



SCHOOL OF SCIENCE AND TECHNOLOGY

CSD3997

**STUDENT PERFORMANCE PREDICTION USING
MACHINE LEARNING**

DATE	20TH APRIL 2018
SUPERVISOR	STEVEN CAMILLERI
STUDENT NAME	NICHOLAS CHUN MARKUS
STUDENT NUMBER	M00598289
CAMPUS	MIDDLESEX UNIVERSITY MALTA

**A THESIS SUBMITTED IN PARTIAL FULFILMENT OF THE REQUIREMENTS FOR
THE DEGREE OF BACHELOR OF SCIENCE (HONS) IN COMPUTER SCIENCE
(SYSTEMS ENGINEERING)**

SCHOOL OF SCIENCE AND TECHNOLOGY

STUDENT NAME NICHOLAS CHUN MARKUS

STUDENT ID M00598289

MODULE NUMBER CSD 3997

I HEREBY CONFIRM THAT THE WORK PRESENTED HERE IN THIS REPORT AND IN ALL OTHER ASSOCIATED MATERIAL IS WHOLLY MY OWN WORK. I CONFIRM THAT THE REPORT HAS BEEN SUBMITTED TO TURNITIN AND THAT THE TURNITIN RESULTS ARE ON CD ATTACHED TO THIS REPORT. I AGREE TO ASSESSMENT FOR PLAGIARISM.

I HEREBY PROVIDE MIDDLESEX UNIVERSITY LIBRARY WITH CONSENT TO DISPLAY AND MAINTAIN THIS WORK WITHIN ITS COLLECTION.

SIGNATURE

DATE

18TH APRIL 2018

ABSTRACT

As a means of understanding the various aspects of students' life that affects their academic performance, a lot of surveys and research are being carried out to determine which aspect of their extra-curricular activities are responsible for affecting them either positive or negative. Being able to predict students' performance is an important tool in every educational environment. It gives a better understanding of the type of students in the educational environment and the possible factors that affect their performances. This report is not any different from other researches done. Although, it takes several features into consideration and then try to use those features to make accurate predictions of a students' performance.

Key Words: Student performance, Artificial Intelligence, Machine Learning, Web Application, Support Vector Machines (SVM), Decision Tree (DT)

ACKNOWLEDGEMENTS

I would like to thank the following people for their help during the writing and production of this thesis.

I want to first of all thank God for helping me throughout the period of writing my thesis. Mr. Steven Camilleri, my thesis mentor, without his help and support, this thesis would not have been possible.

Liza Franco, Middlesex University Malta's Librarian for constantly checking up on me and making sure I'm on the right track. The entire Middlesex University Management and Finally Friends and family for their prayers, support and advice.

TABLE OF CONTENTS**Abstract****Acknowledgements****Table of Contents****Table of Figures****List of Tables****List of Abbreviations****Chapter 1. Introduction**

- 1.1. Introduction
- 1.2. Problem definition
- 1.3. Aims
- 1.4. Objectives
- 1.5. Deliverables
- 1.6. Evaluation
- 1.7. Required Resources
- 1.8. Conclusion

Chapter 2. Literature Review

- 2.1. Introduction
- 2.2. Web application
- 2.3. Artificial Intelligence
- 2.4. Data Science
 - 2.4.1. Prescriptive Analytics
 - 2.4.2. Predictive Analytics
 - 2.4.3. Machine Learning (Predictions)
 - 2.4.4. Machine Learning (Pattern Discovery)
 - 2.4.5. Business Intelligence Vs Data Science
- 2.5. Machine Learning
- 2.6. Data Mining
 - 2.6.1. Classification method
 - 2.6.2. Academic Performance Prediction
 - 2.6.3. Data Preparation

2.7. Conclusion

Chapter 3 - System Requirements

Chapter 4 - Analysis and Design

4.1. Introduction

4.2. Current System

4.3. Weaknesses of Current System

4.4. Conclusion

Chapter 5 - Implementation

5.1. Introduction

5.2. Data gathering

5.3. Algorithm

5.3.1. Decision Tree

5.3.2. Support Vector Machine

5.4. Technologies

5.4.1. Python

5.4.2. HTML

5.4.3. Sqlalchemy

5.4.4. Flask

5.4.5. Bootstrap

5.5. Conclusion

Chapter 6 - Results

6.1. Introduction

6.2. Predictive analysis

Chapter 7 - Evaluation and Testing

7.1. Introduction

7.1.1. Strengths

7.1.2. Weaknesses

7.1.3. Opportunities

7.1.4. Threats

7.1.5. Conclusion

7.2. Testing

Chapter 8 - Conclusion

8.1. Introduction

8.2. Review

8.3. Future work

8.4. conclusion

References/Bibliography

Appendices

Installation Instructions

TABLE OF FIGURES

FIGURE 1: DATABASE CLASS DIAGRAM

FIGURE 2: USE CASE DIAGRAM

FIGURE 3: ACTIVITY DIAGRAM

FIGURE 4: FLOW CHART

FIGURE 5: LOGIN AND REGISTER PAGE

FIGURE 6: DASHBOARD

FIGURE 7: SVM HYPERPLANE (CRISTIANNI & SHAW-TAYLOR, 2000)

FIGURE 8: SCRIPT INITIATING THE APPLICATION

FIGURE 9: INDEX PAGE OF WEB APPLICATION

FIGURE 10: LOGIN PAGE

FIGURE 11: REGISTRATION PAGE

FIGURE 12: DASHBOARD

FIGURE 13: USER MODEL CLASS

FIGURE 14: STUDENT MODEL CLASS

FIGURE 15: STUDENT EDUCATIONAL INFORMATION MODEL CLASS

FIGURE 16.1: SVM ALGORITHM RUNNING

FIGURE 16.2: SVM ALGORITHM PREDICTED RESULT

FIGURE 17: LIST OF REQUIRED PYTHON MODULES

FIGURE 18: COST FUNCTION SCRIPT

FIGURE 19: DECLARATION FORM

LIST OF TABLES

TABLE 1: DIFFERENCE BETWEEN BI AND DS

TABLE 2: SIGN UP

TABLE 4: LOG IN

TABLE 5: VIEW STUDENT (S)

TABLE 6: ADD STUDENT

TABLE 7: IMPLEMENT SVM ALGORITHM

TABLE 8: RUN SVM

TABLE 9: IMPLEMENT DT ALGORITHM

TABLE 10: RUN DT

TABLE 11: CREATE CLASS MODELS

LIST OF ABBREVIATIONS

AI – ARTIFICIAL INTELLIGENCE

IDE – INTEGRATED DEVELOPMENT ENVIRONMENT

HTML – HYPERTEXT MARKUP LANGUAGE

BI – BUSINESS INTELLIGENCE

ML – MACHINE LEARNING

DT – DECISION TREE

SVM – SUPPORT VECTOR MACHINE

DS – DATA SCIENCE

SWOT – STRENGTH, WEAKNESSES, OPPORTUNITIES AND THREATS

ORM – OBJECT-ORIENTED MAPPER

SQL – STRUCTURED QUERY LANGUAGE

GUI – GRAPHICAL USER INTERFACE

OSS – OPEN-SOURCE SOFTWARE

API – APPLICATION PROGRAM INTERFACE

DBAPI – DATABASE APPLICATION PROGRAM INTERFACE

Chapter 1 - Introduction

1.1. Introduction

As web applications grow, so do challenges. The use of Artificial intelligence today has made certain applications to stand out amongst many.

Artificial Intelligence (AI) is the ability of a computer to perform certain activities which are only common to human beings. AI is a term used in the aspect of development of systems that have the capability to think like human.

Artificial means something that is being created with the help of human skills.

Intelligence on the other hand can be said as a capacity of reasoning or learning. Bringing these two terms together, we can also say Artificial intelligence is implementation of human reasoning capability into a developed or designed system.

The Proposed project is a web application system that can handle the prediction of student report to help in knowing the right decision and step to take in helping students get good grades.

Education is an important sector in today's world. As it continues to grow in different ways, finding tools that provides best teaching practices has been a problem. This project aims at predicting student performance base on previous performances and factors surrounding the student. This can possibly help predict a student performance based on the features and factors available so as to provide means of better educational practices.

Applying artificial intelligence, the application will be able to learn and suggest possible solutions to aid good performance of students. The Artificial Intelligence capabilities of the proposed system will enable the system to be able to deduce historic trends of student grades and analyse students' performance with the aid of Machine Learning.

Machine learning is an aspect of artificial intelligence whereby a computer learns to perform certain tasks by studying examples of datasets after which it will carry out those task patterns on datasets it has never encountered. Such computations also involve Data Science. This is the science of processing and analysing data for the purpose of extracting a particular information. An example of an organisation that uses Data Science is Google. They monitor user clicks to enable them to improve the importance of their search engine and also in adverts.

Another aspect of Machine learning is to be able to predict a students' performance base on his/her grades. Although it might be wrong, it will predict and display from the list of datasets the one least amount of error. Because the use of this system might involve large datasets, data mining will help in studying the performance of the students.

1.2. Problem Definition

Manual data retrieval and computations can be exhausting. Using pen and paper or calculator to compute large datasets is time and energy consuming. Sometimes, during computations, errors are made, and you have to start calculations from the beginning. (LinkedIn, 2017) With manual systems, the level of service is dependent on individuals and this puts a requirement on management to run training continuously for staff to keep them motivated and to ensure they are following the correct procedures. It can be all too easy to accidentally switch details and end up with inconsistency in data entry or in hand written orders. This has the effect of not only causing problems with customer service but also making information unable be used for reporting or finding trends with data discovery.

The use of Machine learning to mine this data inputted in the system will reduce the amount of error and also save time.

This project revolves around having a computerized web application capable of providing student's report and analysis using their historic trends and activities.

1.3. Aims

This project aims to design and implement prediction analysis of student performance of an institution with the help of a web application for the interaction between the user and the application. It also aims at providing solutions to tackling the problem of student's school performance with the use of Decision algorithm.

1.4. Objectives

- Develop a web application that enables a school administrator to be able to login, insert student grades, and view their performance.
- Develop a database where data will be stored
- Develop an algorithm to learn the pattern of student grades
- visualization of results generated
- Testing of the whole system

1.5. Deliverables

- Web application
- Machine Learning algorithms for rating of student performance
- Visual presentation of Results

1.6. Evaluation

An evaluation will be carried out on the projects to test the output. The first test the system will undergo will be done by the developer to rectify software bugs and integration problems. The second is the user evaluation. The end-users will test for the user friendliness of the system. These tests will evaluate if the system achieved its intended aim, get feedback and improve on the development.

1.7. Required resources

- Python Development environment (IDE) – For writing of codes and debugging.
- Database – This will serve as a storage space for the data that will be inputted and gotten from the system.
- Server - A server to help in the communication of between the front end and the back end of the web application.

1.8. Conclusion

Having researched on the use of machine learning, one can deduce that it is of great importance in making critical decisions that involves a lot of data and high sensitivity. All these high intensity human processes can be handled by few lines of codes and results will be achieved in little or no time.

Chapter 2 – LITERATURE REVIEW

2.1. Introduction

This section discusses some of the important aspects to be introduced in the project. They form the foundation of understanding the tasks the system will carry out and how to carry them out. Today, a lot of students are exposed to technology and the internet. Some of these internet and technological activities are social media and games etc.

Students spend most of their time on the social media and tend to be eager for new technologies to emerge so as to keep them ahead of their peers. The world is surrounded by technology now that a lot of students sleep and wake up with their devices by their sides (Rosen, 2007). Also, they tend to prefer hanging out more with friends than reading or learning skills that are proven to be productive. The amount of time spent on the internet and hanging out with friends if channeled to education will go a long way to improve students' performance but instead, the education is seen as a barrier for their daily activities and so processing of information is hard (Rosen, 2007). Students tend to have only basic skills in the use of technology and internet but are never able to use that knowledge to solve problems (Kvavik, 2005).

Intake of substances such as alcohol also have negative impact on students. This has proven to be cause of failures in recent years as students get drunk from alcohol and misbehave in class (Lazear, 2001), getting exposed to different types of diseases (Miguel and Kremer, 2004) or dissemination of wrong information (Munshi and Myaux, 2006). Analysis carried out in large university of pairing students who don't drink alcohol with those who take to live together for a period of time, and it was discovered that those with no intake of alcohol that were paired with the ones that do take obtained low grades than those paired with non-drinkers (Kremer and Levy, 2008).

2.2. Web Application

Web application refers to any application which uses the web browser as its client. The application can be of any type, whether single or sophisticated. But as long as it is accessed using a browser, it is therefore a web application. These applications use the web browser to perform various tasks over the internet i.e. if the server and the client are not in the same place. This is because web applications use server-side scripts to handle storage and retrieval of information and client-side scripts to present the information retrieved to the user. Thereby allowing a communication between client and server through mediums of forms, content management systems, etc.

These web applications are usually coded in browser supported languages such as HTML, JavaScript etc. as they rely on the browser to render them executable. The web application will then require a server to manage requests from clients, an application server then performs the task which is being requested and many a times, a database is being developed to store information being exchanged between the client and the server,(Application and Application, 1999).

The web application to be developed will be more of an administrative system. Whereby, the user is the administrator, students don't have access to it.

2.3. Artificial Intelligence

All aspects of learning or features of intelligence can be accurately described in such a way machines can simulate. Attempts will be made in other for the machines to learn different languages(Tutorials Point, 2015), form abstractions and concepts, problems reserved for humans in other to improve themselves (John McCathy). He also mentioned Artificial Intelligence as science and engineering of making intelligent machines, especially intelligent computer programs. It is related to the similar task of using computers to

understand human intelligence, but AI does not have to confine itself to methods that are biologically observable.

Artificial Intelligence techniques enables quick execution of complex programs. Computer programs without AI give solutions to specific questions they are being asked while AI built programs solve generic it's meant to solve. Therefore, it is the ability of a system to comprehend, analyse and retrieve information (Tutorials Point, 2015).

2.4. Data Science

For every activity being carried out today, a lot of data is being generated, be it structured or unstructured. Data science has enabled us to move from an approach where we had reports that tells us about a situation and a picture of what the present condition looks like to a decision making and a predictive approach. so now, we are moving from an approach where we could see the actual/current status vs what actions we need to take forward and it can also help us in making predictions. Below are some examples of why data science is needed.

- Being able to understand the actual requirements of customers of a supermarket from the existing data like the customer's browsing history, purchase history, age and income. With the vast amount and variety of data, you can train models more effectively and recommend the product to your customers with more precision, (Hemant Shamar, 2015). This will go a long way in attracting more business to the Supermarket.
- One way of understanding the role of Data Science in decision making is self-driven cars. The self-driving cars collect live data from sensors, including radars, cameras and lasers to create a map of its surroundings. Based on this data, it takes decisions like when to speed up, when to speed down, when to overtake, where to take a turn.

making use of advanced machine learning algorithms (Hemant Shamar, 2015).

- Data Science can also be used for predictive analytics. Let's take weather forecasting as an example. Data from ships, aircrafts, radars, satellites can be collected and analysed to build models. These models will not only forecast the weather but also help in predicting the occurrence of any natural disasters. It will help in taking appropriate measures beforehand and save lives.

There are several domains in which Data science has been top listed and will continue to dominate as technology grows every day. Some of which are:

- I. Social Media
 - a. Digital marketing
 - b. Sentiment analysis
- II. Marketing
 - a. Upselling
 - b. Cross Selling
 - c. Churn
 - d. Predicting lifetime value of a customer
- III. Travel
 - a. Predicting flight delay
 - b. Pricing
- IV. Automation
 - a. Self-driving cars
 - b. Pilotless air crafts and drones
- V. Credit and Insurance
 - a. Claims prediction
 - b. Fraud and risk detection
- VI. Healthcare
 - a. Disease prediction
 - b. Medication effectiveness

VII. Sales

- a. Discount offering
- b. Demand forecasting

Data Science is the combination of tools, algorithms and machine learning principles with the aim of discovering hidden patterns from raw data. The difference between a data science and a data analysis is, data analysis explains what is happening by processing historic data while a data science uses advanced machine learning algorithms to predict the occurrence of a future event and also explains the processes involved. This means Data science is used in making predictions and decisions with the use of prescriptive analytics, predictive casual analytics and machine learning. The machine learning in this case is used for prediction and pattern discovery.

2.4.1. Prescriptive analytics: This is used to design a model which has the intelligence of making its own decisions and ability to modify it with certain given parameters. It also used to suggest actions and outcomes. An example of this is the self-driven car. The data being gathered from the vehicles can be used to train these self-driven cars. An algorithm will be will be run on the data gotten to make it intelligent. Through this, the car will be able to make decisions like when to turn and when to stop.

2.4.2. Predictive casual analytics: This is used to predict the possibility of a future event. For example, if I provide loan to people, the possibility of them making future payments is a huge concern to me, so I build a model that is capable of performing predictive analytics on the payment history of the customer to predict if their future payments will be on time or not.

2.4.3. Machine Learning (Predictions): This is used mostly in determining future trends. That is to say there is a data on which I can use to train my machines. For example, a

fraud detecting model to monitor fraudulent purchases in in a finance company.

2.4.4. Machine Learning (Pattern Discovery): This type of machine learning is used only if there are no parameters in which predictions can be made. The pattern discovery is used to discover hidden patterns within the dataset to be able to make meaningful predictions.

2.4.5. Business Intelligence (BI) Vs Data Science

Nevertheless, Data science is being misinterpreted to be the same as Business Intelligence (BI). Below are some contrasts to enable have a better understanding of the two.

Business Intelligence finds the hindsight and the insight to describe business trends and basically analysis of previous data. BI allows the collection of data from both internal and external sources. It enables the preparation of the data gotten, run queries on the data and then creates dashboard view of the results for analysis. It also denotes and evaluates the impact of an event.

Data science has to do with what type of data is available and how it is going to occur. It is more of a progressive approach with focus on analysing the current data and predicting future results with the aim of making informed decisions.

Contrast

Feature	Business Intelligence	Data Science
Data source	Structured data e.g. SQL	Structured and Unstructured e.g. logs, text etc.
Approach	Statistics and Visualization	Machine Learning, graph and Statistics
Focus	Past and Present results	Present and Future

	and outcomes	outcomes
Tools	Microsoft BI, R	Weka, R

Table 1. Difference between BI and DS

2.5. Machine Learning

Machine learning teaches computer systems and programs the ability to carry out tasks naturally known to be carried out by humans. Humans and animals learn from experience. Machine learning algorithms use computational methods to gain knowledge from data without relying on a predetermined condition and instruction. The amount of testing data available, determines the algorithms capability to improve their performance of learning (Harrison, 2011).

Due to the rise of Big data, machine learning has been an important tool of solving problems in areas of:

- Natural language processing
- Image processing
- Energy production
- Computational biology etc.

Machine learning algorithms find natural patterns in data to help make better decisions and predictions

2.6. Data mining

Data mining is a technique used in finding patterns from large amounts of information (Ali and Tuteja, 2014). The use of data mining has been helpful in so many areas, such as Health, education (Harwati, Alfiani and Wulandari, 2015) Education is a very important factor needed for a country to develop

(Gaviria, 2002), and the main aim of education institutions is to provide quality education for its students.

One way to provide a high quality educational system, is by being able to predict students' performance and taking actions where necessary to improve the students' performance. This is to show the capabilities of data mining techniques in an institution by providing a data mining model and use of classification to evaluate the students' performance (Baradwaj and Pal, 2012).

The introduction and application of Data mining in an educational system is called Educational Data Mining (EDM). The amount of data being generated today in an educational system are increasing day by day and so, getting direct feed-back can be difficult. Having a predictive system will help in figuring out the students' weak points and make amends.

2.6.1. Classification method

The objective of classification is to assign a predefined class based on a set of known attributes. A feature classification model is that it is being built by part of the training set to be used in training the model. After the model is built, it is then now used in assigning label to new records to the unknown class attributes. To build the models, the Decision Tree (DT) technique is one algorithm which can be used for such classifications amidst many. (Hamasa, Indiradevi and Kizhakkethottam, 2016).

2.6.2. Academic performance prediction

Students' academic information and records are the data being used for prediction (Adelman 1999). The academic information includes certain marks and points needed to be able to proceed further in the institution while academic records include assignment marks, attendance and internal marks. These internal marks are calculated in average using the marks generated from first and second session. From the predictions, students who are at risk and not, to be able to pass will be identified, (Swamy and Hanumanthappa,

2012). The results gotten from the prediction will enable the institution to improve their teaching standards if need be. This will also help in providing a student's expertise for the purpose of employment, (Hamasa, Indiradevi and Kizhakkethottam, 2016).

There are various factors which determine students' academic performance

2.6.3. data preparation

Before any test is run on a data instance, the data has to be cleaned and prepared before uploading it into a system for mining purposes (Ramesh, Parkavi and Ramar, 2013). This is known as data pre-processing.

Chapter 3 - System requirements

3.1. Introduction

This section discusses the requirement specifications needed for the application. Since the system is a software application, the computer will be the hardware requirement needed for the application to run.

3.2. User Characteristics

The administrator is the only person that almost all the rights to access the system. He is the one who has all rights to view the Student and Tutor details and also to modify those details. The administrator also keeps a track of the activity schedules of both the Tutor and the Student. The administrator nevertheless also has his/her limitations i.e. being unable to change a student's grades, (Amit Gandhi).

3.3. Functional Requirements

Functional requirements are those features in which the system is going to provide. Therefore, the system needs to be able to meet the following:

- Read the dataset paths from a folder path.
- Load the dataset into the algorithm operations.
- Provide a list of predictions

3.4. Software Requirements

For the application to be able to meet up with the above-mentioned requirements, some software requirements have to be met:

- Scikit-learn – this is used to implement the machine learning classifier.
- Python – the system is built on python and therefore, runs on a python programming language interpreter. This is being elaborated more in section 5 along with other modules involved.

Chapter 4 – Analysis and Design

4.1. Introduction

A system is not said to be complete if proper analysis is not done to enhance or make corrections based on previous research and development of the system. This section of the report discusses the logical design of the entire application. In section 5, the system or application will be discussed in depth. This project proposes to use SVM classifier, decision tree to and a web application for the predictive analysis of students' performance. This process will be carried out in the following stages; The Training stage, Testing stage and the web application (Fig 5 and 6) to help communicate with the python programming language.

4.2. current system

The current system for the Student Management System deals with maintaining student's grades and other academic activities.

According to the current system, the tutors have to fill in student grades manually, calculate student grades manually and unable to detect when and where a student needs help regarding his / her academic career. The admin also has to manage all the users. He needs to maintain records of all the users, their activity status, submission methods and installation details on paper. The Manual process is more error prone and also slow. Moreover, Students in the academy can interface his/her work area only. But if a predictive application is available then they can communicate whole system. Thus, a simulation of this entire process can be a boon to the Tutors as well as the admin.

4.3. weaknesses of current system

- The present system has certain major disadvantages. A few to be listed can be excessive paperwork, time consuming process flow,

laborious work environment for the school administration and Tutors, difficulty to access historical data and all these problems lead to inefficient working satisfaction of the educational sector causing dissatisfaction in the general public.

- There is lack of transparency in the current system. This being a major drawback in the system needs special attention.
- The problem stated above have certain problems like time consuming process flow for which the Institution may need to change the structure of the process or flow of data integration in certain cases so that the system output can become faster.
- The following listed are the problems or weaknesses of the current system:
 - So much time consume in preparing registers which is having replicated data
 - It is difficult to prepare report for decision making.
 - Attendance related module is not there.

4.4. data modelling and design

- **Class Diagram**

A class diagram describes the static structure of a system. It shows how a system is structured rather than how it behaves. The static structure of a system consists of a number of class diagrams and their dependencies. The main constituents of a class diagram are classes and their relationships: generalization, aggregation, association, and the different kinds of dependencies.

Following diagram represents various classes of the system. The relations between these classes are shown in the next diagram.

System design is the first stage in which the basic approach to solving the problem is selected. During the system designing stage, the overall structure and style of the system are decided. The system architecture is the

overall organization of the system into several components called system. System design deals with transforming the user requirements into a form that is implementable using the programming language. Certain items such as modules, relationships among identified modules, data structures, relationships between the data structures, and algorithms for implementation should be designed during this phase.

As the system designer I made the following design decisions:

- Organize the system into modules
- Organize sub-modules for each module
- Allocate tasks to processors
- Choose an approach to manage data store
- Handle access to global resources
- Choose implementation logic

4.5. Database Design

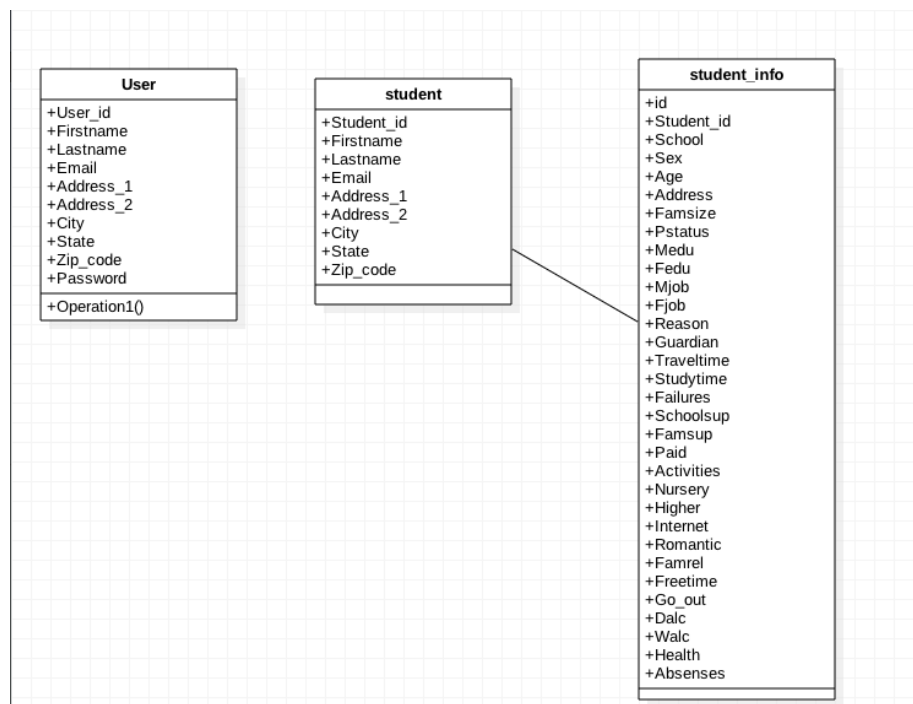


fig 1. Class diagram

4.6. Use Case Diagram

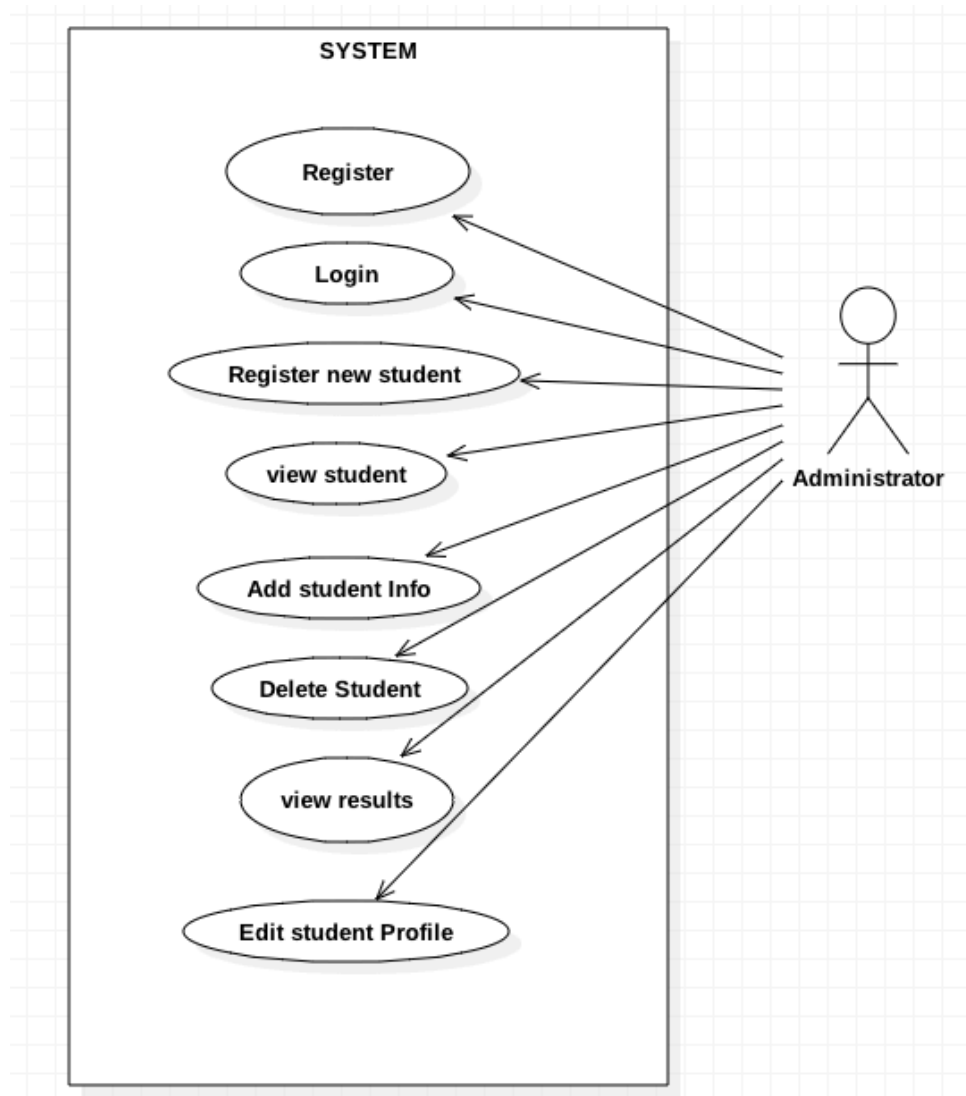


fig 2. Use case

4.7. Activity Diagram

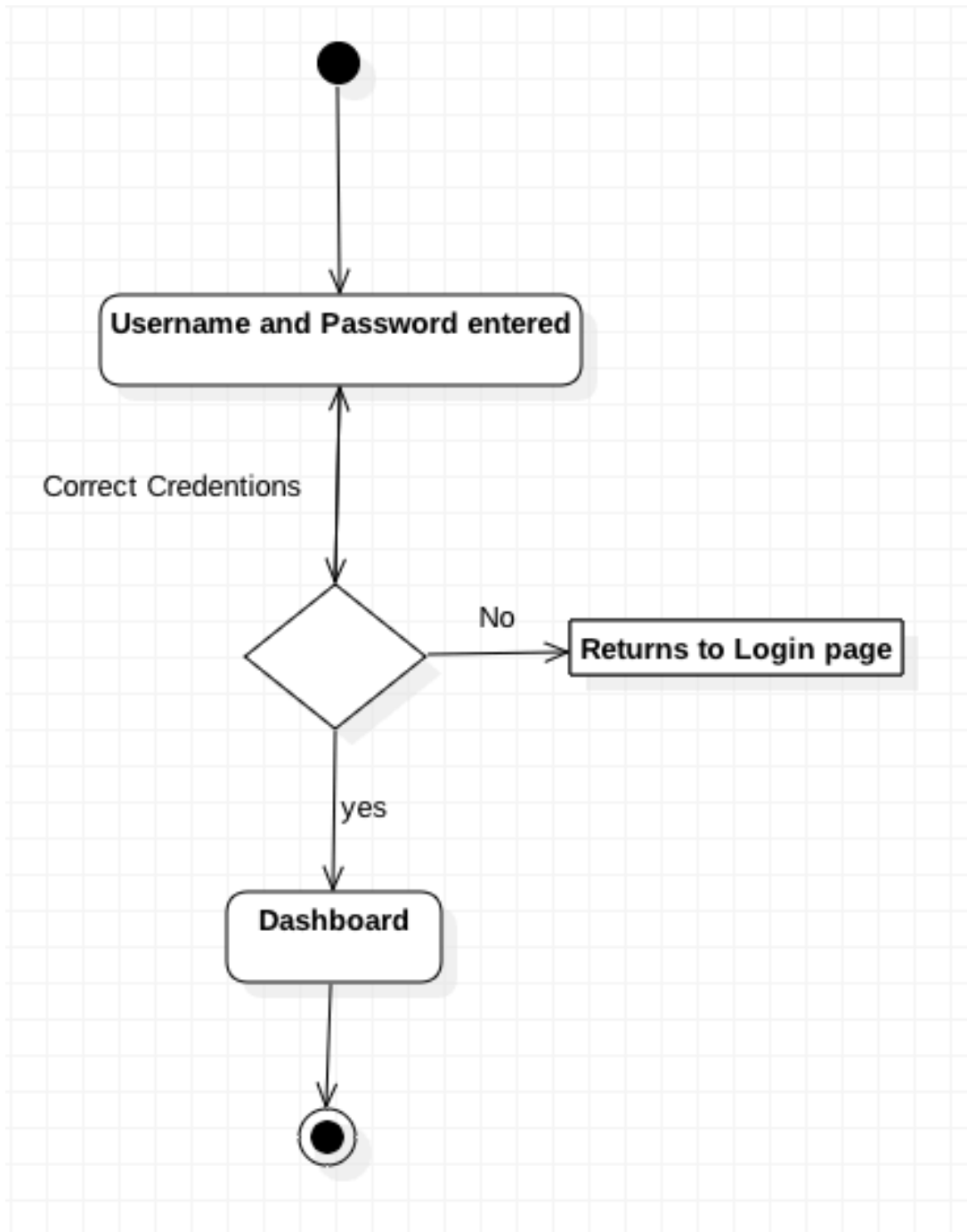


fig 3. Activity diagram

4.8. Flow chart

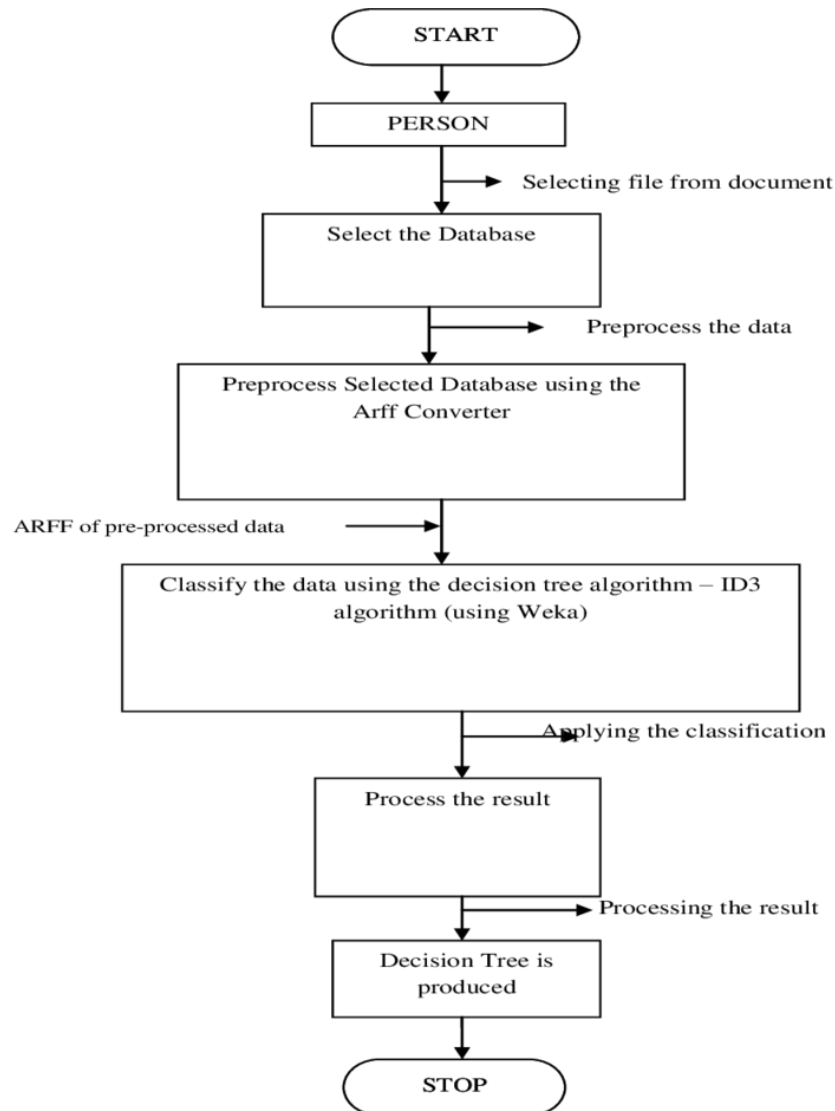


fig 4. Flow chart

The flow chart above describes how data is being collected and processed to form a decision tree for the prediction of our results. The classification is not restricted to Arff converter and Weka. It is subject to change during the process of development.

The diagrams bellow depicts the structure of how the web application is being laid out. The diagram below (fig 5) is the page that shows immediately the application starts.

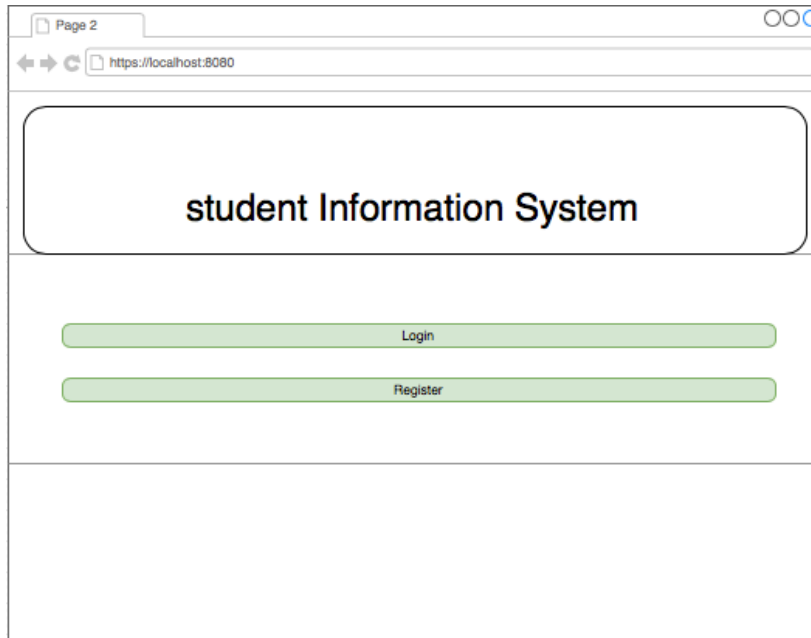


Fig 5. Login and register page

This diagram (fig 6) shows the logical dashboard structure of the system when the user logs in.

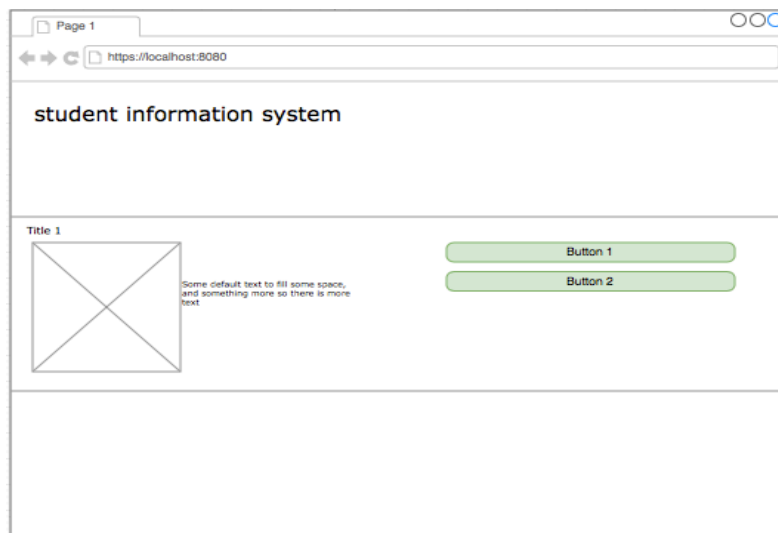


Fig 6 Dashboard

Chapter 5 - Implementation

5.1. Introduction

This section of the report discusses the implementation of the system. It covers the tools components of the system and the various elements of the technologies being used. This chapter will be discussing the algorithms which are being used for our predictive analysis.

5.2. Data Gathering

The data being used for this project is gotten from UCI machine learning repository. The dataset stores information about a secondary school of two Portuguese colleges. The dataset consists of 0-396 instances and 31 attributes which are being discussed in detail below.

Attribute Information:

school - student's school (binary: 'GP' - Gabriel Pereira or 'MS' - Mousinho da Silveira)

sex - student's sex (binary: 'F' - female or 'M' - male)

age - student's age (numeric: from 15 to 22)

address - student's home address type (binary: 'U' - urban or 'R' - rural)

famsize - family size (binary: 'LE3' - less or equal to 3 or 'GT3' - greater than 3)

Pstatus - parent's cohabitation status (binary: 'T' - living together or 'A' - apart)

Medu - mother's education (numeric: 0 - none, 1 - primary education (4th grade), 2 - 5th to 9th grade, 3 - secondary education or 4 - higher education)

Fedu - father's education (numeric: 0 - none, 1 - primary education (4th grade), 2 - 5th to 9th grade, 3 - secondary education or 4 - higher education)

Mjob - mother's job (nominal: 'teacher', 'health' care related, civil 'services'

(e.g. administrative or police), 'at_home' or 'other')

Fjob - father's job (nominal: 'teacher', 'health' care related, civil 'services' (e.g. administrative or police), 'at_home' or 'other')

reason - reason to choose this school (nominal: close to 'home', school 'reputation', 'course' preference or 'other')

guardian - student's guardian (nominal: 'mother', 'father' or 'other')

traveltime - home to school travel time (numeric: 1 - <15 min., 2 - 15 to 30 min., 3 - 30 min. to 1 hour, or 4 - >1 hour)

studytime - weekly study time (numeric: 1 - <2 hours, 2 - 2 to 5 hours, 3 - 5 to 10 hours, or 4 - >10 hours)

failures - number of past class failures (numeric: n if $1 \leq n < 3$, else 4)

schoolsup - extra educational support (binary: yes or no)

famsup - family educational support (binary: yes or no)

paid - extra paid classes within the course subject (Math or Portuguese) (binary: yes or no)

activities - extra-curricular activities (binary: yes or no)

nursery - attended nursery school (binary: yes or no)

higher - wants to take higher education (binary: yes or no)

internet - Internet access at home (binary: yes or no)

romantic - with a romantic relationship (binary: yes or no)

famrel - quality of family relationships (numeric: from 1 - very bad to 5 - excellent)

freetime - free time after school (numeric: from 1 - very low to 5 - very high)

goout - going out with friends (numeric: from 1 - very low to 5 - very high)

Dalc - workday alcohol consumption (numeric: from 1 - very low to 5 - very high)

Walc - weekend alcohol consumption (numeric: from 1 - very low to 5 - very high)

health - current health status (numeric: from 1 - very bad to 5 - very good)

absences - number of school absences (numeric: from 0 to 93)

passed – did student pass (binary: yes or no)

G1- first period grade (numeric: 0 - 93)

G2- second period grade (numeric: 0 - 93)

G3- final grade (numeric: 0 - 93)

The attributes listed above are demographics of the students from the two colleges. Instead of using tests and exams, another dimension is being taken to determine the factors that will affect a student's final grade.

In Portugal, the secondary education system entails of 3 years of schooling, preceding 9 years of basic education followed by higher education. Most students are said to enrol in the public and free education system. There are several courses (e.g. Sciences and Technologies, Arts etc.) that share core subjects such as the Portuguese Language and Mathematics. Like several other countries (e.g. France), a 20-point grading scale is used, where 0 is the lowest grade and 20 is the perfect score. During the school year, students are evaluated in three periods and the last evaluation (G3 of Table 1) corresponds to the final grade. For this project, the G1-G3 will not be used as we will try to predict the students' performances using their demographics.

5.3. Algorithm

5.3.1. Introduction

There are different algorithms available for the predictive analysis of our data and results. For this project, decision tree DT and support vector machines SVM are being used. The choices made at the end of an algorithm depends on the structure of data studied and the final selection criteria is the prediction performance obtained by the model.

5.3.2. Decision Tree (DT)

Decision tree is one of the algorithm used in data mining. It is a tool that handles continuous and non-continuous variables it creates principles by establishing a tree like structured diagram (Chen et al, 2014). It is an algorithm that contains conditional controlled statements that helps in identifying strategies of reaching a desired goal. Decision trees consists of nodes which are categorized into three (Kaminski et al, 2017), namely Decision nodes, Chance nodes and End nodes. Being used as a descriptive tool for calculating conditional probabilities, DT helps in gaining good and accurate predictive analysis being that we are using the student demographics instead of their exams or test scores. It can be used to approximate the output of a function based on its value and slope into decision rules.

The use of decision tree in this project provides a better understanding of the various demographic factors that affects students' grades. It then takes these results and presents it visually on the web application for the user to be able to make decisions appropriately.

5.3.3. Support Vector Machine (SVM)

Support Vector Machines are algorithms that are used for classification (Cristianni & Shawe-Taylor, 2000; Schölkopf & Smola, 2002). An advantage of using SVM's is that they provide unique and optimal solutions when the model parameters are fixed and also, a good performance classification is obtained even with small datasets. It has been reported that classification models obtained using this technique are robust and less subject to the curse of dimensionality and over-fitting (Devos, Downey & Duponchel, 2014). The objective of using this algorithm is to predict the percentage of students that passed in the two colleges. The idea behind the algorithm is to find the optimal hyperplane separating the data. And this is being done by using the formula (Eq. 1):

$$w \cdot X_i + b = 0$$

Where X_i is the i th object of the training set. W , the normal vector to the hyperplane and b , the offset. All these are calculated during the training. The SVM will try to find the hyperplane (Fig.1) that minimises the classification error while maximising the shortest distance from the chosen hyperplane to the nearest training samples of each data. The constrained quadratic optimisation problem is being used for the optimisation of the margin (Eq. 2)

$$\begin{aligned} &\text{minimise } (\|w\|^2 + C \sum_{i=1} \zeta_i) \\ &\text{subject to } \zeta_i + y_i(w \cdot X_i + b) - 1 \geq 0 \text{ with } \zeta_i \geq 0 \end{aligned}$$

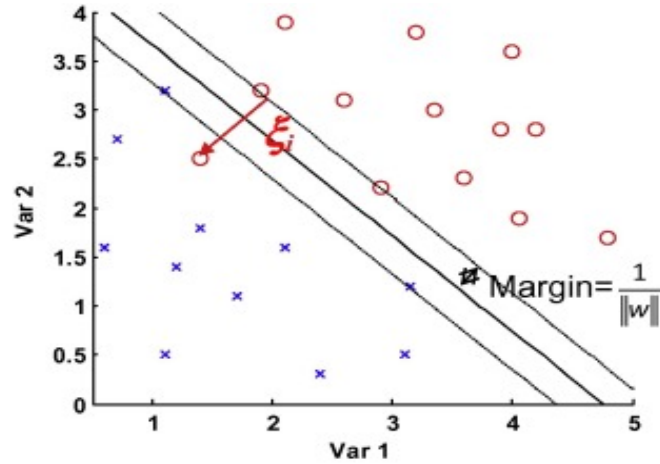


Fig 7. SVM Hyperplane (*Cristianni & Shawe-Taylor, 2000; Schölkopf & Smola, 2002*)

The classification error is represented by distance (ξ_i), of the misclassified sample(i) and the corresponding margin hyperplane. The regularisation meta-parameter (C), controls the trade-off between the two conflicting objectives: when C is small, margin maximisation is emphasised whereas when C is large, error minimisation is predominant.

5.4. Technologies

The technologies here refer to the various tools being used in the project for the development and design of the application.

5.4.1. Python

Python is one of the one of the high-level-programming languages being used all over the world today. It emphasizes on code readability due to the design philosophy it has. The constructs provided by python programming language provides clear programming in large and small scale (Kulman 2012). Python is an Open-source Software (OSS) made available with a license where the developer of the technology gave rights for the software to be used, changed, studied and distributed to anyone for any purpose (St. Laurent & Andrew, 2008). This technology also has a lot of libraries that are being used for the purpose of different types of development globally.

The Use of Python Programming language in this project is the fact that it contains a large comprehensive library for artificial intelligence and machine learning development purposes and can be connected to other technologies like HTML, JAVA, etc. and features automatic memory management and supports multiple programming paradigms such as procedural, imperative and object-oriented (Python Software Foundation, 2012).

In order for the system to communication with the various technologies being implemented, there are couple of dependencies and libraries that need to be installed on the system before the application can run successfully. Or so, these dependencies are being installed in the application folder so that the user will not go through a whole process of downloading them from scratch. All the user needs to do is to open the CMD terminal from the project folder and update the dependencies. Instructions on how to install and update the dependencies will be attached to the application.

For this project, the GUI cannot be executed without the python scripts. This means the html pages are dependent on some python scripts written to activate or call them. They are as follows:

```
# runs the entire application
app = create_app()
options = {"use_reloader":True, "use_debugger":True}
if __name__ == '__main__':
    run_simple(hostname="127.0.0.1", port=8080, application=app, **options)
```

Fig 8: Script that initiates the application.

The codes shown in the above image script initiates the whole application. Without these lines of codes, the application will not run. This last line shows an instruction for the server to run and when it executes, it should start up our application.

5.4.2. HTML

The HTML aspect of the application helps us interact with the application. It is the Graphical User Interface (GUI) for interaction between the user and the system. It is also used to define the structure of our web application and helps us get feedback as to whether our application is producing the right results or not.

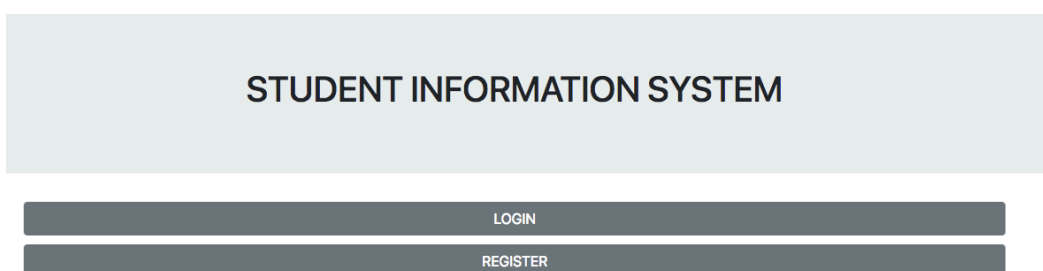
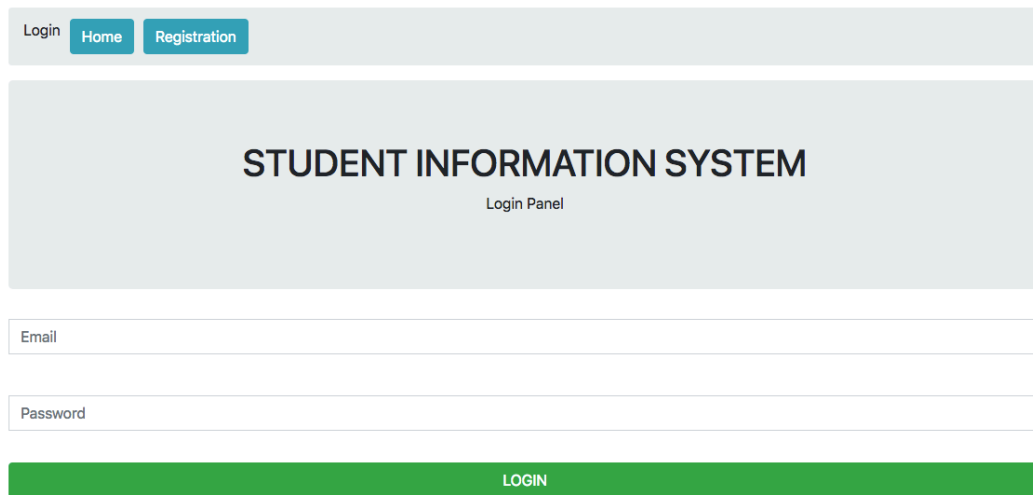


Fig 9: index page of the web application

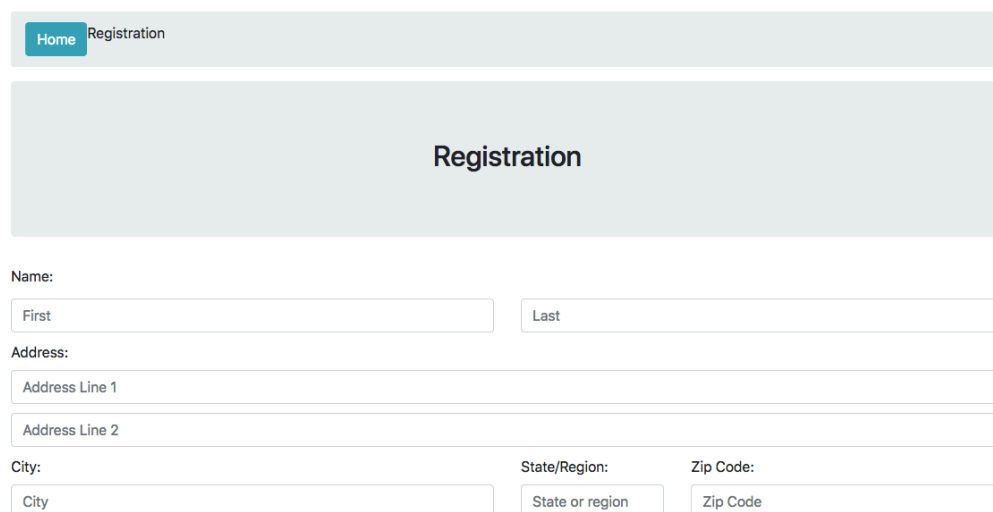
The image above shows the index page of the student management web application. This is the first page the user sees when they start the application. At this stage a user registers into the system if not registered.



The login page features a navigation bar with 'Login', 'Home', and 'Registration' buttons. The main heading is 'STUDENT INFORMATION SYSTEM' with a sub-heading 'Login Panel'. Below these are input fields for 'Email' and 'Password', followed by a green 'LOGIN' button.

Fig 10: Login page

The diagram above shows the login page where the user enters their email and password in order to get access to the application.



The registration page has a navigation bar with 'Home' and 'Registration' buttons. The main heading is 'Registration'. The form includes input fields for 'Name' (First and Last), 'Address' (Address Line 1 and Address Line 2), 'City', 'State/Region', and 'Zip Code'.

Fig 11: Registration page

The diagram (Fig 10) above shows the registration page of the system. Whoever is going to use the application must register first before having access.

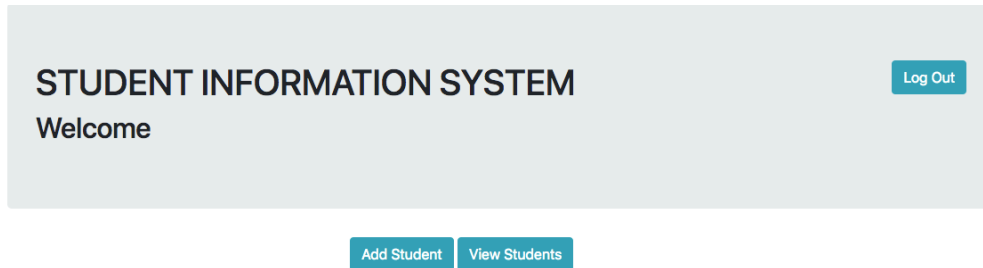


Fig 12: Dashboard

The diagram above shows the user dashboard. Here, the admin can add students and view students currently are registered in the system.

5.4.3. Sqlalchemy

Sqlalchemy is a toolkit mostly used in the python environment. It is an object relational mapper (ORM) that has the power and flexibility of SQL (Sqlalchemy). The ORM in Sqlalchemy provides mapper patterns, where classes created are mapped to the database in open ended, multiple ways - allowing the object model and database schema to develop in a clean way from the beginning. the Sqlalchemy makes available tools and components for assisting in database development at different level. It provides a consistent façade over python DBAPI. Sqlalchemy brings the usage of different databases and interfaces as consistent as possible. What this means is, it searches for a middle ground between all these database tools being used and finds a consistent feature and uses it. At the same time, it doesn't take away the advantages in which those other tools provide.

5.4.4. Flask

The flask technology is a tool that is dependent on Sqlalchemy. This contains a declarative extension which allows declaration of models. These models contain classes being set by the user of the carious tables to be created in the database (Sqlalchemy). The models for the application are:

```

20 #creates a user table
21 class User(db.Model, CRUD):
22     __tablename__ = 'users'
23     id = db.Column(db.Integer, primary_key=True)
24     firstname = db.Column(db.String(64))
25     lastname = db.Column(db.String(64))
26     email = db.Column(db.String(100), unique=True, index=True)
27     address_1 = db.Column(db.String(200))
28     address_2 = db.Column(db.String(200))
29     city = db.Column(db.String(200))
30     state = db.Column(db.String(200))
31     zip_code = db.Column(db.String(200))
32     password_hash = db.Column(db.String(128))
33     date_created = db.Column(db.DateTime(), default=datetime.utcnow)
34     last_seen = db.Column(db.DateTime(), default=datetime.utcnow)
35

```

Fig 13: user model class

This model creates a user table in our database.

```

60 #creates a student table
61 class Student(db.Model, CRUD):
62     __tablename__ = 'student'
63     id = db.Column(db.Integer, primary_key=True)
64     firstname = db.Column(db.String(64))
65     lastname = db.Column(db.String(64))
66     email = db.Column(db.String(100), unique=True, index=True)
67     address_1 = db.Column(db.String(200))
68     address_2 = db.Column(db.String(200))
69     city = db.Column(db.String(200))
70     state = db.Column(db.String(200))
71     zip_code = db.Column(db.String(200))
72     date_created = db.Column(db.DateTime(), default=datetime.utcnow)
73     student_profile = db.relationship('StudentProfile', backref="student", cascade="all, delete-orphan",
74

```

Fig 14: student model class

This creates a student table that stores a student's basic information e.g. name, email address etc.

```

81  #creates a student profile table
82  class StudentProfile(db.Model, CRUD):
83      __tablename__ = 'studentprofile'
84      id = db.Column(db.Integer, primary_key=True)
85      student_id = db.Column(db.Integer, db.ForeignKey('student.id'))
86      school = db.Column(db.String(64))
87      sex = db.Column(db.String(64))
88      age = db.Column(db.String(64))
89      address = db.Column(db.String(225))
90      famsize = db.Column(db.String(64))
91      pstatus = db.Column(db.String(64))
92      medu = db.Column(db.String(64))
93      fedu = db.Column(db.String(64))
94      mjob = db.Column(db.String(64))
95      fjob = db.Column(db.String(64))
96      reason = db.Column(db.String(64))
97      guardian = db.Column(db.String(64))
98      traveltime = db.Column(db.String(64))
99      studytime = db.Column(db.String(64))
100     failures = db.Column(db.String(64))
101     schoolsup = db.Column(db.String(64))
102     famsup = db.Column(db.String(64))
103     paid = db.Column(db.String(64))
104     activities = db.Column(db.String(64))
105     nursery = db.Column(db.String(64))
106     higher = db.Column(db.String(64))
107     internet = db.Column(db.String(64))
108     romantic = db.Column(db.String(64))
109     famrel = db.Column(db.String(64))
110     freetime = db.Column(db.String(64))
111     goout = db.Column(db.String(64))
112     dalc = db.Column(db.String(64))
113     walc = db.Column(db.String(64))
114     health = db.Column(db.String(64))
115     absences = db.Column(db.String(64))

```

Fig 15: student educational information Model class

5.4.5. Bootstrap

The bootstrap technology is used to add fancy designs to the web page. It comes with embedded CSS, JavaScript and jQuery libraries which can be activated by calling out their classes and ID's.

Chapter 6 - Results

6.1. Introduction

This section shows and discusses the results gotten from using the data inputted into the application with representations

6.2. Predictive Analysis

After the data was inputted into the algorithm, the results and percentage accuracy of students that passed are seen as follows:

```

/Users/nicholas/Desktop/StudentS/env/bin/python /Users/nicholas/Desktop/StudentS/src/classifier.py
[INPUT REQUIRED] Please input a value for the cost function (decimal): 0.001

=====
[OPTION] SHOW PREDICTIONS [y]/n ? : y
[INFO] predicted [ yes ]...expected...[ no ]
[INFO] predicted [ yes ]...expected...[ yes ]
[INFO] predicted [ no ]...expected...[ no ]
[INFO] predicted [ yes ]...expected...[ yes ]
[INFO] predicted [ yes ]...expected...[ yes ]
[INFO] predicted [ yes ]...expected...[ yes ]
[INFO] predicted [ yes ]...expected...[ no ]
[INFO] predicted [ yes ]...expected...[ yes ]
[INFO] predicted [ yes ]...expected...[ yes ]
[INFO] predicted [ yes ]...expected...[ no ]
[INFO] predicted [ yes ]...expected...[ yes ]
[INFO] predicted [ yes ]...expected...[ yes ]
[INFO] predicted [ yes ]...expected...[ no ]
[INFO] predicted [ yes ]...expected...[ no ]
[INFO] predicted [ yes ]...expected...[ yes ]
[INFO] predicted [ yes ]...expected...[ yes ]
[INFO] predicted [ yes ]...expected...[ yes ]
[INFO] predicted [ yes ]...expected...[ yes ]
[INFO] predicted [ yes ]...expected...[ no ]
[INFO] predicted [ yes ]...expected...[ yes ]
[INFO] predicted [ yes ]...expected...[ yes ]
[INFO] predicted [ yes ]...expected...[ yes ]
[INFO] predicted [ yes ]...expected...[ yes ]
[INFO] predicted [ yes ]...expected...[ yes ]
[INFO] predicted [ yes ]...expected...[ yes ]
[INFO] predicted [ yes ]...expected...[ yes ]
[INFO] predicted [ yes ]...expected...[ yes ]
[INFO] predicted [ yes ]...expected...[ no ]
[INFO] predicted [ yes ]...expected...[ yes ]
[INFO] predicted [ yes ]...expected...[ no ]
[INFO] predicted [ yes ]...expected...[ yes ]
[INFO] predicted [ no ]...expected...[ yes ]
[INFO] predicted [ yes ]...expected...[ yes ]
[INFO] predicted [ yes ]...expected...[ yes ]
[INFO] predicted [ yes ]...expected...[ yes ]
[INFO] predicted [ yes ]...expected...[ yes ]

```

Fig 16.1 svm algorithm running


```

1 alembic==0.9.8
2 click==6.7
3 Flask==0.12.2
4 Flask-Migrate==2.1.1
5 Flask-Script==2.0.6
6 Flask-SQLAlchemy==2.3.2
7 itsdangerous==0.24
8 Jinja2==2.10
9 Mako==1.0.7
10 MarkupSafe==1.0
11 marshmallow==2.15.0
12 numpy==1.14.2
13 python-dateutil==2.7.0
14 python-editor==1.0.3
15 scikit-learn==0.19.1
16 six==1.11.0
17 sklearn==0.0
18 SQLAlchemy==1.2.5
19 Werkzeug==0.14.1
20

```

Fig 17. List of required python modules

The results gotten the machine learning classifier needs the above-mentioned python modules to be instantiated before it can carry out its task. These are some of the Libraries contained within the python programming language that reduces the amount of work done by simply calling them up using the word “Import”.

```

cost = float(input("[INPUT REQUIRED] Please input a value for the cost function (decimal):"))
model = LinearSVC(C=cost, penalty='l2', loss="squared_hinge")
model.fit(data_train, label_train)
return model

```

Fig 18. Cost function script

The cost function script is being used to measure how badly models are performing. This process can be done by running the model repeatedly to

compare the predictions (Cornor 2017). This is being carried out because there are instances where the hyperplane isn't straight, and the idea is to find the point that is more closest to the hyperplane with lowest error rate to produce a high accuracy of prediction. There are instances where these operations are done manually (small scale dataset) or automatically (large dataset) and the one with the highest accuracy of prediction will be sorted out and used as model for Testing the remaining dataset

Chapter 7 - Evaluation and Testing

7.1. Evaluation

7.1.1. Introduction

The Purpose of Evaluation is check the develop system's results. This is an important aspect of the development as it helps in making decisions based on the outcome of the tests carried out on the system. This comes handy in regard to future work and re-designing of the system. For the Evaluation, SWOT analysis is being used. This is a tool used for the development of strategies and measurement of the strengths, weaknesses, opportunities and threats of the system.

7.1.2. Strengths

- The system is easy to use. No skill is required before the user can start using it.
- The application/system provides different features which integrates across other applications
- It creates a platform for addition features
- It is scalable

7.1.3. Weaknesses

- It requires update whenever new models are added.
- Can occupy a lot of space as it is being used everyday

7.1.4. Opportunities

- Improvement in the application layout
- Implement Tutors administrative pages

7.1.5. Threats

- The web application might be prone to external hack attempts.
- Issues with database will require the restart of web application and re-installation of some dependencies.

7.2. Testing

On completion of the application, a complete system testing as carried out. Below are the test cases used.

Test	Sign up
Description	User will try to sign up from the sign-up page using valid credentials
Expected Result	Sign up successful
Outcome	User account created
Pass or Fail	Pass

Table 2-Test Case 1

Test	Log in
Description	User will try and log into the system using credentials provided during registration
Expected Result	Logs into dashboard
Outcome	Login successful
Pass or Fail	pass

Table 3-Test Case 2

Test	View student
Description	User will try to view list of students currently registered into the system
Expected Result	List of students is shown
Outcome	List of student shows
Pass or Fail	pass

Table 4-Test Case 3

Test	Add student
Description	User will try to add new student into the system
Expected Result	Student added successfully
Outcome	Student adding was successful
Pass or Fail	pass

Table 5-Test Case 4

Test	Implement SVM (support vector machine) algorithm
Description	SVM should be implemented in the application
Expected Result	SVM classifier codes
Outcome	Verified SVM classifier
Pass or Fail	Pass

Table 6 -Test Case 5

Test	Run SVM
Description	SVM should run and predict percentage pass
Expected Result	SVM to run
Outcome	SVM was initiated successfully
Pass or Fail	pass

Table 7 -Test Case 6

Test	Implement DT (decision tree) algorithm
Description	DT should be implemented in the application
Expected Result	DT classifier codes
Outcome	DT classifier codes not available
Pass or Fail	Fail

Table 8 -Test Case 7

Test	Run DT
Description	DT should run and predict percentage pass
Expected Result	DT to run
Outcome	SVM wasn't initiated
Pass or Fail	Fail

Table 9 -Test Case 8

Test	Run entire application
Description	Entire application should be able to start
Expected Result	Web application
Outcome	Well application started successfully
Pass or Fail	Pass

Table 10 -Test Case 9

Test	Create class Models
Description	Classes should be able to be created in our models
Expected Result	Created classes
Outcome	Classes created successfully
Pass or Fail	Pass

Table 11- -Test Case 10

Chapter 8 - Conclusion and Future Work

8.1. Introduction

This section of the report summarizes the thesis and proposes future improvement of the web application developed.

8.2. Review

Throughout the course of this dissertation, it can be said that web applications play a vital role in software application development, machine learning and data classification. Getting data for classification and prediction for analysis and pattern recognition purposes can be sometimes tedious. Introduction of student web application provides a friendly interface and environment where data can be collected and sent into a classification model for analysis and prediction.

8.3. Future work

A lot of research is still being carried out pertaining student's behavior and performance. For the future redevelopment of this project, I will suggest the introduction of tools that provides a better understanding of how a student will perform through their interaction in class and with their peers.

8.4. Conclusion

Software engineering development when carried out by multiple individuals produces more accurate results and calculations through proper planning and adequate use of resources could go a long way in achieving part of the future work suggested.

REFERENCE/BIBLIOGRAPHY

- Asif, R., Merceron, A., Ali, S. & Haider, N. 2017, "Analysing undergraduate students' performance using educational data mining", *COMPUTERS & EDUCATION*, vol. 113, pp. 177-194.
- Adelman, C. 1999. Answers in the Toolbox. Academic Intensity, Attendance Patterns and bachelor's Degree Attainment.
- Artificial intelligence (AI) 2017. *Britannica Academic*. Retrieved 26 October 2017, from <http://academic.eb.com.ezproxy.mdx.ac.uk/levels/collegiate/article/artificial-intelligence/9711>
- Amos Ndegwa 2016 <https://www.maxcdn.com/one/visual-glossary/web-application> Khosla, V. (2015) 'Predicting a Student ' s Performance', pp. 1–6.
- [Chen, Suduan](#); [Yeong-Jia, James Goo](#); [Zone-De Shen](#), (2014). *The Scientific World Journal*.
- Conor McDonald, 2017. Machine Learning Fundamentals (I). Cost functions and gradient decent.
- Cristianni and Shawe-Taylor, 2000 N. Cristianni, J. Shawe-Taylor Support vector machines and other kernel-based learning methods Cambridge University Press (2000)
- Devos, O., Downey, G. & Duponchel, L. 2014, "Simultaneous data pre-processing and SVM classification model selection based on a parallel genetic algorithm applied to spectroscopic data of olive oils", *Food chemistry*, vol. 148, pp. 124-130.
- Hemant Sharma. 2017, https://www.edureka.co/blog/what-is-data-science/?utm_source=youtube&utm_medium=youtube-description&utm_campaign=youtube-data-science
- <https://www.sqlalchemy.org/>
- <https://archive.ics.uci.edu/ml/datasets/student+performance>
- <https://www.python.org/about/>
- <http://flask-sqlalchemy.pocoo.org/2.3/>

- <https://docs.djangoproject.com/en/2.0/topics/db/models/>
- <http://www.diveintopython3.net/serializing.html>
- <https://www.sqlalchemy.org/>
- https://www.researchgate.net/profile/Marcel_Mizzi2/publication/280577231_Point_of_Sale_in_The_Cloud/links/55bb900508ae092e9660ab1a.pdf
- https://www.researchgate.net/profile/Kadaruddin_Aiyub
- <https://link.springer.com/content/pdf/10.1007/s10462-018-9620-8.pdf>
- <http://digitalcommons.unl.edu/cgi/viewcontent.cgi?article=2908&context=libphilprac>
- <http://www.skepticalscience.com/climate-models.htm>
- <http://spotidoc.com/doc/276059/developing-methods-for-machine-learning-algorithms-hala-h...>
- <https://ecommons.usask.ca/bitstream/handle/10388/7283/MANGLAI-THESIS.pdf?sequence=1&isAllowed=y>
- <https://www.scribd.com/doc/96668480/Leadership-Theories>
- <http://gradschool.unc.edu/academics/thesis-diss/guide/ordercomponents.html>
- <https://www.quora.com/What-is-the-difference-between-Data-Analytics-Data-Analysis-Data-Mining-Data-Science-Machine-Learning-and-Big-Data-1>
- <https://www.scribd.com/doc/53949573/Project-Report-Format-for-Final-MCA>
- <https://www.scribd.com/document/45159015/Global-and-Transnational-Strategy-Walmart-Full-Report-1>
- <https://www.scribd.com/document/166495414/Jawad-Nagi>
- http://flipkarma.com/media_dir/main_documents/Major%2520Final.pdf
- <http://www.123helpme.com/search.asp?text=Artificial+Intelligence>
- <https://www.encyclopedia.com/science-and-technology/computers-and-electrical-engineering/computers-and-computing/artificial>
- <https://intelligence.org/ie-faq/>

- <https://www.scribd.com/doc/109329863/AI-application-in-Military>
- <https://www.sciencedirect.com/science/article/pii/B9780128037324000015>
- https://en.wikibooks.org/wiki/Cognition_and_Instruction/Print_version
- <https://reportgarden.com/2016/11/08/marketing-client-reporting/>
- <https://www.linkedin.com/pulse/what-7-disadvantages-manual-system-richard-breitmeyer>
- <https://www.coursehero.com/file/24457923/Week-7-Discussion-ACC-100docx/>
- <https://www.microsoft.com/en-us/research/project/project-hawaii/>
- <http://www.123helpme.com/search.asp?text=heuristic>
- <https://www.hindawi.com/journals/bmri/2015/801436/>
- <http://www.inderscienceonline.com/doi/full/10.1504/IJPOM.2013.053155>
- https://help.sap.com/doc/fiori_bs2013/1.0%202016-01/en-US/bf/d1b053a647e842e10000000a4450e5/content.htm
- http://www.academia.edu/1057005/Putting_the_pieces_of_the_puzzle_together_Using_Nvivo_for_a_literature_review
- https://busyteacher.org/classroom_activities-vocabulary/modern_technologies-worksheets/
- http://www.elon.edu/eweb/predictions/expertsurveys/2012survey/future_generation_AO_2020_anon.xhtml
- <https://www.ukessays.com/essays/criminology/impact-of-drug-and-substance-abuse-criminology-essay.php>
- <https://issuu.com/world.bank.publications/docs/9781464803420>
- <https://stacks.iop.org/help/-topic=jargon/>
- <https://quizlet.com/143448742/distributed-systems-flash-cards/>
- http://taggedwiki.zubiaga.org/new_content/e42d397bef8887ed4f3cb66a4432d7a5
- <https://docs.oracle.com/en/database/oracle/oracle-database/18/cncpt/introduction-to-oracle-database.html>

- <https://www.codeproject.com/Articles/275876/Simple-Practical-introduction-to-Asynchronous-Java>
- <https://quizlet.com/94797523/chapter-4-introduction-to-cognitive-science-and-cognitive-informatics-flash-cards/>
- https://issuu.com/retailnews/docs/retail_news_april_2015.
- <https://www.newyorker.com/magazine/2013/11/25/auto-correct>
- <https://www.edureka.co/blog/what-is-data-science/>
- <https://www.mantralabsglobal.com/blogs/author/vidushi/>
- http://uni-obuda.hu/journal/Ivancsy_Vajk_5.pdf
- <https://link.springer.com/article/10.1007/s00146-017-0730-7>
- <https://www.quora.com/What-is-data-science>
- <https://www.wikihow.com/Make-Decisions>
- <http://www.gamesauce.biz/2017/05/12/predictive-analytics-games/>
- <https://arxiv.org/pdf/1802.01029.pdf>
- <https://machinelearningmastery.com/a-tour-of-machine-learning-algorithms/>
- <http://jhb13competenciesverification.wikispaces.com/file/view/Mosters+are+Due.doc>
- <https://www.cio.com/article/3088692/big-data/simplify-the-way-your-business-handles-complex-data.html>
- https://wikivisually.com/wiki/Waldorf_education
- <http://www3.dsi.uminho.pt/pcortez/student.pdf>
- http://www.academia.edu/3114409/Using_data_mining_to_predict_secondary_school_student_performance
- <https://doi.acm.org/10.1145/860435.860443>
- <http://www.csun.edu/~twang/595DM/Slides/Week4.pdf>
- <https://www.sciencedirect.com/science/article/pii/S0168169910000062>
- <http://home.ubalt.edu/ntsbarsh/Business-stat/stat-data/Forecast.htm>
- <https://www.kent.ac.uk/courses/undergraduate/141/business-information-technology-with-a-year-in-industry>
- <http://www.w3.org/Graphics/SVG/IG/resources/svgprimer.html>

- <http://www.mdpi.com/1424-8220/15/6/13705/htm>
- https://www.researchgate.net/publication/278793328_Detection_of_subjects_and_brain_regions_related_to_Alzheimer's_disease_using_3D_MRI_scans_based_on_eigenbrain_and_machine_learning
- <https://www.sciencedirect.com/science/article/pii/B9781907568275500044>
- <https://wenku.baidu.com/view/283b0e8a84868762caaed541.html>
- <http://www.enggjournals.com/ijet/docs/IJET17-09-03S-008.pdf>
- <http://www.tandfonline.com/doi/full/10.1080/10408340600969486?scroll=top&needAccess=true>
- <https://wenku.baidu.com/view/f5036521192e45361066f52d.html>
- <https://www.scribd.com/document/45609229/Burgess-PhD-Thesis-A-Computational-Methodology-for-Modelling-the-Dynamics-of-Statistical-Arbitrage-Copy>
- <https://citrusbits.com/python-vs-node-js-which-is-best-for-startup/>
- <https://www.sourcecodester.com/python/11361/python-simple-password-generator.html>
- https://en.wikipedia.org/wiki/Open_source
- <http://journals.ametsoc.org/doi/10.1175/BAMS-D-13-00240.1>
- <http://ruby-doc.com/docs/ProgrammingRuby/>
- <http://www.dotnetcurry.com/angularjs/1400/angular-application-architecture>
- <https://www.csitquestion.com/visual-basic-tutorial/graphical-user-interface>
- <https://www.manualslib.com/manual/234720/Edimax-Br-6215srg.html>
- https://www.globalknowledge.com/en-gb/courses/Applications_Development/Applications_Development_and_Programming/GK2351
- https://www.reddit.com/r/programming/comments/ki83r/or_m_is_an_anti_pattern/
- http://www.archimuse.com/publishing/archival_methods/

- <https://planet.gentoo.org/archives/2014/2014-11-06.html>
- <http://www.ling.helsinki.fi/kit/2008s/clt231/nltk-0.9.5/doc/en/book.html>
- <http://www.ai.mit.edu/publications/pubsDB/pubs.html>
- <https://quizlet.com/35429493/qdc1-flash-cards/>
- <https://www.projectsmart.co.uk/swot-analysis.php>
- <https://www.slideshare.net/guestc44c6788/independent-study-guide>
- <https://www.pubfacts.com/author/Yeong-Jia+James+Goo>
- <http://dl.acm.org/citation.cfm?id=1224073>
- <https://www.deepdyve.com/lp/elsevier/simultaneous-data-pre-processing-and-svm-classification-model-bclAXRnb9R>
- <https://pdfs.semanticscholar.org/0572/e4cb844a1ca04bd3c555e38accab84e11c4b.pdf>
- https://wikivisually.com/wiki/Decision_tree
- <http://ithreads.pbworks.com/f/bookreviewR.pdf>
- <http://booktele.com/file/advantages-of-manual-data-input>
- <http://www2.imm.dtu.dk/pubdb/public/publications.php>
- <http://www.worldcat.org/title/me-myspace-and-i-parenting-the-net-generation/oclc/440045194>
- <https://www.wur.nl/nl/project/Europese-samenwerking-dier-1.htm>
- https://www.researchgate.net/profile/Hossain_Amran/publication/296695210_STUDENT_ALCOHOL_CONSUMPTION_presentation/links/56d8762008aee1aa5f801dd3.pdf?origin=publication_detail
- <https://max.book118.com/html/2018/0225/154564359.shtm>
- <http://spvryan.org/Issue1Volume1/12.pdf>
- <http://www.worldscientific.com/worldscibooks/10.1142/6604>
- <https://www.noexperiencenecessarybook.com/WvEYp/ijcsi-publication-2013-international-journal-of-computer-science.html>
- <https://prezi.com/4dh-rgnicdba/oss-foss-floss/>
- <http://dl.acm.org/citation.cfm?id=2405519>
- https://www.researchgate.net/profile/Dilek_Hakkani-Tur
- <https://cacm.acm.org/magazines/2015/7/188732-exascale-computing->

and-big-data/fulltext

- <https://www.jove.com/visualize?author=Richard+H+Myers>
- https://www.researchgate.net/profile/Imas_Sitanggang
- <http://hypatia.teiath.gr/xmlui/bitstream/handle/11400/10440/Improving%20Quality%20of%20Educational%20Processes.pdf?sequence=1>
- https://www.researchgate.net/profile/Qaysar_Mahdy/publication/280009082_Web_page_Classification_by_Using_Neural_Networks/links/55a60f9408ae81aec91388fd/Web-page-Classification-by-Using-Neural-Networks.pdf
- <https://dl.acm.org/citation.cfm?id=2355573.2356420&coll=DL&dl=GUIDE>
- <http://www.oracle.com/us/products/applications/tutor-upk/oracleupk-techspecs12-1-2186665.pdf>
- <https://www.scribd.com/document/272634408/Hotel-Management-System-final>
- http://download.microsoft.com/download/8/F/A/8FABACD7-803E-40FC-ADF8-355E7D218F4C/SQL_Server_2012_Security_Best_Practice_Whitepaper_Apr2012.docx
- <https://link.springer.com/article/10.1007/s00766-015-0240-4>
- [https://www.revolvy.com/main/index.php?s=Python%20\(programming%20language\)&item_type=topic](https://www.revolvy.com/main/index.php?s=Python%20(programming%20language)&item_type=topic)
- http://www.planning.salford.ac.uk/__data/assets/pdf_file/0005/20696/20100908-Best-Practices-and-Strategies-in-Organizational-Structure-and-Design-Membership.pdf
- http://ipa.math.uni-heidelberg.de/dokuwiki/Papers/nair2013_survey_to_fusion.pdf
- <https://www.simplilearn.com/prince2-foundation-exam-questions-free-practice-test>
- http://www.who.int/entity/medicines/areas/quality_safety/quality_assurance/s20084en.ppt?ua=1

- https://www.slideshare.net/amit_gandhi/documentation-35795570
- <https://www.nap.edu/read/1120/chapter/9>
- <https://www.scribd.com/document/97272768/Chapter-2>
- http://resources.docs.salesforce.com/200/20/en-us/sfdc/pdf/salesforce_case_implementation_guide.pdf
- <https://www.scribd.com/document/336616090/Documentation-on-Warehouse-Management-Sy-docx>
- <https://www.scribd.com/document/81205042/Uml>
- <https://www.scribd.com/document/259382698/Class-and-Interaction-Diagrams-pdf>
- <http://nptel.ac.in/courses/106105087/pdf/m07L16.pdf>
- http://www.jot.fm/issues/issue_2004_09/article1/
- http://users.ece.cmu.edu/~koopman/des_s99/requirements_specs/
- https://en.wikipedia.org/wiki/Systems_design
- <https://www.scribd.com/document/50683633/software-manual-2>
- <https://www.slideshare.net/MujahedAhmed1/clinic-database-and-software-management-system>
- <https://docs.microsoft.com/en-us/biztalk/core/design-phase-recommendations>
- http://www.ijeit.com/Vol%205/Issue%2011/IJEIT1412201605_08.pdf
- <https://www.sciencedirect.com/science/article/pii/S0957417417302221>
- <https://www.scribd.com/document/347780012/Machine-Learning-for-Designers>
- John McCathy (2007) <http://www.formal.stanford.edu/jmc/whatisai/node1.html>
- Kirschner, P.A. & Karpinski, A.C. 2010, "Facebook® and academic performance", *Computers in Human Behavior*, vol. 26, no. 6, pp. 1237-1245.
- Kamiński, B.; Jakubczyk, M.; Szufel, P. (2017). "A framework for sensitivity analysis of decision trees". *Central European Journal of Operations Research*. doi:10.1007/s10100-017-0479-6

- Kremer, M. & Levy, D. 2008, "Peer effects and alcohol use among college students", *The journal of economic perspectives*, vol. 22, no. 3, pp. 206;189; -206.
- Kulman Dave (2012). "A Python Book: Beginning Python, Advanced Python, and Python Exercises".
- Kvavik, R. (2005). Convenience, communications, and control: How students use technology. In D. Oblinger, & J. Oblinger (Eds.), *Educating the net generation (chapter 7)* [e-book]
- LinkedIn. (2017). What are the 7 disadvantages to a manual system? [online] Available at: <https://www.linkedin.com/pulse/what-7-disadvantages-manual-system-richard-breitmeyer> [Accessed 27 Oct. 2017].
- **Lazear, Edward.** 2001. "Educational Production." *Quarterly Journal of Economics*, 116(3): 777– 803.
- Louridas, P. & Ebert, C. 2016, "Machine Learning", *IEEE Software*, vol. 33, no. 5, pp. 110-115.
- L.D. Rosen (2007) Palgrave Macmillan, New York Me, Myspace, and I: Parenting the next generation
- **Miguel, Edward, and Michael Kremer.** 2004. "Worms: Identifying Impacts on Education and Health in the Presence of Treatment Externalities." *Econometrica*, 72(1): 159–217.
- Martin-Collado, D., Diaz, D., Mäki-Tanila, A., Colinet, F., Duclos, D., Hiemstra, S.J., Gandini, G. & EURECA Consortium 2013, "The use of SWOT analysis to explore and prioritize conservation and development strategies for local cattle breeds", *Animal*, vol. 7, no. 6, pp. 885-894.
- **Munshi, Kaivan, and Jacques Myaux.** 2006. "Social Norms and the Fertility Transition." *Journal of Development Economics*, 80(1): 1–38.
- Machine learning 2017. *Britannica Academic*. Retrieved 26 October 2017, from <http://academic.eb.com.ezproxy.mdx.ac.uk/levels/collegiate/article/machine-learning/474180>

- Nations, Daniel. "Web Applications". *About.com*. Retrieved 25 December 2017.
- P. Cortez and A. Silva (April 2008). Using Data Mining to Predict Secondary School Student Performance. In A. Brito and J. Teixeira Eds., *Proceedings of 5th Future Business Technology Conference (FUBUTEC 2008)* pp. 5-12, Porto, Portugal, EUROSIS, ISBN 978-9077381-39-7.
- Rokach, L. & Maimon, O. 2015, *Data mining with decision trees: theory and applications*, 2nd edition, World Scientific, Hackensack.
- Raval M Kalyani (2012) *Data Mining Techniques*, International Journal of Advanced Research in Computer Science and Software Engineering Volume 2, Issue 10.
- Rokach, L. & Maimon, O. 2015, *Data mining with decision trees: theory and applications*, 2nd edn, World Scientific, Hackensack.
- Ramesh, V., Parkavi, p., and Ramar, k. (2013). Predicting student performance: a statistical and data mining approach. *International journal of computer applications*, 63(8), 35-39.
- St. Laurent, Andrew M. (2008). *Understanding Open Source and Free Software Licensing*. O'Reilly Media. p. 4. ISBN 9780596553951.
- Tutorials Point (2015) 'Artificial intelligence - Intelligent systems', p. 62.
- Vani Kholasa (2015), Predicting a student's performance
- Witten, I.H., Frank, E. & Hall, M. 2011, *Data mining: practical machine learning tools and techniques*, 3rd edn, Morgan Kaufmann, Amsterdam.
- Yu, H., Hsieh, C., Chang, K. & Lin, C. 2012, "Large Linear Classification When Data Cannot Fit in Memory", *ACM Transactions on Knowledge Discovery from Data (TKDD)*, vol. 5, no. 4, pp. 1-23.

APPENDICES

Appendix A.

School of Science & Technology,
Department of Computer Science Ethics Committee



Form D: Declaration Form

This form should be given to your supervisor along with your project proposal. It must also be included in your Project Report.

Student Project: Ethical Approval Request			
Name: NICHOLAS CHUN MARKUS	Student ID: M00598289	Date: 22/03/2018	
Supervisor: STEVEN CAMILLERI			
Title STUDENT PERFORMANCE PREDICTION USING MACHINE LEARNING			
<u>Ethical Approval Statement:</u>			
Declaration A			
(i)	I have studied the Ethical Approval section.		
(ii)	I have established that my study does not require additional human participation.		
(iii)	I agree to re-apply for approval if the nature or goals of my project change.		
<input checked="" type="checkbox"/>			
Declaration B			
Project Goals involving human participation:			

(i)	I have studied the Ethical Approval section.		
(ii)	My study involves human participation through		
	o observation		
	o questioning.		
(iii)	Participants will be selected without coercion (see Chart 1).		
(iv)	I will obtain informed consent (see Chart 2) from each participant using Form C.		
(v)	I have arrangements in place for the protection of personal data (see Chart 3).		
(vi)	I agree to re-apply for approval if the nature or goals of my project change.		
<input type="checkbox"/>			
Declaration C			
My project does not fulfil the conditions for fast track Ethical Approval and I am applying separately to the Ethics Committee			
<input type="checkbox"/>			
Note: to make an application to the Ethics Committee, you need to complete Form E - Application for Ethical Approval & Form C - Informed Consent Form (Download from: http://tinyurl.com/mdx-ethics)			
Student Signature.....		Date..... 22/03/2018.....	
Supervisor's Signature.....		Date..... 22/03/2018.....	

Fig 19. Declaration Form

Appendix B

INSTALLATION INSTRUCTION

The application comes with a prebuilt environment of python 3 with the required modules and libraries needed to run.

To run the application, follow the instructions bellow

- Open terminal in the project folder
- Run “source env/bin/activate”
- Next, you run “python run.py”
- Open browser and run <http://localhost:8080>

For the machine learning classifiers

- Open project folder – src – run “classifier.py”
- Enter a cost function of your choice and press enter.