GUJARAT TECHNOLOGICAL UNIVERSITY

Chandkheda, Ahmedabad

Affiliated

Babaria Institute of Technology

A

Project Report

On

Result Prediction & Analysis

UDP Project

Prepared as a part of the requirements for the subject of

PROJECT - I

B. E. IV, Semester – VIII

(Computer Science and Engineering)

Submitted by:

Group:

Sr.	Name of Student	Enrolment No.
1.	Sarth Desai	140050107022
2.	Arunuday Ganju	140050107029
3.	Perin Shah	140050107543
4.	Sanjana Udar	140050107556

Dr. Avani R. Vasant

(Faculty Guide)

Dr. Avani R. Vasant

Head of the Department

(2017-2018)

Academic year

A

Project Report
On

Result Prediction & Analysis

Guided by
Dr. Avani R. Vasant

Prepared by

Student Name	Enrolment Number
1. **Sarth Desai**	**140050107022**
2. **Arunuday Ganju**	**140050107029**
3. **Perin Shah**	**140050107543**
4. **Sanjana Udar**	**140050107556**

Babaria Institute of Technology
Department of Computer Science and Engineering
At: Varnama, Ta: Vadodara, Dist: Vadodara, Pin: 391240

B A B A R I A
INSTITUTE OF TECHNOLOGY

CERTIFICATE

This is to certify that the project report entitled **Result Prediction & Analysis** is prepared and presented by **Sarth Desai, Arunuday Ganju, Perin Shah and Sanjana Udar** bearing Enrolment Nos. 140050107022, 140050107029, 140050107543 and 140050107556 respectively 4th Year of **B.E (Computer Science & Engineering)** and their work is satisfactory.

.

Dr. Avani R. Vasant
Guide Name

Dr. Avani R. Vasant
Head of Department

Babaria Institute of Technology
Department of Computer Science and Engineering
At: Varnama, Ta: Vadodara, Dist: Vadodara, Pin: 391240

INDEX

Acknowledgement

When a person strives to achieve big goals, he alone cannot manage it all. There is a need of large contribution of many people. There are many direct and indirect contributions which may have remained invisible.

We would like to take this opportunity to thank our mentor, Dr. Avani R. Vasant, Head of Department, who guided us throughout the process. We are highly indebted to Avani mam for her guidance and constant supervision, as well as for providing necessary information regarding the project.

We would also like to convey our sincere gratitude and indebtedness to Prof. Parth Goel for his unabated enthusiasm in calculating into us and rendering his expert guidance throughout every milestone of our project.

Last but not the least, we would like to express our gratitude towards our parents and faculty members of BITS EDU CAMPUS, for their kind support at all stages of the work and encouragement which helped us in this project. It is beyond the reach of our language to express our sincere appreciation to our parents, for their infinite inspiration and sacrifices for the benefit of our academic career.

1. INTRODUCTION

1.1. Problem summary

Education is one of the most important industries in today's world, and while it continues to develop in meaningful ways, finding the best practices for teaching and the best evaluation tools is still a difficult problem.

The ability to predict a student's performance on a given concept is an influential tool for the Education industry; it allows for understanding what types of students there are and what are key-concepts that help shape the perception of another. These are important factors for educators to know to constantly modify and improve educating tools (such as text books and lecture plans).

This project aims to predict a student's performance or score on the concept of Machine Learning, and their own performance on any previous semester grades. This can naturally lead into the prediction of score of the student by using the dataset analysis provided.

1.2. Aims and Objectives

Over the years, the educational systems have been improving drastically. Result prediction has become a frequent practice for the improvement of the institution and for the betterment of students. Predicting results by the institutions provides plentiful time to students for improving themselves. This helps in improving the student's academic performance and the ranking of the educational institution.

Currently the student's academic performance is predicted by considering their assessments and behavioural pattern during class hours. This approach of prediction being followed is crude and does not give accurate results. In today's context result prediction plays a dominant role in not only for ranking of educational institutions but can also be used for much wider applications.

This project aims to build a predictor, by finding similar student's performance and score for a given student and predicting their result on a concept they have not explored based on past achievement.

1.3. Problems specifications

Many experiments have already been done to show that things like location, wealth, and students who are better at some subjects will perform better than their counterparts, but this project aims to move away from these evaluations to find more a different result.

While it is known that increasing the features and attributes of the model further increases the preciseness of the predicted result. Hence, to achieve more accurate result we are using all the data of students to find pattern with the current data of student and predicting their result of the required subject based on the generalized clustering pattern we find from the integration of current and past student records.

1.4. Literature review

Many people have proposed various systems that analyse and predict the behavioural and academic pattern of students. In the year of 2014, Naren J [1] proposed a system that specifies the classification techniques for predicting the career options and to predict the violent behaviour prevalent among students. A response sheet was used to gather details regarding the background information, reaction of a student when irritated and the interests of a student. These parameters were mined to find the corresponding behavioural pattern of a student. However, additional research on student mindsets and learning is needed to predict the patterns efficiently. The response sheet must be created in such a way that it avoids false values and other possible complications. Nominal number of attributes is used which in turn does not give accurate values.

Authors Nitya Upadhyay, Vinodini Katiyar [2] have proposed a system for predicting the academic performance and the behaviour of a student. Various techniques such as Naïve Bayesian Classification, Multilayer Perceptron, ID3 and J48 were used to analyse the attributes. However only a provisional analysis on which technique would give accurate results was specified but not on how they need to be implemented.

1.5. Plan of work

In this section, we are going to cite a conceptual model of our system and explain the various aspects of it in a systematic manner.

A. Data Gathering

To predict the students, result we need to accumulate large amount of data. Larger amount of data results in more precise predictions. Hence, data gathering plays a vital role. The educational organizations are conscientious for collecting and providing these data.

There are two types of important data required for creating efficient dataset:

- Data of previous students
- Data of current student

The data of past students refers to the collective records of the previous pass out batches which are stored to predict current student results. Current student data refers to the group of students for whom the prediction process is to be carried out.

B. Attribute Selection

There is large amount of student data available which is not being utilized efficiently. Attribute selection is one of the most important steps. It is important to extract substantial attributes from gathered data.

The attributes are classified as:

- Quantitative attributes
- Qualitative attributes

Quantitative data are those which have a definitive measure. These maybe in the form of numbers. The data attributes that can be used here are Internal Assessments, Attendance percentage, etc.

Qualitative data are those which are cannot be represented in the form of numbers. It rudimentarily defines some qualities or characteristics. This kind of qualitative data are converted into numerical form by using questionnaires.

C. Implementation

There are numerous numbers of algorithms available which can be used for the prediction process. Some of these algorithms are neural networks, naïve Bayes, k-Nearest Neighbour, decision tree, etc.

The main step before the prediction process is to check all the past data of students are available. Only with the help of present and past dataset of student's performance record the prediction process is possible. Once the data is available, the required attributes for result prediction are taken into consideration.

1.6. Materials and tools required

A. PC (Supporting python 3 and its libraries)

B. Student Performance Dataset from Kaggle

C. Machine Learning

Machine learning is a core sub-domain of artificial intelligence as it enables computers to get into a mode of self-learning without being explicitly programmed. When exposed to new data, computer program is enabled to learn, grow, change, and develop by themselves.

ML Algorithms used in this project:

- Perceptron: In machine learning, the perceptron is an algorithm for supervised learning of binary classifiers (functions that can decide whether an input, represented by a vector of numbers, belongs to some specific class or not). It is a type of linear classifier, i.e. a classification algorithm that makes its predictions based on a linear predictor function combining a set of weights with the feature vector.
- Support Vector Machine(SVM-SVC): A support vector machine constructs a hyperplane or set of hyperplanes in a high or infinite-dimensional space, which can be used for classification, regression, or other tasks like outlier's detection.
- StandardScaler: It standardize the features by removing the mean and scaling to unit variance. Centering and scaling happen independently on each feature by computing the relevant statistics on the samples in the training set. Mean and standard deviation are then stored to be used on later data using the transform method.
- Logistic Regression: Logistic regression falls under the category of supervised learning; it measures the relationship between the categorical dependent variable and one or more independent variables by estimating probabilities using a logistic/sigmoid function. In spite of the name 'logistic regression', this is not used for regression problem where the task is to predict the real-valued output.
- Xgboost: It is short for 'Extreme Gradient Boosting'. It is an open-source software library which provides the gradient boosting framework. It is used for

supervised learning problems, where we use the training data (with multiple features) xi to predict a target variable yi.

D. Python

Python is a general-purpose language interactive, object-oriented, and high-level programming language; which means it can be used to build models, which will be made easy with the right tools and python libraries.

Professionally, Python is great for backend web development, data analysis, artificial intelligence, and scientific computing.

Beyond Python there are many open source libraries generally used to facilitate practical machine learning. In general, these are the main scientific Python libraries we use when performing machine learning tasks:

- NumPy – Useful for creating N-dimensional array objects.
- Pandas – Python data analysis library includes structures like dataframes, etc.
- Matplotlib – 2D plotting library
- SciKit-Learn – the ML algorithms used for data analysis and data mining tasks.
- SciPy – It contains modules for linear algebra, optimization, integration, and statistics.

2. Design: Analysis, Design Methodology and Implementation Strategy

2.1. Observation Matrix (AEIOU Summary)

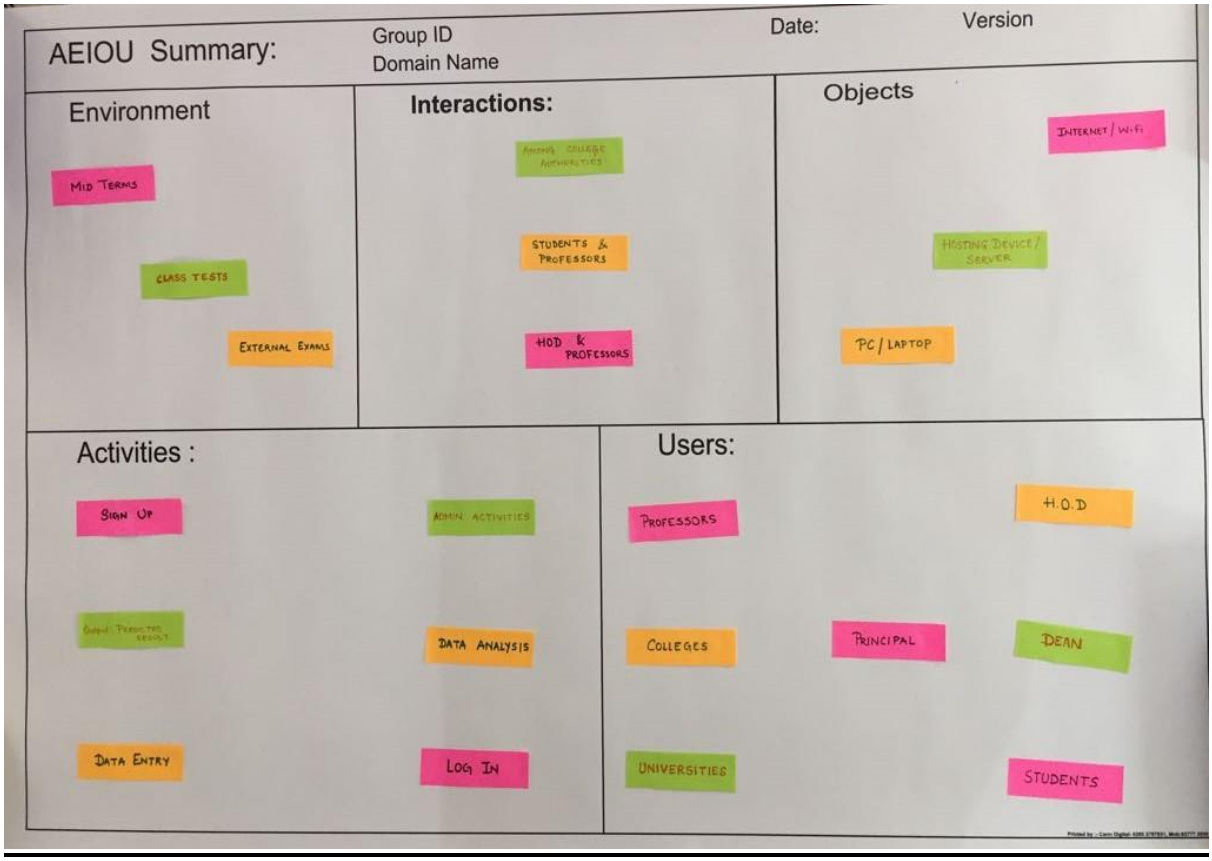


Fig. 2.1 Observation Matrix Canvas

A. Activities

We find a few activities taking place in our project field of Result Prediction and classification using machine learning. The possible users will be seen indulging in the following activities as per the scenario. The general impressions and observations that we could take note of are as follows:

- Sign Up
- Log In
- Data Entry
- Data Analysis
- Admin Activities

- Output: Predicted Result

B. Environment

We also wanted to study the environment around the user to efficiently develop a tool for the user to use. So, the general impressions and observations, which include the style, materials and the atmosphere around the user are observed. These are listed below:

- Mid Terms
- Class Tests
- External Exams

In this canvas, we wanted to observe and take note of the environment the user was in. This is also important as we will be better able to understand the situations the user may face and the scenarios that could occur or are occurring. Hence, we can be able to optimise our project accordingly. We noticed that the environment was modern and a bit chaotic.

C. Interactions

We took note of all the communication and interaction that was occurring at the place of interest. Some of these are:

- Among College Authorities
- Student and Professors
- HOD and Faculties

The elements, features and special notes that were noticeable at the place of interest is listed below:

- Mobile phone
- Email
- In-person

In this canvas we tried to look for an interaction scene that concerned our project. User interaction was the one of our interest.

D. Objects

To understand our problem better we needed to know what the objects are involved in the surroundings we want to work in. And noticed the following objects.

We also must note the components that were involved while we were observing the scenario.

Some of our general impressions were: -

- Internet/Wifi
- Hosting Device/Server
- PC/Laptop

We also took note of how the objects were related to the activities. These were co-related to the observations we took previously in our activity canvas.

This canvas was very beneficial to us as we could decide now what the objects we could add here were and what were the already existing objects we could work with in our project.

E. Users

Now it was important to know the users so that we could make our project more personalised to our users and understand what needs to be fixed for them. While observing the user, the key observations about the user, that we noted are listed as follows: -

- Professors
- Colleges
- Universities
- Principal
- HOD
- Dean
- Students

2.2. Ideation Canvas

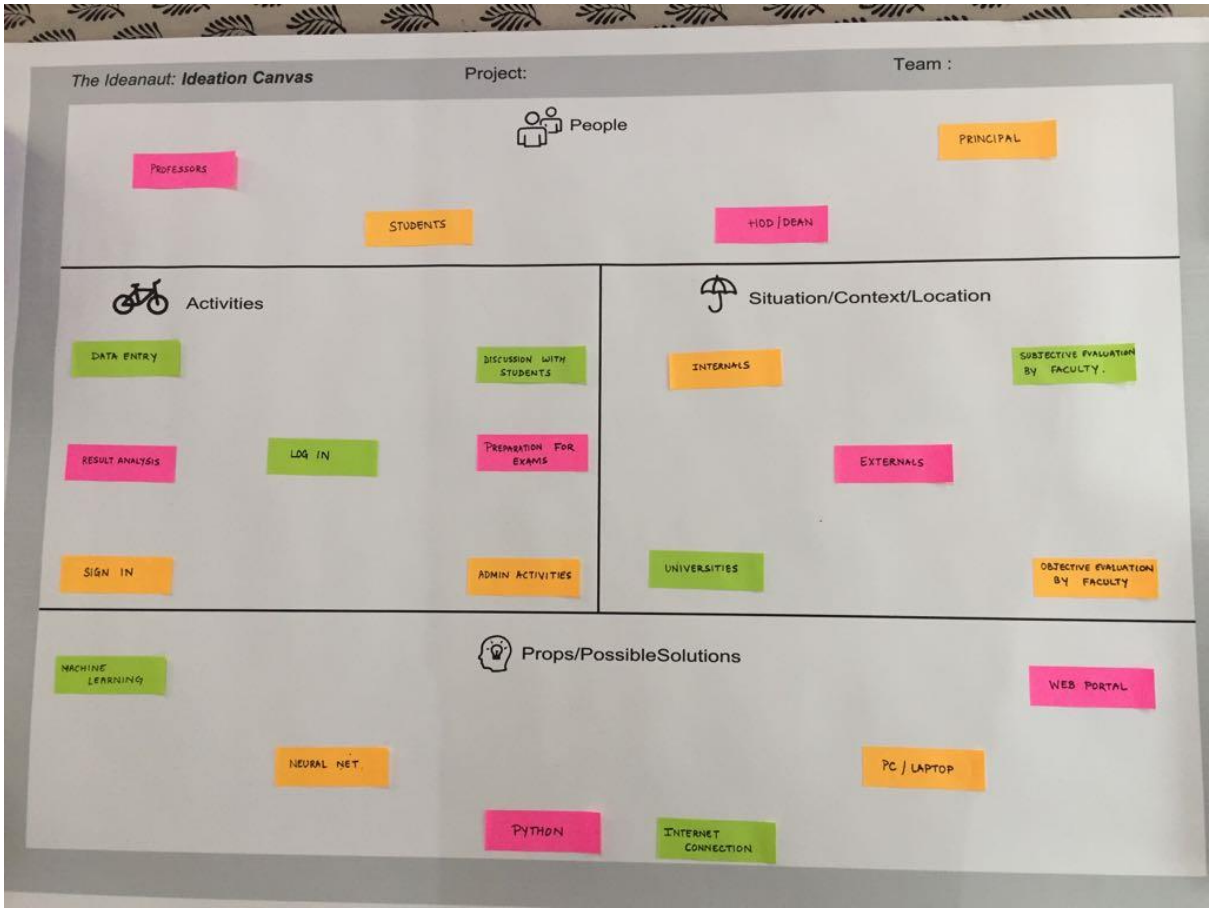


Fig. 2.2 Ideation Canvas

A. People

In this canvas, we listed the people who are going to use this project:

- Professors
- Students
- HOD/Dean
- Principal

B. Activities

After that we listed the activities performed by the above-mentioned people and the Situation and context in which the people are to be assumed.

- Data Entry
- Result Analysis
- Sign In

- Log In
- Discussion with Students
- Preparation for Exams
- Admin Activities

C. Situation/Context/Location

Lastly, we listed the all the situation /context /location for this system:

- Internals
- Externals
- Universities
- Subjective evaluation by faculty
- Objective evaluation by faculty

D. Props/Possible Solutions

Lastly, we listed the all the props/possible solutions for this system:

- Machine learning
- Neural net
- Python
- Internet Connection
- PC/Laptop
- Web Portal

2.3.Product Development Canvas

This canvas consists of eight important aspects of the product that we have decided to create for the well fare of humans.

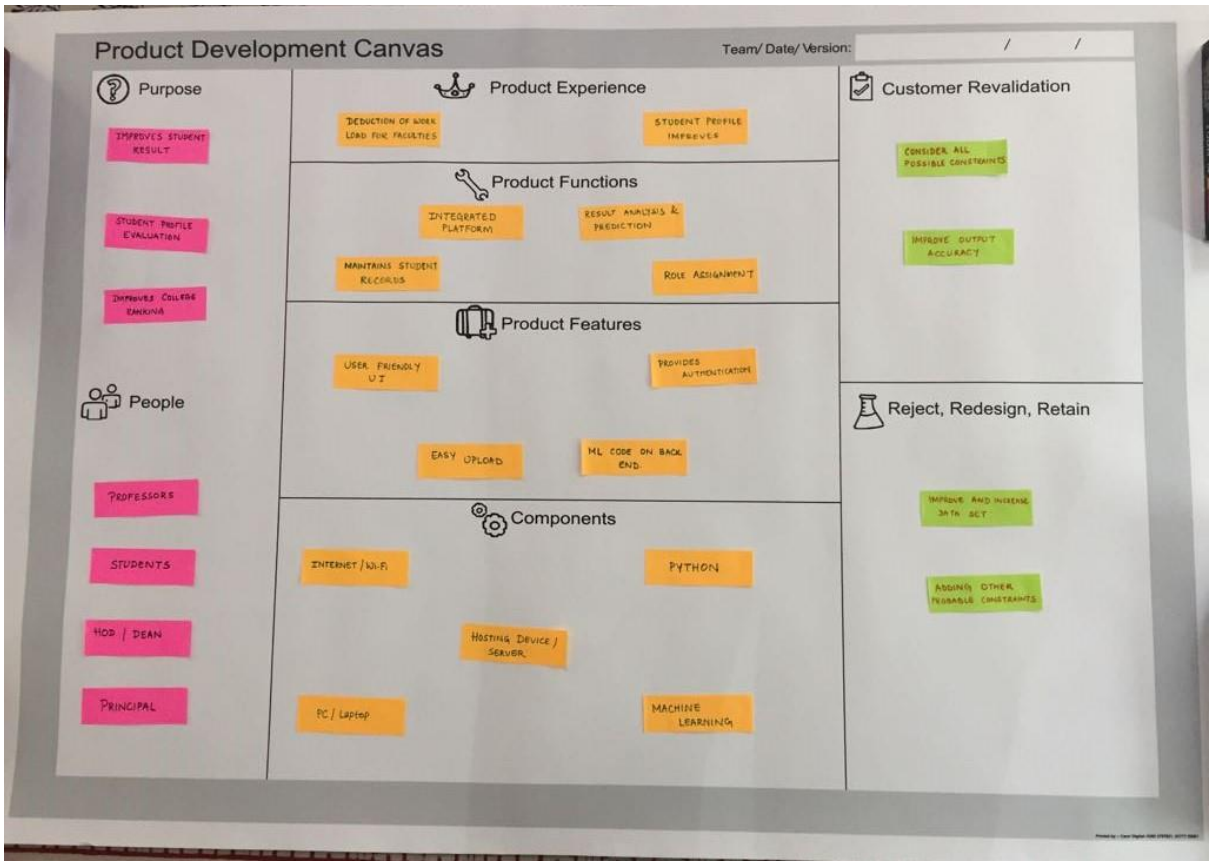


Fig. 2.3 Product Development Canvas

These eight aspects are:

A. Purpose

- Improve Student Result
- Student Profile Evaluation
- Improves College Ranking

B. People

- Professors
- Students
- HOD/Dean
- Principal

C. Product Experience

- Deduction of Work Load for Faculties
- Student Profile Improves

D. Product Functions

- Integrated Platform
- Result Analysis and Prediction
- Maintains Student Records
- Role Assignment

E. Product Features

- User Friendly UI
- Provides Authentication
- Easy Upload
- ML code on the Back End

F. Components

- Internet/Wifi
- Pc/Laptop
- Hosting Device/Server
- Python
- Machine Learning

G. Customer Revalidation

- Consider all Possible Constraints
- Improve Output Accuracy

H. Reject/Redesign/Revalidate

- Improve and Increase Data Set
- Adding Other Probable Constraints

2.4.Empathy Mapping Canvas

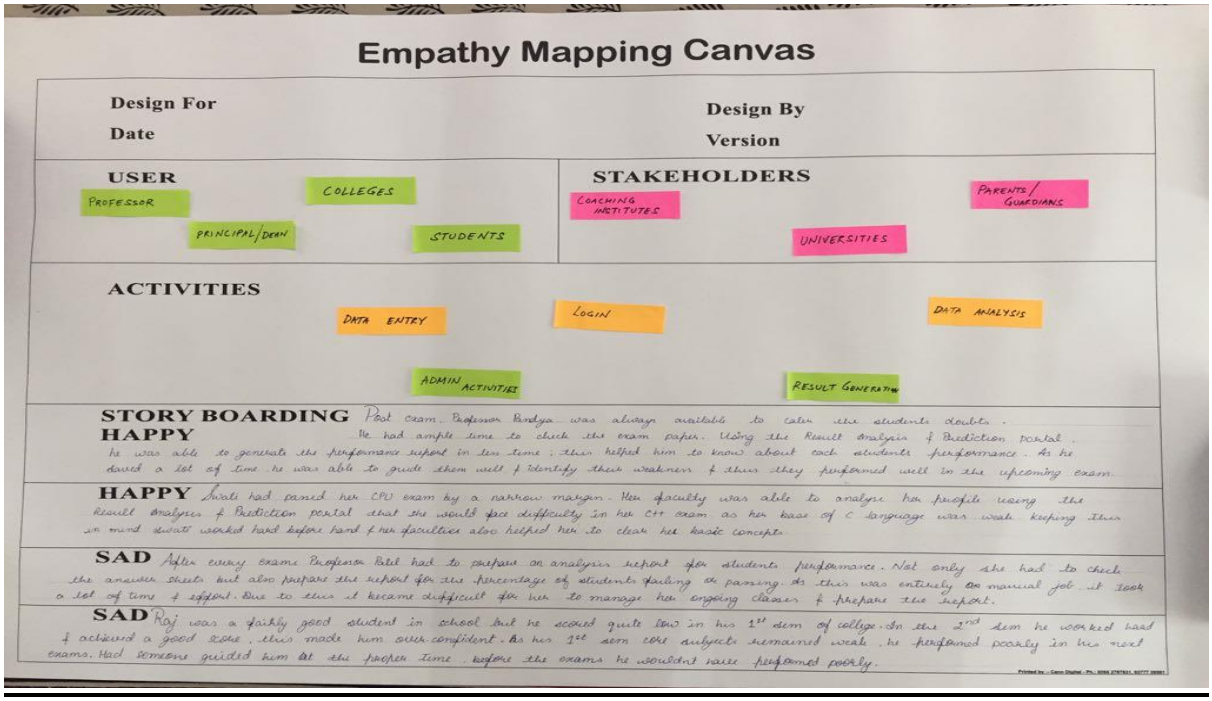


Fig. 2.4 Empathy Mapping Canvas

A. User

- Professors
- Principal
- Dean
- Colleges
- Students

B. Stakeholders

- Coaching Institutes
- Universities
- Parents

C. Activities

- Data Entry
- Admin Activities
- Login
- Result Generation

- Data Analysis

D. Story Boarding

- Happy

  Post exam, Professor Pandya was always available to clear the students doubts. He had ample time to check the exam paper. Using the Result Analysis and Prediction Portal, he was able to generate the performance report in less time; this helped him to know about each student's performance. As he saved a lot of time, he was able to guide them well and identify their weakness and thus they performed well in the upcoming exam.

- Happy

  Swati had passed her CPU exam by a narrow margin. Her faculty could analyse her profile using the Result Analysis and Prediction Portal that she would face difficulty in her C++ exam as her base of C language was weak keeping this in mind Swati worked hard before hand and her faculties also helped her to clear her basic concepts.

- Sad

  After every exam Professor Patel had to prepare an analysis report for student's performance. Not only she had to check the answer sheets but also prepare report for the percentage of students failing or passing. As this was entirely a manual job, it took a lot of time and effort. Due to this it became difficult for her to manage her ongoing classes and prepare the report.

- Sad

  Raj was a good student in school, but he scored quite low in his 1st sem of college. In the 2nd sem he worked hard and achieved a good score, this made him over-confident. As his 1st sem core subjects remained weak, he performed poorly in his next exams. Had someone guided him at the proper time before the exams he would've performed better.

2.5.Business Model Canvas

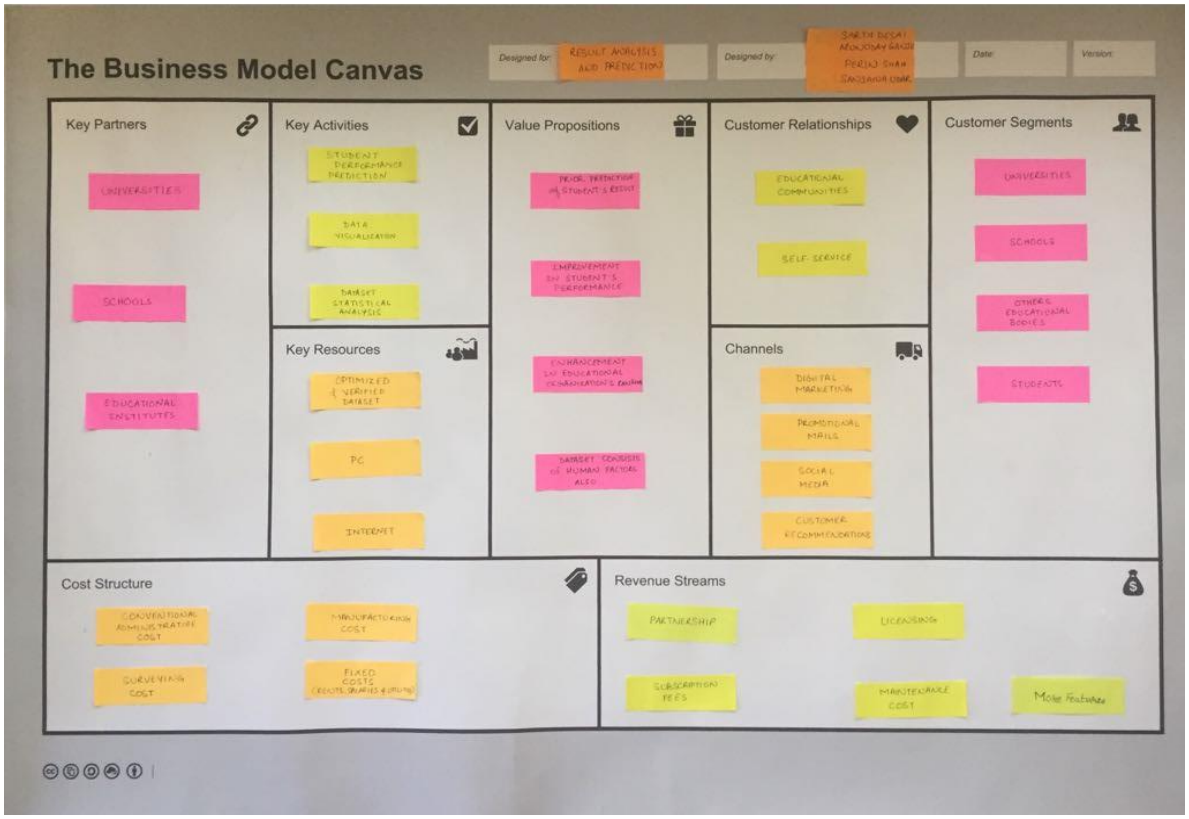


Fig. 2.5 Business Model Canvas

A. Key partners
   - Universities.
   - Schools.
   - Educational Institutes.

B. Key activity
   - Student Performance Prediction
   - Data Visualization
   - Dataset Statistical Analysis

C. Key resources
   - Optimized and Verified Dataset
   - PC
   - Internet

D. Value proposition
   - Prior Prediction of Student's Result.
   - Improvement in Student's Performance.
   - Enhancement in Education Institute's ranking.
   - Dataset Consists of Human Factors also.

E. Customer relationship
    - Educational Communities.
    - Self Service

F. Channels
    - Digital Marketing.
    - Promotional Mails.
    - Social Media.
    - Customer Recommendation.

G. Customer segment
    - Universities
    - Schools
    - Other Educational Bodies
    - Students

H. Cost structure
    - Conventional Administrative Cost
    - Surveying Cost
    - Manufacturing Cost
    - Fixed Costs (Rents, Salaries and Utilities)

I. Revenue stream
    - Partnership
    - Licensing
    - Subscription Fee
    - Maintenance Cost
    - More Features

2.6.Diagrams

A. Flowchart

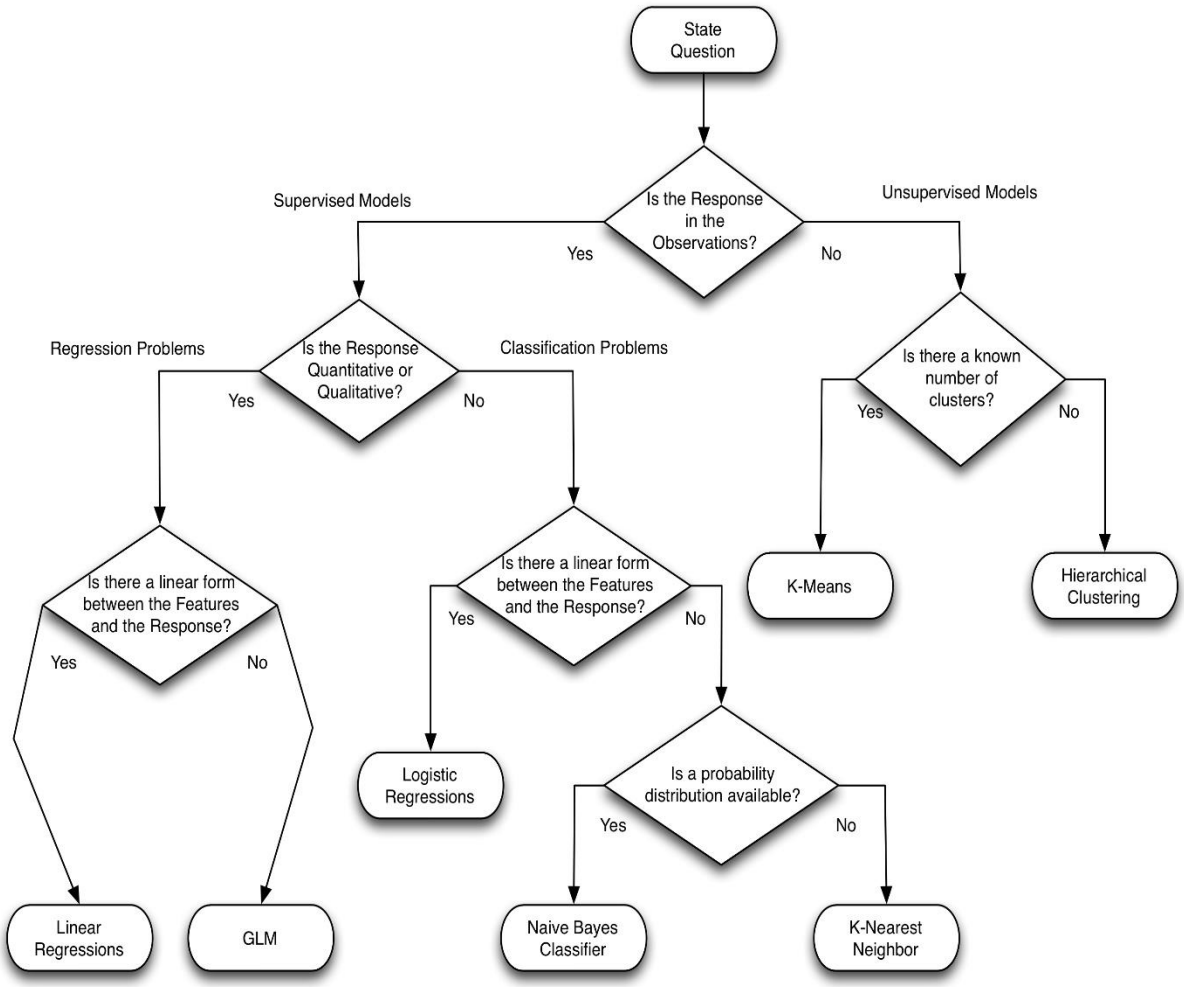


Fig. 2.6.1 Machine Learning flowchart

B. Single Layer Perceptron

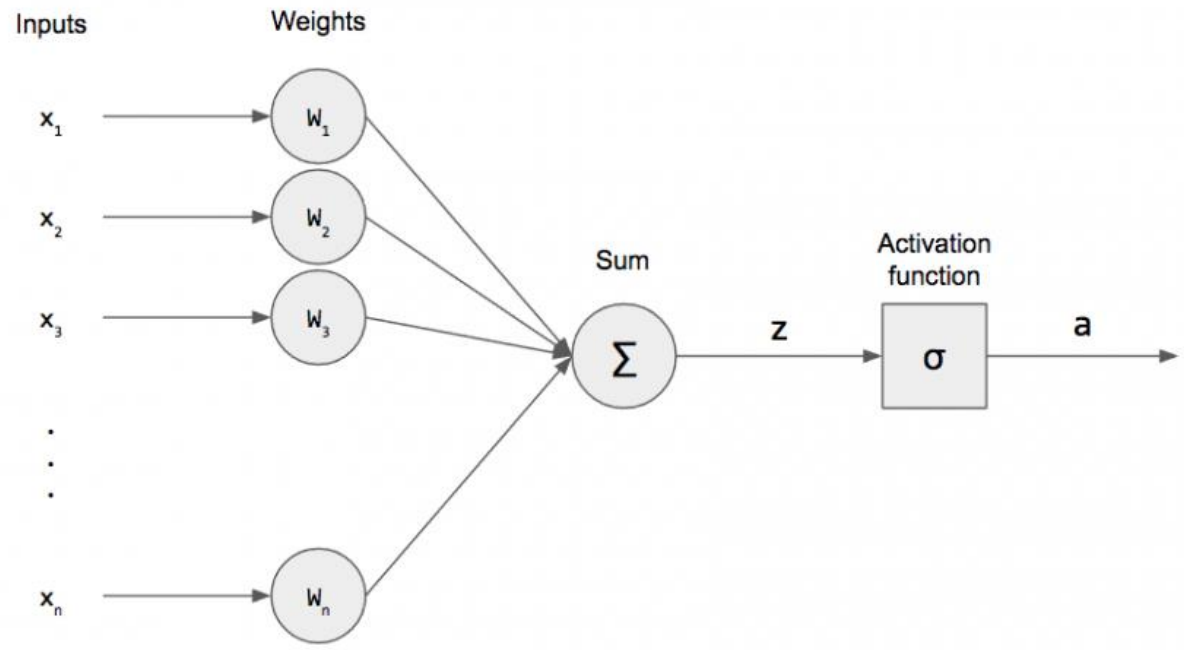


Fig. 2.6.2 Single Layer Perceptron

C. Flow of Support Vector Machine

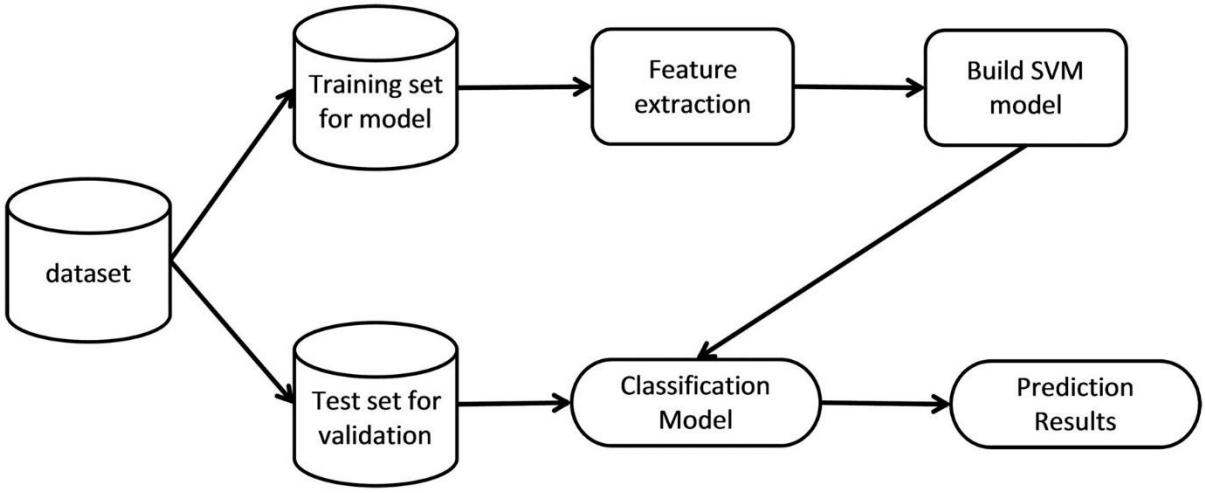


Fig. 2.6.3 Flowchart of SVM

3. Implementation

3.1. Modules and Flow

These are the following modules of the system:

A. Dataset: As per the requirement of the system, it needed real time and correct data to work upon so that the prediction accuracy is as high as possible and reflects the real world. we have adapted a dataset from Kaggle which was created using real time data of the students nearby in that area.

B. Constraint list: These are the various constraints on which the accuracy of the system is dependent.

C. Constraints hierarchy: These are the factors on which the prediction is dependent. Constraint hierarchy is the order on which the prediction is dependent in respective order. So higher the position of a constraint in the order, higher is the dependency of the prediction on that constraint.

D. Statistical Analysis: This consists of various graphs and heat maps. These graphs reflect the result of statistical analysis performed of the Dataset such as inter constraint dependency etc.

E. Algorithmic Results: It displays the constraint matrix of each and every algorithm used for making the prediction. After the matrix the predicted result for the respective algorithm is displayed.

3.1.1. Basic work flow of application:

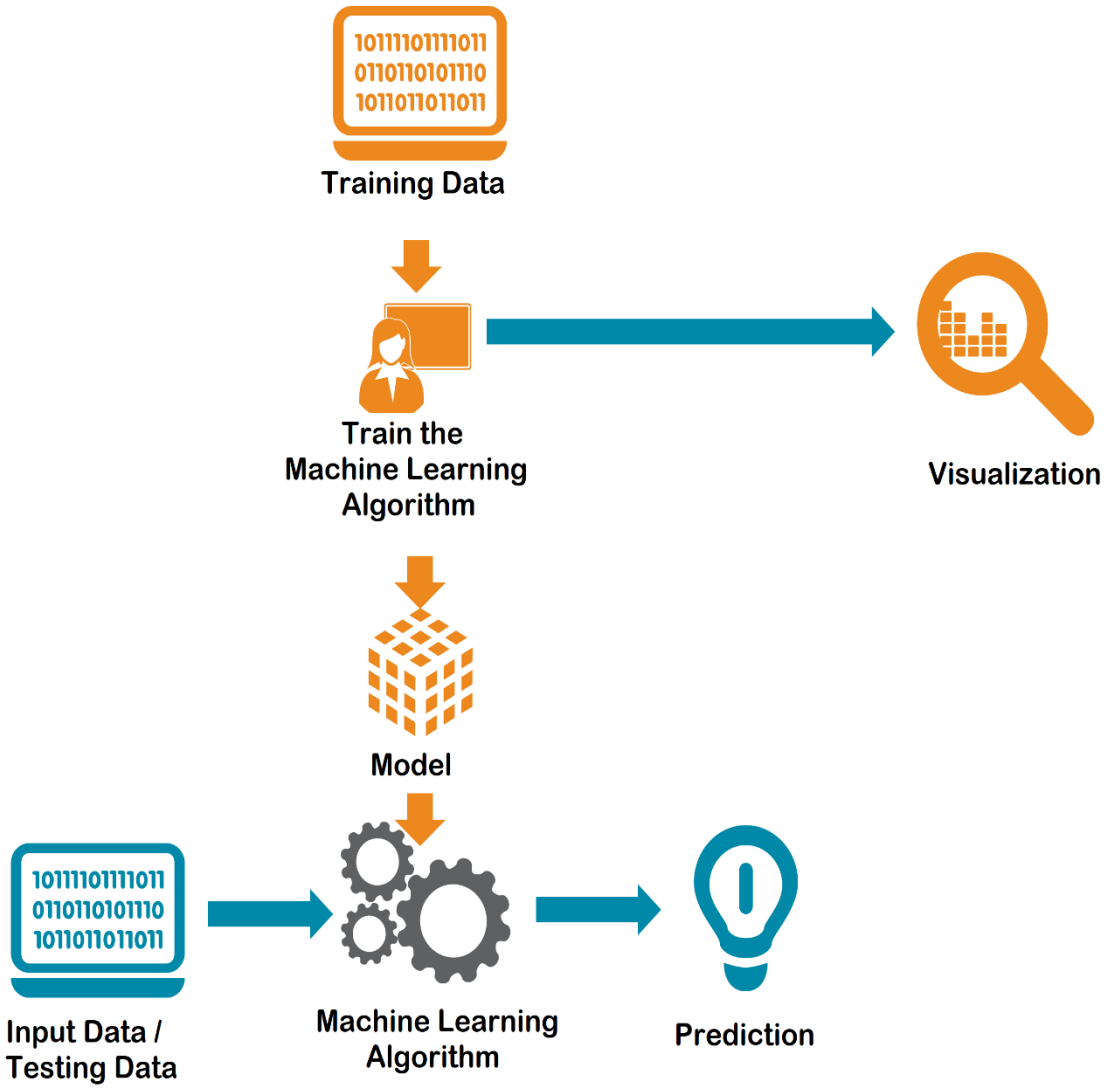


Fig 3.1.1 Flow diagram of the program

3.2. Data Dictionary

3.2.1. Student Performance Dataset

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q
1	gender	NationalITy	PlaceofBirth	StageID	GradeID	SectionID	Topic	Semester	Relation	raisedhands	VisITedResources	AnnouncementsView	Discussion	ParentAnsweringSurvey	ParentschoolSatisfaction	StudentAbsenceDays	Class
2	M	KW	KuwaIT	lowerlevel	G-04	A	IT	F	Father	15	16	2	20	Yes	Good	Under-7	M
3	M	KW	KuwaIT	lowerlevel	G-04	A	IT	F	Father	20	20	3	25	Yes	Good	Under-7	M
4	M	KW	KuwaIT	lowerlevel	G-04	A	IT	F	Father	10	7	0	30	No	Bad	Above-7	L
5	M	KW	KuwaIT	lowerlevel	G-04	A	IT	F	Father	30	25	5	35	No	Bad	Above-7	L
6	M	KW	KuwaIT	lowerlevel	G-04	A	IT	F	Father	40	50	12	50	No	Bad	Above-7	M
7	F	KW	KuwaIT	lowerlevel	G-04	A	IT	F	Father	42	30	13	70	Yes	Bad	Above-7	M
8	M	KW	KuwaIT	MiddleSchool	G-07	A	Math	F	Father	35	12	0	17	No	Bad	Above-7	L
9	M	KW	KuwaIT	MiddleSchool	G-07	A	Math	F	Father	50	10	15	22	Yes	Good	Under-7	M
10	F	KW	KuwaIT	MiddleSchool	G-07	A	Math	F	Father	12	21	16	50	Yes	Good	Under-7	M
11	F	KW	KuwaIT	MiddleSchool	G-07	B	IT	F	Father	70	80	25	70	Yes	Good	Under-7	M
12	M	KW	KuwaIT	MiddleSchool	G-07	A	Math	F	Father	50	88	30	80	Yes	Good	Under-7	H
13	M	KW	KuwaIT	MiddleSchool	G-07	B	Math	F	Father	19	6	19	12	Yes	Good	Under-7	M
14	M	KW	KuwaIT	lowerlevel	G-04	A	IT	F	Father	5	1	0	11	No	Bad	Above-7	L
15	M	lebanon	lebanon	MiddleSchool	G-08	A	Math	F	Father	20	14	12	19	No	Bad	Above-7	L
16	F	KW	KuwaIT	MiddleSchool	G-08	A	Math	F	Mum	62	70	44	60	No	Bad	Above-7	H
17	F	KW	KuwaIT	MiddleSchool	G-06	A	IT	F	Father	30	40	22	66	Yes	Good	Under-7	M
18	M	KW	KuwaIT	MiddleSchool	G-07	B	IT	F	Father	36	30	20	80	No	Bad	Above-7	M
19	M	KW	KuwaIT	MiddleSchool	G-07	A	Math	F	Father	55	13	35	90	No	Bad	Above-7	M
20	F	KW	KuwaIT	MiddleSchool	G-07	A	IT	F	Mum	69	15	36	96	Yes	Good	Under-7	M
21	M	KW	KuwaIT	MiddleSchool	G-07	B	IT	F	Mum	70	50	40	99	Yes	Good	Under-7	H
22	F	KW	KuwaIT	MiddleSchool	G-07	A	IT	F	Father	60	60	33	90	No	Bad	Above-7	M
23	F	KW	KuwaIT	MiddleSchool	G-07	B	IT	F	Father	10	12	4	80	No	Bad	Under-7	M
24	M	KW	KuwaIT	MiddleSchool	G-07	A	IT	F	Father	15	21	2	90	No	Bad	Under-7	M
25	M	KW	KuwaIT	MiddleSchool	G-07	A	IT	F	Father	2	0	2	50	No	Bad	Above-7	L
26	M	KW	KuwaIT	MiddleSchool	G-07	B	IT	F	Father	0	2	3	70	Yes	Good	Above-7	L
27	M	KW	KuwaIT	MiddleSchool	G-07	A	IT	F	Father	8	7	30	40	Yes	Good	Above-7	L
28	M	KW	KuwaIT	MiddleSchool	G-07	B	IT	F	Father	19	19	25	40	Yes	Bad	Under-7	M
29	M	KW	KuwaIT	MiddleSchool	G-08	A	Arabic	F	Father	25	15	12	33	No	Bad	Above-7	L

Fig 3.2.1.1 Student Performance Dataset

Here we are using the Student's Academic Performance Dataset (xAPI-Edu-Data) which consists of 480 unique records and 17 different attributes. The data is collected using a learner activity tracker tool, which called experience API (xAPI). The xAPI is a component of the training and learning architecture (TLA) that enables to monitor learning progress and learner's actions.

Data Description:

- Data Set Characteristics: Multivariate
- Number of Instances: 480
- Area: E-learning, Education, Predictive models, Educational Data Mining
- Attribute Characteristics: Integer/Categorical
- Number of Attributes: 17
- Date: 2016-11-8
- Associated Tasks: Classification
- Missing Values; No
- File formats: .csv

Source:

Elaf Abu Amrieh, Thair Hamtini, and Ibrahim Aljarah, The University of Jordan, Amman, Jordan, http://www.Ibrahimaljarah.com www.ju.edu.jo

```
(480, 17)


gender : (Categorical) 2 unique value(s) - ['F' 'M']
NationalITy : (Categorical) 14 unique value(s) - ['Egypt' 'Iran' 'Iraq' 'Jordan' 'KW' 'Lybia' 'Morocco' 'Palestine'
 'SaudiArabia' 'Syria' 'Tunis' 'USA' 'lebanon' 'venzuela']
PlaceofBirth : (Categorical) 14 unique value(s) - ['Egypt' 'Iran' 'Iraq' 'Jordan' 'KuwaIT' 'Lybia' 'Morocco' 'Palestine'
 'SaudiArabia' 'Syria' 'Tunis' 'USA' 'lebanon' 'venzuela']
StageID : (Categorical) 3 unique value(s) - ['HighSchool' 'MiddleSchool' 'lowerlevel']
GradeID : (Categorical) 10 unique value(s) - ['G-02' 'G-04' 'G-05' 'G-06' 'G-07' 'G-08' 'G-09' 'G-10' 'G-11' 'G-12']
SectionID : (Categorical) 3 unique value(s) - ['A' 'B' 'C']
Topic : (Categorical) 12 unique value(s) - ['Arabic' 'Biology' 'Chemistry' 'English' 'French' 'Geology' 'History' 'IT'
 'Math' 'Quran' 'Science' 'Spanish']
Semester : (Categorical) 2 unique value(s) - ['F' 'S']
Relation : (Categorical) 2 unique value(s) - ['Father' 'Mum']
raisedhands : (Continuous) range of values -  [ 0 to 100]  { 82  values}
VisITedResources : (Continuous) range of values -  [ 0 to 99]  { 89  values}
AnnouncementsView : (Continuous) range of values -  [ 0 to 98]  { 88  values}
Discussion : (Continuous) range of values -  [ 1 to 99]  { 90  values}
ParentAnsweringSurvey : (Categorical) 2 unique value(s) - ['No' 'Yes']
ParentschoolSatisfaction : (Categorical) 2 unique value(s) - ['Bad' 'Good']
StudentAbsenceDays : (Categorical) 2 unique value(s) - ['Above-7' 'Under-7']
Class : (Categorical) 3 unique value(s) - ['H' 'L' 'M']
```

Fig. 3.2.1.2 Dataset attributes

```
Feature ranking:
1. feature 13 ParentAnsweringSurvey (0.136028)
2. feature 9 raisedhands (0.130080)
3. feature 10 VisITedResources (0.125761)
4. feature 12 Discussion (0.109986)
5. feature 11 AnnouncementsView (0.092082)
6. feature 6 Topic (0.065035)
7. feature 15 Class (0.050670)
8. feature 4 GradeID (0.047136)
9. feature 2 PlaceofBirth (0.046140)
10. feature 1 NationalITy (0.044464)
11. feature 8 Relation (0.030515)
12. feature 16 Failed (0.030388)
13. feature 5 SectionID (0.024950)
14. feature 3 StageID (0.024619)
15. feature 0 gender (0.018015)
16. feature 14 StudentAbsenceDays (0.017072)
17. feature 7 Semester (0.007060)
```

Fig. 3.2.1.3 Attributes ranking

3.3.Screenshots

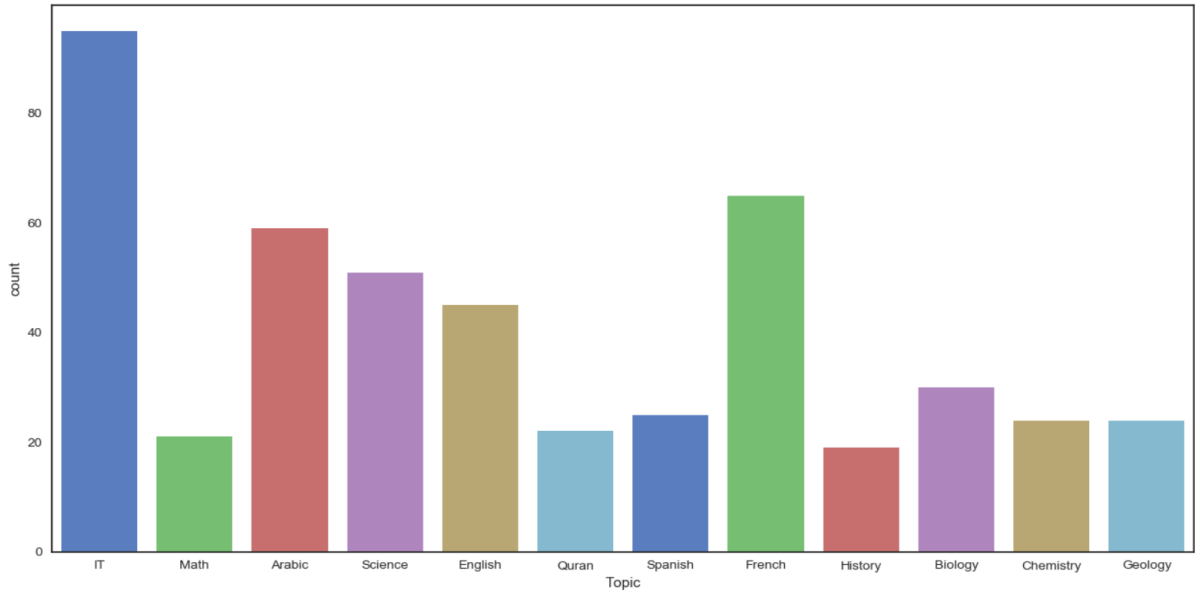


Fig 3.3.1 Count vs Topic

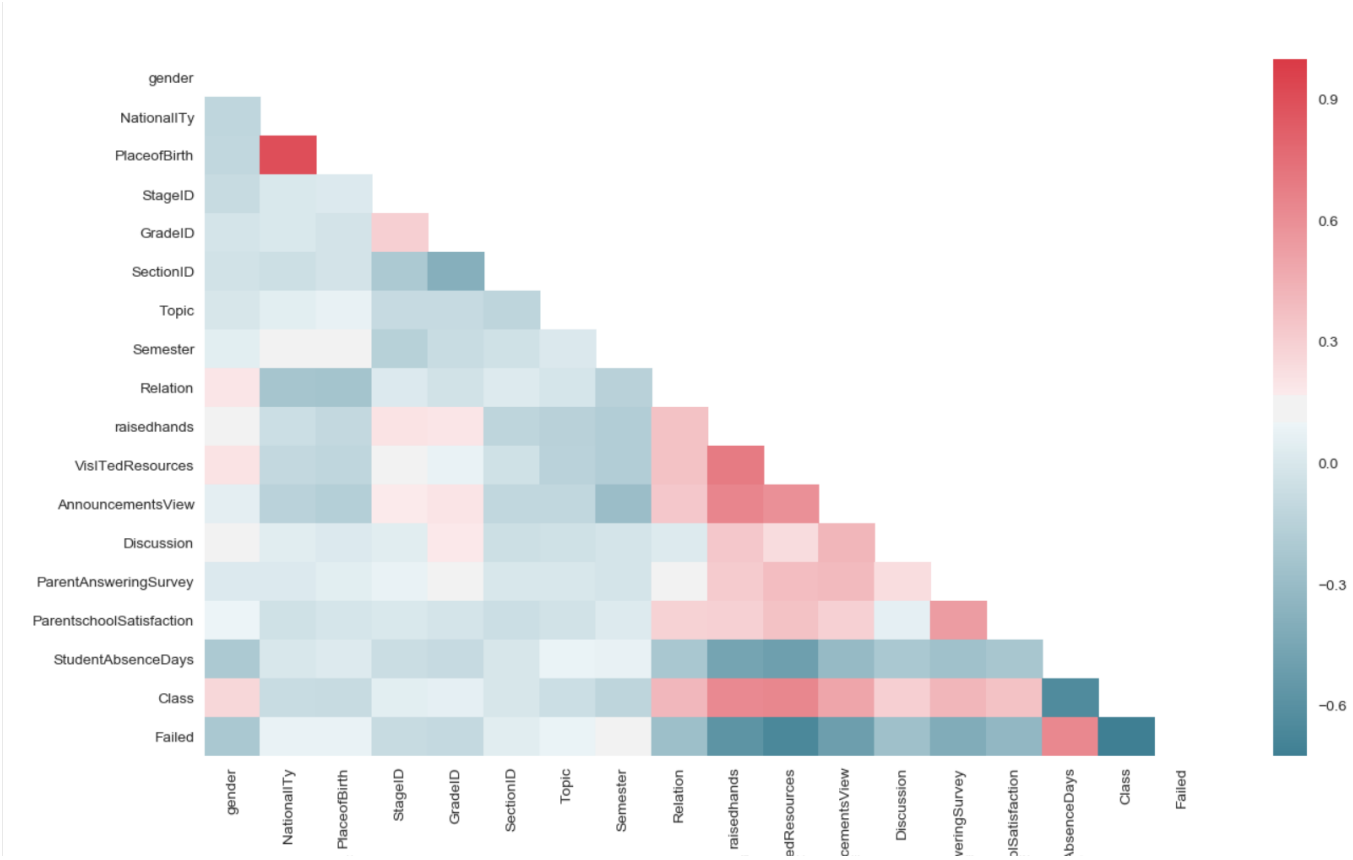


Fig 3.3.2 Attributes heatmap

```
----------------------------Perceptron-------------------------------
             precision    recall  f1-score   support

          H       0.55      0.61      0.58        28
          L       0.75      0.80      0.77        30
          M       0.55      0.47      0.51        38

avg / total       0.61      0.61      0.61        96

Misclassified samples: 37
Accuracy: 0.61
```

Fig 3.3.3 Perceptron

```
-------------------------------SVC_linear-----------------------------
             precision    recall  f1-score   support

          H       0.70      0.55      0.61        55
          L       0.74      0.85      0.79        33
          M       0.56      0.62      0.59        56

avg / total       0.65      0.65      0.64       144

Misclassified samples: 51
Accuracy: 0.65


-------------------------------SVC_rbf--------------------------------
             precision    recall  f1-score   support

          H       0.63      0.49      0.55        35
          L       0.82      0.68      0.74        34
          M       0.64      0.77      0.70        66

avg / total       0.68      0.67      0.67       135

Misclassified samples: 44
Accuracy: 0.67


-------------------------------SVC_poly-------------------------------
             precision    recall  f1-score   support

          H       0.55      0.51      0.53        35
          L       0.77      0.88      0.82        34
          M       0.70      0.67      0.68        66

avg / total       0.68      0.68      0.68       135

Misclassified samples: 43
Accuracy: 0.68
```

Fig 3.3.4 Support Vector Machine (linear, rbf, poly)

```
-------------------------SC---------------------------------
             precision    recall  f1-score   support

          H       0.70      0.57      0.63        28
          L       0.79      0.96      0.86        23
          M       0.73      0.73      0.73        45

avg / total       0.73      0.74      0.73        96

Misclassified samples: 25
Accuracy: 0.74


Mean Accuracy: 0.68 (+/- 0.08)
```

Fig 3.3.5 StandardScaler

```
----------------------Logistic Regression----------------------


Prediction:
 ['L' 'M' 'M' 'M' 'M' 'H' 'M' 'L' 'M' 'H' 'L' 'M' 'L' 'M' 'M' 'L' 'H' 'M'
 'H' 'H' 'M' 'L' 'L' 'M' 'H' 'M' 'M' 'M' 'H' 'H' 'M' 'L' 'L' 'L' 'H' 'M'
 'H' 'H' 'M' 'M' 'H' 'M' 'L' 'H' 'M' 'L' 'M' 'H' 'L' 'M' 'M' 'L' 'M' 'L'
 'M' 'L' 'H' 'L' 'H' 'L' 'H' 'L' 'H' 'L' 'H' 'M' 'M' 'M' 'L' 'M' 'M' 'H'
 'H' 'H' 'M' 'M' 'L' 'M' 'H' 'L' 'H' 'H' 'H' 'H' 'H' 'M' 'M' 'L' 'H' 'M'
 'M' 'M' 'L' 'M' 'L' 'M']


             precision    recall  f1-score   support

          H       0.69      0.71      0.70        28
          L       0.92      0.86      0.89        28
          M       0.71      0.72      0.72        40

avg / total       0.77      0.76      0.76        96

Accuracy:  0.760416666667
```

Fig. 3.3.6 Logistic Regression

```
----------------------------XGBoost-------------------------------


Prediction:
 ['M' 'L' 'L' 'M' 'M' 'M' 'H' 'L' 'L' 'M' 'H' 'M' 'M' 'L' 'M' 'H' 'M' 'L'
 'M' 'H' 'L' 'H' 'M' 'M' 'M' 'H' 'M' 'M' 'M' 'L' 'M' 'L' 'H' 'M' 'M' 'M'
 'H' 'H' 'M' 'M' 'M' 'L' 'M' 'H' 'M' 'M' 'H' 'M' 'L' 'L' 'M' 'L' 'H' 'L'
 'M' 'M' 'L' 'L' 'H' 'M' 'M' 'M' 'H' 'H' 'H' 'M' 'M' 'M' 'L' 'L' 'L' 'L'
 'M' 'L' 'M' 'L' 'M' 'L' 'M' 'M' 'H' 'M' 'M' 'L' 'L' 'H' 'M' 'H' 'M' 'H'
 'M' 'H' 'H' 'H' 'M' 'M']


             precision    recall  f1-score   support

          H       0.78      0.64      0.71        28
          L       0.88      0.96      0.92        23
          M       0.79      0.84      0.82        45

avg / total       0.81      0.81      0.81        96

Accuracy:  0.8125
```

Fig. 3.3.7 Extreme Gradient Boosting

4. Summary

After successful completion of the project system, user will experience easy way to understand the dataset that is provided to the system. User will get the information about the student's performance on the basis of the data bestowed to the framework. User can also scrutinize the data by statistical and analytical reports and visualization by using different types of bar graphs and heat maps that are produced by the system.

4.1 Advantages of the System

A. All the work that was done earlier manually will now be done by the system saving a lot of valuable time.
B. Students can be characterized according to their performance so that it will be easier for the faculties to help them respectively.
C. Predicted result will help the students to work hard beforehand on their weak subjects so that their result will improve.
D. The system will not only help the students to improve their performance but also be a catalyst in improving the college ranking.

4.2 Unique Features

A. It perfectly describes the dataset such that user can easily deduce the gist.
B. The faculty can upload the data of a single student or a group of students through the medium of an excel sheet.
C. It presents the prediction with high amount of accuracy.
D. Every attributes are meticulously scrutinized and envisaged.

4.3 Conclusion and Scope of further Work

Conclusion

The aim of the project is prediction of the student's efficacy result on the basis of result analysis of other students. This system would be a great help to students and faculties. It would help them to improve and score high in the examination. Not only that it would be a great help to the faculties and college authorities to help them to device new methods to help and guide the students of have a tendency in scoring less in the upcoming exams. Also, it would help the administrator/user to determine the integrity of the dataset.

Scope

In the future, we will implement some more modules in our system to make our artifice more user friendly and convenient. We will implement Graphical User Interface using Tkinter and other various supported GUI modules of python, which will also make our system interaction easy and efficient for the user.

5. REFERENCES

[1] Naren J, “Application of Data Mining in Educational Database for Predicting Behavioral Patterns of the Students” IJCSIT, Vol.5 No.03 2014.

[2] Nitya Upadhyay, Vinodini Katiyar, “A Survey on the International Journal of Computer Applications Technology and Research, Vol.3 No.11, 2014

[3] U. bin Mat, N. Buniyamin, P. M. Arsad, R. Kassim, An overview of using academic analytics to predict and improve students’ achievement:A proposed proactive intelligent intervention, in: Engineering Education (ICEED), 2013 IEEE 5th Conference on, IEEE, 2013, pp.126–130.

[4] Shreenath Acharya, Madhu N,”Discovery of students’ academic patterns using data mining techniques” IJCSE,Vol. 4 No. 06 June 2012.

[5] T. Mishra, D. Kumar, S. Gupta, “Mining students’ data for prediction performance”, in: Proceedings of the 2014 Fourth International Conference on Advanced Computing & Communication Technologies, ACCT ’14, IEEE Computer Society, Washington, DC, USA, 2014, pp. 255–262.

[6] Diego Garca-Saiz and Marta Zorrilla, “Comparing classification methods for predicting distance students' performance”:MLR: Workshop and Conference Proceedings 17 (2011) 26-32 2nd Workshop on
Applications of Pattern Analysis.

[7] G.Narasinga Rao, Srinivasan Nagaraj, “A Study on the Prediction of Student’s Performance by applying straightline regression analysis using the method of least squares”:G.Narasinga Rao et al. / International Journal of Computer Science Engineering (IJCSE).

[8] Vaneet Kumar, Dr. Vinod Sharma, “Student’s Examination Result Mining: A Predictive Approach” International Journal of Scientific & Engineering Research, Vol.3, No.11 November 2014.

[9] Amrieh, E. A., Hamtini, T., & Aljarah, I. (2016). Mining Educational Data to Predict Student’s academic Performance using Ensemble Methods. International Journal of Database Theory and Application, 9(8), 119-136.

[10] Amrieh, E. A., Hamtini, T., & Aljarah, I. (2015, November). Preprocessing and analyzing educational data set using X-API for improving student's performance. In Applied Electrical Engineering and Computing Technologies (AEECT), 2015 IEEE Jordan Conference on (pp. 1-5). IEEE.

[11] M Georgiopoulos, I Dagher, G L Heileman, and G Bebis. Properties of learning of a fuzzy ART variant. Neural networks, 837-850, 1999.

[12] Hong, Richang and Wang, Meng and Gao, Yue and Tao, Dacheng and Li, Xuelong and Wu, Xindong. Image annotation by multiple-instance
learning with discriminative feature mapping and selection, IEEE transactions on cybernetics, 44, 5, 669-680, IEEE, 2014.

[13] Krebs, Hermano I and Krams, Michael and Agrafiotis, Dimitris K and DiBernardo, Allitia and Chavez, Juan C and Littman, Gary S and
Yang, Eric and Byttebier, Geert and Dipietro, Laura and Rykman, Avrielle and others. Robotic measurement of arm movements after stroke
establishes biomarkers of motor recovery, Stroke, 45, 1, 200-204, Am Heart Assoc, 2014.

[14] Wu, Xindong and Zhu, Xingquan and Wu, Gong-Qing and Ding, Wei. Data mining with big data, IEEE transactions on knowledge and data
engineering, 26, 1, 97-107, IEEE, 2014.

[15] Biggio, Battista and Fumera, Giorgio and Roli, Fabio. Security evaluation of pattern classifiers under attack, IEEE transactions on knowledge
and data engineering, 26, 4, 984-996, IEEE, 2014.

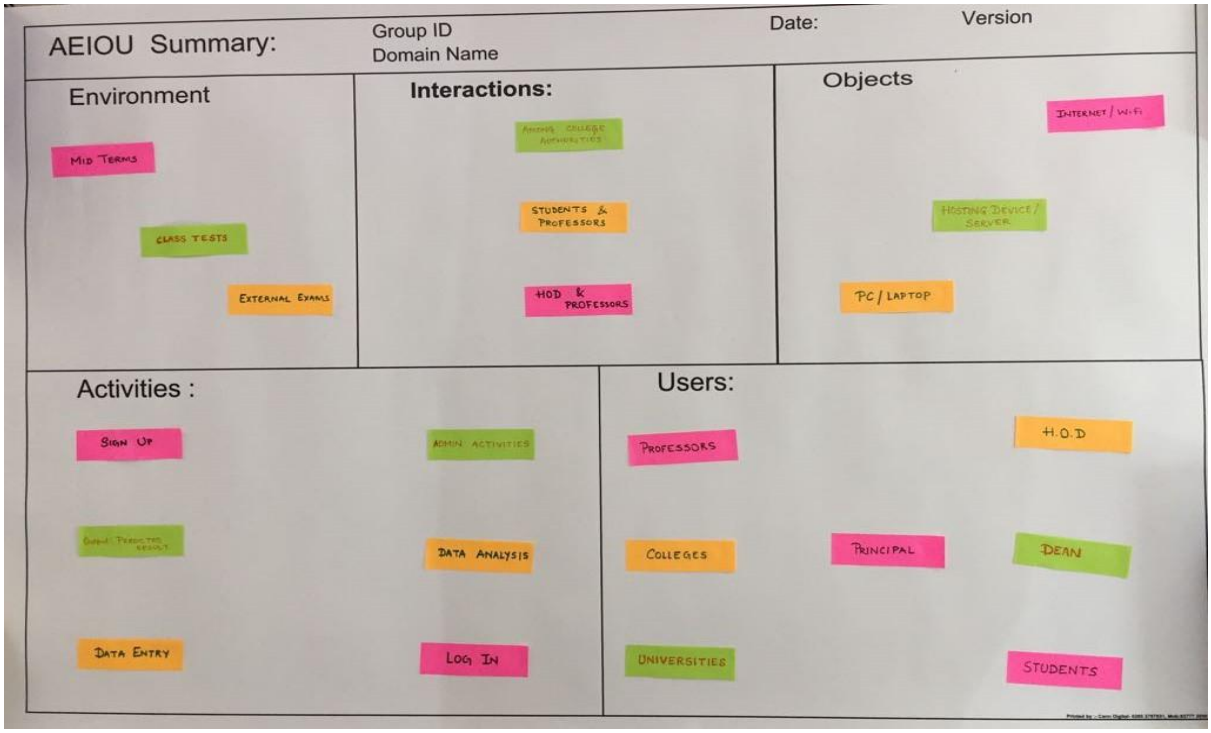
[16] Cloud based intelligent system for delivering health care as a service, Kaur, Pankaj Deep and Chana, Inderveer, Computer methods and
programs in biomedicine, 113, 1, 346-359, Elsevier, 2014.

[17] White, Ryen W. Interactions with search systems. Cambridge University Press, 2016.

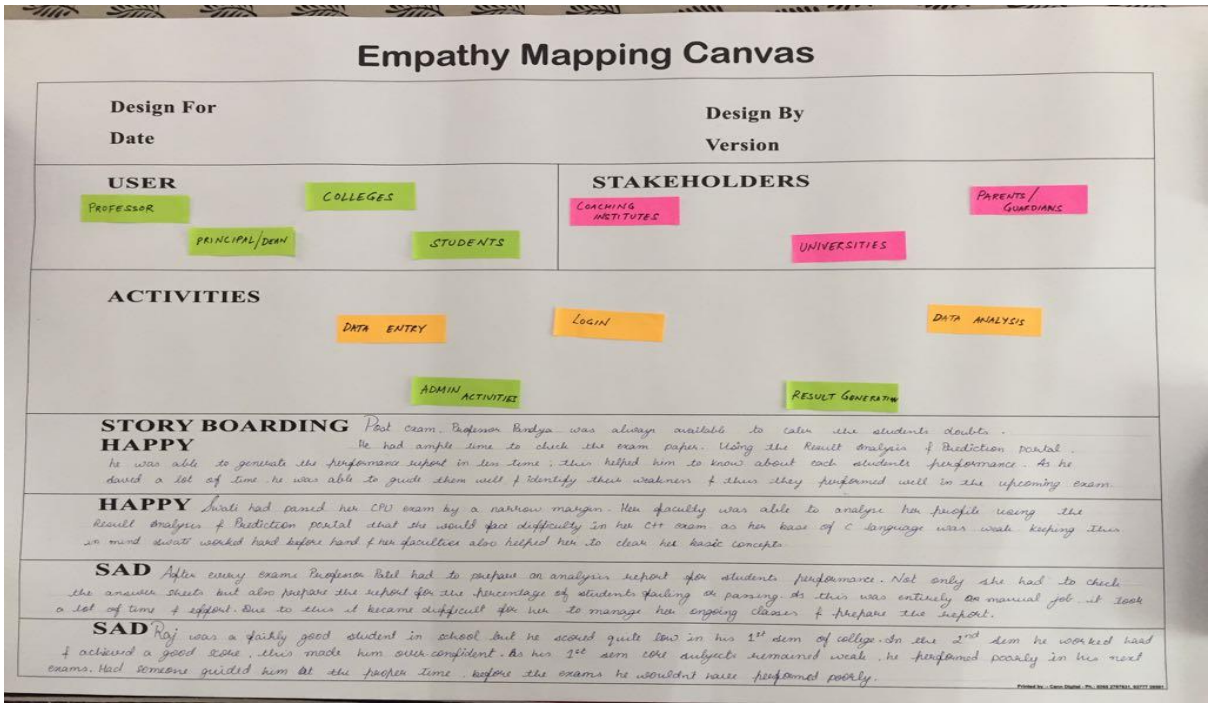
Appendix

- **Design Engineering Canvas (DEC)**

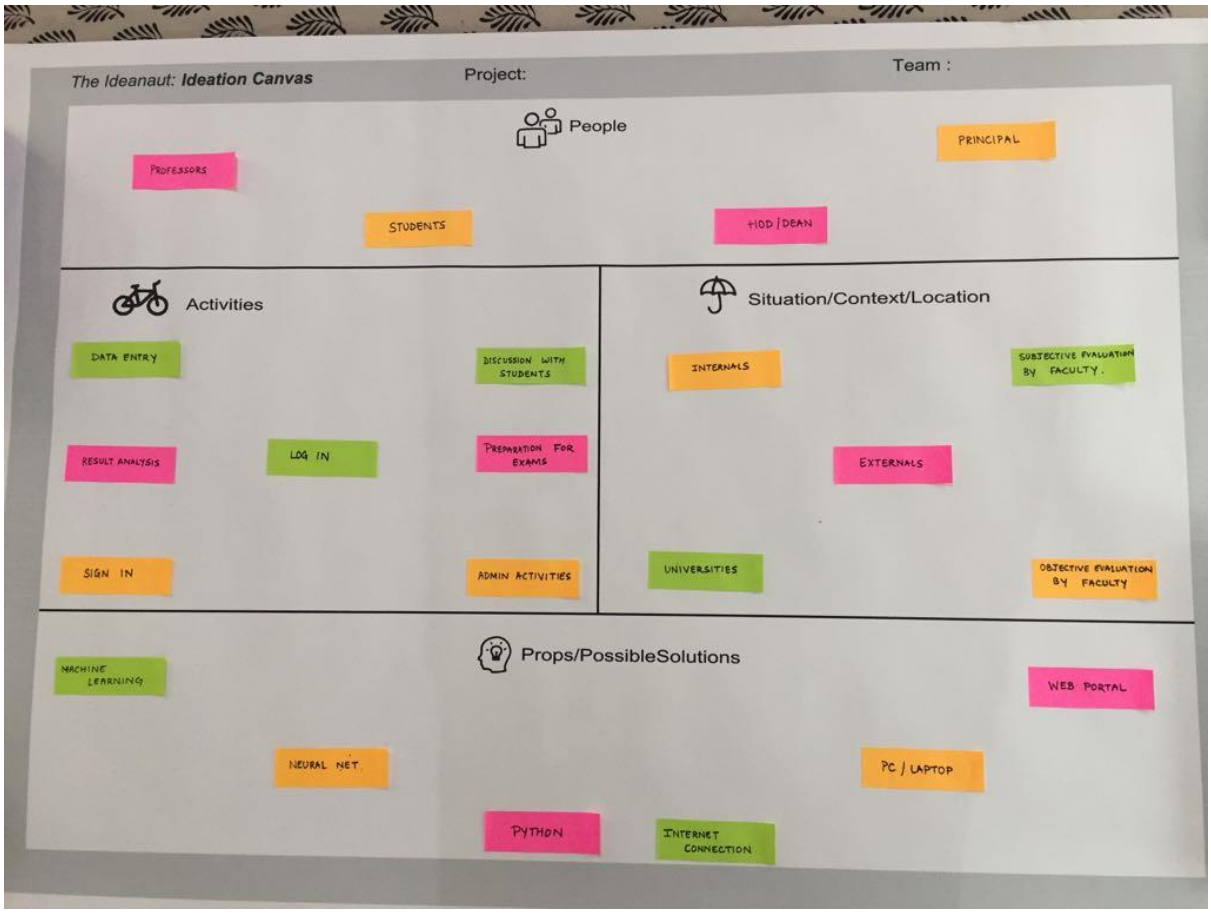
AEIOU Summary Canvas



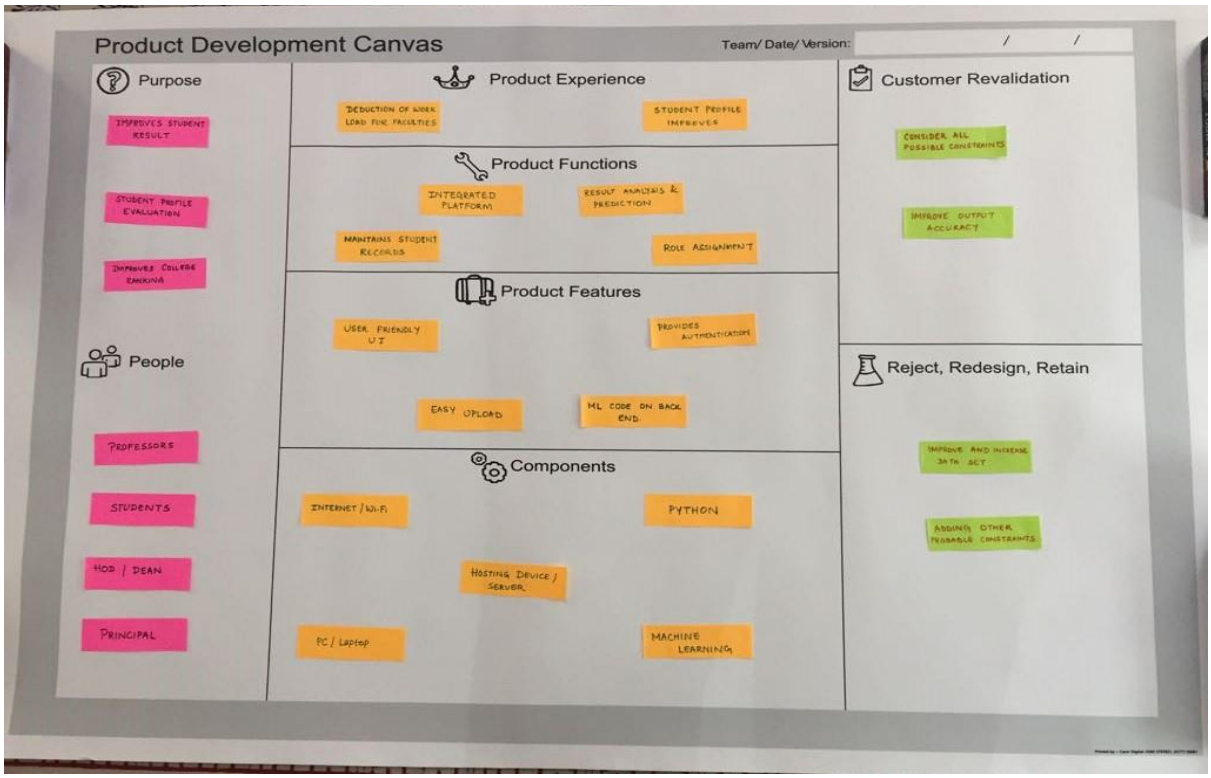
Empathy Summary Canvas



Ideation Canvas



Product Development Canvas



Business Model Canvas