Department of Computer Science
FAST National University of Computer and Emerging Sciences
Karachi Campus

Software Design Specification

Predictive Analytics on the Academic Record of NUCES

Version: 1.0

Supervisor	Dr Jawwad Ahmed Shamsi
Project Team	Obaid ur Rehman (17k-3848) Areeka Aijaz (17k-3913) Tooba Shahid (17k-3731)
Submission Date	14th December 2020

Document Sign-Off

Version	Sign-off Authority	Sign-off Date
1.0	Supervisor	14 Dec 2020

Document Information

Category	Information
Customer	FAST NUCES
Project	Predictive Analytics on Academic Record of NUCES
Document	Software Design Specification
Document Version	1.0
Status	Initial Version
Author	Project Team
Approver	Project Supervisor
Issue Date	14th Dec 2020
Document Location	
Distribution	Supervisor & FYP Coordinators

Definition of Terms, Abbreviation and Acronyms

Term	Description
NUCES	National University Of Computer and Emerging Sciences
FYP	Final Year Project
ANOVA	Analysis of Variance
KNN	K nearest neighbours
SVM	Support Vector Machine
ERD	Entity Relationship Diagram
MERN	MongoDB, Express, React, Node

Table of Contents

1. Introduction

1.1. Purpose

This document is the written description of the software product that we are going to develop as a part of our FYP . This document gives a description of the design of the system for clear understanding of what is to be built and how it is to be built. This document also serves the purpose of overall guidance of the development team.

1.2. Intended Audience

- User

  The user of this project will get an idea of the functionalities of the system. The users of this project may be different academic institutions.

- Project Team

  Project team will have a clear understanding of the user requirements, functionalities, functional and non functional requirements.

- Project Instructor

  The instructor will be able to easily identify the proposed system together with the scope, requirements and functionalities that will help the instructor to grade the project.

1.3. Document Conventions

- Main headings

  Font: Calibri (Body)
  Face: Bold, underline
  Size: 24

- Text Subheadings

  Font: Calibri (Body)
  Face: Bold

Size: 20

- Description
  Font: Calibri (Body)
  Face: None
  Size: 14
- Tables and images
  Font: Calibri (Body)
  Face: None
  Size: 22

1.4. Project Overview

Predictive Analytics is the process of using past data to make future predictions . The past data is used to capture important trends with the help of a mathematical model and the model is then used to make predictions on current data . We aim to perform predictive analytics on the academic record of NUCES . Our project will use the past academic record of NUCES to make insights and find out correlations between different attributes and then build a predictive model , based on the statistical inference , to make predictions on the current data. Finally we aim to integrate our analytics with a system , developing a fully functional web portal . The web portal will support the features such as importing data , query processing on data , visualizing data and predicting results .

1.5. Scope

Student Performance Prediction Systems are not a popular thing in our culture , however , if used essentially , they are a powerful and significant need. Our project aims to develop a website for educational institutions that they can use for the purpose of predicting the performance of their enrolled students. We aim to enable the educational institutions to use systems to import their data , perform different visualizations and get various insights about their data by query processing and data analytics. Not only this but we look forward to providing them a system which will help them predict the performance of their students

beforehand and hence can improve their educational performance and success rates.

2. Design Considerations

2.1. Assumption and Dependencies

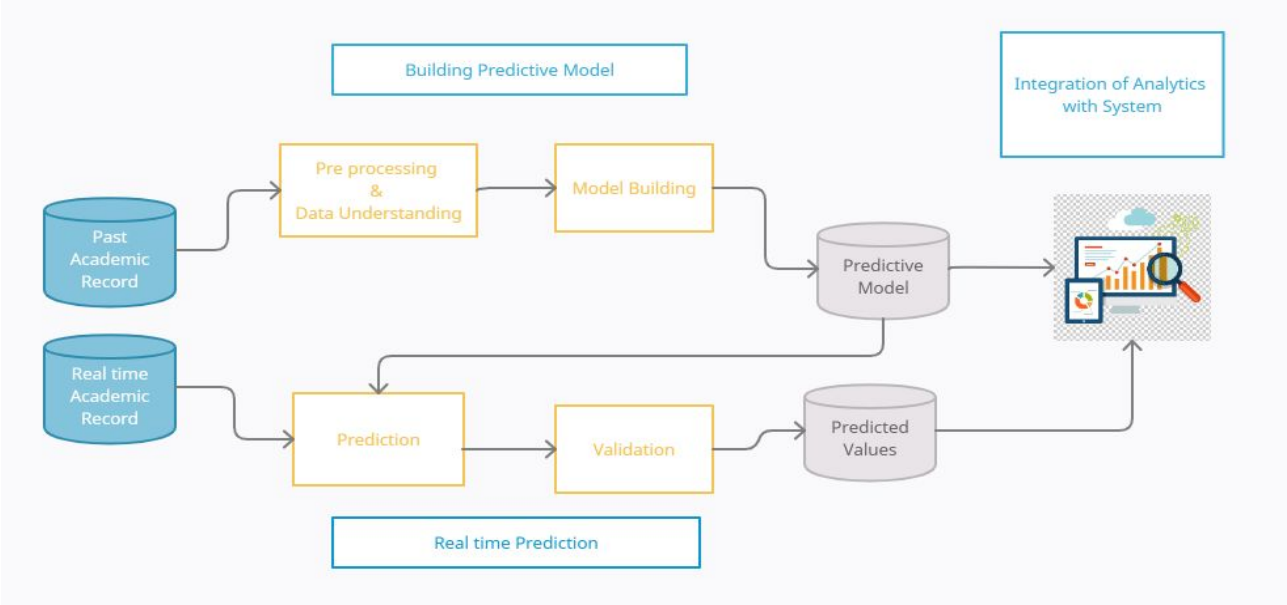
- While making predictions , the system assumes that the future will continue to be like the past. Therefore the system is dependent on the data being fed into the system , and assumes that important attributes for prediction are not missing , false or vague.
- The design is dependent on the format (template) of the academic data .

2.2. Risks and volatile areas

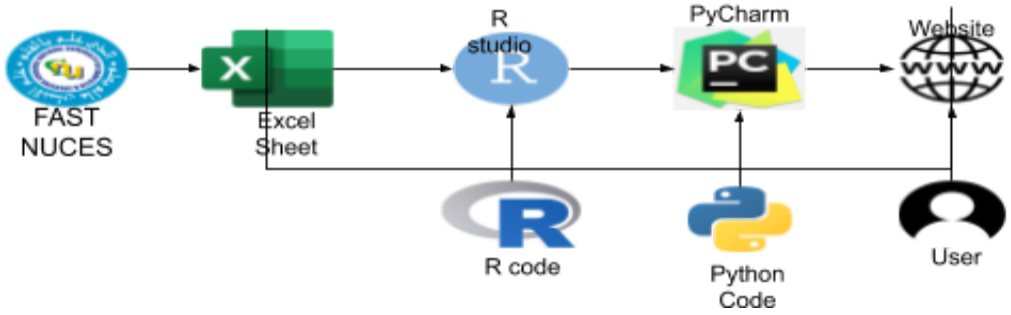
There is a risk in the system if the data dictionary is not compatible with the system assumed format.

3. System Architecture

3.1. System Level Architecture



3.2. Software Architecture



4. Design Strategy

The design strategy will follow the below mentioned steps:

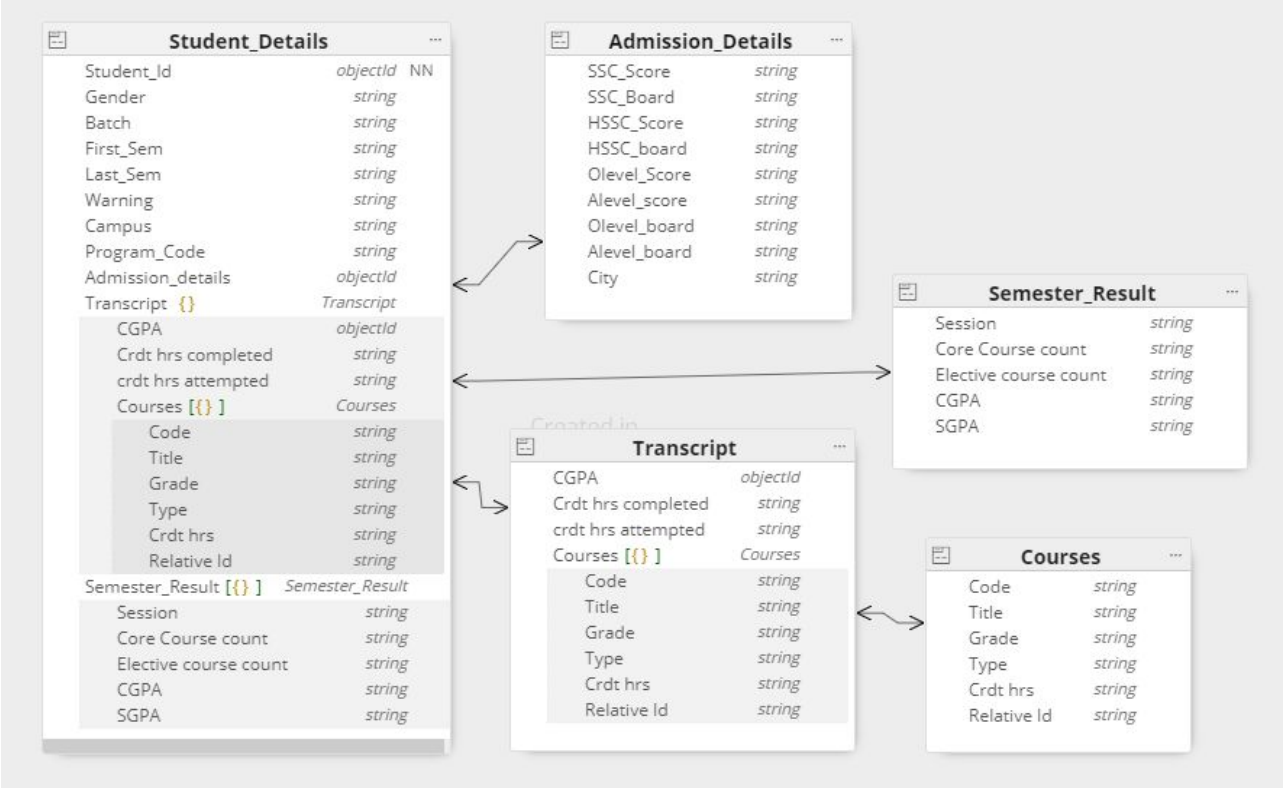
- A data import feature will be implemented in a web portal , with a user friendly interface .
- The data will be uploaded by a user (admin of an educational institution) , to the website.
- The data will then be loaded into R studio.
- Queries will be written , in R , to transform the data
- The data will be visualized using R script .
- Statistical Inference and Data Analytics will be performed on the data , to get useful insights.
- Different statistical models like ANOVA , correlation , linear regression etc will be used.
- The analytics will then be loaded into PyCharm.
- The analytics will be used to build a predictive model in Python.
- Different models including KNN , Decision Tree , SVM etc will be used.
- The dashboard will be developed for the website.
- The analytics will then be integrated with a website .
- The model performance comparison results will also be integrated into the website.
- The predictive model will also be integrated into the website.
- A query processing feature will also be added to the website.
- The website will be developed using MERN stack .

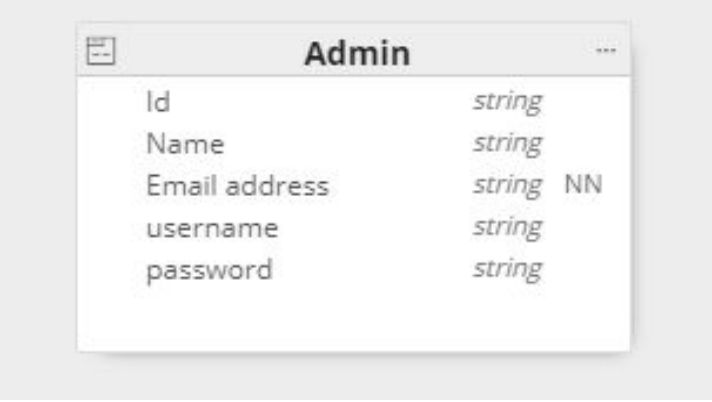
5. Detailed System Design

5.1. Database Design

In our project , the database we will use is MongoDB , which is a NoSQL database.

5.1.1. ERD





5.1.2. Data Dictionary

Students Data

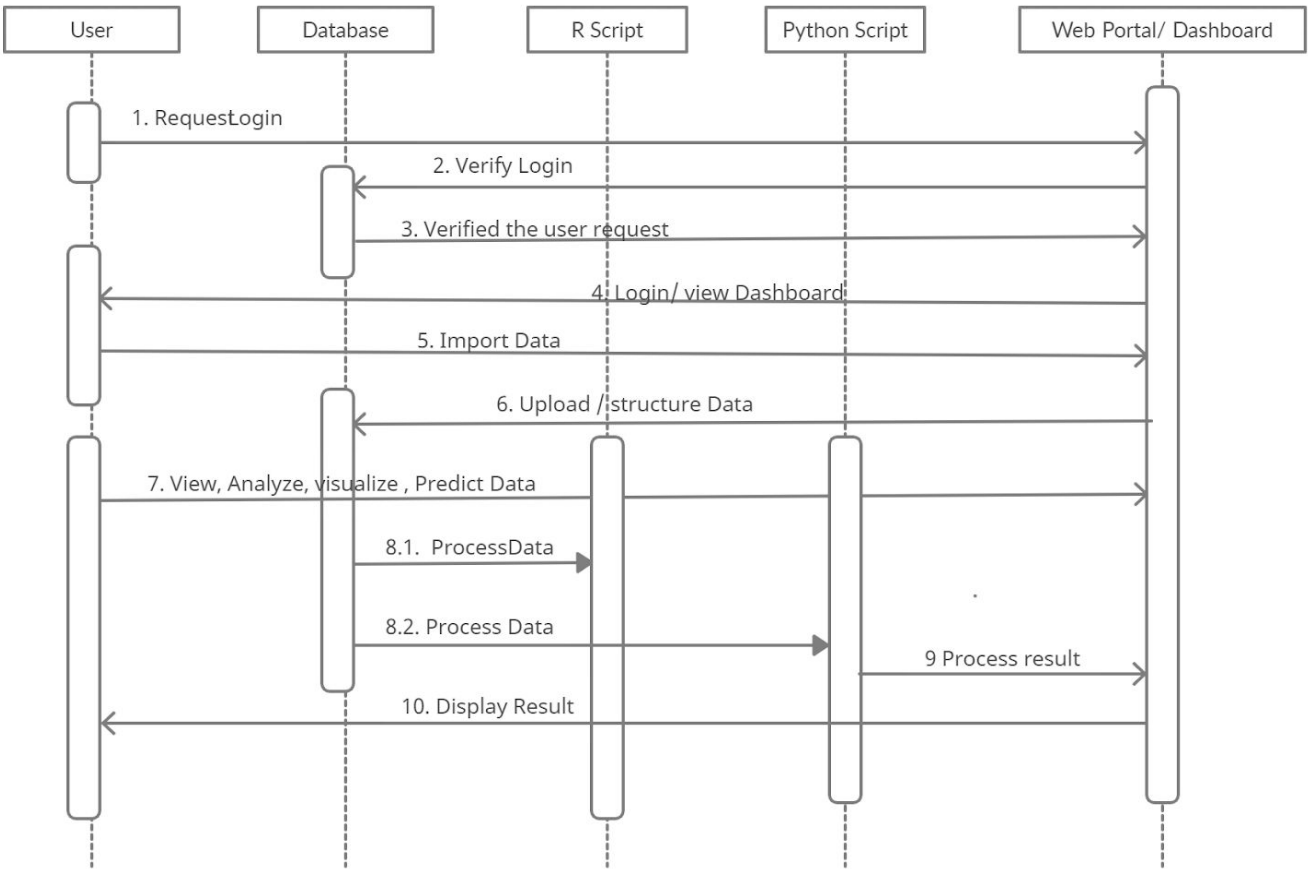
```
Student_Details{
        Student_Id,
        Gender,
        Batch,
        First_Sem,
        Last_Sem,
        Campus,
        Program_Code,
        Warning_Count,
        Admission_Details: {
                SSC_Score,
                SSC_board,
                HSSC_Score,
                HSSC_board,
                Olevel_Score,
                Alevel_Score,
                City
        },
        Transcript: {
                CGPA,
                Credit_hours Attempted,
                Credit_hours Completed,
        },
        Courses Enrolled: [
                {
                        Course_Code,
                        Course_Name,
                        Credit hours,
                        Grade,
                        Course_type,
                        Grade_Point,
                        Relation_Id
                }
        ],
        Semester_Details: [
                {
                        Session,
                        CGPA,
                        SGPA,
                        Core course count,
                        Elective course count
                },
        ]
}
```

Admin Details

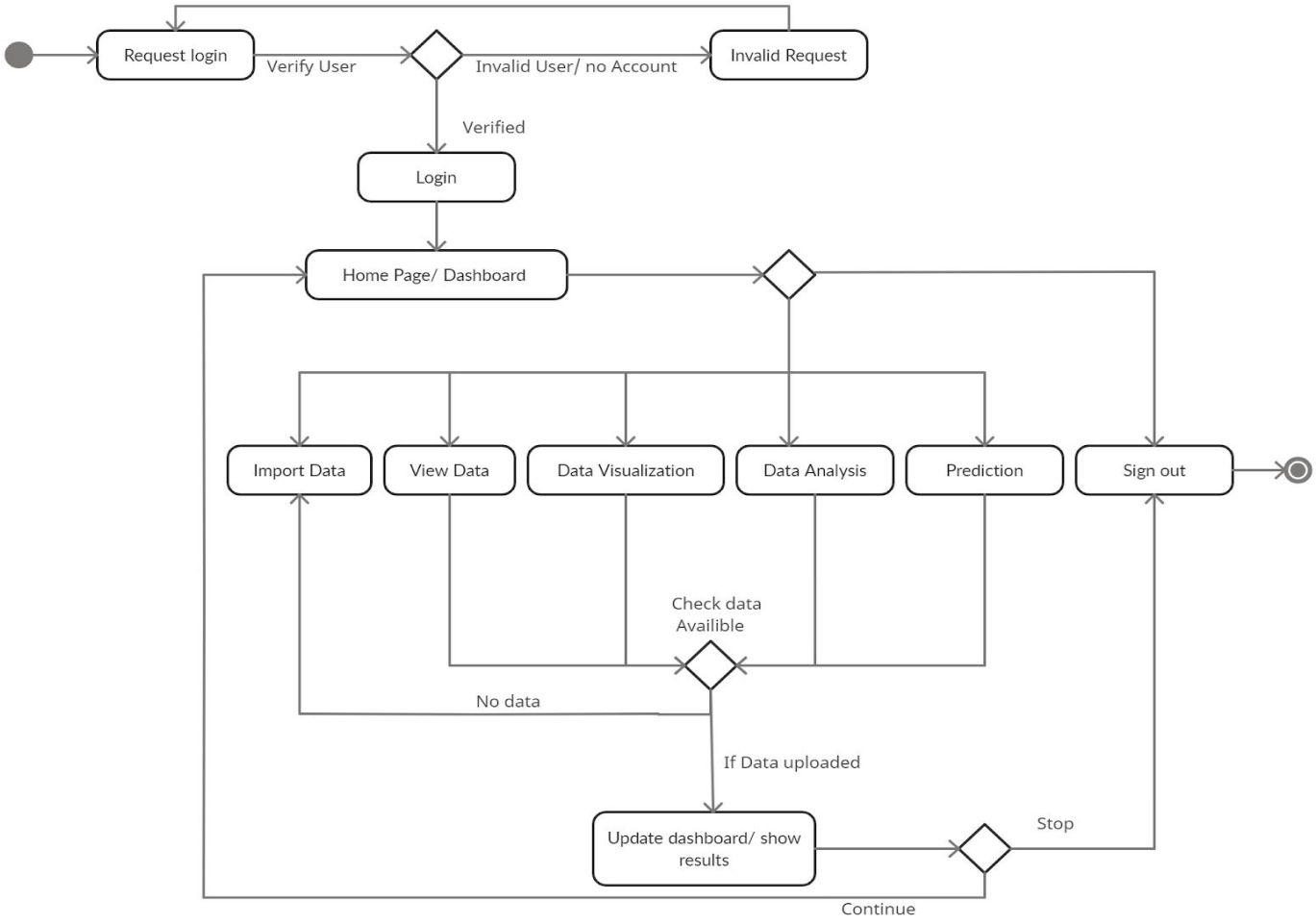
```
Admin_Details{
        Id,
        Name,
        Username,
        Password,
        Email_address
}
```

5.2. Application Design

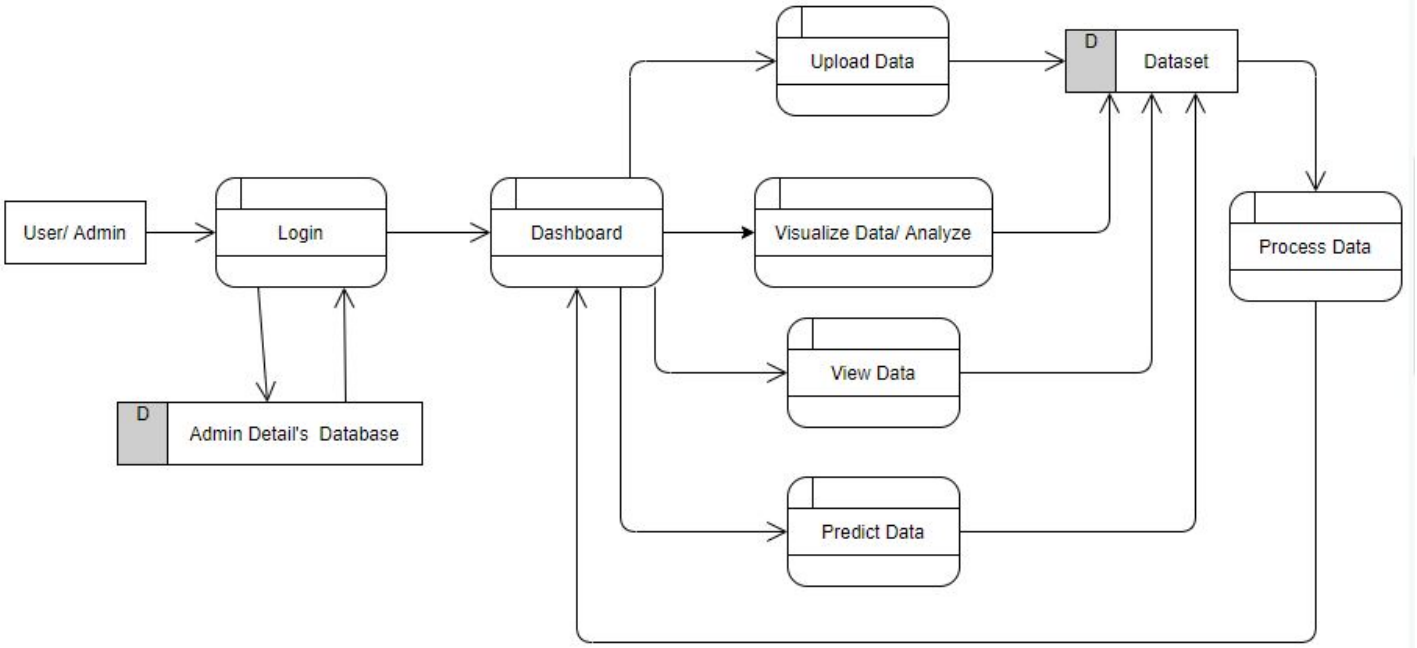
5.2.1. Sequence Diagram



5.2.2. State Diagram



5.2.3. Data Flow Diagram (Level 1)



References

1. Hajra Waheed , Saeed-Ul Hassan , Julie Hardman , Salem Alelyani , Raheel Nawaz , '"Predicting academic performance of students from VLE big data using deep learning models" , Volume 104, March 2020
2. Eyman Alyahyan , Dilek , "Predicting academic success in higher education: literature review and best practices" , International Journal of Educational Technology in Higher Education , 10 February 2020
3. Muhammad Yunus Iqbal Basheer, Sofianita Mutalib, Nurzeatul Hamimah Abdul Hamid, Shuzlina Abdul-Rahman, Ariff Md Ab Malik , "Predictive analytics of university student intake using supervised methods" , IAES International Journal of Artificial Intelligence , Vol. 8, No. 4, December 2019, pp. 367~374
4. Sonali Rawat , "Predictive Analytics for Placement of Student- A Comparative Study" , International Research Journal of Engineering and Technology (IRJET) , Volume: 06 Issue: 06 , June 2019
5. Ashish Dutt , Maizatul Akmar Ismail , Tutut Herawan , "A Systematic Review on Educational Data Mining" , IEEE Access , Volume: 5 , January 2017

Appendix A
Terminologies

A.1 Statistical Inference
Statistical inference is the process through which inferences about a population are made based on certain statistics calculated from a sample of data drawn from that population .

A.2 Data Analytics
Data analytics is a process of inspecting, cleansing, transforming, and modeling data with the goal of discovering useful information, informing conclusions, and supporting decision-making

A.3 NoSQL Database
NoSQL databases (aka "not only SQL") are non tabular, and store data differently than relational tables.

Appendix B

Statistical Models

A statistical model is a combination of inferences based on collected data and population understanding used to predict information in an idealized form. There are different types of statistical models known as tests that can be used to analyze data.

B.1 ANOVA

Anova , short for Analysis of Variance , ANOVA is a statistical technique that assesses potential differences in a scale-level dependent variable by a nominal-level variable having 2 or more categories.

B.2 Correlation

In statistics, correlation or dependence is any statistical relationship, whether causal or not, between two random variables or bivariate data. When two sets of data are strongly linked together we say they have a High Correlation. The word Correlation is made of Co- (meaning "together"), and Relation. Correlation is Positive when the values increase together, and. Correlation is Negative when one value decreases as the other increases.

B.3 Linear Regression

In statistics, linear regression is a linear approach to modelling the relationship between a scalar response and one or more explanatory variables.

Appendix C

Predictive Models

Predictive Model uses known results to create, process, and validate a model that can be used to forecast future outcomes.

C.1 KNN

The k-nearest neighbors (KNN) algorithm is a simple, easy-to-implement supervised machine learning algorithm that can be used to solve both classification and regression problems.

C.2 Decision Tree

A decision tree is an approach to predictive analysis that can help you make decisions. As the name goes, it uses a tree-like model of decisions. Though a commonly used tool in data mining for deriving a strategy to reach a particular goal, it's also widely used in machine learning.

C.3 SVM

In machine learning, support-vector machines are supervised learning models with associated learning algorithms that analyze data used for classification and regression analysis