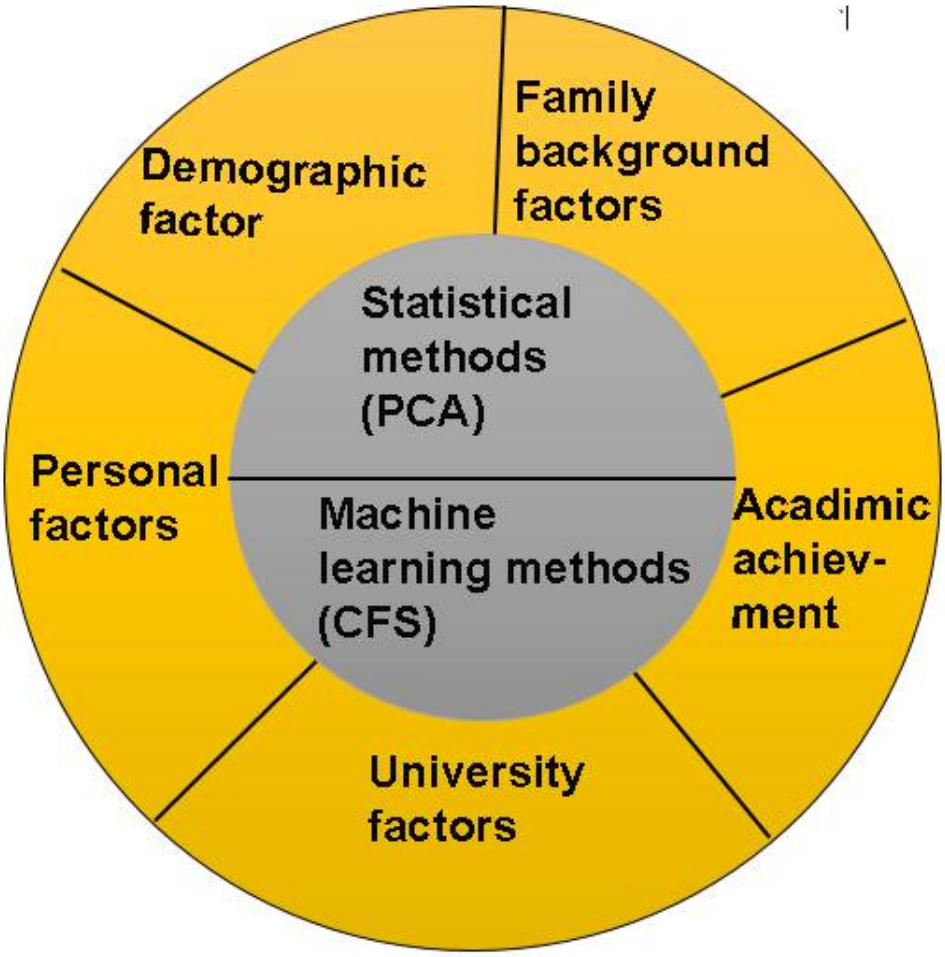


Figure 4.2: The basic local influential factors in Palestine

## 4.2 Developing the Proposed Classification Model

One of the main goals in this research is developing classification models to help high school students to select university specialization. We used PART classifier through WEKA environment and tuning the parameters throw WEKA to obtain optimal classification model. (k-fold) cross validation was used for evaluation to ensure each training set (k-1) times and once for testing. A cross-validation is applied to evaluate performance predictive model. 4 folds for training the model and 1 fold for testing. By applying (5 folds) cross-validation the accuracy of our proposed model was improved as shown in the Table 4.4. Our proposed model was built based on students instances at the undergraduate level, who belong to all university years, from the first year to fifth-year students. This sample of students instances used as a training data to teach and train our classification model. We found 628 valid students instance. And 7 class label (major). The accuracy of the development model was 67.7% when split training data into 5 folds using cross-validation. We improved the accuracy to 72.1% when split training data into 9 folds. Then we make experiments by reducing the number of classes (majors) and tuning the cross-validation into 5 folds as well we used filters to get rid of the irrelevant features then the accuracy improved to 77.5%. According to the experiments and results, we noted that whenever reduced the number of the classes, then the accuracy will be improved.

### 4.2.1 Improved the Proposed Model

To improve our proposed model we making features engineering by puts a hypothesis, for example, we merging two similar features into one feature and calculate the average weight for them. But the accuracy was not improved

a lot.

Table 4.4: The accuracy of the development model vs number of folds

| Number of folds | Accuracy |
|---|---|
| 3 folds | 76.9% |
| 4 folds | 73.2% |
| 5 folds | 77.4% |
| 6 folds | 74.8% |
| 7 folds | 75% |
| 8 folds | 73.9% |
| 9 folds | 76.1% |
| 10 folds | 73.4% |

Figure 4.3 illustrated part of the code was used to build the PART classifier using WEKA environment.

```
package javaapplication6;
import java.io.BufferedReader;
import java.io.FileReader;
import weka.core.Instances;
import weka.classifiers.rules.PART;
public class JavaApplication6 {

   public static void main(String[] args, Instances instance12) throws Exception {
       PART part1=new PART();
Instances instance1;
    instance1 = new Instances(new BufferedReader(new FileReader("C:\\Program Files\\Weka-3-8\\Lubaba-Data.arff")));
 //set class index to the last attribute
   System.out.println(instance1);
instance1.setClassIndex(instance1.numAttributes()-1);
part1.setUnpruned(false);

part1.setNumFolds(3);
part1.setMinNumObj(2);
part1.setReducedErrorPruning(false);
part1.setConfidenceFactor(0.2f);
part1.setBinarySplits(true);
part1.buildClassifier(instance1);
part1.setSeed(1);
    part1.getClass();
    part1.buildClassifier(instance1);

 System.out.println(part1);
   }
```

Figure 4.3: Building PART classifier by using Java code

### 4.2.2 Rules Extractions

It is represented as a map to help students choosing their future specialization. The extracted rules represented in an understandable language built

using classifier PART. After developing our classification model we extracted these rules. The goals of these rules are to understand the conditions of enrollment in specialization at the university. Each rule is a pattern includes a condition or a set of conditions and factors to meet with each other to represent a rule. The following tables illustrated the extraction rules that help students in choosing their specialization in the future.

Table 4.5: The Rules of Studying Arts

| |
|---|
| 1- The student can study this specialization if track = literary<br>AND English Language Avg = v.good AND track = literary AND skills = multi- Languages |
| 2- The student can study Arts if track = literary AND skills = different Cultures |
| 3- The student can study Arts if track = literary AND English-Language Avg = excellent |
| 4-The student can study Arts if track = literary AND Tawjihi Avg = 90-94.9 |
| 5- The student can study Arts if track = literary AND Tawjihi Avg = 85-89.9 |

According to the extracted rules in the Table 4.5, we concluded that the students who are familiar with language skills have interests in learning about other culture, and their Tawjihi literary stream average ranges between (85-94.9), then our proposed model found there was compatibility between this feature and the specialization, so they were advised by the proposed model to study Arts specialization.

At engineering specialization, we noted that the students who were distinguished with problem-solving skills, designing skills, and communication, so they were advised to study engineering and or the subjects related to technology, English language, Arabic language and geography-history. Furthermore, our proposed model encouraged the males to enroll in one of these subjects, we noted that the students who did experiments at the childhood stage, as fixing devices, making scientific experiments or driving the car at adulthood, can study engineering. Table 4.6 illustrated the extracted rules of studying engineering specialization.

According to the extracting rules from our proposed model, we found that

Table 4.6: The Rules of Studying Engineering

| |
|---|
| 1-The student can study Engineering if Problem-Solving Skill = moderate<br>AND gender = male<br>AND breeding Animal At Child= no AND scientific EXP at Child = chemical Exp AND the preferred Subject = other |
| 2- The student can study Engineering if skills = designing |
| 3- The student can study Engineering if gender = male AND breeding Animal At Child = no<br>AND scientific EXP at Child = chemical Exp AND Problem Solving Skill = moderate<br>AND the preferred Subject=technology AND tawjihi Avg = 80-84.9 |
| 4- The student can study Engineering if gender = male<br>AND breeding Animal At Child = no<br>AND scientific EXP at Child = chemical Exp AND Problem Solving Skill = moderate<br>AND the preferred Subject = Scientist Subject |
| 5- The student can study Engineering if the preferred Subject = geography-history<br>AND gender = male |
| 6- The student can study Engineering if track= Scientific AND<br>skills = communication |
| 7- The student can study Engineering if the preferred Subject = English Language AND gender = male |
| 8- The student can study Engineering if the preferred Subject = Arabic Language |
| 9- The student can study Engineering if driving Car = no |
| 10- The student can study Engineering if the preferred Subject = geography-history<br>AND gender = male |
| 11- The student can study Engineering if the preferred Subject = technology |
| 12- The student can study Engineering if fixing Devices = yes |

students with Tawjihi average ranges between (75-79.9), and the university near the residential place, the outputs of this group advised them to study agriculture. As well if they prefer the subjects of science, health and technology and they did biological or physical experiments at the childhood stage, so the proposed model advised them to study agriculture. This feature, whether combined or separated, affects in the selection of the agricultural major, Table 4.7 illustrates the rules that help to study agriculture

Table 4.7: Rules of Studying Agricultural

| |
|---|
| 1- The student can study Agricultural if breeding Animal At Child = yes<br>AND scientific EXP at Child = physical Exp |
| 2- The student can study Agricultural if breeding Animal At Child = yes<br>AND scientific EXP at Child = bio Exp AND the preferred Subject = Scientist Subject |
| 3- The student can study Agricultural if scientific EXP at Child = bio Exp<br>AND preferred Major Before = Eng Major |
| 4- The student can study Agricultural if scientific EXP at Child = bio Exp<br>AND the Reason To Choose University = nearness |
| 5- The student can study Agricultural if scientific EXP at Child = bio Exp<br>AND the Reason To Choose University = future Jop Fits |
| 6- The student can study Agricultural if the preferred Subject = health Subj |
| 7- The student can study Agricultural if the preferred Subject = technology<br>AND Tawjihi Avg = 75-79.9 |

Based on the extracted rules in Table 4.8, we found that students chose the literary track at secondary stage and have skills of presentation, problem-

solving and analysis-inference, and if the students Tawjihi average ranges between (70-74.9), then our proposed model advised them to study education.

Table 4.8: The Rules to studying Education

| |
|---|
| 1- The student can study Education if track = literary AND skills = analysis-inference |
| 2- The student can study Education if track = literary AND skills = Problem Solving THEN STUDY Education |
| 3- The student can study Education if track = literary AND skills = presentation |
| 4-The student can study Education if track = literary AND Tawjihi Avg = 70-74.9 |
| 5- The student can study Education if track = literary |

According to the extracted rules in the Table 4.9, we concluded that student who studied at the scientific track and preferred the subjects of maths or science, did biological experiments or bred animal at the childhood stage, such students were advised by the proposed model to study science.

Table 4.9: The Rules of Studying Sciences

| |
|---|
| 1- The student can study Sciences if breeding Animal At Child = yes<br>AND the preferred Subject = Scientist Subject |
| 2- The student can study Sciences if breeding Animal At Child = yes AND the preferred Subject = maths |
| 3- The student can study Sciences if scientific EXP at Child = bio Exp<br>AND preferred Major Before = pharmacy AND breeding Animal At Child = no |
| 4-The student can study Sciences if track = Scientific AND scientific EXP at Child = physical Exp |
| 5- The student can study this specialization if preferred Major Before = IT AND Animal breeding At Child = yes |
| 6-The student can study Sciences if the preferred Subject = maths |

We concluded from Table 4.10that the students who preferred science subjects and the average of Tawjihi ranges between (90-94.5) can study pharmacy.

Table 4.10: The Rules of Studying Pharmacy

| |
|---|
| 1-The student can study Pharmacy if the preferred Subject = other<br>AND Tawjihi Avg = 90-94.9 |
| 2- The student can study Pharmacy if the preferred Subject = Scientist Subject |

As shown by the proposed model, studying Computer Sciences and Information technology for the students who prefer management, technology or English language, in addition to females with Tawjihi average between (90-94.9). also, the students who are distinguished with persuasion or fixing devices were encouraged to study Computer Sciences and Information

Technology. Table 4.11 contains the extracted rules that help high school students to choose.

Table 4.11: The Rules of Studying Computer Sciences and Information Technology

| |
|---|
| 1- The student can study CS &IT if skills = persuasion<br>AND Problem Solving Skill = few |
| 2- The student can study CS &IT if the preferred Subject = management |
| 3- The student can study CS &IT if the preferred Subject = technology<br>AND breeding Animal At Child = no AND skills = persuasion AND gender = male |
| 4- The student can study CS &IT if the preferred Subject = technology<br>AND gender = female AND Tawjihi Avg = 90-94.9 |
| 5- The student can study CS &IT if preferred Major Before = sciences |
| 6- The student can study CS &IT if prefer Major Before = arts |
| 7- The student can study CS &IT if prefer Major Before = other<br>AND the Reason To Choose University = good Reputation |
| 8- The student can study CS &IT if prefer Major Before = pharmacy |
| 9- The student can study CS &IT if the preferred Subject = other AND fixing Devices = yes |
| 10- The student can be studying CS &IT if the preferred Subject = technology AND gender = female |
| 11- The student can study CS &IT if the preferred Subject = English Language |

## 4.3 Testing our Prediction Model (Model Verification)

After building our proposed model we tested it by using other students instances with an unknown class, they wanted to know the future specialization selection at the university. The number of students was 70. The accuracy testing of the developing model was 73.7%.

## 4.4 Performance Measurement

### 4.4.1 Confusion Matrix

Based on the results as extracted from the developing classifier, Table 4.12 showed:

1. There are 177 students study Engineering. After training the classification model, 137 students were classified into engineering correctly.

But 40 students incorrectly classified.

2. There are 115 students studied Information Technology and Computer engineering (IT-CE) the classification model classified 85 students correctly to IT-CE. But 30 students incorrectly classified.

3. There are 81 students studied Science, the classification model classified correctly 32 students to Science. But 49 students incorrectly classified.

4. There are 60 students at the Agriculture, 20 were classified correctly, whereas 40 were not.

5. There are 61 students study Pharmacy. The classification model classifies them correctly to 30 at Pharmacy. But 31 students incorrectly classified.

6. There are 72 students who studied Education. The classification model classified 62 students correctly at Education but 10 students incorrectly classified.

7. There are 63 students who study Arts, the classification model classified them into 57 of Art, but 6 students incorrectly classified.

From the previous confusion matrix results, we noted that in some majors the classification model miss classified them because some of the majors are similar in the features. For this reason, the classification model may miss classified it. Table 4.12 represents the confusion matrix with n*n size. It is provided with a description of classifying the examples into class labels based on a predictive model.

Table 4.12: The results of confusion matrix for our prediction model

| A | B | C | D | E | F | G | Class label |
|---|---|---|---|---|---|---|---|
| 137 | 24 | 5 | 1 | 10 | 0 | 0 | A= Engineering |
| 16 | 85 | 9 | 0 | 5 | 0 | 0 | B=IT and CE |
| 5 | 14 | 32 | 15 | 15 | 0 | 0 | C= Sciences |
| 2 | 7 | 23 | 20 | 8 | 0 | 0 | D= Agricultural |
| 5 | 10 | 9 | 7 | 30 | 0 | 0 | E= Pharmacy |
| 0 | 1 | 0 | 0 | 0 | 62 | 9 | F= Education |
| 1 | 0 | 0 | 0 | 0 | 5 | 57 | G= Arts |

### 4.4.2 Area Under Roc Curve

The area under roc curve makes a description of the examples that classified correctly in the proposed prediction model. Whenever the ROC curve for the classifier is nearest to 1 value, then the average of the AUC is well. The AUC value ranged between [0-1]. Suppose we have AUC=1, AUC of the class label $=x; x > 0; x < 1$ then $ErrorRate = AUC - x$. The error rate value represents the proportion that the classifier failed in classifying the training set correctly. Based on the Table 4.13 show the results of the AUC for each class (major). The highest value of AUC for Arts major (AUC, 0.986), the next was for Education (AUC, 0.974). Then Engineering (AUC, 0.926) Then IT&CE (AUC, 0.879). The next major is Agriculture (AUC, 0.829) and the Pharmacy (AUC, 0.824) and the lowest AUC was 0.818 for Sciences major. So this value nearest to 1 value and we can depend on it to predict the suitable major in the future dataset. From the results, we note that there is a relationship between AUC value and the number of students in each major. Whenever the number of students in the class is big, then the AUC will be the best.

Table 4.13: Area Under Roc curve for the predication model using PART

| Class | TP Rate | FP Rate | Precision | ROC Area |
|---|---|---|---|---|
| Engineering | 0.774 | 0.064 | 0.825 | 0.926 |
| IT and CS | 0.739 | 0.109 | 0.603 | 0.897 |
| Sciences | 0.395 | 0.084 | 0.410 | 0.818 |
| Agricultural | 0.333 | 0.040 | 0.465 | 0.829 |
| Pharmacy | 0.492 | 0.067 | 0.441 | 0.824 |
| Education | 0.861 | 0.009 | 0.925 | .0974 |
| Arts | 0.905 | 0.016 | 0.864 | 0.986 |
| ROC Average | | | | 0.899 |

### 4.4.3 The ROC Under Curve for the Majors

The Receiver Operating Characteristics is a graphical representation used to measure the performance of the classifier. The ROC curve contains two coordinates which are sensitivity and specificity [39]. In figures 4.5, 4.6 and 4.9 represent the best values of AUC for (Arts,0.986), (Education,0.974) and (Engineering,0.926) respectively.

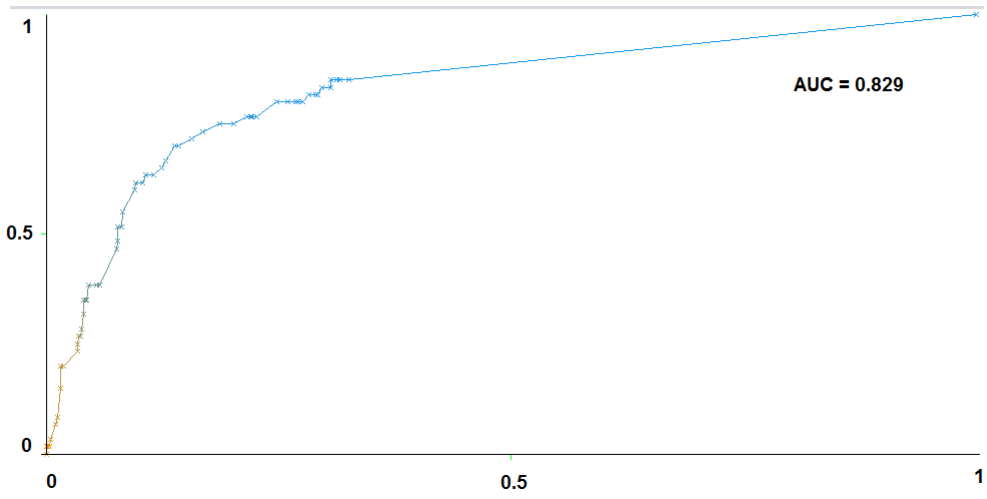


Figure 4.4: ROC Curve for Agricultural major

Then the AUC value for (Computer Sciences and Information Technology, 0.879) as illustrated in figure 4.7. In the figures 4.9, 4.8 the AUC values are (Pharmacy,0.824), (Arts,0.986) and the last AUC value for (Sciences,0.818)

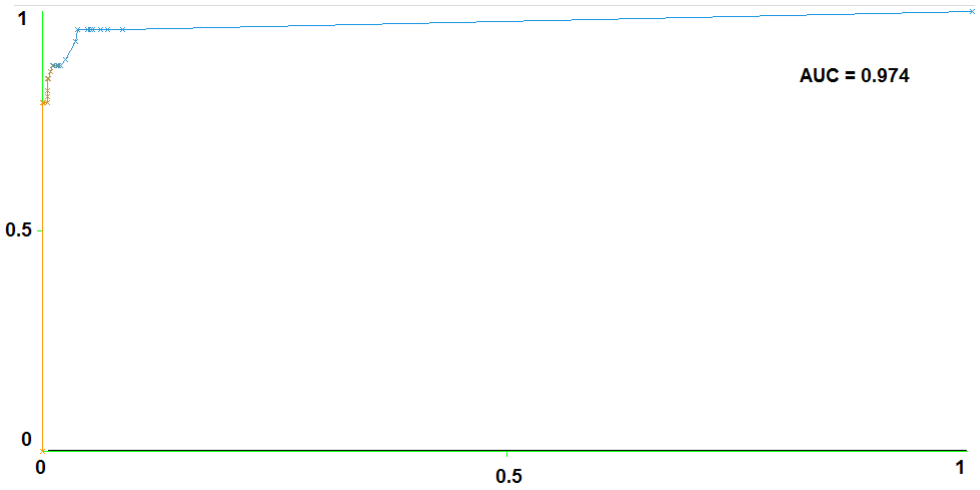


Figure 4.5: ROC Curve for Education major

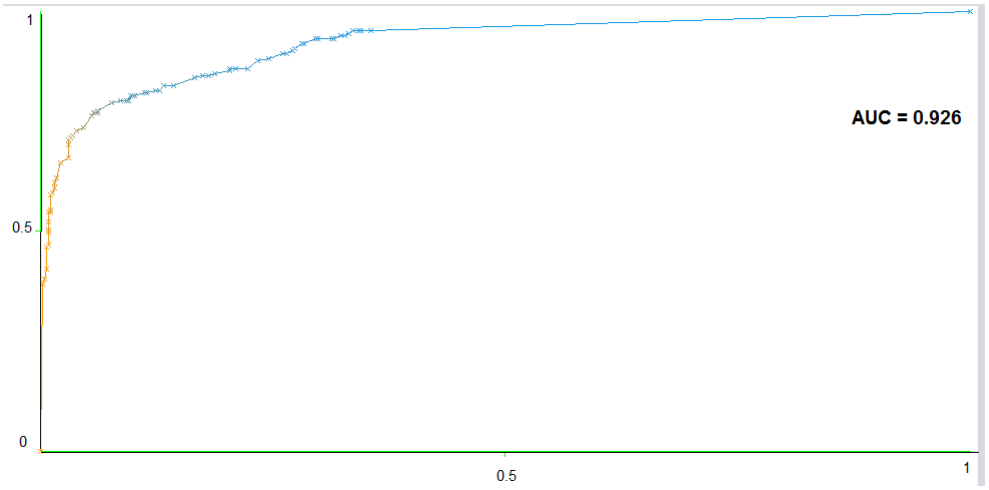


Figure 4.6: ROC Curve for Engineering major

illustrated in the figure 4.10.

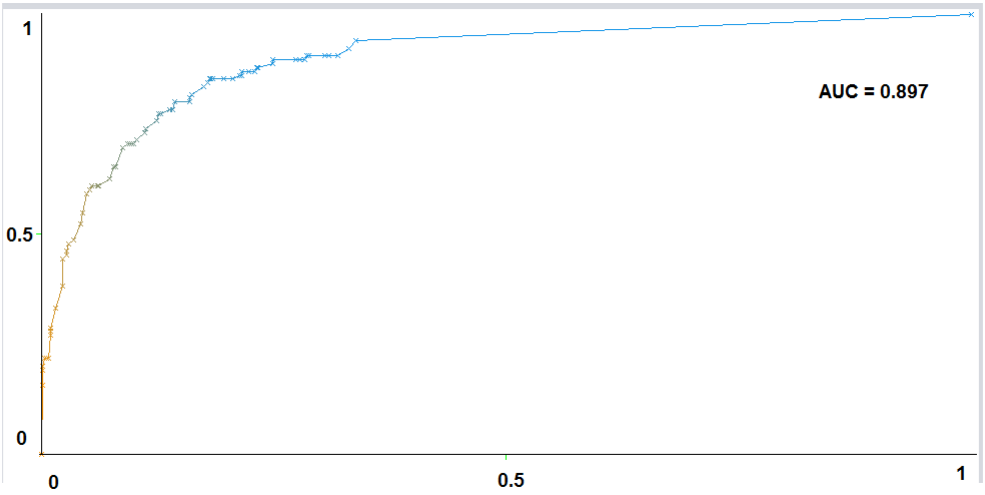


Figure 4.7: ROC Curve for IT-CE major

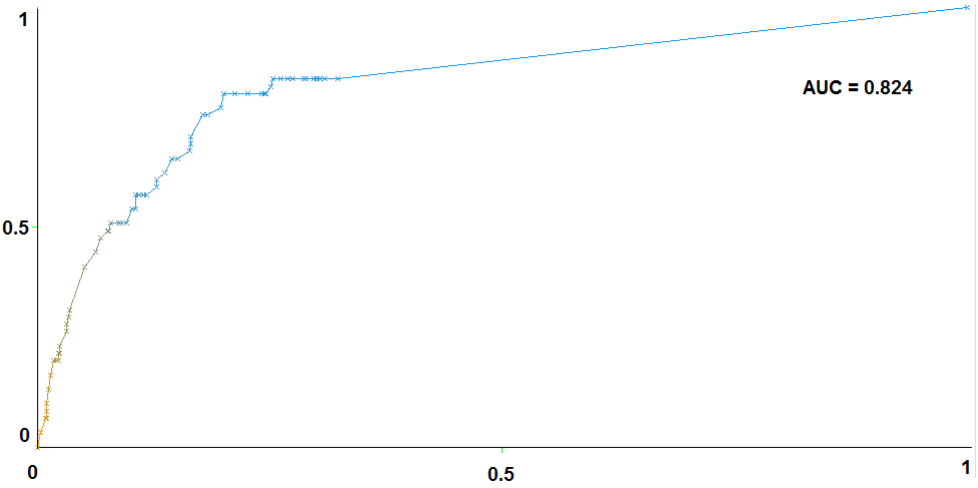


Figure 4.8: ROC Curve for Pharmacy major

Table 4.15 illustrated the comparison of our proposed model before reduced number of majors and after that.

From the results of the Table 4.15, we found that the accuracy of our model with 10 classes is (ACC,43.2%), (AUC average, 0.788). But with reduced number of classes (majors) to 7 classes; the accuracy of AUC was improved to (ACC, 72.1%), (AUC average, 0.908). While the majors were

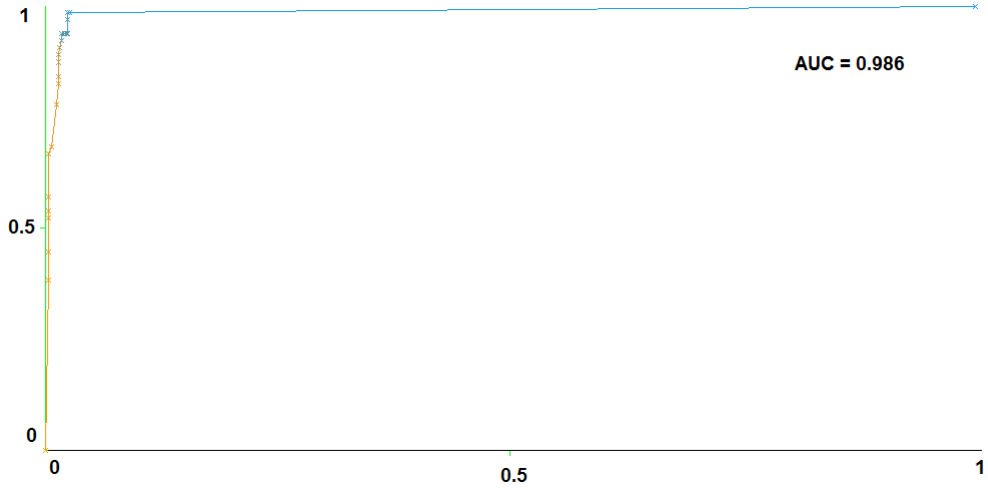


Figure 4.9: ROC Curve for Art major

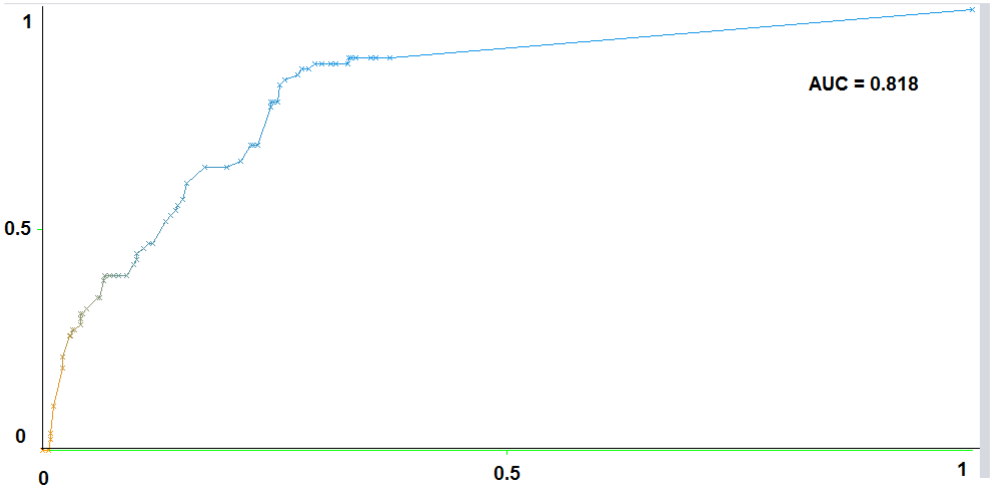


Figure 4.10: ROC Curve for Sciences major

Table 4.14: Confusion matrix for 6 class label

| A | B | C | D | E | F | Class label |
|---|---|---|---|---|---|---|
| 151 | 15 | 4 | 7 | 0 | 0 | A = Engineering |
| 16 | 81 | 10 | 8 | 0 | 0 | B = IT-CE |
| 4 | 20 | 108 | 9 | 0 | 0 | C = Sciences |
| 4 | 10 | 20 | 27 | 0 | 0 | D = Pharmacy |
| 0 | 0 | 0 | 0 | 62 | 10 | E = Education |
| 0 | 1 | 0 | 0 | 4 | 58 | F = Arts |

Table 4.15: Comparing PART results with changing number of classes

| Number of rules | Number of classes | Accuracy | ROC Curve average |
|---|---|---|---|
| 49 | 6 | 77.4% | 0.926 |
| 52 | 7 | 72.1% | 0.908 |
| 76 | 10 | 43.7% | 0.788 |

reduced to 6, the model building became more accurate than before with (ACC, 77.4%) and the performance of AUC was (0.926). Based on our experiments and results, and by reducing the number of classes from 10 classes then to 7 classes then to 6, the ACC and AUC were improved.

## 4.5 Comparing our Proposed Model with Other Approach

In this experiment, we are comparing our proposed model with another classifier to validate performance. We used the RIPPER classifier for comparison [50]. In the Tables below, we found that the extracted rules of RIPPER method were less detailed and contained few features and fewer rules than PART method. Based on the rules, each rule had one repeated idea at each time, which cannot be relied on to be a guideline to help the high school for future specialization selection. Tables 4.16, 4.17 4.18, 4.19, 4.20 show the extracted rules using RIPPER approach.

Table 4.16: The Rules of Studying Agricultural

| |
|---|
| 1- The student can study Agricultural if animal breeding at child=yes<br>and making scientific experiments at child = bio experiments |
| 2- The student can study Agricultural if making scientific experiments at child = Physical experiments<br>and animal breeding at child=yes |
| 3- The student can study Agricultural if animal breeding at child=yes and the preferred subject = technology<br>and making scientific experiments at child= chemical experiments<br>and the student was attend to study nursing. |
| 4- The student can study Agricultural if the reason to choose university = learning efficiency |

According to the results shown in Tables 4.21 and 4.22, the classifier PART with (ACC, 77.4%) is better than the classifier with RIPPER [50]

Table 4.17: The Rules of Studying Pharmacy

| |
|---|
| 1- The student can studying Pharmacy if the preferred subject = scientist subject and Tawjihi average = 95-100 |
| 2-The student can studying Pharmacy if Tawjhi Average = (90-94.5)<br>and make scientific experiments at child = bio experiments |

Table 4.18: The Rules of Studying Arts

| |
|---|
| 1- The student can study Arts if the track = literary<br>and English Language average = excellent |
| 2- The student can study Arts if the track = literary<br>and skills = knowledge different culture |
| 3- The student can study Arts if the track = literary<br>and Tawjihi average = (90-94.9) |
| 4- The student can study Arts if the track = literary<br>and Tawjihi average = (85-89.9)<br>and the preferred subject = English language |
| 5- The student can study Arts if the track = literary<br>and Tawjihi average = (95-100) |
| 6- The student can study Arts if the track = literary<br>and Tawjihi average = (85-89.9) and the students was attend to studying nursing |

Table 4.19: The Rules of Studying Sciences

| |
|---|
| 1- The student can study Sciences if animal breeding at child = yes |
| 2- The student can study Sciences if gender = female |
| 3- The student can study this specialization if animal breeding at child = yes<br>and the preferred subject = maths |

Table 4.20: The Rules of Studying Computer Sciences and Information Technology

| |
|---|
| 1- The student can study CS & IT if animal breeding at child = yes |
| 2- The student can study CS & IT if gender = female |
| 3- The student can study CS & IT if animal breeding at child = yes and the preferred subject = maths |
| 4- The student can study CS & IT if<br>the preferred subject = technology and skills = persuasion and gender = female |
| 5- The student can study CS & IT<br>if Problem solving skill = few |
| 6- The student can study CS & IT if skills = persuasion<br>and (the preferred subject = technology and Tawjihi average = (70-74.9) |
| 7- The student can study CS & IT if skills = persuasion<br>and Tawjihi average = (75-79.5) |
| 8- The student can study CS & IT if Problem solving skill = Vsignifi) |
| 9- The student can study CS & IT if Problem solving skill = signify |
| 10- The student can study CS & IT if gender = female and Tawjihi average = (80-84.5)<br>and the preferred subject = Arabic Language the preferred subject = management |

(ACC, 71%) for 6 classes (majors). And the ACC of PART for 7 classes is better than RIPPER classifier.

Table 4.21: The result of comparison between PART and RIPPER for 7 classes (majors)

| **Classifier** | **Number of rules** | **Accuracy** |
|---|---|---|
| PART | 52 | 67.7% |
| RIPPER | 27 | 64.7% |

Table 4.22: The result of comparison between PART and RIPPER for 6 classes (majors)

| **Classifier** | **Number of rules** | **Accuracy** |
|---|---|---|
| PART | 49 | 77.4% |
| RIPPER | 20 | 71% |

### 4.5.1 Confusion Matrix

In this section we explain Confusion matrix for RIPPER classifier as illustrated in Table 4.23

1. are 177 students who studied Engineering. After training the classification model, 153 students were classified correctly to study engineering, but 24 students incorrectly classified.

2. There are 115 students who study Information Technology and Computer engineering (IT-CE), 72 classified by the model to study IT-CE, whereas, 43 students were incorrectly classified.

3. There are 81 students who study Sciences, but the classification model classified 27 students to study Science, but 45 students failed to be classified for the same.

4. There are 60 students who study Agriculture, only 40 students were correctly classified.

5. There are 61 students who study Pharmacy, but there are 48 students incorrectly classified.

6. There are 72 students who study Education, but the classification model classified 60 at Education, but 12 students were incorrectly classified.

7. There are 63 students who study Arts. The classification model classified 57 of Arts, but 6 students were incorrectly classified.

Table 4.23: Confusion matrix for 7 class label by using RIPPER algorithm

| A | B | C | D | E | F | G | Class label |
|---|---|---|---|---|---|---|---|
| 153 | 18 | 1 | 1 | 4 | 0 | 0 | A = Engineering |
| 40 | 72 | 1 | 0 | 2 | 0 | 0 | B = IT-CE |
| 30 | 8 | 27 | 11 | 5 | 0 | 0 | C = Sciences |
| 18 | 6 | 16 | 20 | 0 | 0 | 0 | D = Agricultural |
| 36 | 5 | 3 | 4 | 13 | 0 | 0 | E = Pharmacy |
| 0 | 0 | 0 | 0 | 0 | 60 | 12 | F = Education |
| 0 | 0 | 0 | 0 | 0 | 6 | 57 | G = Arts |

## 4.6 Model Evaluation

To evaluate our proposed model we will compare it with other models in terms of a number of features, the number of classes, and the accuracy.

Radaida et al. proposed classification model in [5], the researcher found that the accuracy of the building model is a very good rate 87.9%. This model was proposed to support high school students to choose the suitable track at second secondary stage (Tawjihi), they depend on the student's cumulative grade point average (CGPA) at 8th, 9th, 10th grades then there existed one feature which is the student marks, and the number of a class label (track) is 4 classes so the results of this model were very good because the number of classes were suitable with the number of features. Then will did not found overfitting problem. This model was built by C4.5 tree represented as a decision tree to map the conditions. Figure 4.12 illustrated the generated decision tree. Another study proposed a classification model to help high-school students in the selection of undergraduate programs by Jirapanthong [4], this model building using C4.5. The variables (features) contained educational status: Third-year and Fourth-year, high School performance, high school GPA, old major, student demographics, gender. The number of the feature is compatible with the number of a class label (specialization)and the number of students samples around 4000 students so the probability to face overfitting problem is low. The accuracy of the building model after doing nine experiments in the first case was 87.9%. In the second case is 83.2%. The decision tree of the building classification model shows in figure 4.11

In contrast, our proposed model developed by using PART classifier logically different from C4.5. C4.5 it is a classification tree. But PART is classified based on predictive association rules. Our data set contains 52 features and 6 class label Some what we faced overfitting problem and the accuracy of our model is 77.4% so it is a good result. With comparing the accuracy of the previous classification models with the results of our model we concluded that our model is more comprehensive and flexible because it is directed to a

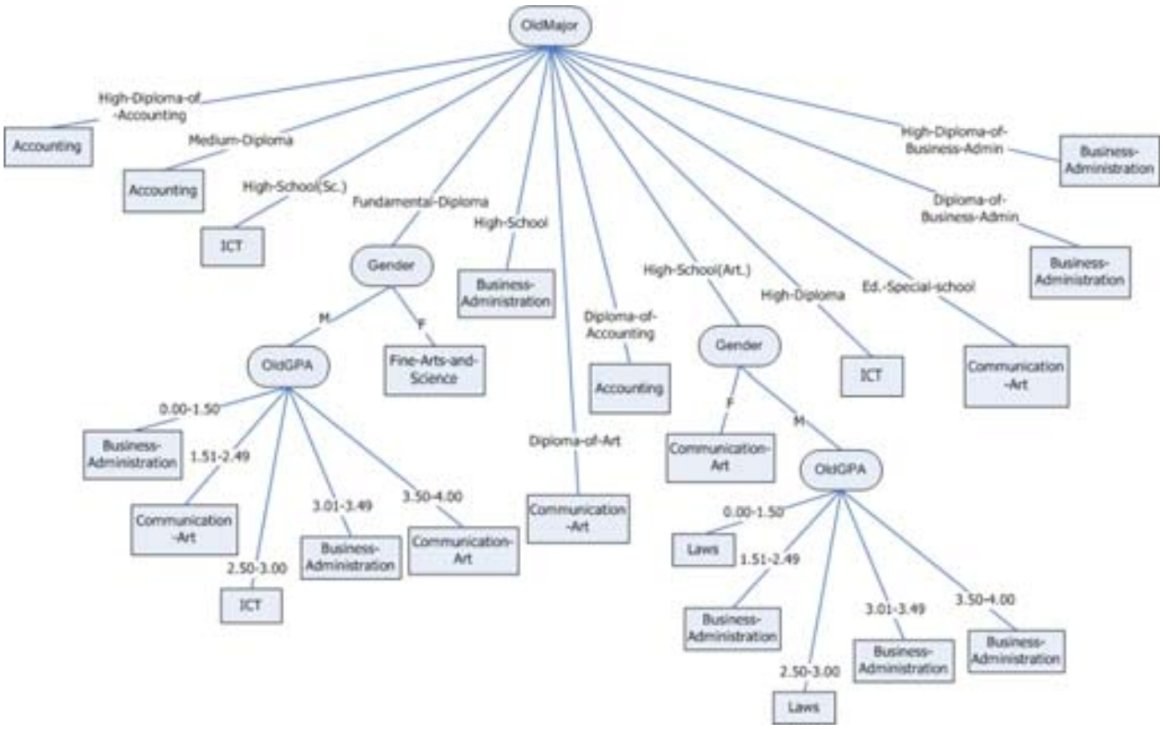


Figure 4.11: The generated decision tree [4]

wider range they are university students thus we take into account the factors that may affect with students specialization selection and we extracted the local influential factors in specialization selection at Palestine.

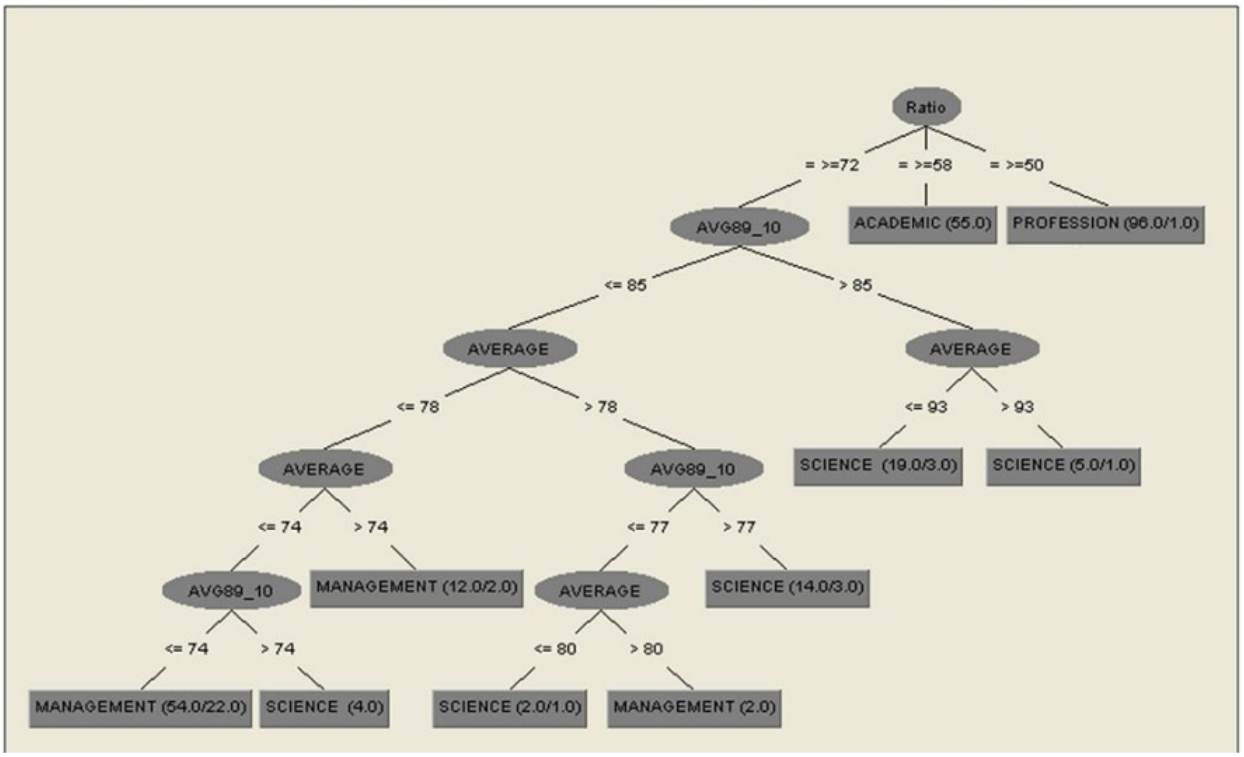


Figure 4.12: The generated decision tree [5]

# Chapter 5

# Conclusions and Recommendations

## 5.1 Introduction

This thesis has proposed a novel intelligent model for suitable university specialization selection using classification algorithm. The proposed model was developed using rule-based classifier (PART). The influential factors in university specialization selection taking into account during developing our proposed model.

## 5.2 Conclusions

One of the objectives of our research is to identify the influential factors affected the specialization selection based on literature review. Hence, we studied the influential factors and summarized them into the following factors; social factors, future Prospect factors, personal factors, grades and academic achievement, family background factors, financial factors, university factors and demographic factors. Each factor contains sub factors which we studied

it in details.

The second objective of our thesis is to identify the local influential factor in Palestine. To achieve this objective, we used reduce dimensionality technique which contains two methods, feature selection, and feature extraction. In the feature selection method, we used the filter Correlated-based Feature Selection (CFS) to identify the important factors which have a significant impact and lower error rates. In feature extraction method, we used Principle Component Analysis (PCA) to find the factors with the significant correlation between other and extract components include the correlated factors. The results of the research argued that there are five influential factors in Palestine which are family background factor, demographic factor, personal factor, university factor and academic achievement factor.

The third objective of our research is developing an intelligent model, which can help high school students to predict what specialization suitable for them in future. The accuracy of the model was 77.4%. Then the model was tested with a sample of students who were not involved in the training set. The accuracy of the model for the testing set was 73.2%, so the developing model was considered to be a good model.

## 5.3 Recommendations for the Ministry of Education in Palestine:

The adoption of an intelligent system in the Palestinian universities helps secondary students in determining their academic specialization commensurate with their scientific abilities and interests. Thus graduating them with high scientific efficiency, which helps to create jobs in the future, and reducing unemployment among graduates. As a result, economic conditions

were improved at the national level. In addition to mention progress in other areas such as health and agriculture because the basis of that is education. The key to prosperity of the homeland is the youth. So we tried to help and guide them to choose their specialization closest to their personality, interests, abilities and also commensurate with their circumstances, then leading them to creativity.

## 5.4 Recommendations for the Researchers:

Adoption of this system in universities must be examined for several years and compare the results to obtain more accurate results and credibility. We recommend the size of the training sample in thousands and then adopted as a system according to the accuracy of the results.

## 5.5 Recommendations for the Palestinian Universities:

We believe that this system should be in the universities because it is a map that helps secondary school students in determining the appropriate university disciplines and thus graduate students with high scientific efficiency and this is a certificate of excellence for the university.

## 5.6 Recommendation for the Students:

The use of such a model by students is considered a guide to provide them with the best and most appropriate specialization because it takes into account the factors surrounding the student that affect the choice of the stu-

dent as economic, social, demographic and personal. This intelligent system enables the student to understand his personality, abilities, interests and economic conditions in a deeper way, and to define his academic and professional objectives.

## 5.7 Future Work

In the future, it is possible to use other algorithm than PART as CPAR (Classification based on Prediction Association Rules) to developing this intelligent model, because this algorithm logically works as PART. Both of them are rule -based classification algorithms. To improve our proposed model to be more flexible and more intelligent by allowing saved input data sets and using it to train the model incrementally as illustrates in figure 5.1. Hence, the new data set can incrementally be learned from the old data set. This issue can be done by using the incremental learning algorithm.

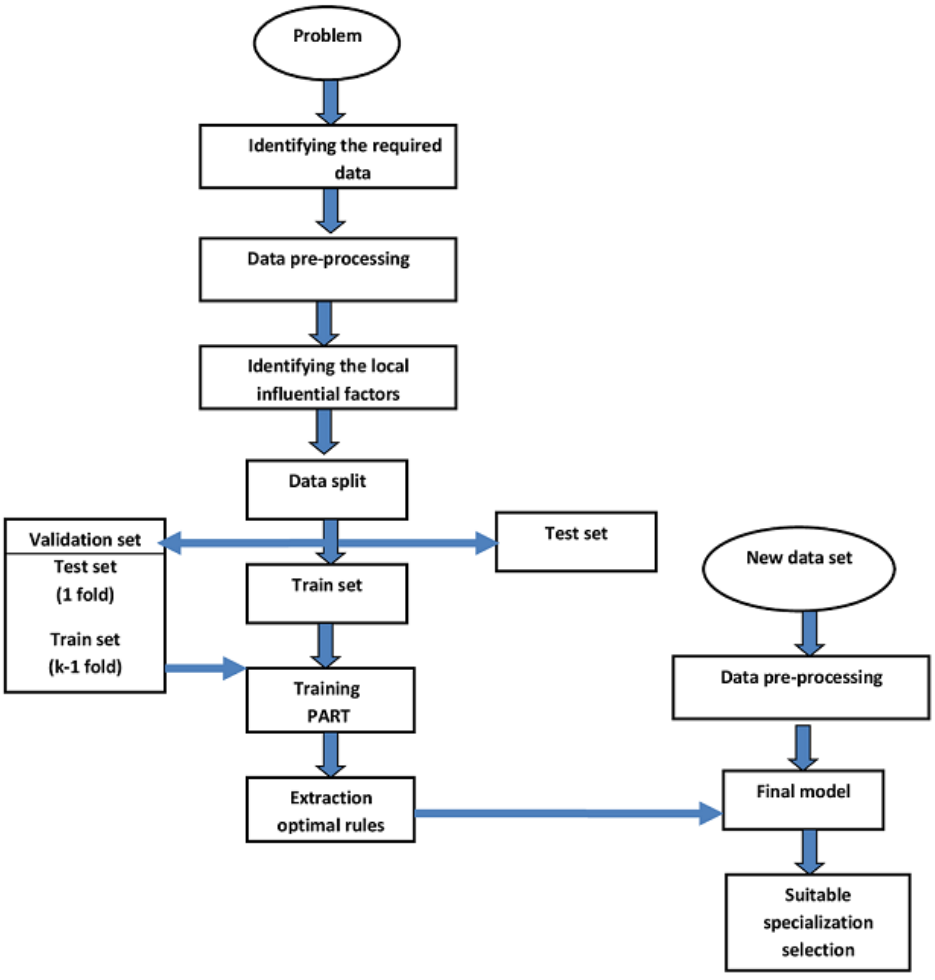


Figure 5.1: The proposed classification model

# Bibliography


[1] Cesar Vialardi Sacin, Javier Bravo Agapito, Leila Shafti, and Alvaro Ortigosa. Recommendation in higher education using data mining techniques. In *Educational Data Mining 2009*, 2009.

[2] R Sathya and Annamma Abraham. Comparison of supervised and unsupervised learning algorithms for pattern classification. *International Journal of Advanced Research in Artificial Intelligence*, 2(2):34–38, 2013.

[3] Mark A Hall. *Correlation-based feature selection for machine learning*. PhD thesis, The University of Waikato, 1999.

[4] Waraporn Jirapanthong. Classification model for selecting undergraduate programs. In *Natural Language Processing, 2009. SNLP'09. Eighth International Symposium on*, pages 89–95. IEEE, 2009.

[5] Q Al-Radaideh, Ahmad Al Ananbeh, and E Al-Shawakfa. A classification model for predicting the suitable study track for school students. *Int. J. Res. Rev. Appl. Sci*, 8(2):247–252, 2011.

[6] Kelechi U Lazarus and C Ihuoma. The role of guidance counsellors in the career development of adolescents and young adults with special needs. *British Journal of Arts and Social Sciences*, 2(1):51–62, 2011.

[7] Ahmad Nahar Al-Rfou. Factors that influence the choice of business major evidence from jordan. *Journal of Business and Management*, 8(2):104–8, 2013.

[8] Nancy T Pascual. Factors affecting high school students career preference: A basis for career planning program. *International Journal of Science: Basics and Applied Research*, 16(1):1–14, 2014.

[9] Ming-Shang Su, Ta-Chun Chang, Chin-Chang Wu, and Chin-Wen Liao. Factors affecting the student career decision-making of junior high school students in central taiwan area. *International Journal of Information and Education Technology*, 6(11):843, 2016.

[10] Aamir Sarwar and Rizwana Masood. Factors affecting selection of specialization by business graduates. *Science International*, 27(1), 2015.

[11] Stephen R Porter and Paul D Umbach. College major choice: An analysis of person–environment fit. *Research in Higher Education*, 47(4):429–449, 2006.

[12] Aniva Kartika. The effect of counseling in self and career development to enhance students awareness of their personal and study problems, 2008.

[13] NCC Lawrence K. Jones, Ph.D. *Choosing a College Major Based on Your Personality*, 2012 (accessed February 3, 2016). `http:https://www.montclair.edu/media/montclairedu/careerservices/pdfs/personality_college_major_match_research.pdf`.

[14] Treesha Beatrice Estaol De Mesa Katherine Denise Evangelista Martin Fritz Gerald Golamco Mestosamente Sharee Sonido Zaguirre Ma. Karlson Reye Alba, Kenneth Eduard Castillo Bertol. https://www.coursehero.com/file/12317190/the-factors-that-affect-students-decision-in-choosing-their-college-courses/, September 2016.

[15] Robert D Reason. Student variables that predict retention: Recent research and new developments. *Naspa Journal*, 40(4):172–191, 2003.

[16] M Beikzadeh and Naeimeh Delavari. A new analysis model for data mining processes in higher educational systems. *On the proceedings of the 6th Information Technology Based Higher Education and Training*, pages 7–9, 2005.

[17] Zlatko Kovacic. Early prediction of student success: Mining students' enrolment data. 2010.

[18] P Adinarayana Reddy, D Uma Devi, and E Mahadeva Reddy. A study of the vocational education preferences and interests of the indian undergraduate students. *Bulgarian Journal of Science and Education Policy (BJSEP)*, 5(1):94–114, 2011.

[19] Paulo Cortez and Alice Maria Gonçalves Silva. Using data mining to predict secondary school student performance. 2008.

[20] Yiming Ma, Bing Liu, Ching Kian Wong, Philip S Yu, and Shuik Ming Lee. Targeting the right students using data mining. In *Proceedings of the sixth ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 457–464. ACM, 2000.

[21] `http://www.vocabulary.com.` Retrieved at May 12,2016.

[22] Nealcheng Xeng Thao. *Examining family and community influences on the attitudes to education and career aspirations of Hmong/Mong high school students.* PhD thesis, University of Minnesota, 2009.

[23] `http://http://www.dictionary.com.` Retrieved at May 20,2016.

[24] David Kim, F Scott Markham, and Joseph D Cangelosi. Why students pursue the business degree: A comparison of business majors across universities. *Journal of Education for Business*, 78(1):28–32, 2002.

[25] Duyen Q Nguyen. The essential skills and attributes of an engineer: A comparative study of academics, industry personnel and engineering students. *Global J. of Engng. Educ*, 2(1):65–75, 1998.

[26] Amaury Nora. The role of habitus and cultural capital in choosing a college, transitioning from high school to higher education, and persisting in college among minority and nonminority students. *Journal of Hispanic Higher Education*, 3(2):180–208, 2004.

[27] J Breuker, R Dieng-Kuntz, N Guarino, JN Kok, J Liu, R López de Mántaras, R Mizoguchi, M Musen, and N Zhong. Frontiers in artificial intelligence and applications. *Frontiers in artijicial intelligence and applications*, 21, 1994.

[28] Sotiris B Kotsiantis, I Zaharakis, and P Pintelas. Supervised machine learning: A review of classification techniques, 2007.

[29] Pádraig Cunningham, Matthieu Cord, and Sarah Jane Delany. Supervised learning. *Machine learning techniques for multimedia*, pages 21–49, 2008.

[30] Noelia Sánchez-Maroño, Amparo Alonso-Betanzos, and María Tombilla-Sanromán. Filter methods for feature selection–a comparative study. *Intelligent Data Engineering and Automated Learning-IDEAL 2007*, pages 178–187, 2007.

[31] Vipin Kumar and Sonajharia Minz. Feature selection: A literature review. *SMART COMPUTING REVIEW*, 4(3):211–229, 2014.

[32] Manfur Sodik Anqori, Zaenal Arifin, and Dwi Sri Handayani. Relationship with the body position of lower back pain in cipta nugraha santosa building workers. *Jurnal Ilmu Kesehatan STIKES Duta Gama Klaten*, 6(1), 2014.

[33] E Alpaydin. Introduction to machine learning. 2-nd edition, 2010.

[34] Steven L Salzberg. C4. 5: Programs for machine learning by j. ross quinlan. morgan kaufmann publishers, inc., 1993. *Machine Learning*, 16(3):235–240, 1994.

[35] William W Cohen. Fast effective rule induction. In *Proceedings of the twelfth international conference on machine learning*, pages 115–123, 1995.

[36] Eibe Frank and Ian H Witten. Generating accurate rule sets without global optimization. 1998.

[37] Hazwani Rahmat, Aida Mustapha, Masniza Shaheeda Md Said, and Noor Afiza Amit. Measuring differences in accuracy, compactness, and speed between c 4. 5 and cpar in classification. *Journal of Communication and Computer*, 9(1):42–46, 2012.

[38] Robert C Holte. Very simple classification rules perform well on most commonly used datasets. *Machine learning*, 11(1):63–90, 1993.

[39] Andrew P Bradley. The use of the area under the roc curve in the evaluation of machine learning algorithms. *Pattern recognition*, 30(7):1145–1159, 1997.

[40] Ron Kohavi et al. A study of cross-validation and bootstrap for accuracy estimation and model selection. In *Ijcai*, volume 14, pages 1137–1145. Stanford, CA, 1995.

[41] Paula Lagares Barreiro and Justo Puerto Albandoz. Population and sample. sampling techniques. *Management Mathematics for European Schools MaMaEusch (994342-CP-1-2001-1-DECOMENIUS-C21*, 2001.

[42] Deanne Katherine Bird. The use of questionnaires for acquiring information on public perception of natural hazards and risk mitigation-a review of current knowledge and practice. *Natural Hazards and Earth System Sciences*, 9(4):1307, 2009.

[43] Mark Hall, Eibe Frank, Geoffrey Holmes, Bernhard Pfahringer, Peter Reutemann, and Ian H Witten. The weka data mining software: an update. *ACM SIGKDD explorations newsletter*, 11(1):10–18, 2009.

[44] `http://www.cs.waikato.ac.nz/ml/weka/` Retrieved at March 12,2017.

[45] `https://docs.google.com.`

[46] http://www.hebron.edu, September 2016.

[47] https://www.ppu.edu/p/sites/default/files/ppu%20guide%202016-2017.pdf, February 2017.

[48] Arthur D Chapman. *Principles and methods of data cleaning*. GBIF, 2005.

[49] Ian H Witten. Data mining with weka. *Class Lesson, Department of Computer Science University of Waikato, New Zealand.*, 2013.

[50] William W Cohen. Fast effective rule induction. In *Proceedings of the twelfth international conference on machine learning*, pages 115–123, 1995.

**An Appendix**

Table 5.1: Predictions on user test set

| Example number | unknown class label | actual predicted | error prediction |
|---|---|---|---|
| 1 | ? | IT-CE | 0.9 |
| 2 | ? | IT-CE | 1 |
| 3 | ? | IT-CE | 1 |
| 4 | ? | IT-CE | 1 |
| 5 | ? | EngMajor | 1 |
| 6 | ? | IT-CE | 1 |
| 7 | ? | IT-CE | 0.625 |
| 8 | ? | EngMajor | 1 |
| 9 | ? | EngMajor | 1 |
| 10 | ? | Pharmacy | 0.5 |
| 11 | ? | EngMajor | 1 |
| 12 | ? | EngMajor | 1 |
| 13 | ? | Agriculture | 0.5 |
| 14 | ? | IT-CE | 0.625 |
| 15 | ? | EngMajor | 1 |
| 16 | ? | EngMajor | 1 |
| 17 | ? | EngMajor | 1 |
| 18 | ? | Pharmacy | 0.5 |
| 19 | ? | Agriculture | 0.5 |
| 20 | ? | Agriculture | 0.5 |
| 21 | ? | Agriculture | 0.5 |
| 22 | ? | Education | 0.4 |
| 23 | ? | Sciences | 0.594 |
| 24 | ? | Education | 0.4 |
| 25 | ? | Education | 0.4 |
| 26 | ? | Arts | 0.571 |
| 27 | ? | Sciences | 0.594 |
| 28 | ? | Pharmacy | 0.4 |
| 29 | ? | Sciences | 0.594 |
| 30 | ? | Pharmacy | 0.882 |
| 31 | ? | Agriculture | 0.5 |
| 32 | ? | Agriculture | 0.5 |
| 33 | ? | Sciences | 0.594 |
| 34 | ? | Sciences | 0.594 |
| 35 | ? | Agriculture | 0.833 |
| 36 | ? | Pharmacy | 0.5 |
| 37 | ? | Pharmacy | 0.5 |
| 38 | ? | Agriculture | 0.667 |
| 39 | ? | Agriculture | 0.5 |
| 40 | ? | Agriculture | 0.75 |
| 41 | ? | Sciences | 0.594 |
| 42 | ? | Arts | 0.857 |
| 43 | ? | Pharmacy | 0.875 |
| 44 | ? | Agriculture | 0.75 |
| 45 | ? | Agriculture | 0.5 |
| 46 | ? | Pharmacy | 0.45 |
| 47 | ? | Pharmacy | 0.882 |
| 48 | ? | Pharmacy | 0.45 |
| 49 | ? | Pharmacy | 0.875 |
| 50 | ? | Arts | 0.905 |
| 51 | ? | Arts | 1 |
| 52 | ? | Arts | 0.857 |
| 53 | ? | Arts | 0.857 |
| 54 | ? | Arts | 1 |
| 55 | ? | Sciences | 0.667 k2 |
| 56 | ? | Arts | 1 |
| 57 | ? | Arts | 1 |
| 58 | ? | Sciences | 0.667 |
| 59 | ? | Education | 0.667 |
| 60 | ? | Sciences | 0.594 |
| 61 | ? | Arts | 1 |
| 62 | ? | Arts | 0.857 |
| 63 | ? | Education | 0.4 |
| 64 | ? | Education | 0.667 |
| 65 | ? | Education | 0.667 |
| 66 | ? | Education | 0.7 |
| 67 | ? | Education | 0.4 |

**طلب توزيع استمارة**

**حضرة السيد عميد الدراسات العليا المحترم**
**بواسطة منسق برنامج ماجستير المعلوماتية المحترم/ة**

يرجى العلم بأن الطالبة لبابة طميزه تعمل على بحث رسالة الماجستير بعنوان " نظام خبير لمساعدة طلبة الثانوية العامة في اختيار التخصص المناسب لهم في الجامعة باستخدام خوارزميات الذكاء الاصطناعي" وهي بحاجة لجمع معلومات من طلبة التخصصات المختلفة في جامعة الخليل و البوليتكنك من خلال توزيع استمارة الكترونية وورقية بين طلاب الجامعتين.
وعليه يرجى من حضرتكم التكرم بمخاطبة الجهات المختلفة وذات العلاقة للسماح للطالبة بتوزيع الاستمارات (الكترونياً وورقياً) لطلبة التخصصات المختلفة للسنوات الثانية والثالثة والرابعة (تغطي التخصصات).

**وتقبلوا فائق الاحترام**

**توقيع مشرفا الرسالة**
**د. غسان شاهين**
**د. رضوان طهبوب**

**رابط الاستمارة الالكترونية**
https://docs.google.com/forms/d/e/1FAIpQLScx_4oL9CZoiWBmCeyhzIeS8syiK4YRecY5eeZIsMM64Ic16g/viewform?usp=send_form

# استمارة بحث علمي

**عزيزي/تي الطالب/ة تحية طيبة وبعد أرجو من حضرتكم تعبئة هذه الاستمارة بعنوان (نظام خبير يساعد طلبة الثانوية العامة في اختيار التخصص المناسب لهم في الجامعة باستخدام خوارزميات الذكاء الاصطناعي) حيث أن نتائج هذه الاستمارة ستكون سرية وتهدف فقط الى البحث العلمي.**

**أولا : قسم المعلومات العامة**

1. الجنس أ. ☐ ذكر ب. ☐ أنثى
2. عدد أفراد الأسرة ( رقما).....
3. مكان السكن الأصلي:

أ. ☐ شمال الضفة الغربية (نابلس، سلفيت، جنين، قلقيلية، طوباس ، طولكرم)

ب. ☐ وسط الضفة الغربية (اريحا، القدس، رام الله)

ت. ☐ جنوب الضفة الغربية ( الخليل ،بيت لحم)

ث. ☐ داخل الخط الأخضر

4. الفرع:

أ. ☐ العلمي ب. ☐ الأدبي ت. ☐ التجاري

ث. ☐ الصناعي ج. ☐ الفندقي ح. ☐ المهني

5. المستوى:

أ. ☐ سنة أولى ب. ☐ سنة ثانية ج. ☐ سنة ثالثة د. ☐ سنة رابعة ه. ☐ سنة خامسة

1

**ثانيا: القسم الخاص بكل فرع (علمي -أدبي - تجاري- صناعي - مهني - فندقي)**

**(هذه الأسئلة خاصة بكل فرع يجيبها الطالب حسب فروعه في الثانوية العامة)**

**خاص بالفرع العلمي**

6. قدر علاماتك في مساق الرياضيات في مرحلة الثانوية ( العاشر – الحادي عشر- التوجيهي)

أ. ☐ مقبول ب.☐ متوسط ت. ☐ جيد ☐ث. جيد جدا ☐ ج. ممتاز ☐ ح. مقصر

7. قدّر علاماتك في مساق تكنولوجيا المعلومات في مرحلة الثانوية ( العاشر – الحادي عشر- التوجيهي)

أ. ☐ مقصر ب.☐ مقبول ت.☐ متوسط ث.☐ جيد ج. ☐ جيد جدا ح.☐ ممتاز

8. قدّر علاماتك في مساق الفيزياء في مرحلة الثانوية ( العاشر – الحادي عشر- التوجيهي)

أ. ☐ مقصر ب.☐ مقبول ت. ☐ متوسط ث.☐ جيد ☐ ج. جيد جدا ح.☐ ممتاز

9. قدّر علاماتك في مساق الكيمياء في مرحلة الثانوية ( العاشر – الحادي عشر- التوجيهي)

أ. ☐ مقصر ب.☐ مقبول ت.☐ متوسط ث.☐ جيد ج. ☐ جيد جدا ح.☐ ممتاز

10. قدّر علاماتك في مساق الأحياء في مرحلة الثانوية ( العاشر – الحادي عشر- التوجيهي)

أ.☐ مقصر ب☐ مقبول ت☐ متوسط ث☐ جيد ج.☐ جيد جدا ح.☐ ممتاز

11. قدّر علاماتك في مساق التربية الفنية خلال المراحل(الابتدائية- الأساسية- الثانوية)

أ. ☐ مقصر ب.☐ مقبول ت.☐ متوسط ث.☐ جيد ج. ☐ جيد جدا ح.☐ ممتاز

2

**Commented [l1]:**
90- 100 = ممتاز.
80 - 89 = جيد جداً.
70 - 79 = جيد.
60 – 69 = متوسط.
50 - 59 = مقبول.
49 فما دون = مقصر

12. قدّر علاماتك في مساق اللغة العربية في مرحلة الثانوية ( العاشر – الحادي عشر- التوجيهي)

أ. ☐ مقصر ب. ☐ مقبول ت. ☐ متوسط ث. ☐ جيد ج. ☐ جيد جدا ح. ☐ ممتاز

**خاص بالفرع الأدبي**

13. قدّر علاماتك في مساق اللغة العربية في مرحلة الثانوية ( العاشر – الحادي عشر- التوجيهي)

أ. ☐ مقصر ب. ☐ مقبول ت. ☐ متوسط ث. ☐ جيد ج. ☐ جيد جدا ح. ☐ ممتاز

14. قدر علاماتك في مساق اللغة الانجليزية في مرحلة الثانوية ( العاشر – الحادي عشر- التوجيهي)

أ. ☐ مقصر ب. ☐ مقبول ت. ☐ متوسط ث. ☐ جيد ج. ☐ جيد جدا ح. ☐ ممتاز

15. قدّر علاماتك في مساق الجغرافيا في مرحلة الثانوية ( العاشر – الحادي عشر- التوجيهي)

أ. ☐ مقصر ب. ☐ مقبول ت. ☐ متوسط ث. ☐ جيد ج. ☐ جيد جدا ح. ☐ ممتاز

16. قدّر علاماتك في مساق التاريخ في مرحلة الثانوية ( العاشر – الحادي عشر- التوجيهي)

أ. ☐ مقصر ب. ☐ مقبول ت. ☐ متوسط ث. ☐ جيد ج. ☐ جيد جدا ح. ☐ ممتاز

17. قدر علاماتك في مساق القضايا المعاصرة في مرحلة الثانوية ( العاشر – الحادي عشر- التوجيهي)

أ. ☐ مقصر ب. ☐ مقبول ت. ☐ متوسط ث. ☐ جيد ج. ☐ جيد جدا ح. ☐ ممتاز

18. قدّر علاماتك في مساق التربية الإسلامية في مرحلة الثانوية ( العاشر – الحادي عشر- التوجيهي)

أ. ☐ مقصر ب. ☐ مقبول ت. ☐ متوسط ث. ☐ جيد ج. ☐ جيد جدا ح. ☐ ممتاز

3

19. قدّر علاماتك في مساق الصحة والبيئة في مرحلة الثانوية ( الابتدائية- الأساسية-الثانوية)

أ. ☐ مقصر ب. ☐ مقبول ت. ☐ متوسط ث. ☐ جيد ج. ☐ جيد جدا ح. ☐ ممتاز

**الفرع التجاري:**

20. قدّر علاماتك في مساق الإدارة في مرحلة الثانوية ( العاشر – الحادي عشر- التوجيهي)

أ. ☐ مقصر ب. ☐ مقبول ت. ☐ متوسط ث. ☐ جيد ج. ☐ جيد جدا ح. ☐ ممتاز

21. قدّر علاماتك في مساق المحاسبة في مرحلة الثانوية ( العاشر – الحادي عشر- التوجيهي)

أ. ☐ مقصر ب. ☐ مقبول ت. ☐ متوسط ث. ☐ جيد ج. ☐ جيد جدا ح. ☐ ممتاز

22. قدّر علاماتك في مساق القضايا المعاصرة في مرحلة الثانوية ( العاشر – الحادي عشر- التوجيهي)

أ. ☐ مقصر ب☐ مقبول ت. ☐ متوسط ث☐ جيد ج. ☐ جيد جدا ح. ☐ ممتاز

23. قدّر علاماتك في مساق الرياضيات في مرحلة الثانوية ( العاشر – الحادي عشر- التوجيهي)

أ. ☐ مقصر ب. ☐ مقبول ت☐ متوسط ث☐ جيد ج. ☐ جيد جدا ح.☐ ممتاز

**الفرع الصناعي**

24. قدّر علاماتك في مساق الرياضيات في مرحلة الثانوية ( العاشر – الحادي عشر- التوجيهي)

أ. ☐ مقصر ب. ☐ مقبول ت. ☐ متوسط ث. ☐ جيد ج. ☐ جيد جدا ح. ☐ ممتاز

25. قدّر علاماتك في مساق الفيزياء في مرحلة الثانوية ( العاشر – الحادي عشر- التوجيهي)

أ. ☐ مقصر ب. ☐ مقبول ت. ☐ متوسط ث. ☐ جيد ج. ☐ جيد جدا ح. ☐ ممتاز

4

26.قدّر علاماتك في مساق علم الصناعة في مرحلة الثانوية ( العاشر – الحادي عشر- التوجيهي)

أ. ☐ مقصر ب.☐ مقبول ت.☐ متوسط ث.☐ جيد ج. ☐ جيد جدا ح.☐ ممتاز

**ثالثا: قسم الأسئلة العامة (يجيب عليها جميع الطلاب من مختلف التخصصات):**

27.يقدر معدلك التراكمي في الجامعة في الفترة؟

أ.☐ 70-65 ب.☐ 75- 70 ت.☐80-75 ث.☐ 85-80 ج. ☐ 90-85 ح☐ 95-90

28.ما أكثر المواضيع التي تفضلها وتستمتع بها؟

أ.☐ التكنولوجيا والحاسوب ب. ☐ العلوم ( فيزياء-كيمياء –أحياء) ت. ☐ الرياضيات

ث.☐ الصحة والبيئة ج. ☐ اللغة العربية وآدابها ح. ☐ إدارة الأعمال والمحاسبة

خ. ☐ الجغرافيا والتاريخ د. ☐ اللغة الانجليزية وآدابها ذ. ☐ غير ذلك

29.ما أكثر القدرات العقلية والذهنية التي تعتقد انك تتمتع بها؟

أ. ☐ الحفظ ب. ☐ الفهم والاستدلال ج. ☐ الاسترجاع والتذكر د.☐ التحليل المنطقي ه. ☐ الحفظ والفهم

30. كيف تقيم مهارتك باستخدام الحاسوب **قبل التحاقك** بتخصصك الحالي؟

أ. ☐ ممتازة ب. ☐ جيدة ت. ☐ متوسطة ث☐ ضعيفة

31.ما أهم هواياتك؟

أ.☐ القراءة والمطالعة ب. ☐ مشاهدة التلفاز ☐ ت. تصفح الانترنت

ث☐ الرياضة ج.☐ لعب الألعاب الرقمية ☐ ح. لعب الألعاب التقليدية

خ☐ الكتابة د. ☐ برمجة الحاسوب وتصليحه ☐ غير ذلك أذكرها.....

32.يقع معدلك في الثانوية العامة (التوجيهي) في الفترة:

5

**ثالثا: قسم الأسئلة العامة (يجيب عليها جميع الطلاب من مختلف التخصصات):**

1. يقدر معدلك التراكمي في الجامعة في الفترة؟

أ. ☐ 69.9-65 ب. ☐ 74.5- 70 ت. ☐ 79.9-75 ث. ☐ 84.5-80 ج. ☐ 89.9-85

ح. ☐ 94.5-90 خ. ☐ 100-95

2. ما أكثر المواضيع التي تفضلها وتستمتع بها؟

أ. ☐ التكنولوجيا والحاسوب ب. ☐ العلوم ( فيزياء-كيمياء –أحياء) ت. ☐ الرياضيات

ث. ☐ الصحة والبيئة ج. ☐ اللغة العربية وآدابها ح. ☐ إدارة الأعمال والمحاسبة

خ. ☐ الجغرافيا والتاريخ د. ☐ اللغة الانجليزية وآدابها ذ. ☐ غير ذلك

3. ما أكثر القدرات العقلية والذهنية التي تعتقد انك تتمتع بها؟

أ. ☐ الحفظ ب. ☐ الفهم والاستدلال ج. ☐ الاسترجاع والتذكر د. ☐ التحليل المنطقي ه. ☐ الحفظ والفهم

4. كيف تقيم مهارتك باستخدام الحاسوب **قبل التحاقك** بتخصصك الحالي؟

أ. ☐ ممتازة ب. ☐ جيدة ت. ☐ متوسطة ث. ☐ ضعيفة

5. ما أهم هواياتك؟

أ. ☐ القراءة والمطالعة ب. ☐ مشاهدة التلفاز ت. ☐ تصفح الانترنت

ث. ☐ الرياضة ج. ☐ لعب الألعاب الرقمية ح. ☐ لعب الألعاب التقليدية

خ. ☐ الكتابة د. ☐ برمجة الحاسوب وتصليحه ه. ☐ غير ذلك أذكرها.....

6. يقع معدلك في الثانوية العامة (التوجيهي) في الفترة:

أ. ☐ 69.9-65 ب. ☐ 74.5- 70 ت. ☐ 79.9-75 ث. ☐ 84.5-80 ج. ☐ 89.9-85

ح. ☐ 94.5-90 خ. ☐ 100-95

7. ما **السبب الرئيسي** لاختيارك **لجامعتك الحالية؟**

أ. ☐ لقربها من سكني ب. ☐ كفاءتها بالتعليم ت. ☐ سمعتها الجيدة

ج. ☐ تخصصاتها موافقة لمهنتي المستقبلية ح. ☐ إمكانية الحصول على منحة

خ. ☐ لتشجيع أصدقائي وزملائي ث. ☐ إمكانية إكمال الدراسات العليا

ث. ☐ التجارة والفنادق والمطاعم ج. ☐ النقل والتخزين والاتصالات ح. ☐ قطاع التعليم

خ. ☐ قطاع الصحة والخدمات الأخرى د. ☐ لا يعمل ذ. ☐ غير ذلك

40. ما **السبب الرئيسي** لاختيارك **تخصصك الحالي** ؟

أ. ☐ حلمك منذ الطفولة ب. ☐ رغبة الوالدين والأهل ت. ☐ لاعتقادك بأنه مطلوب في سوق العمل

ث. ☐ تشجيع أصدقائي وزملائي ج. ☐ مناسب لميولك واهتماماتك ح. ☐ ملائم لقدراتك العقلية والفكرية

خ. ☐ معدلك التوجيهي ألزمك بهذا التخصص د. ☐ المكانة الاجتماعية التي يعطيك اياها التخصص

41. ما الدخل الشهري للأسرة ( العملة بالشيكل)؟

أ ☐ 2500-1500 ب. ☐ 3500-2500 ت. ☐ 4500-3500

ث. ☐ 5500-4500 ج. ☐ 6500- 5500 ح. ☐ فوق 6500

42. ما أكثر المهارات الفكرية التي تتمتع بها؟

أ. ☐ القدرة على الإدارة والتنظيم ب. ☐ التفكير المنطقي ت. ☐ مهارة حل المشكلات

ث. ☐ القدرة على التواصل مع الآخرين ج. ☐ القدرة على التحليل والاستنتاج

43. ما هو الدور الذي كنت تلعبه وتتقمصه في طفولتك (الأقرب إلى شخصيتك)؟

أ. ☐ الطبيب ب. ☐ المعلم ت. ☐ المتعلم ث. ☐ المدير ج ☐ الباحث

ح. ☐ صحفي خ. ☐ لاعب كرة قدم د. ☐ مهندس ذ. ☐ محامي ر. ☐ الشرطي

ع. ☐ مزارع غ. ☐ طيار ق. ☐ تاجر ك. ☐ فنان ل. ☐ ممثل/ مغني

44. ما مدى رضاك عما حققته إلى الآن؟

أ. ☐ بشكل كبير جدا ب. ☐ بشكل كبير ج. ☐ بشكل متوسط ث. ☐ بشكل قليل ج. ☐ غير راض

7

45. ما الألعاب المفضلة لديك في طفولتك؟

أ. ☐ الألعاب الاستكشافية والاستقصائية ب. ☐ العاب تمثيل الأدوار ت. ☐ الألعاب الرياضية الترويحية ( الألعاب الشع

ث. ☐ الألعاب الفنية(الرسم..الخ) ج. ☐ الألعاب الثقافية(مثل الكتابة، المطالعة،المسرح الخ) ح. ☐ الألعاب التركيبية

46. ما هو تخصصك الحالي؟**(الانتقال الى تخصصك الحالي من القسم الرابع)**

أ. ☐ تكنولوجيا معلومات وأنظمة الحاسوب ب. ☐ علوم إدارية ونظم المعلومات( إدارة أعمال ،علوم مالية و محاسبة)
ت. ☐ العلوم(رياضيات –فيزياء-كيمياء-أحياء) ث. ☐ التربية (علم نفس-أصول تربية) ج. ☐ علوم صيدلية (صيدلة ، علوم
ح. ☐ الحقوق والعلوم السياسية خ. ☐ التمريض د. ☐ الشريعة
ذ. ☐ هندسة حاسوب ر. ☐ هندسة ميكانيكا ع. ☐ هندسة كهربائية
غ. ☐ هندسة معمارية ومدنية ق. ☐ الآداب (لغة عربية ، لغة انجليزية، سياحة وآثار ،جغرافيا ،تاريخ ،إعلام)
ف. ☐ الزراعة

47. ما التخصص الذي كنت ترغب به **قبل تخصصك الحالي** ؟

أ. ☐ تخصصي الحالي ك. ☐ غير ذلك

ب. ☐ تكنولوجيا معلومات وأنظمة الحاسوب ل. ☐ علوم إدارية ونظم المعلومات( إدارة أعمال ،علوم مالية و محاسبة)
ت. ☐ العلوم(رياضيات –فيزياء-كيمياء-أحياء) ث. ☐ التربية (علم نفس-أصول تربية) ج. ☐ علوم صيدلية (صيدلة ، علوم
ح. ☐ الحقوق والعلوم السياسية خ. ☐ التمريض د. ☐ الشريعة
ذ. ☐ هندسة حاسوب ر. ☐ هندسة ميكانيكا ع. ☐ هندسة كهربائية
غ. ☐ هندسة معمارية ومدنية ق. ☐ الآداب (لغة عربية ، لغة انجليزية، سياحة وآثار ،جغرافيا ،تاريخ ،إعلام)
ف. ☐ الزراعة

**في حال كان هناك تخصص كنت ترغب به سابقا الانتقال إلى أسئلة ذلك التخصص مما يلي** :

**رابعا: قسم الأسئلة الخاصة بكل تخصص:**

**1-هندسة الحاسوب**

48. هل لديك مهارة في نمذجة المشكلات باستخدام الحاسوب (Problem Solving)؟

أ. ☐ بشكل كبير جدا ب. ☐ بشكل كبير ت. ☐ بشكل متوسط ث. ☐ بشكل قليل ج. [

8

49. هل **<u>السبب الرئيسي</u> <u>لالتحاقك بهذا التخصص:</u>** هو يقينك بأن هذا عصر العولمة والثورة التكنولوجية وضرورة مجاراتها ؟

أ. ☐ نعم ب. ☐ لا

**2-الهندسة المعمارية والمدنية**

50. كيف تقيم مهارتك في التواصل مع الفريق والعمل الجماعي؟

أ. ☐ غير مقبولة ب. ☐ مقبولة ج. ☐ جيدة د. ☐ ممتازة

51. ما أكثر المهارات التي تتميز بها؟ (اختر اكثر من واحدة حسب ما ينطبق)

أ. ☐ مهارات العرض والتقديم ب. ☐ التواصل مع الآخرين ج. ☐ مهارات العمل مع الفريق

د. ☐ مهارة التصميم ه. ☐ مهارة التصميم و. ☐ مهارات حل المشكلات

ي. ☐ مهارات الإدارة والقيادة

3- **الهندسة الميكانيكية**

52. هل كنت ترغب بتصليح الأجهزة والمعدات قبل التحاقك بهذا التخصص؟

أ. ☐ نعم ب. ☐ لا

53. هل سبق لك وان قدت سيارة في المرحلة الثانوية وما قبلها ؟

أ. ☐ نعم ب. ☐ لا

**4-الهندسة الكهربائية**

54. هل سبق وان أصلحت جهاز الكتروني أو ميكانيكي في طفولتك؟

أ. ☐ نعم ب. ☐ لا

55. هل كنت ترغب بتصليح الأجهزة والمعدات قبل التحاقك بهذا التخصص؟

أ. ☐ نعم ب. ☐ لا

9

**5-الآداب**

56. ما أكثر المهارات التي تتمتع بها هي:

أ. ☐ مهارة الإلقاء والخطابة ب. ☐ الإقناع ج. ☐ مهارة التحدث بلغات مختلفة د. ☐ الإلمام بـ

57. هل سبق وان زرت معلما سياحيا وأثريا معروفا؟

أ. ☐ نعم ب. ☐ لا

58. تفضل مشاهدة الأفلام والبرامج الوثائقية ب :

أ. ☐ باللغة العربية ب. ☐ اللغة الانجليزية ج. ☐ لا أشاهدها

59. ما أكثر المواضيع التي تستمتع بها؟

أ. ☐ المواضيع التاريخية ب. ☐ المواضيع الدينية ت. ☐ المواضيع العامة

ث. ☐ المواضيع السياسية ج. ☐ القضايا الإنسانية ح. ☐ القضايا الأخلاقية

**6-العلوم الإدارية ونظم المعلومات**

60. بماذا تتميز شخصيتك ؟

أ. ☐ القدرة على التخطيط والتنفيذ ب. ☐ القدرة على إدارة المشاريع

ج. ☐ الاطلاع على الثقافات المختلفة د. ☐ تمتلك رغبة في التغيير والإبداع

61. أي من المهارات التالية تتمتع بها؟

أ. ☐ القدرة على الإقناع ب. ☐ مهارات البيع والشراء ج. ☐ تمتلك مهارات التسويق

62. أي من القضايا التالية تهتم بها ؟

أ. ☐ القضايا الأخلاقية ب. ☐ القضايا البيئية

ج. ☐ القضايا السياسية د. ☐ القضايا الإنسانية

10

**7-العلوم ( رياضيات –فيزياء-كيمياء-أحياء)**

63. ما هي التجربة العلمية التي قمت بها في طفولتك بهدف الاستكشاف؟

أ. ☐ خلط مواد كيميائية معا ب. ☐ عمل تجربة فيزيائية
ج. ☐ قمت بتشريح كائن حي د. ☐ غير ذلك

64. هل سبق وأن قمت بزراعة نبته في طفولتك؟

أ. ☐ نعم ب. ☐ لا

**65.** هل سبق وأن قمت بتربية حيوان في طفولتك؟

أ. ☐ نعم ب. ☐ لا

**8-الصيدلة**

66. ما التجربة العلمية التي قمت بها في طفولتك بهدف الاستكشاف؟

أ. ☐ خلط مواد كيميائية معا ب. ☐ عمل تجربة فيزيائية
ج. ☐ قمت بتشريح كائن حي د. ☐ غير ذلك

**9-الزراعة**

67. هل سبق وأن قمت بزراعة نبته في طفولتك؟

أ. ☐ نعم ب. ☐ لا

68. هل سبق وأن قمت بتربية حيوان في طفولتك؟

أ. ☐ نعم ب. ☐ لا

69. ما التجربة العلمية التي قمت بها في طفولتك بهدف الاستكشاف؟

أ. ☐ خلط مواد كيميائية معا ب. ☐ عمل تجربة فيزيائية

ت. ☐ قمت بتشريح كائن حي ث. ☐ غير ذلك

11

**10-تكنولوجيا المعلومات وأنظمة الاتصالات**

70. هل لديك مهارة في نمذجة المشكلات باستخدام الحاسوب (Problem Solving)؟

أ. ☐ بشكل كبير جدا ب. ☐ بشكل كبير ج. ☐ بشكل متوسط د. ☐ بشكل قليل ه. ☐ لا

71. هل **<u>السبب الرئيسي لالتحاقك بهذا التخصص:</u>** هو يقينك بأن هذا عصر العولمة والثورة التكنولوجية وضرورة مجاراتها ؟

أ. ☐ نعم ب. ☐ لا

**11-الشريعة**

72. ما أكثر المهارات التي تتميز بها؟

أ. ☐ العرض والتقديم ب. ☐ القدرة على الإقناع ج. ☐ مهارة الإلقاء والخطابة

د. ☐ القدرة على حل المشكلات ه. ☐ القدرة على التحليل والاستنتاج

73. ما أكثر القضايا التي تهتم بها؟

أ. ☐ القضايا الدينية ب. ☐ القضايا الإنسانية وحقوق الإنسان ج. ☐ القضايا السياسية د☐ القضايا البيئية ه. ☐

**ب. 12-كلية الحقوق والعلوم السياسية:**

**74.** ما أكثر القضايا التي تهتم بها؟

أ. ☐ القضايا الدينية ب. ☐ القضايا الإنسانية وحقوق الإنسان

ج. ☐ القضايا السياسية د. ☐ القضايا البيئية

75. ما أكثر المهارات التي تتميز بها؟

أ. ☐ العرض والتقديم ب. ☐ القدرة على الإقناع

ج. ☐ القدرة على التحليل والاستنتاج د. ☐ القدرة على حل المشكلات

12

**13-كلية التربية**

76. ما أكثر المهارات التي تتميز بها؟

ب. ☐ العرض والتقديم ب. ☐ القدرة على الإقناع

ج. ☐ القدرة على التحليل والاستنتاج د. ☐ القدرة على حل المشكلات

77. هل تمتلك القدرة على ملاحظة سلوك الأشخاص وتحديد شخصياتهم قبل التحاقك بهذا التخصص؟

أ. ☐ نعم ب. ☐ لا

78. هل كنت تمتلك القدرة على تحليل شخصية من أمامك قبل التحاقك بهذا التخصص؟

أ. ☐ بشكل كبير جدا ب. ☐ بشكل كبير ت. ☐ بشكل متوسط ث. ☐ بشكل قليل ج. ☐ لا أمتلك ا

**14-كلية التمريض**

79. ما أكثر المهارات التي تتميز بها؟

أ. ☐ العرض والتقديم ب. ☐ القدرة على الإقناع

ج. ☐ القدرة على التحليل والاستنتاج د. ☐ القدرة على حل المشكلات

80. ما التجربة العلمية التي قمت بها في طفولتك بهدف الاستكشاف؟

ب☐ خلط مواد كيميائية معا ب. ☐ عمل تجربة فيزيائية

ت. ☐ قمت بتشريح كائن حي ث. ☐ غير ذلك

**شكرا لحسن تعاونكم**

13