State of Palestine
An-Najah National University
Faculty of Information Technology
Computerized Information Systems

دولـــــــــــة فلســـــــــطين
جامعة النـجـاح الـوطنـيـة
كليـة تكنولوجيا المعلومات
قسم نظم المعلومات المحوسبة


# The Correlation Between the Academic Achievements Of Students In Both School And University Stages

*Students:*

| | |
|---|---|
| Ahd Hinnawi | 11420877 |
| Shatha Abulhawa | 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

# 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 2nd 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. **10th grade marks:**
2. **11th grade marks:**
3. **12th 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

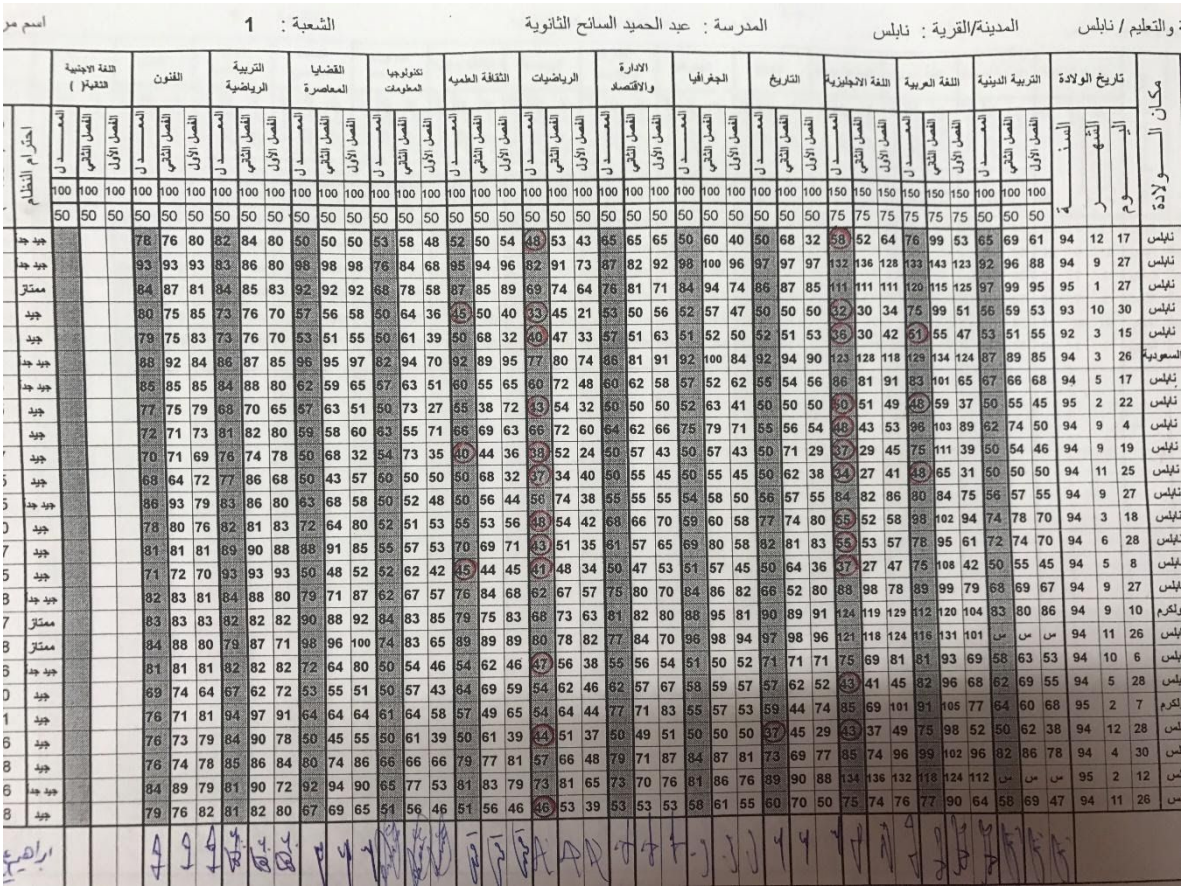



*Figure 1: Paper copy of students' school grades*

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.

<u>Difficulties encountered during this process</u>

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.

| كلية العلوم الانسانية | |
|---|---|
| التاريخ 10321 | |
| تاريخ الجزيرة العربية قبل الإسلام | 10321122 |
| تاريخ الدولة الأموية | 10321124 |
| تاريخ المغرب في العصر الإسلامي | 10321125 |
| تاريخ الأندلس | 10321126 |
| تاريخ الدولة العباسية | 10321227 |
| تاريخ الأيوبيين والمماليك | 10321230 |
| فكر سياسي إسلامي | 10321231 |
| تاريخ الدولة العثمانية | 10321335 |
| تاريخ فلسطين الحديث | 10321440 |
| تاريخ القدس الحديث والمعاصر | 10321441 |
| | |
| السياحة والاثار 10316 | |
| تاريخ فلسطين والأردن في العصور القديمة | 10316114 |
| التخطيط السياحي | 10316220 |
| إدارة المواقع السياحية والأثرية | 10316221 |
| السياحة في فلسطين | 10316222 |
| آثار اليونان والرومان* | 10316225 |
| آثار بيزنطيّة | 10316226 |
| لغة قديمة | 10316337 |
| تاريخ وآثار 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:

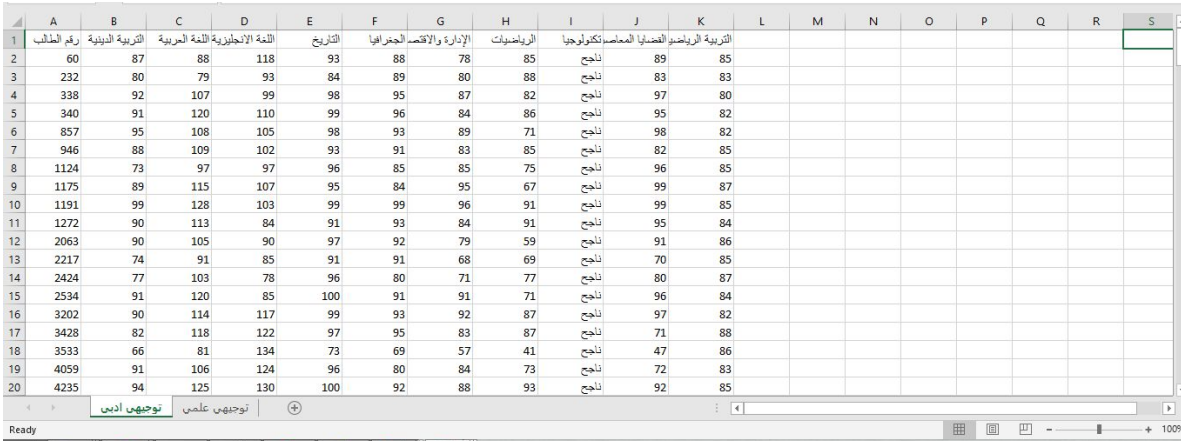
| | A | B | C | D | E | F | G | H | I | J | K | L | M | N | O | P | Q | R | S | T | U |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | rand | 10201101 | 10201102 | 10201107 | 10201232 | 10201264 | 10201321 | 10201341 | 10201347 | 10206392 | 10211101 | 10211102 | 10211109 | 10211201 | 10211212 | 10211241 | 10211242 | 10211302 | 10211321 | 10216302 | 1021630 |
| 2 | 33 | | | | | | | | | | 75 | 65 | | 60 | | | | | | | |
| 3 | 44 | | | | | | | | | | | | | | | | | | | | |
| 4 | 47 | | | | | | | | | | | | 63 | | | | | | | | |
| 5 | 60 | | | | | | | | | | | | 60 | | | | | | | | |
| 6 | 79 | | | | | | | | | | | | | | | | | | | | |
| 7 | 133 | | | | | | | | | | | | 75 | | | | | | | | |
| 8 | 150 | | | | | | | | | | | | | | | | | | | | |
| 9 | 185 | | | | | | | | | | | | | | | | | | | | |
| 10 | 222 | | | | | | | | | | | | | | | | | | | | |
| 11 | 225 | | | | | | | | | | | | | | | | | | | | |
| 12 | 232 | | | | | | | | | | | | | | | | | | | | |
| 13 | 271 | | | | | | | | | | | | 62 | | | | | | | | |
| 14 | 338 | | | | | | | | | | | | 88 | | | | | | | | |
| 15 | 340 | | | | | | | | | | | | | | | | | | | | |
| 16 | 467 | | | | | | | | | | | | 60 | | | | | | | | |
| 17 | 499 | | | | | | | | | | 82 | 86 | | 78 | | | | | | | |
| 18 | 534 | | | | | | | | | | 88 | 93 | | 78 | | | | | | | |
| 19 | 541 | | | | | | | | | | 80 | 80 | | 75 | | | | | | | |
| 20 | 560 | | | | | | | | | | 85 | 74 | | 70 | | | | | | | |
| 21 | 592 | | | | | | | | | | | | | | | | | | | | |

Sheet1

Ready

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

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



| | A | B | C | D | E | F | G | H | I | J | K |
|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | رقم الطالب | التربية الدينية | اللغة العربية | اللغة الانجليزية | التاريخ | الجغرافيا | الإدارة والاقتصا | الرياضيات | تكنولوجيا | الفضايا المعاصر | التربية الرياضي |
| 2 | 60 | 87 | 88 | 118 | 93 | 88 | 78 | 85 | ناجح | 89 | 85 |
| 3 | 232 | 80 | 79 | 93 | 84 | 89 | 80 | 88 | ناجح | 83 | 83 |
| 4 | 338 | 92 | 107 | 99 | 98 | 95 | 87 | 82 | ناجح | 97 | 80 |
| 5 | 340 | 91 | 120 | 110 | 99 | 96 | 84 | 86 | ناجح | 95 | 82 |
| 6 | 857 | 95 | 108 | 105 | 98 | 93 | 89 | 71 | ناجح | 98 | 82 |
| 7 | 946 | 88 | 109 | 102 | 93 | 91 | 83 | 85 | ناجح | 82 | 85 |
| 8 | 1124 | 73 | 97 | 97 | 96 | 85 | 85 | 75 | ناجح | 96 | 85 |
| 9 | 1175 | 89 | 115 | 107 | 95 | 84 | 95 | 67 | ناجح | 99 | 87 |
| 10 | 1191 | 99 | 128 | 103 | 99 | 99 | 96 | 91 | ناجح | 99 | 85 |
| 11 | 1272 | 90 | 113 | 84 | 91 | 93 | 84 | 91 | ناجح | 95 | 84 |
| 12 | 2063 | 90 | 105 | 90 | 97 | 92 | 79 | 59 | ناجح | 91 | 86 |
| 13 | 2217 | 74 | 91 | 85 | 91 | 91 | 68 | 69 | ناجح | 70 | 85 |
| 14 | 2424 | 77 | 103 | 78 | 96 | 80 | 71 | 77 | ناجح | 80 | 87 |
| 15 | 2534 | 91 | 120 | 85 | 100 | 91 | 91 | 71 | ناجح | 96 | 84 |
| 16 | 3202 | 90 | 114 | 117 | 99 | 93 | 92 | 87 | ناجح | 97 | 82 |
| 17 | 3428 | 82 | 118 | 122 | 97 | 95 | 83 | 87 | ناجح | 71 | 88 |
| 18 | 3533 | 66 | 81 | 134 | 73 | 69 | 57 | 41 | ناجح | 47 | 86 |
| 19 | 4059 | 91 | 106 | 124 | 96 | 80 | 84 | 73 | ناجح | 72 | 83 |
| 20 | 4235 | 94 | 125 | 130 | 100 | 92 | 88 | 93 | ناجح | 92 | 85 |



*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.

<u>Data preprocessing included:</u>

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

Which went as follows:

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

| Student Id | Math | Tech | 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 |

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.

| Student Id | Math | Technology | English | Physics | Chemistry | Biology | Arabic | Religion | Science | Geographic | History | Administration and 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

| Student Id | Math | Technology | English | Data Structure | Design and Analysis of Algorithms | Computer Organization and Assembly Language | Linear Algebra for Computer Science | Calculus I | Calculus II |
|---|---|---|---|---|---|---|---|---|---|
| 1 | 67 | 65 | 55 | 50 | 53 | 51 | 55 | 58 | 59 |

| Student Id | Math | Technology | English | UNI | Course Number | Course Name |
|---|---|---|---|---|---|---|
| 1 | 67 | 65 | 55 | 50 | 10671210 | Data Structure |
| 1 | 67 | 65 | 55 | 53 | 10671212 | Design and Analysis of Algorithms |
| 1 | 67 | 65 | 55 | 51 | 10671243 | Assembly Language |
| 1 | 67 | 65 | 55 | 55 | 10671244 | Linear Algebra |
| 1 | 67 | 65 | 55 | 58 | 10211101 | Calculus I |
| 1 | 67 | 65 | 55 | 59 | 10211102 | Calculus II |
| 1 | 67 | 65 | 55 | 62 | 10671321 | Computer Architecture |
| 1 | 67 | 65 | 55 | 60 | 10671353 | Database Systems Design |
| 1 | 67 | 65 | 55 | 61 | 10671102 | Principles of Programming II |
| 1 | 67 | 65 | 55 | 62 | 10671204 | Web Programming |

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:

| Memorization | Perception | Language | Analysis |
|---|---|---|---|

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

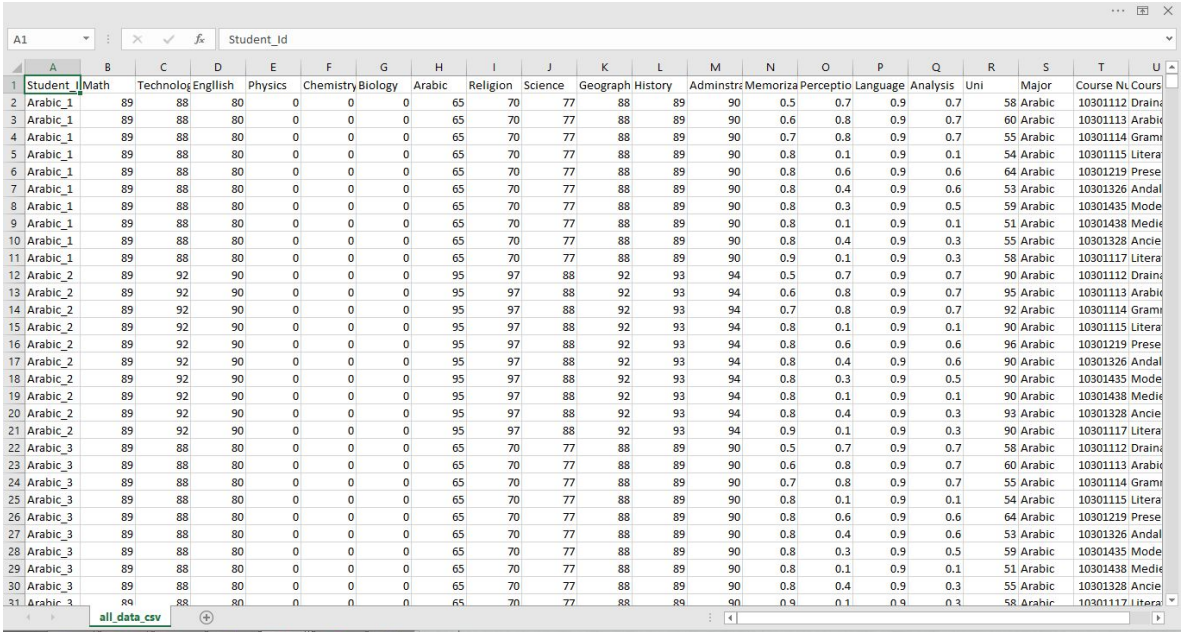
# CHAPTER FOUR

# *MODEL BUILDING*

## *4.1 Datasets*

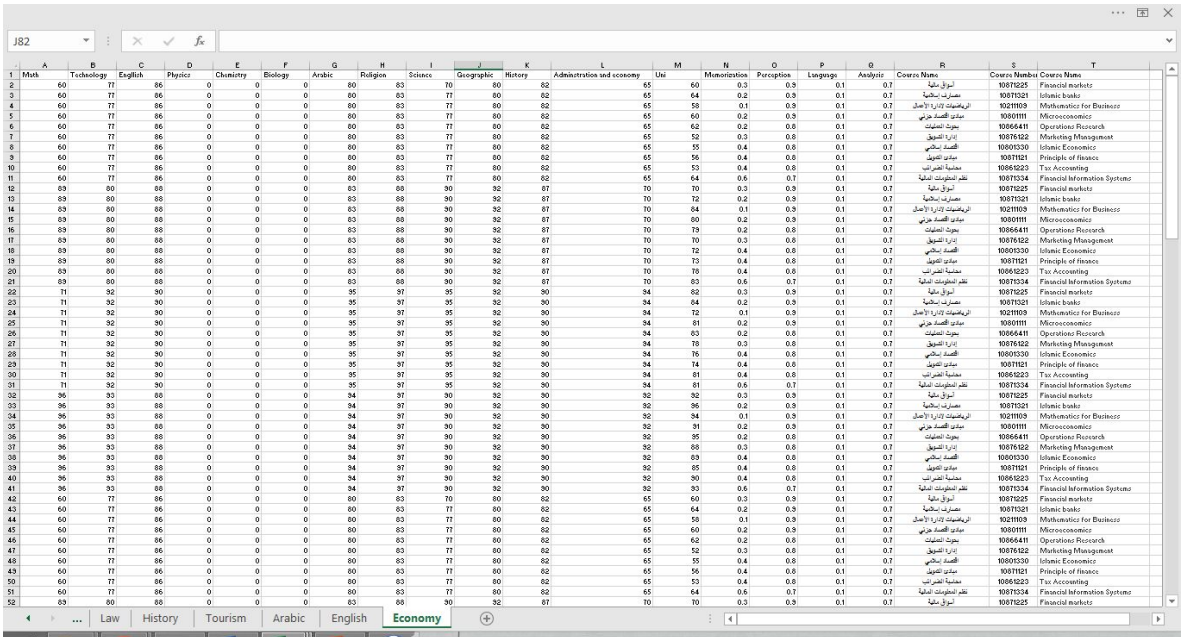
To build up the real picture of what we want,  we needed to check different datasets, and different collections of information and choose the best
So, we went through 2 datasets:

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



| Student_I | Math | Technolog | Englisch | Physics | Chemistry | Biology | Arabic | Religion | Science | Geograph | History | Adminstra | Memoriza | Perceptio | Language | Analysis | Uni | Major | Course Nu | Cours |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Arabic_1 | 89 | 88 | 80 | 0 | 0 | 0 | 65 | 70 | 77 | 88 | 89 | 90 | 0.5 | 0.7 | 0.9 | 0.7 | 58 | Arabic | 10301112 | Draina |
| Arabic_1 | 89 | 88 | 80 | 0 | 0 | 0 | 65 | 70 | 77 | 88 | 89 | 90 | 0.6 | 0.8 | 0.9 | 0.7 | 60 | Arabic | 10301113 | Arabic |
| Arabic_1 | 89 | 88 | 80 | 0 | 0 | 0 | 65 | 70 | 77 | 88 | 89 | 90 | 0.7 | 0.8 | 0.9 | 0.7 | 55 | Arabic | 10301114 | Grami |
| Arabic_1 | 89 | 88 | 80 | 0 | 0 | 0 | 65 | 70 | 77 | 88 | 89 | 90 | 0.8 | 0.1 | 0.9 | 0.1 | 54 | Arabic | 10301115 | Litera |
| Arabic_1 | 89 | 88 | 80 | 0 | 0 | 0 | 65 | 70 | 77 | 88 | 89 | 90 | 0.8 | 0.6 | 0.9 | 0.6 | 64 | Arabic | 10301219 | Prese |
| Arabic_1 | 89 | 88 | 80 | 0 | 0 | 0 | 65 | 70 | 77 | 88 | 89 | 90 | 0.8 | 0.4 | 0.9 | 0.6 | 53 | Arabic | 10301326 | Andal |
| Arabic_1 | 89 | 88 | 80 | 0 | 0 | 0 | 65 | 70 | 77 | 88 | 89 | 90 | 0.8 | 0.3 | 0.9 | 0.5 | 59 | Arabic | 10301435 | Mode |
| Arabic_1 | 89 | 88 | 80 | 0 | 0 | 0 | 65 | 70 | 77 | 88 | 89 | 90 | 0.8 | 0.1 | 0.9 | 0.1 | 51 | Arabic | 10301438 | Medie |
| Arabic_1 | 89 | 88 | 80 | 0 | 0 | 0 | 65 | 70 | 77 | 88 | 89 | 90 | 0.8 | 0.4 | 0.9 | 0.3 | 55 | Arabic | 10301328 | Ancie |
| Arabic_1 | 89 | 88 | 80 | 0 | 0 | 0 | 65 | 70 | 77 | 88 | 89 | 90 | 0.9 | 0.1 | 0.9 | 0.3 | 58 | Arabic | 10301117 | Litera |
| Arabic_2 | 89 | 92 | 90 | 0 | 0 | 0 | 95 | 97 | 88 | 92 | 93 | 94 | 0.5 | 0.7 | 0.9 | 0.7 | 90 | Arabic | 10301112 | Draina |
| Arabic_2 | 89 | 92 | 90 | 0 | 0 | 0 | 95 | 97 | 88 | 92 | 93 | 94 | 0.6 | 0.8 | 0.9 | 0.7 | 95 | Arabic | 10301113 | Arabic |
| Arabic_2 | 89 | 92 | 90 | 0 | 0 | 0 | 95 | 97 | 88 | 92 | 93 | 94 | 0.7 | 0.8 | 0.9 | 0.7 | 92 | Arabic | 10301114 | Grami |
| Arabic_2 | 89 | 92 | 90 | 0 | 0 | 0 | 95 | 97 | 88 | 92 | 93 | 94 | 0.8 | 0.1 | 0.9 | 0.1 | 90 | Arabic | 10301115 | Litera |
| Arabic_2 | 89 | 92 | 90 | 0 | 0 | 0 | 95 | 97 | 88 | 92 | 93 | 94 | 0.8 | 0.6 | 0.9 | 0.6 | 96 | Arabic | 10301219 | Prese |
| Arabic_2 | 89 | 92 | 90 | 0 | 0 | 0 | 95 | 97 | 88 | 92 | 93 | 94 | 0.8 | 0.4 | 0.9 | 0.6 | 90 | Arabic | 10301326 | Andal |
| Arabic_2 | 89 | 92 | 90 | 0 | 0 | 0 | 95 | 97 | 88 | 92 | 93 | 94 | 0.8 | 0.3 | 0.9 | 0.5 | 90 | Arabic | 10301435 | Mode |
| Arabic_2 | 89 | 92 | 90 | 0 | 0 | 0 | 95 | 97 | 88 | 92 | 93 | 94 | 0.8 | 0.1 | 0.9 | 0.1 | 90 | Arabic | 10301438 | Medie |
| Arabic_2 | 89 | 92 | 90 | 0 | 0 | 0 | 95 | 97 | 88 | 92 | 93 | 94 | 0.8 | 0.4 | 0.9 | 0.3 | 93 | Arabic | 10301328 | Ancie |
| Arabic_2 | 89 | 92 | 90 | 0 | 0 | 0 | 95 | 97 | 88 | 92 | 93 | 94 | 0.9 | 0.1 | 0.9 | 0.3 | 90 | Arabic | 10301117 | Litera |
| Arabic_3 | 89 | 88 | 80 | 0 | 0 | 0 | 65 | 70 | 77 | 88 | 89 | 90 | 0.5 | 0.7 | 0.9 | 0.7 | 58 | Arabic | 10301112 | Draina |
| Arabic_3 | 89 | 88 | 80 | 0 | 0 | 0 | 65 | 70 | 77 | 88 | 89 | 90 | 0.6 | 0.8 | 0.9 | 0.7 | 60 | Arabic | 10301113 | Arabic |
| Arabic_3 | 89 | 88 | 80 | 0 | 0 | 0 | 65 | 70 | 77 | 88 | 89 | 90 | 0.7 | 0.8 | 0.9 | 0.7 | 55 | Arabic | 10301114 | Grami |
| Arabic_3 | 89 | 88 | 80 | 0 | 0 | 0 | 65 | 70 | 77 | 88 | 89 | 90 | 0.8 | 0.1 | 0.9 | 0.1 | 54 | Arabic | 10301115 | Litera |
| Arabic_3 | 89 | 88 | 80 | 0 | 0 | 0 | 65 | 70 | 77 | 88 | 89 | 90 | 0.8 | 0.6 | 0.9 | 0.6 | 64 | Arabic | 10301219 | Prese |
| Arabic_3 | 89 | 88 | 80 | 0 | 0 | 0 | 65 | 70 | 77 | 88 | 89 | 90 | 0.8 | 0.4 | 0.9 | 0.6 | 53 | Arabic | 10301326 | Andal |
| Arabic_3 | 89 | 88 | 80 | 0 | 0 | 0 | 65 | 70 | 77 | 88 | 89 | 90 | 0.8 | 0.3 | 0.9 | 0.5 | 59 | Arabic | 10301435 | Mode |
| Arabic_3 | 89 | 88 | 80 | 0 | 0 | 0 | 65 | 70 | 77 | 88 | 89 | 90 | 0.8 | 0.1 | 0.9 | 0.1 | 51 | Arabic | 10301438 | Medie |
| Arabic_3 | 89 | 88 | 80 | 0 | 0 | 0 | 65 | 70 | 77 | 88 | 89 | 90 | 0.8 | 0.4 | 0.9 | 0.3 | 55 | Arabic | 10301328 | Ancie |
| Arabic_3 | 89 | 88 | 80 | 0 | 0 | 0 | 65 | 70 | 77 | 88 | 89 | 90 | 0.9 | 0.1 | 0.9 | 0.3 | 58 | Arabic | 10301117 | Litera |

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



| | A | B | C | D | E | F | G | H | I | J | K | L | M | N | O | P | Q | R | S | T |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | Math | Technology | English | Physics | Chemistry | Biology | Arabic | Religion | Science | Geographic | History | Administration and economy | Uni | Memorization | Perception | Languages | Analysis | Course Name | Course Number | Course Name |
| 2 | 60 | 77 | 86 | 0 | 0 | 0 | 80 | 83 | 70 | 80 | 82 | 65 | 60 | 0.3 | 0.3 | 0.1 | 0.7 | أسواق مالية | 10871225 | Financial markets |
| 3 | 60 | 77 | 86 | 0 | 0 | 0 | 80 | 83 | 77 | 80 | 82 | 65 | 64 | 0.2 | 0.3 | 0.1 | 0.7 | مصارف إسلامية | 10871321 | Islamic banks |
| 4 | 60 | 77 | 86 | 0 | 0 | 0 | 80 | 83 | 77 | 80 | 82 | 65 | 59 | 0.1 | 0.3 | 0.1 | 0.7 | الرياضيات لإدارة الأعمال | 10211103 | Mathematics for Business |
| 5 | 60 | 77 | 86 | 0 | 0 | 0 | 80 | 83 | 77 | 80 | 82 | 65 | 60 | 0.2 | 0.3 | 0.1 | 0.7 | مبادئ اقتصاد جزئي | 10801111 | Microeconomics |
| 6 | 60 | 77 | 86 | 0 | 0 | 0 | 80 | 83 | 77 | 80 | 82 | 65 | 62 | 0.2 | 0.3 | 0.1 | 0.7 | بحوث العمليات | 10866411 | Operations Research |
| 7 | 60 | 77 | 86 | 0 | 0 | 0 | 80 | 83 | 77 | 80 | 82 | 65 | 52 | 0.3 | 0.3 | 0.1 | 0.7 | إدارة التسويق | 10876122 | Marketing Management |
| 8 | 60 | 77 | 86 | 0 | 0 | 0 | 80 | 83 | 77 | 80 | 82 | 65 | 55 | 0.4 | 0.3 | 0.1 | 0.7 | اقتصاد إسلامي | 10801330 | Islamic Economics |
| 9 | 60 | 77 | 86 | 0 | 0 | 0 | 80 | 83 | 77 | 80 | 82 | 65 | 56 | 0.4 | 0.3 | 0.1 | 0.7 | مبادئ التمويل | 10871121 | Principle of finance |
| 10 | 60 | 77 | 86 | 0 | 0 | 0 | 80 | 83 | 77 | 80 | 82 | 65 | 53 | 0.4 | 0.3 | 0.1 | 0.7 | محاسبة الضرائب | 10861223 | Tax Accounting |
| 11 | 60 | 77 | 86 | 0 | 0 | 0 | 80 | 83 | 77 | 80 | 82 | 65 | 64 | 0.6 | 0.7 | 0.1 | 0.7 | نظم المعلومات المالية | 10871334 | Financial Information Systems |
| 12 | 89 | 80 | 88 | 0 | 0 | 0 | 83 | 88 | 90 | 92 | 87 | 70 | 70 | 0.3 | 0.3 | 0.1 | 0.7 | أسواق مالية | 10871225 | Financial markets |
| 13 | 89 | 80 | 88 | 0 | 0 | 0 | 83 | 88 | 90 | 92 | 87 | 70 | 72 | 0.2 | 0.3 | 0.1 | 0.7 | مصارف إسلامية | 10871321 | Islamic banks |
| 14 | 89 | 80 | 88 | 0 | 0 | 0 | 83 | 88 | 90 | 92 | 87 | 70 | 64 | 0.1 | 0.3 | 0.1 | 0.7 | الرياضيات لإدارة الأعمال | 10211103 | Mathematics for Business |
| 15 | 89 | 80 | 88 | 0 | 0 | 0 | 83 | 88 | 90 | 92 | 87 | 70 | 60 | 0.2 | 0.3 | 0.1 | 0.7 | مبادئ اقتصاد جزئي | 10801111 | Microeconomics |
| 16 | 89 | 80 | 88 | 0 | 0 | 0 | 83 | 88 | 90 | 92 | 87 | 70 | 79 | 0.2 | 0.3 | 0.1 | 0.7 | بحوث العمليات | 10866411 | Operations Research |
| 17 | 89 | 80 | 88 | 0 | 0 | 0 | 83 | 88 | 90 | 92 | 87 | 70 | 70 | 0.3 | 0.3 | 0.1 | 0.7 | إدارة التسويق | 10876122 | Marketing Management |
| 18 | 89 | 80 | 88 | 0 | 0 | 0 | 83 | 88 | 90 | 92 | 87 | 70 | 72 | 0.4 | 0.3 | 0.1 | 0.7 | اقتصاد إسلامي | 10801330 | Islamic Economics |
| 19 | 89 | 80 | 88 | 0 | 0 | 0 | 83 | 88 | 90 | 92 | 87 | 70 | 73 | 0.4 | 0.3 | 0.1 | 0.7 | مبادئ التمويل | 10871121 | Principle of finance |
| 20 | 89 | 80 | 88 | 0 | 0 | 0 | 83 | 88 | 90 | 92 | 87 | 70 | 70 | 0.4 | 0.3 | 0.1 | 0.7 | محاسبة الضرائب | 10861223 | Tax Accounting |
| 21 | 89 | 80 | 88 | 0 | 0 | 0 | 83 | 88 | 90 | 92 | 87 | 70 | 83 | 0.6 | 0.7 | 0.1 | 0.7 | نظم المعلومات المالية | 10871334 | Financial Information Systems |
| 22 | 71 | 92 | 90 | 0 | 0 | 0 | 95 | 97 | 95 | 92 | 90 | 94 | 82 | 0.3 | 0.3 | 0.1 | 0.7 | أسواق مالية | 10871225 | Financial markets |
| 23 | 71 | 92 | 90 | 0 | 0 | 0 | 95 | 97 | 95 | 92 | 90 | 94 | 64 | 0.2 | 0.3 | 0.1 | 0.7 | مصارف إسلامية | 10871321 | Islamic banks |
| 24 | 71 | 92 | 90 | 0 | 0 | 0 | 95 | 97 | 95 | 92 | 90 | 94 | 72 | 0.1 | 0.3 | 0.1 | 0.7 | الرياضيات لإدارة الأعمال | 10211103 | Mathematics for Business |
| 25 | 71 | 92 | 90 | 0 | 0 | 0 | 95 | 97 | 95 | 92 | 90 | 94 | 81 | 0.2 | 0.3 | 0.1 | 0.7 | مبادئ اقتصاد جزئي | 10801111 | Microeconomics |
| 26 | 71 | 92 | 90 | 0 | 0 | 0 | 95 | 97 | 95 | 92 | 90 | 94 | 63 | 0.2 | 0.3 | 0.1 | 0.7 | بحوث العمليات | 10866411 | Operations Research |
| 27 | 71 | 92 | 90 | 0 | 0 | 0 | 95 | 97 | 95 | 92 | 90 | 94 | 78 | 0.3 | 0.3 | 0.1 | 0.7 | إدارة التسويق | 10876122 | Marketing Management |
| 28 | 71 | 92 | 90 | 0 | 0 | 0 | 95 | 97 | 95 | 92 | 90 | 94 | 76 | 0.4 | 0.3 | 0.1 | 0.7 | اقتصاد إسلامي | 10801330 | Islamic Economics |
| 29 | 71 | 92 | 90 | 0 | 0 | 0 | 95 | 97 | 95 | 92 | 90 | 94 | 74 | 0.4 | 0.3 | 0.1 | 0.7 | مبادئ التمويل | 10871121 | Principle of finance |
| 30 | 71 | 92 | 90 | 0 | 0 | 0 | 95 | 97 | 95 | 92 | 90 | 94 | 81 | 0.4 | 0.3 | 0.1 | 0.7 | محاسبة الضرائب | 10861223 | Tax Accounting |
| 31 | 71 | 92 | 90 | 0 | 0 | 0 | 95 | 97 | 95 | 92 | 90 | 94 | 81 | 0.6 | 0.7 | 0.1 | 0.7 | نظم المعلومات المالية | 10871334 | Financial Information Systems |
| 32 | 96 | 93 | 88 | 0 | 0 | 0 | 94 | 97 | 90 | 92 | 90 | 92 | 92 | 0.3 | 0.3 | 0.1 | 0.7 | أسواق مالية | 10871225 | Financial markets |
| 33 | 96 | 93 | 88 | 0 | 0 | 0 | 94 | 97 | 90 | 92 | 90 | 92 | 96 | 0.2 | 0.3 | 0.1 | 0.7 | مصارف إسلامية | 10871321 | Islamic banks |
| 34 | 96 | 93 | 88 | 0 | 0 | 0 | 94 | 97 | 90 | 92 | 90 | 92 | 94 | 0.1 | 0.3 | 0.1 | 0.7 | الرياضيات لإدارة الأعمال | 10211103 | Mathematics for Business |
| 35 | 96 | 93 | 88 | 0 | 0 | 0 | 94 | 97 | 90 | 92 | 90 | 92 | 91 | 0.2 | 0.3 | 0.1 | 0.7 | مبادئ اقتصاد جزئي | 10801111 | Microeconomics |
| 36 | 96 | 93 | 88 | 0 | 0 | 0 | 94 | 97 | 90 | 92 | 90 | 92 | 95 | 0.2 | 0.3 | 0.1 | 0.7 | بحوث العمليات | 10866411 | Operations Research |
| 37 | 96 | 93 | 88 | 0 | 0 | 0 | 94 | 97 | 90 | 92 | 90 | 92 | 88 | 0.3 | 0.3 | 0.1 | 0.7 | إدارة التسويق | 10876122 | Marketing Management |
| 38 | 96 | 93 | 88 | 0 | 0 | 0 | 94 | 97 | 90 | 92 | 90 | 92 | 83 | 0.4 | 0.3 | 0.1 | 0.7 | اقتصاد إسلامي | 10801330 | Islamic Economics |
| 39 | 96 | 93 | 88 | 0 | 0 | 0 | 94 | 97 | 90 | 92 | 90 | 92 | 85 | 0.4 | 0.3 | 0.1 | 0.7 | مبادئ التمويل | 10871121 | Principle of finance |
| 40 | 96 | 93 | 88 | 0 | 0 | 0 | 94 | 97 | 90 | 92 | 90 | 92 | 90 | 0.4 | 0.3 | 0.1 | 0.7 | محاسبة الضرائب | 10861223 | Tax Accounting |
| 41 | 96 | 93 | 88 | 0 | 0 | 0 | 94 | 97 | 90 | 92 | 90 | 92 | 93 | 0.6 | 0.7 | 0.1 | 0.7 | نظم المعلومات المالية | 10871334 | Financial Information Systems |
| 42 | 60 | 77 | 86 | 0 | 0 | 0 | 80 | 83 | 70 | 80 | 82 | 65 | 60 | 0.3 | 0.3 | 0.1 | 0.7 | أسواق مالية | 10871225 | Financial markets |
| 43 | 60 | 77 | 86 | 0 | 0 | 0 | 80 | 83 | 77 | 80 | 82 | 65 | 64 | 0.2 | 0.3 | 0.1 | 0.7 | مصارف إسلامية | 10871321 | Islamic banks |
| 44 | 60 | 77 | 86 | 0 | 0 | 0 | 80 | 83 | 77 | 80 | 82 | 65 | 59 | 0.1 | 0.3 | 0.1 | 0.7 | الرياضيات لإدارة الأعمال | 10211103 | Mathematics for Business |
| 45 | 60 | 77 | 86 | 0 | 0 | 0 | 80 | 83 | 77 | 80 | 82 | 65 | 60 | 0.2 | 0.3 | 0.1 | 0.7 | مبادئ اقتصاد جزئي | 10801111 | Microeconomics |
| 46 | 60 | 77 | 86 | 0 | 0 | 0 | 80 | 83 | 77 | 80 | 82 | 65 | 62 | 0.2 | 0.3 | 0.1 | 0.7 | بحوث العمليات | 10866411 | Operations Research |
| 47 | 60 | 77 | 86 | 0 | 0 | 0 | 80 | 83 | 77 | 80 | 82 | 65 | 52 | 0.3 | 0.3 | 0.1 | 0.7 | إدارة التسويق | 10876122 | Marketing Management |
| 48 | 60 | 77 | 86 | 0 | 0 | 0 | 80 | 83 | 77 | 80 | 82 | 65 | 55 | 0.4 | 0.3 | 0.1 | 0.7 | اقتصاد إسلامي | 10801330 | Islamic Economics |
| 49 | 60 | 77 | 86 | 0 | 0 | 0 | 80 | 83 | 77 | 80 | 82 | 65 | 56 | 0.4 | 0.3 | 0.1 | 0.7 | مبادئ التمويل | 10871121 | Principle of finance |
| 50 | 60 | 77 | 86 | 0 | 0 | 0 | 80 | 83 | 77 | 80 | 82 | 65 | 53 | 0.4 | 0.3 | 0.1 | 0.7 | محاسبة الضرائب | 10861223 | Tax Accounting |
| 51 | 60 | 77 | 86 | 0 | 0 | 0 | 80 | 83 | 77 | 80 | 82 | 65 | 64 | 0.6 | 0.7 | 0.1 | 0.7 | نظم المعلومات المالية | 10871334 | Financial Information Systems |
| 52 | 89 | 80 | 88 | 0 | 0 | 0 | 83 | 88 | 90 | 92 | 87 | 70 | 70 | 0.3 | 0.3 | 0.1 | 0.7 | أسواق مالية | 10871225 | Financial markets |

Law | History | Tourism | Arabic | English | Economy

## *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 independant and the target variables.

Mechanism of Linear Regression

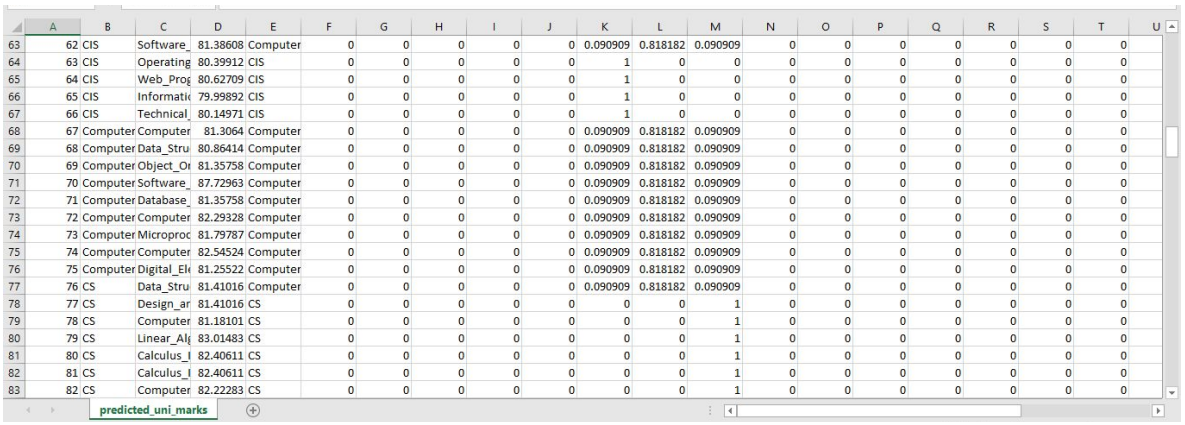
What linear regression does is, that it gives a straight line ( when plott in 2d) and that line is aimed to represent the underlying data.

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.



| | A | B | C | D | E | F | G | H | I | J | K | L | M | N | O | P | Q | R | S | T | 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 | Informatic | 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 | Computer | Computer_ | 81.3064 | Computer | 0 | 0 | 0 | 0 | 0 | 0.090909 | 0.818182 | 0.090909 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | |
| 69 | 68 | Computer | Data_Stru | 80.86414 | Computer | 0 | 0 | 0 | 0 | 0 | 0.090909 | 0.818182 | 0.090909 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | |
| 70 | 69 | Computer | 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 | Computer | Software_ | 87.72963 | Computer | 0 | 0 | 0 | 0 | 0 | 0.090909 | 0.818182 | 0.090909 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | |
| 72 | 71 | Computer | Database_ | 81.35758 | Computer | 0 | 0 | 0 | 0 | 0 | 0.090909 | 0.818182 | 0.090909 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | |
| 73 | 72 | Computer | Computer_ | 82.29328 | Computer | 0 | 0 | 0 | 0 | 0 | 0.090909 | 0.818182 | 0.090909 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | |
| 74 | 73 | Computer | Microproc | 81.79787 | Computer | 0 | 0 | 0 | 0 | 0 | 0.090909 | 0.818182 | 0.090909 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | |
| 75 | 74 | Computer | Computer_ | 82.54524 | Computer | 0 | 0 | 0 | 0 | 0 | 0.090909 | 0.818182 | 0.090909 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | |
| 76 | 75 | Computer | Digital_El | 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_Alg | 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 | |

predicted_uni_marks

## *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).

$$\mathrm{MSE} = \frac{1}{n}\sum_{i=1}^{n}(Y_i - \hat{Y}_i)^2.$$

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.

| | A | B | 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 | | | | | |

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.