State of Palestine An-Najah National University Faculty of Information Technology Computerized Information Systems دول\_\_\_\_\_ ة فلس\_\_\_\_ طين جامعة النجاح الوطنية كلية تكنولوجيا المعلومات قسم نظم المعلومات المحوسبة



# THE CORRELATION BETWEEN THE ACADEMIC ACHIEVEMENTS OF STUDENTS IN BOTH SCHOOL AND UNIVERSITY STAGES

Students:

Ahd Hinnawi Shatha Abulhawa 11420877 11421356

*Supervisor:* Dr. Hamed Abdelhaq A project report submitted in partial fulfillment of the requirements for Graduation Project course in Computer Information Systems department.

> An-Najah National University 2018

# Acknowledgement

We would like to show our gratitude to **Dr. Hamed Abadelhaq** who provided insight and expertise that greatly assisted our research by sharing his pearls of wisdom with us during the course of this research.

Ahd Hinnawi Shatha Abu Alhawa

# Table of Contents

List of Figures	5
List of Tables	6
Abstract	7
Chapter 1	8
Introduction	8
1.1 Problem Specification	8
1.2 Project Scope	8
1.3 Project Requirements	8
1.4 Goals and Objectives	9
Chapter Two	9
Methodology of work	10
2.1 Introduction	10
2.2 Current System	10
2.3 Proposed System and Analysis	11
2.4 Methodology	12
Data History "Grades gathered from schools": -	12

# List of Figures

Figure 1: Paper copy of students' school grades	13
Figure 2: Majors with courses IDs and names	14
Figure 3: University Courses and grades as sent by the computer center	15
Figure 4: School Grades with random numbers linked to university grades as se	nt by
the computer center	15

# Abstract

A critical issue facing universities in Palestine is allocating students' majors by the factor of the final high school grade (Tawjihi), therefore many students choose their majors according to their final grade rather than what they are actually interested in.

The project is based on a case study of the subsequent performance at An-Najah National University of students of different majors.

By tracking their performance for the years 2010-2014, using data mining methods, we were able to estimate the likelihood of success of subsequent students based on their high school grades rather than the final grade.

# Chapter 1

# INTRODUCTION

# 1.1 Problem Specification

As the number of students graduated from secondary education has increased, the number of applicants at Universities increased. Therefore, universities aimed to make acceptance fair, legitimate and stable, depending on the final high school grade.

Nevertheless, this method was not convenient for many students because they were constrained based on their final grade not on their abilities or interests, the thing that led to a serious issue of academic failure.

# 1.2 Project Scope

The project will represent a new means of identifying students' acceptance process. Using data mining methods, we aim to find the relevance between a sample of 2010-2014 students' high school grades and their current university academic achievement.

Based on these results, the acceptance process will change as follows:

Universities will require each grade in every subject in the latest 3 years of high school and sincerely care about their tendencies throughout an ability test that they have to go through.

#### Time and Budget:

The project shall be accomplished within 4 months, and restricted to a budget of 80,000\$.

# 1.3 Project Requirements

- 1. R GUI and RStudio
- 2. Excel Sheets in CSV format.

# 1.4 Goals and Objectives

- Changing the acceptance process.
- Provide an opportunity to the students to choose their majors based on their abilities and interests.
- Increase the academic achievement rate.

# CHAPTER TWO

# Methodology of work

This chapter will introduce a background for the current system and also discuss the proposed new system, the advantages and disadvantages of the proposed system will be identified. Finally, the development methodology for the proposed system will be shown.

## 2.1 Introduction

Our system manages to include what the applicants will need in a way that corresponds their interests, it also provides good and feasible solution for existing problems in the old system, in this chapter we are going to compare between old and new system, show the characteristics of both, and give reasonable views on why would our system be more efficient than the old one.

#### 2.2 Current System

The current way is allocating students' majors by the factor of the final high school grade (Tawjihi).

## Disadvantages of the currents system:

#### 1. Academic Failure

Because the current system depends on the final grade, many students rely on this year only to achieve good marks, therefore in some cases they achieve higher marks than their actual ability, the thing that guides them to choose a major that does not suit their capability, and that was very clear in their university marks, the thing that lead to a serious academic failure.

## 2. Inadequate

The current system depends on the demand and acceptance, so there are many students who achieve high marks and had a desire to enter a particular major, but because of this process they will not be able to join it, even though they may be qualified for that major.

# 2.3 Proposed System and Analysis

## Advantages of the Proposed system:

#### 1. Adequate

Unlike the old system, this system allows the participants to choose their university specialization based on the marks of 3 consecutive years, allowing them to prove their worth over these years and not one year to determine their fate.

#### 2. Choosing Depending on Students' Desires:

In addition to students' marks, the system will allow them to test their abilities to discover their skills and choose what they like to specialize in.

# 3. Resulting in good academic achievement and reducing unemployment rate

As students choose what they'd like to major in, their academic score will increase by means of their own decisions, resulting in everyone working in what they love rather than what they have to.

## Disadvantages of the Proposed system:

## 1. May not be that much accurate

Because of the lack of truthfulness we may face in gathering student's scores in some schools, and that will return to the  $2^{nd}$  point of the disadvantages of the currents system.

# 2.4 Methodology

Data mining, RStudio and the working mechanisms:

#### Data Set"Grades gathered from schools":

#### The data history for all listed grades in the consists of main sections which are:

- 1. 10<sup>th</sup> grade marks:
- 2. 11<sup>th</sup> grade marks:
- 3. 12<sup>th</sup> grade marks:
- 4. University marks:
- 5. Other Categories:

# CHAPTER THREE

# Working Mechanism

Through research and exploration, the data were arranged to be suitable to be used for RStudio.

## 3.1 Data Collection:

Here is the process where we have gathered and collected targeted variables that we would use for our study, which took place in two stages:

## 3.1.1 Collecting school grades

Obtaining high school grades was a difficult and long stage, which led to our delay in moving to the second stage of data collection.

#### Difficulties encountered during this process

- 1. Intensive privacy measures by schools and the Ministry of Education on students' grades.
- 2. The lack of computerized copies in schools, which forced us to transfer all these Data manually as shown in the picture below

ىم مر	-1								1	:	عبة	الش								وية	الثاذ	سائح	د الس	حميد	بد ال	-	سة	مدر،	IL				ں س	نابل	: 4	قريا	] /ā.:	لمدي	J		J	نابله	ليم ا	: والتع
		نيية (	ة الاجا تقية(	113 A	ù	القنو		: پة	التربية رياض	a	يا رة	القضار لمعاص	1	وبا نات	تكلولو. المطوه		لعلميه	تلغة ا	<u>نٹ</u>	يات	ارياض	JI	ية سلا	الادار الاقتم	,	الغيا	الجغر		ы.	التار	a,	الجليز	للغة اا	1 2,	العري	اللغة	رة	ة الديد	تريپا	n 3	الولادة	اريخ	5	Y
احترام ا	1		الفصل الثقي	الفصل الأول	المعادل	الفصل الثقي	الفصل الأول	المعادل	الفصل الثاني	الفصل الأول	المعد ل	الغصل الثاثي	الفصل الأول	المعادر	القصل الثاتي	الفصل الأول	1140 - E [	الغصل الثاثي	الفصل الأول	المعادل	الفصل الثاتي	الفصل الأول	المع د ل	الفصل الثقى	الفصل الأول	1 D	11000 ILINO	111000 180D	H2 1 H4C	1100 (1110)	11	111-1112	اللعمل الأدل	1111	الفصل الثلار	الفصل الأول	llan - [	الفصل الثاتر.	111-11-11-1	1	المدز	11	7	
لنظام	1 1 1 1	00 50	100 50	100 50	100 50	100 50	100 50	100	100 50	100	100	100	100	100	100	100	100 1	100	100 50	100	100	100	100	100 50	100 50	100 1 50 (	00 1	00 1 50 5	00 1	00 1	00 1	50 1	50 15	0 15 5 7	0 15	0 15	0 10	0 10	0 10	00		Y	62	ولادة
- 44	*		-		7B	76	80	82	84	80	50	50	50	53	58	48	52	50	54	48)	53	43	65	65	65	50	50 4	40	50 6	58	32	8)5	2 6	4 7	6 9	9 5:	3 6	5 6	9 6	1	94	12	17	ئايلس
د جد	#	1			93	93	93	83	86	80	98	98	98	76	84	68	95	94	96	82	91	73	87	82	92	98	00 5	96	37 5	97 9	97 1	32 1	36 13	28 13	3 14	13 12	3 9	2 96	6 8	8	94	9	27	تابلس
ىتتر					84	87	81	84	85	83	92	92	92	68	78	58	87	85	89	69	74	64	76	81	71	84	94 7	74	86 8	37 1	85	11 1	11 1	11 13	20 1	15 12	5 9	7 99	9 9	5	95	1	27	ئاپلس
-					80	75	85	73	76	70	57	56	58	50	64	36	45)	50	40	33	45	21	53	50	56	52	57 4	47	50 5	50	50	32)	30 3	4 7	5 9	9 5	1 5	6 5	9 5	3	93	10	30	ئاپلس
يد					79	75	83	73	76	70	53	51	55	50	61	39	50	68	32	40	47	33	57	51	63	51	52	50	52 3	51	53	36)	30 4	2 6	1):	5 4	7 5	3 5	1 5	15	92	3	15	ئايلس
جد	<del>44</del>				88	92	84	86	87	85	96	95	97	82	94	70	92	89	95	77	80	74	86	81	91	92	100	84	92 9	94	90	23 1	28 1	18 1	29 1	34 1:	24 8	7 8	9 8	\$5	94	3	26	لسعوديا
4	44				85	85	85	84	88	80	62	59	65	57	63	51	60	55	65	60	72	48	60	62	58	57	52	62	55	54	56	36 8	31 9	1 8	3 1	01 6	5 6	7 6	6 6	8	94	5	17	نىس
يد	÷				77	75	79	68	70	65	57	63	51	50	73	27	55	38	72	43)	54	32	50	50	50	52	63	41	50	50	50	0	51 4	19		59 3	17 5	0 5	5 4	15	95	2	22	تايلس
4	÷				72	71	73	81	82	80	59	58	60	63	55	71	66	69	63	66	72	60	64	62	66	75	79	71	55	56	54		43 :	15		03 8	19 10	2 1	4	46	94	9	4	نابلين
4	4			_	70	71	69	76	74	78	50	68	32	54	73	35	40	44	36	38	52	24	50	57	43	50	5/	43	50	62	29	3	29 4	11	3	85	21	50 4	50	50	94	11	25	تابلس
5 4	*			-	68	64	72	77	86	68	50	43	57	50	50	50	50	68	32	20	74	40	50	55	45	50	58	45	56	57	55	RA	82 1	86	30	84	75	56 5	57	55	94	9	27	نابلس
<del>بد</del> ز	**				86	93	79	83	86	80	63	68	58	50	52	48	50	50	44	20	54	30	68	66	70	59	60	58	77	74	80	55	52	58	38	102	94	74	78	70	94	3	18	نابلس
	**	100		-	78	80	76	82	81	83	12	04	80	24	57	53	70	69	71	43	51	35	61	57	65	69	80	58	82	81	83	55	53	57	78	95	61	72	74	70	94	6	28	نابلس
	**				16	81	81	89	90	02	60	18	52	52	62	42	45	44	45	41	48	34	50	47	53	51	57	45	50	64	36	37)	27	47	75	108	42	50	55	45	94	5	8	الملس ا
	**	- 22			11	12	10	80	33	80	79	71	87	62	67	57	76	84	68	62	67	57	75	80	70	84	86	82	66	52	80	88	98	78	89	99	79	68	69	67	94	9	27	ابلس
	***			-	83	83	83	87	82	82	90	88	92	84	83	85	79	75	83	68	73	63	81	82	80	88	95	81	90	89	91	124	119	129	112	120	104	83	80	86	94	9	10	ولكرم
R	1			-	84	88	80	79	87	71	98	96	100	74	83	65	89	89	89	80	78	82	77	84	70	96	98	94	97	98	96	121	118	124	116	131	101	J	v	5	94	11	26	بلس
	-			-	81	81	81	82	2 82	82	72	64	80	50	54	46	54	62	46	47	56	38	55	56	54	51	50	52	71	71	71	75	69	81	81	93	69	58	63	53	94	10	6	بلس
				+	69	74	64	67	7 62	72	53	55	51	50	57	43	64	69	59	54	62	46	62	57	67	58	59	57	57	62	52	E	41	45	82	96	68	62	69	55	94	5	28	بلس
1	us.			-	76	71	81	94	1 97	91	64	64	64	61	64	58	57	49	65	54	64	44	77	71	83	55	57	53	59	44	74	85	69	101	91	105	77	64	60	68	95	2	7	لكرم
6	uş.			-	76	73	79	84	4 90	78	50	45	55	50	61	39	50	61	39	4	51	37	50	49	51	50	50	50	37	45	29	33	37	49		98	52	50	62	38	94	12	2 28	- und
8 .	44			-	76	74	78	8	5 86	84	80	74	86	66	66	66	79	77	81	57	66	48	79	71	87	84	87	81	73	69	77	85	74	96	99	102	96	82	86	78	94	4	30	1 cm
6 4	- 49			-	84	89	79	8	1 90	72	92	94	90	65	77	53	81	83	79	73	81	65	73	70	76	81	86	76	89	90	88	134	136	132		124	112	0	0	0	95	2	12	
8	47			-	79	76	82	81	1 82	80	67	69	65	51	56	46	51	56	46	46	53	39	53	53	53	58	61	55	60	70	50	75	74	76	11	90	64	158	65	41	94	_ P	1 2	10
6-A	1	T			At		14	A	et c	TO	i	9 1	1	Sal	King	X	1000	1	1	A	4	Ag	T	f F		- 1	1	61	4	4		1	A	1		A B		A	11	SA.	S			

Figure 1: Paper copy of students' school grades

3. The problem of linking students' marks in the three grades, because some schools end at grade 10 and then their students move to other schools.

## **3.1.2** Collecting university grades

As the first stage, we have encountered some problems and difficulties which led to a long waiting period of time.

## Difficulties encountered during this process

- 1. Intensive privacy measures by the University on students' grades.
- 2. The significant decline in the number of students who were admitted for the obtained.

For security and privacy reasons, the computer center in the university did not agree to give us all student marks in all courses, so we had to choose some of the courses to be sent to the computer center,

10 subjects were selected for each major as follows:

- Some majors, courses were chosen by Assistant professors of each department
- Some other majors have been chosen from the junior and senior years.

لوم الانسانية	كلية الع
خ 10321	التاري
تاريخ الجزيرة العربية قبل الإسلام	10321122
تاريخ الدولة الأموية	10321124
تاريخ المغرب في العصىر الإسلامي	10321125
تاريخ الأندلس	10321126
تاريخ الدولة العباسية	10321227
تاريخ الأبوبيين والمماليك	10321230
فكر سياسى إسلامى	10321231
تاريخ الدولة العتمانية	10321335
تاريخ فلسطين الحديث	10321440
تاريخ القدس الحديث والمعاصير	10321441
والاثار 10316	السياحة
تاريخ فلسطين والأردن في العصور القديمة	10316114
التخطيط السياحي	10316220
إدارة المواقع السياحية والأثرية	10316221
السياحة في فلسطين	10316222
أثار اليونان والرومان*	10316225
آتار بيزنطيّة	10316226
لغة قديمة	10316337
تاريخ وأتار القدس	10316440
الإرشاد السياحي	10316442
السياحة باللغة الانجليزية *	10316336

Figure 2: Majors with courses IDs and names

Majors with chosen courses were sent to the computer center in addition to the names and school grades. And the returned result was by given random numbers as shown in the picture bellow:



Figure 3: University Courses and grades as sent by the computer center

And their linked student's school marks with same linked random numbers:

4	A	В	C	D	E	F	G	н	1	J	ĸ	L	M	N	0	Р	Q	R	
1	رقم الطالب	التربية الدينية	رية اللغة العربية	اللغة الانجليز	التاريخ	ء ا <b>ل</b> جغر افيا	الإدارة والاقتص	الرياضيات	س تکنولو جيا	والقضباوا المعاص	التربية الرياضا							1	
2	60	87	88	118	93	88	78	85	ناجح	89	85								
3	232	80	79	93	84	89	80	88	ناجح	83	83								
1	338	92	107	99	98	95	87	82	ناجح	97	80								
	340	91	120	110	99	96	84	86	ناجح	95	82								
	857	95	108	105	98	93	89	71	ناجح	98	82								
	946	88	109	102	93	91	83	85	ناجح	82	85								
	1124	73	97	97	96	85	85	75	ناجح	96	85								
	1175	89	115	107	95	84	95	67	ناجح	99	87								
)	1191	99	128	103	99	99	96	91	ناجح	99	85								
	1272	90	113	84	91	93	84	91	ناجح	95	84								
2	2063	90	105	90	97	92	79	59	ناجح	91	86								
5	2217	74	91	85	91	91	68	69	ناجح	70	85								
£.	2424	77	103	78	96	80	71	77	ناجح	80	87								
5	2534	91	120	85	100	91	91	71	ناجح	96	84								
5	3202	90	114	117	99	93	92	87	ناجح	97	82								
7	3428	82	118	122	97	95	83	87	ناجح	71	88								
3	3533	66	81	134	73	69	57	41	ناجح	47	86								
9	4059	91	106	124	96	80	84	73	ناجح	72	83								
0	4235	94	125	130	100	92	88	93	ناجح	92	85								
A	E.	جیهی ادبی	ہی علمی توج	) توجيو	+)						1.4	T							

Figure 4: School Grades with random numbers linked to university grades as sent by the computer center

## 3.2 Data Preparation

The second stage: creation of the training, for our analysis task, preprocessing data into a suitable format was an important consideration. This is a little time-consuming because it involved manual classification of vignettes from a series of samples into several classes. The number of classes to include in the training set depends on the number of school classes plus other variables as would be explained below.

Data preprocessing included:

- Data editing:
- Data cleansing
- Data reduction
- Data wrangling

Which went as follows:

	mark f	or each sub	oject.					
STUDENT ID	Матн	Тесн	English	PHYSICS	CHEMISTRY	BIOLOGY	ARABIC	RELIGION
STUDENT ID	Math	Tech	English	Physics	Chemistry	Biology	Arabic	Religion
1	44	58	50	82	70	83	65	78
1	59	60	52	83	81	89	66	70
1	63	70	54	89	88	85	80	84
AVG:	55	63	52	84	79	85	70	77

1. We have calculated the average between the 3 school grades to obtain one mark for each subject.

- 2. We merged students marks in both, Scientific and Literary stream
- 3. That way we end up with empty values for columns of physics, chemistry and biology for the Literary stream, and empty values for columns of history, geography and science for the Scientific stream, which we all gave values of 50s.

Stud	М	Теснио	Engl	Рнү	Снемі	BIOL	Ara	Relig	SCIE	GEOGRA	Ніят	Administr
ENT	ATH	LOGY	ISH	SICS	STRY	OGY	BIC	ION	NCE	РНІС	ORY	ATION AND
lo												ECONOMY
1	69	65	55	60	70	72	80	83	50	50	50	89
2	72	92	95	50	50	50	88	86	65	66	78	83

- 4. We made a match between the random numbers given for school and university grades.
- 5. As a result, we end up with 10 target variables (university marks) for every raw,

and that was solved by duplicating each raw by the number of available university marks to have one target variable for each raw

STUDEN	Мат	TECHNOLOG	ENGLIS	DATA	DESIGN AND	COMPUTER	LINEAR	CALCULU	CALCULU
t Id	н	Y	н	STRUCTUR	ANALYSIS OF	Organizatio	ALGEBRA	s I	s II
				Е	Algorithm	N AND	FOR		
					s	ASSEMBLY	COMPUTE		
	_					LANGUAGE	R SCIENCE		
1	67	65	55	50	53	51	55	58	59

STUDENT	Матн	TECHNOLOGY	ENGLISH	UNI	COURSE	Course Name
lo					NUMBER	
1	67	65	55		1067121	Data Structure
				50	0	
1	67	65	55		1067121	Design and Analysis of
				53	2	Algorithms
1	67	65	55		1067124	Assembly Language
				51	3	
1	67	65	55		1067124	Linear Algebra
				55	4	
1	67	65	55		1021110	Calculus I
				58	1	
1	67	65	55		1021110	Calculus II
				59	2	
1	67	65	55		1067132	<b>Computer Architecture</b>
				62	1	
1	67	65	55		1067135	Database Systems Design
				60	3	
1	67	65	55		1067110	Principles of Programming II
				61	2	
1	67	65	55		1067120	Web Programming
				62	4	

6. That led us to another problem which is having the same independent variables leading to different dependent variables, which is not reasonable for the regression model, and that was solved by adding 4 more variables:

		1	
MEMORIZATION	Perception	Language	ANALYSIS

These variables were manually entered and each variable is fixed for each course.

# CHAPTER FOUR

# Model Building

## 4.1 Datasets

To build up the real picture of what we want, we needed to check different datasets, and different collections of information and choose the best So, we went through 2 datasets:

Δ1	• : ×	V 6	Stu	lent Id																
A	B	c	D	E	F	G	Н	E. F	l lj	к	L	M	N	0	Р	0	R	S	Т	U
Student I	Math Tech		gllish	Physics	Chemist		Arabic	Religion	Science	Geograph	History	Adminstra	Memoriza	Perceptio	Language	Analysis	Uni	Major	Course N	Cours
Arabic 1	89	88	80	(	)	0	0	5 70	77	88	8	9 90	0.5	0.7	0.9	0.7	5	Arabic	10301112	2 Draina
Arabic 1	89	88	80	(	)	0	0	55 70	77	88	8	9 90	0.6	0.8	0.9	0.7	6	Arabic	10301113	Arabi
Arabic 1	89	88	80	(	)	0	0	55 70	77	88	8	9 90	0.7	0.8	0.9	0.7	55	Arabic	10301114	Gram
Arabic 1	89	88	80	(	)	0	0	55 70	77	88	8	9 90	0.8	0.1	0.9	0.1	54	Arabic	10301115	Litera
Arabic 1	89	88	80	(	)	0	0	55 70	77	88	8	9 90	0.8	0.6	0.9	0.6	6	Arabic	10301219	Prese
Arabic 1	89	88	80	(	)	0	0	5 70	77	88	8	9 90	0.8	0.4	0.9	0.6	5	Arabic	10301326	Andal
Arabic 1	89	88	80	(	)	0	0	55 70	77	88	8	9 90	0.8	0.3	0.9	0.5	5	Arabic	10301435	Mode
Arabic 1	89	88	80	(	)	0	0	55 70	77	88	8	9 90	0.8	0.1	0.9	0.1	5	Arabic	10301438	Media
0 Arabic_1	89	88	80	(	)	0	0	55 70	77	88	8	9 90	0.8	0.4	0.9	0.3	5	Arabic	10301328	Ancie
1 Arabic_1	89	88	80	(	)	0	0	55 70	77	88	8	9 90	0.9	0.1	0.9	0.3	5	Arabic	10301117	Litera
2 Arabic_2	89	92	90	(	)	0	0 !	95 97	7 88	92	9	3 94	0.5	0.7	0.9	0.7	9	Arabic	10301112	2 Drain
3 Arabic_2	89	92	90	(	)	0	0 !	95 97	7 88	92	9	3 94	0.6	0.8	0.9	0.7	9	Arabic	10301113	Arabi
4 Arabic_2	89	92	90	(	)	0	0 !	95 97	7 88	92	9	3 94	0.7	0.8	0.9	0.7	9	Arabic	10301114	Gram
5 Arabic_2	89	92	90	(	)	0	0 !	95 97	7 88	92	9	3 94	0.8	0.1	0.9	0.1	9	Arabic	10301115	Litera
6 Arabic_2	89	92	90	(	)	0	0 !	95 97	7 88	92	9	3 94	0.8	0.6	0.9	0.6	9	Arabic	10301219	Prese
7 Arabic_2	89	92	90	(	)	0	0	95 97	7 88	92	9	3 94	0.8	0.4	0.9	0.6	9	Arabic	10301326	i Anda
8 Arabic_2	89	92	90	(	)	0	0 !	95 97	7 88	92	9	3 94	0.8	0.3	0.9	0.5	9	Arabic	10301435	Mode
9 Arabic_2	89	92	90	(	)	0	0 !	95 97	7 88	92	9	3 94	0.8	0.1	0.9	0.1	9	Arabic	10301438	Medie
0 Arabic_2	89	92	90	(	)	0	0 !	95 97	7 88	92	9	3 94	0.8	0.4	0.9	0.3	9	Arabic	10301328	Ancie
1 Arabic_2	89	92	90	(	)	0	0 1	95 97	7 88	92	9	3 94	0.9	0.1	0.9	0.3	9	Arabic	10301117	Litera
2 Arabic_3	89	88	80	(	)	0	0	55 70	77	88	8	9 90	0.5	0.7	0.9	0.7	5	Arabic	10301112	2 Draina
3 Arabic_3	89	88	80	(	)	0	0	55 70	77	88	8	9 90	0.6	0.8	0.9	0.7	6	Arabic	10301113	Arabi
4 Arabic_3	89	88	80	(	)	0	0	55 70	77	88	8	9 90	0.7	0.8	0.9	0.7	5	Arabic	10301114	Gram
5 Arabic_3	89	88	80	(	)	0	0	55 70	77	88	8	9 90	0.8	0.1	0.9	0.1	5	Arabic	10301115	Litera
6 Arabic_3	89	88	80	(		0	0	55 70	77	88	8	9 90	0.8	0.6	0.9	0.6	6-	Arabic	10301219	Prese
7 Arabic_3	89	88	80	(	)	0	0	55 70	77	88	8	9 90	0.8	0.4	0.9	0.6	5	Arabic	10301326	Anda
8 Arabic_3	89	88	80	(	)	0	0 1	55 70	77	88	8	9 90	0.8	0.3	0.9	0.5	5	Arabic	10301435	Mode
9 Arabic_3	89	88	80	(	)	0	0	55 70	77	88	8	9 90	0.8	0.1	0.9	0.1	5	Arabic	10301438	Medi
0 Arabic_3	89	88	80	(	)	0	0	55 70	77	88	8	9 90	0.8	0.4	0.9	0.3	5	Arabic	10301328	Ancie
1 Arahic 3	PR all data cm	88	80	(	1	0	0	55 70	77	88	8	90 90	0.9	0.1	0.9	0.3	5	Arahir	10301117	Litera

1. A complete dataset for one model that includes all majors

# 2. A dataset that for 24 different model- each major was built a different regression model

																		*** 图
J82		: ×	V .	fx														
		6	l n	F	L F	6	н Г. Г	1	×	1	м	N		P	0	в	\$	т
1 Math	Tachaolo	gy English	Physics	Chanistry	Biology	Arabic Rol	gion Science	Geographi	c History	Administration and sconomy	Uni	Memorization	Perception	Lunguage	Analyzia	Course Name	Course Numbe	Course Name
2	60	n	86	0	0 0	0 80	83	70	80 8	2 6	5 6	0 0.3	0.9	0.1	0.7	المواق مالية	10871225	Financial markets
3	60	77	86	0	0 0	08 0	83	77	80 8	2 6	5 6	4 0.2	0.3	0.1	0.7	بمبارث إبالانية	10871321	Islamic banks
4	60	77	86	0	0 0	0 80	83	77	80 8	2 6	5 5	8 0.1	0.9	0.1	0.7	الرياضيات لإدارة الأصال	10211109	Mathematics for Business
5	60	n	86	0	0 (	08 0	83	11	80 8	2 6	5 6	0.2	0.9	0.1	0.7	مبلاى اقتصاد حزني	10801111	Microeconomics
6	60	77	86	0	0 0	08 C	83	77	80 8	2 6	5 6	2 0.2	8.0	0.1	0.7	بحرث الصليات	10866411	Operations Research
7	60	77	86	0	0 0	08 0	83	77	80 8	2 6	5 5	2 0.3	0.8	0.1	0.7	إدارة السويق	10876122	Marketing Management
8	60	n	86	0	0 (	08 0	83	17	80 8	2 6	5 5	0.4	0.8	0.1	0.7	فصد إساني	10801330	Islamic Economics
3	60	11	86	0	0 0	08 0	83	77	80 8	2 6	5 5	6 0.4	0.8	0.1	0.7	مبلائ الصويل	10871121	Principle of finance
10	60	77	86	0	0 (	080	83	77	80 8	2 6	5 5	3 0.4	0.8	0.1	0.7	محامية الضبراقب	10861223	Tax Accounting
11	60	n	86	0	0 0	08 0	83	11	80 8	2 6	5 6	4 0.6	0.7	0.1	0.7	لظم المعلومات المالية	10871334	Financial Information Systems
12	83	80	88	0	0 0	5 83	88	30	32 8	1	0 7	0 0.3	0.9	0.1	0.7	البواق مكية	10871225	Financial markets
10	03	00	00	0	0 0	83	88	30	3e 8	1	0 7	× 0.2	0.9	0.1	0.7	المنارف إسلامها	108/0321	icionic posks
14	03	00	00	0	0 0	83	00	30	32 8		0 8	4 0.1	0.9	0.1	0.7	الرياشيك لإثاره الإهبان	10211103	Mathematics for Dusiness
15	83	80	88	0	0 0	3 83	88	30	32 8		0 8	0 0.2	0.3	0.1	0.7	ميدن التغسد جزيي	10801111	Microsconomics
10	0.0	00	00	0	0 0	00	00	30	36 0		0 1	3 0.2	0.0	0.1	0.1	i du di citta	10000411	Operations Research
11	03	00	00	0	0 0	0.00	00	30	32 0		0 7	0 0.3	0.0	0.1	0.7	الدرة الشواق	10076122	Marketing Management
10	0.0	80	00	0	0 0	3 63	00	90	92 0	2 2	0 7	2 0.6	0.0	0.1	0.1	Justin alia	10001330	Descripto of firmers
00	03	80	00	0	0 0	3 63	00	30	96 0 90 9		0 7	0.6	0.8	0.1	0.1	and the second second	10841002	Principle of mance
20	00	00	00	0	0 0	000	00	30	02 0		0 1	0.0.0	0.0	0.1	0.1	2.4. de al contente de la	10001223	Tax Accounting
22	21	90	90	0	0 0	0 00	97	30	92 0 93 9		4 0	0.0	0.1	0.1	0.1	J.a. k. l	10011336	Financial mediate
02		90	80	0			67		90 9			4 0.0	0.5	0.1	0.1	1.0.1.1.1.	10011227	The second second
04	71	92	30	0	0 0	2	97	95	02 0 90 9		4 7	• 0.2	0.3	0.1	0.1	A white official state	10011321	Mathematics for Dusings
	24	92	90	0	0 0	) 95	97	91	92 9			. 0.1	0.0	0.1	0.7	Statistication of the state	1020100	AS
06	71	92	30	0	0 0	3 95	97	95	90 9		4 8	0.2	0.5	0.1	0.7	مندو منتخذ فراني	10866411	Operations Received
07	74	92	90	ů.	0 0	0.00	97	30	90 9		4 7	9 0.2	0.0	0.1	0.7	Sec. Stat. And	10076100	Marketing Management
28	21	92	30	0	0 0	95	97	95	92 9			5 0.6	0.0	0.1	0.7	And Lock	10801330	Idamic Economics
23	71	92	30	ů.	0 0	3 95	97	26	90 9			4 0.4	0.0	0.1	0.7	Judicts	50871121	Principle of figures
30	21	32	30	0	0 0	3 35	97	35	32 3	n	4	0.4	0.0	0.1	0.7	ومليبة الفيراني	10861223	Tay Accombing
31	11	92	90	0	0 0	95	97	95	92 9			0.6	0.7	0.1	0.7	2.0.alt. the shealt alti	10871334	Figure in Information Surfame
32	26	33	00	0	0 0	1 34	37	30	32 3		2 3	2 0.3	0.3	0.1	0.7	المالية الملا	10871225	Financial markets
33	96	93	88	0	0 0	94	97	90	92 9	9 9	2 9	6 0.2	0.9	0.1	0.7	مميار ف التلامية	10871321	Iclanic banks
34	35	33	88	0	0 0	94	97	30	92 3	9 9	2 3	4 0.1	0.9	0.1	0.7	ال بافتيات لايا، والأصل	10211103	Mathematics for Business
35	36	33	88	0	0 0	34	37	30	32 3	0 3	2 3	0.2	0.3	0.1	0.7	سلان الصباد جزئي	10801111	Microeconomics
36	96	93	88	0	0 0	94	97	90	92 9	9 9	2 3	0.2	0.8	0.1	0.7	بحوث الصليات	10866411	Operations Research
37	96	33	88	0	0 0	94	97	90	92 9	9 9	2 8	8 0.3	0.8	0.1	0.7	إدارة الشويق	10876122	Marketing Management
38	36	33	88	0	0 0	34	37	30	32 3	0 9	2 8	0.4	0.8	0.1	0.7	افسد بنمى	10801330	Islanic Economics
39	96	93	88	0	0 0	94	97	90	92 9	9 9	2 8	0.4	0.8	0.1	0.7	مبلاي المويل	10871121	Principle of finance
40	36	33	88	0	0 0	94	97	90	92 9	9 9	2 3	0 0.4	0.8	0.1	0.7	محاسبة الطنيرالاب	10861223	Tax Accounting
41	36	33	88	0	0 0	34	37	30	32 3	0 9	2 3	3 0.6	0.7	0.1	0.7	نظم المعلومات المالية	10871334	Financial Information Systems
42	60	77	86	0	0 0	0 80	83	70	80 8	2 6	5 6	0 0.3	0.9	0.1	0.7	المواقى ماتهة	10871225	Financial markets
43	60	77	86	0	0 0	08 0	83	77	80 8	2 6	5 6	4 0.2	0.9	0.1	0.7	الصارف إسلامهة	10871321	Islamic banks
44	60	n	86	0	0 0	08 0	83	77	80 8	2 6	5 5	8 0.1	0.9	0.1	0.7	الرياشيك لإداره الأصل	10211103	Mathematics for Business
45	60	77	86	0	0 0	0 80	83	77	80 8	2 6	5 6	0 0.2	0.9	0.1	0.7	مبلاي اقصباد حزتي	10801111	Microeconomics
46	60	77	86	0	0 0	08 0	83	77	80 8	2 6	5 6	2 0.2	0.8	0.1	0.7	بحرث المليات	10866411	Operations Research
47	60	77	86	0	0 0	08 0	83	77	80 8	2 6	5 5	2 0.3	0.8	0.1	0.7	إدارة الشويق	10876122	Marketing Management
48	60	77	86	0	0 0	0 80	83	77	80 8	2 6	5 5	5 0.4	0.8	0.1	0.7	الأعداد إسلامي	10801330	Islanic Economics
43	60	n	86	0	0 0	08 0	83	77	80 8	2 6	5 5	6 0.4	0.8	0.1	0.7	مبلاى الفويل	10871121	Principle of finance
50	60	77	86	0	0 0	080	83	77	80 8	2 6	5 5	3 0.4	0.8	0.1	0.7	محلبية الخنراتي	10861223	Tax Accounting
51	60	n	86	0	0 0	080	83	77	80 8	2 6	5 6	4 0.6	0.7	0.1	0.7	نظم المعلومات المالهة	10871334	Financial Information Systems
52	83	80	88	0	0 0	0 83	68	30	32 8	1 1	0 7	0 0.3	0.3	0.1	0.7	البواق مائية	10871225	Financial markets
	F	Law H	listory	Tourism	Arabic	English	Economy	· (+	)			1						Þ
				-														

## 4.2 Regression

For regression, the machine learning algorithm that we have used is simple linear regression, the reason behind that is that there is a linear relationship between the independent and the target variables.

#### Mechanism of Linear Regression

What linear regression does is, that it gives a straight line (when plott in 2d) and that line is aimed to represent the underlying data.

Linear regression gave us the equation as under

#### Predicted= aX1 + bX2 + c

For two independent/prediction variables X1 and X2 and for some parameters a,b and c that are estimated by the regression model.

Regression model assumes the random values of a,b,c (described above) and put them in the equation written above and obtain the predicted value.

4	А	В	С	D	E	F	G	H	1	J	ĸ	L	м	N	0	P	Q	R	S	Т	U
63	62	CIS	Software_	81.38608	Computer	0	0	0	0	0	0.090909	0.818182	0.090909	0	0	0	0	0	0	0	
64	63	CIS	Operating	80.39912	CIS	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	
65	64	CIS	Web_Prog	80.62709	CIS	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	
66	65	CIS	Informatio	79.99892	CIS	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	
67	66	CIS	Technical	80.14971	CIS	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	
68	67	Compute	Computer	81.3064	Computer	0	0	0	0	0	0.090909	0.818182	0.090909	0	0	0	0	0	0	0	
69	68	Compute	Data_Stru	80.86414	Computer	0	0	0	0	0	0.090909	0.818182	0.090909	0	0	0	0	0	0	0	1
70	69	Compute	Object_Or	81.35758	Computer	0	0	0	0	0	0.090909	0.818182	0.090909	0	0	0	0	0	0	0	
71	70	Compute	Software_	87.72963	Computer	0	0	0	0	0	0.090909	0.818182	0.090909	0	0	0	0	0	0	0	
72	71	Compute	Database	81.35758	Computer	0	0	0	0	0	0.090909	0.818182	0.090909	0	0	0	0	0	0	0	
73	72	Compute	Computer	82.29328	Computer	0	0	0	0	0	0.090909	0.818182	0.090909	0	0	0	0	0	0	0	
74	73	Compute	Microproc	81.79787	Computer	0	0	0	0	0	0.090909	0.818182	0.090909	0	0	0	0	0	0	0	
75	74	Compute	Computer	82.54524	Computer	0	0	0	0	0	0.090909	0.818182	0.090909	0	0	0	0	0	0	0	
76	75	Compute	Digital_Ele	81.25522	Computer	0	0	0	0	0	0.090909	0.818182	0.090909	0	0	0	0	0	0	0	
77	76	CS	Data_Stru	81.41016	Computer	0	0	0	0	0	0.090909	0.818182	0.090909	0	0	0	0	0	0	0	
78	77	CS	Design_ar	81.41016	CS	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	
79	78	CS	Computer	81.18101	CS	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	
80	79	CS	Linear_Al	83.01483	CS	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	
81	80	CS	Calculus_I	82.40611	CS	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	
82	81	CS	Calculus_I	82.40611	CS	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	
83	82	CS	Computer	82.22283	CS	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	
	10	predic	ted uni ma	rks	(+)								1								Þ

#### 4.3MSE

After obtaining the predicted value, it is compared with the actual value and difference is calculated which is called the loss.

The same procedure is performed for all data points. And the loss is calculated for each point.

The loss for each point is squared and then added and then we take square root of the added loss. That is called mean square error (MSE).

$$ext{MSE} = rac{1}{n}\sum_{i=1}^n (Y_i - \hat{Y_i})^2.$$

The purpose of regression model is to find the optimum values for its parameters (in this example a,b,c) so that the mse should be minimum.

So, regression does all the process (explained above) iteratively by changing the values of a,b,c in every iteration since the aim of machine learning models is to find the minimum loss.

Then after some iterations, the process stops. The stopping criteria may be different. For example

1. The number of iterations

2. The difference in MSE loss of this iteration with previous iteration is not greater than a preset threshold..etc

At the end, we have optimum values of a,b,c parameters. So that when we put these values in our regression equation along with prediction variables, the equation gives us a prediction that is very near to the original value.

1	A	В	C	D	E
65	54.69108742	60	28.185		
66	58.27843728	62	13.850		
67	56.69813238	58	1.695		
68	55.98066241	59	9.116		
69	54.69108742	54	0.478		
70	56.91617981	51	35.001		
71	55.51496331	50	30.415		
72	90.16647656	90	0.028		
73	91.92586984	96	16.599		
74	92.89511068	92	0.801		
75	88.76937924	93	17.898		
76	92.3567291	97	21.560		
77	90.7764242	90	0.603		
78	90.05895423	91	0.886		
79	88.76937924	93	17.898		
80	90.99447163	90	0.989		
81	89.59325512	90	0.165		
82					
83		MSE	8.680		
84					
85		in the second			
	4 F.	Sheet1	(+)		

# CHAPTER FIVE

# Conclusion and Recommendation

# 5.1 Future Work

- Cooperating with the Ministry of Labor to obtain the results of the examination of abilities, skills, competencies and tendencies. And use them as attributes in the regression model to increase accuracy
- Include all university Majors

## 5.2 Conclusion

Given the criteria for our case study, We found that a student's High School marks in some particular subjects is a strong predictor of their University performance relating each major, we have investigated the predicted marks for university students to find the relation between each major and it's correlated subject at high school.

As a result, it turned out that there is a linear relation between each major and some particular high school class marks.

Nevertheless a number of other factors may play a significant role on their achievement that can't be included in our study, such as side pressures, the nature of the social environment, etc.