

# Grade Prediction of Educational Institute Using ML Algorithm

Sushmita Gaonkar

## ABSTRACT

A massive amount of data is reproduced across numerous pursuits such as education, medical science, defenses, social media, and so on and so forth. Machine Learning (ML) and Data Mining (DM) are techniques that can be used to identify and improve the hidden patterns automatically through experience seen as a subset of Artificial intelligence. One of the key areas of this application is Educational Data Mining(EDM) which uses ML and statistics to extract large repositories of data associated with learning activities. These learning management systems are majorly used to predict college grades. The proposed model is built to predict the future grade of colleges and universities, established on the current activities they execute. Machine learning algorithms are found to be very practical and effective. It is the most valuable under circumstances where the individual doesn't have an adequate amount of knowledge. ML algorithm predicts the future based on the input given to it, it investigates and analyzes given input data. They are trained based on it and infer a hypothesis /theory. The proposed model has used the Random Forest regression (RFR) algorithm which will help colleges to priorly know the grades and if these grades are less than what they anticipated they can improve them by enriching the current activities.

**Keywords :** Machine Learning, Data Mining, AI, Random forest algorithm, Predictions, Educational Data mining, hypothesis

## INTRODUCTION

Large amount of data is produced across a variety of fields. The Big Data is used to collect and organize data and valuable data need to be extracted. We need a large amount of data to analyze the useful information. In order to get that we use Data Mining and Machine Learning Techniques to build a model which can analyze the patterns and give the required prediction or the results. Educational Data Mining (EDM) collects the data about the organization and learns their environment to provide a useful model to identify the patterns from the available data.

Educational organization is always a concern about their grades and ranking of colleges, so by using Machine Learning and Data Mining techniques we

have developed a model which can help them to improve their college grades.

The prediction models will not only give the prediction about the grades of the college, but will also provide real time advice to resolve their difficulties. The key difference between human and computer is the ability to think and learn from experience. Humans learn from the experiences at other hand Computers execute human-made algorithms to get the knowledge, get trained and make the predictions. Machine learning algorithms are found to be very effective to particular type of learning. They are mostly useful in circumstances where the human doesn't have sufficient amount of knowledge.

ML algorithm predict the future based on the input given to it, it explores given input data and get trained on it and produce hypotheses.

### PROBLEM DEFINITION

Performance measurement not only plays important role in business world but also in the educational world. These quantities can be used to get the success, in the world of education by improving the quality of education. In order to get the accreditation grade to college, organization need to improve the quality of the education.

Accreditation has seven criteria's

- (1) Vision, mission, goals and objectives, as well as the achievement of the strategy;
- (2) Teachers ' governance, leadership, management systems, and quality assurance;
- (3) Students and graduates;
- (4) Human resources
- (5) Curriculum, learning, and academic atmosphere;
- (6) Financing, means and infrastructure, as well as information systems;
- (7) Research and devotion to society, and cooperation.

### PROJECT PURPOSE

In order to get Accreditation Grades College has to get idea where their college grades stands. To get the prior knowledge about the college status we are using machine learning models to predict the future grades of the college so that they can improve based on the predicted output.

### FUNCTIONAL REQUIREMENTS

In software engineering software components or software system is defined in the functional requirement. A function can be described as set of input data, behavior and the output. Functional requirements includes technical details, data

manipulation, calculations and processing, and also specifies what the major goal of the system or model is.

Here, the system perform the following tasks:

- Model has to get response from user
- Model has to predict the grades
- Model has to suggests the improvement in the act ivies carried out in college

### NON-FUNCTIONAL REQUIREMENTS

In non-functional requirement we are not concern about the specific behavior or the functional behavior. Rather work on how the system operates

In Other words non-functional requirements are "quality attributes", "constraints", "quality of service requirements", "quality goals", and "non-behavioral requirements".

Some of the quality attributes are as follows:

#### ACCESSIBILITY:

Accessibility is describing the degree at which a model, device, product, service, or environment can be accessed by the people.

In our case authorized organization members can use the model User interface is simple, efficient and easy to use.

#### MAINTAINABILITY:

Maintainability is an important module in software engineering once the project is developed it need to be maintained.

- Correct defects
- Update the versions
- Meet new requirements

#### SCALABILITY:

System can work normally when given huge about of data, as well as when data size is less.

## DESIGN GOALS

To build a model that predict the grades of the college based on the given data-set. Which will help college to improve their grades or the ranks of the college.

## INPUT/OUTPUT PRIVACY

Data is collected from the college self-study report which submitted by respective colleges so data is not manipulated and results are given to authorize members only.

## SYSTEM ARCHITECTURE

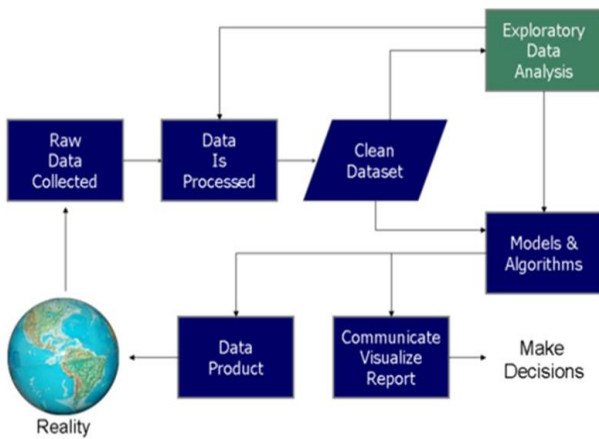


Figure 1: System Architecture

## IMPLIMENTATION

To develop the prediction model following steps need to be carried out

1. Problem Statement
2. Collection of data
3. Preprocessing
4. Data Mining
5. Machine Learning Algorithm
6. Result Evaluation
7. Knowledge flow representation

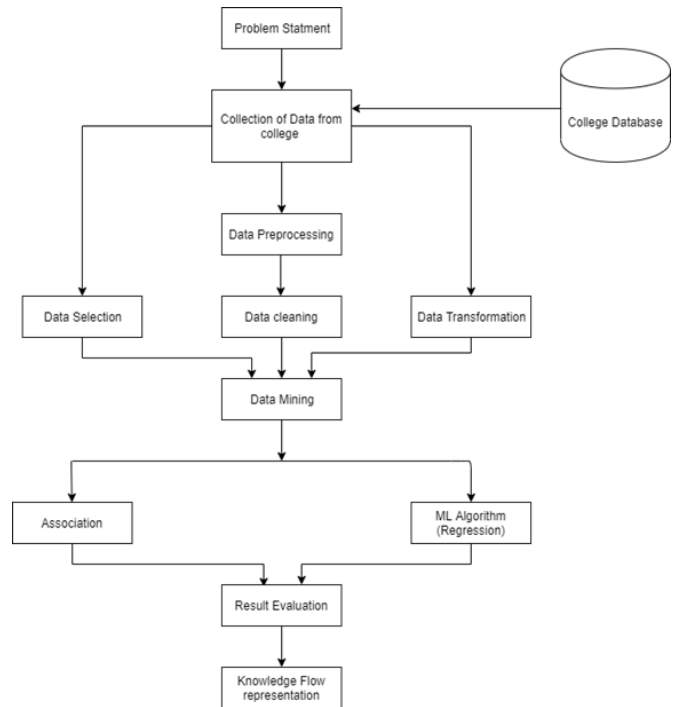


Figure 2: data flow

### Problem Statement:

Here we are looking forward to predict the college future grades based on their current activities they perform. Which will help the organization to improve their grade by improving the current activities.

### Data Set

Once we get the problem statement, the next process is to collect the relevant data. We have collected the data from the internet sources, where colleges upload their self-study report (SSR).Based on the SSR which college has uploaded we have accessed the needed data.

### Data Pre-processing

Once we collected the raw data we need to process it, before we do any analysis on that data. Usually data from such sources will be very huge, messy, duplicate data can be present, there can be some missing values. In order to get the accurate analysis we had to cross check the data.

The common errors which we came across:

- Missing values
- Unreverent data for our analysis such as address and the names of the students and staff etc.
- Invalid entries

### Data Mining

Once we get the data cleaned and transform it to required format. We had to closely examine the factors which will directly affect the result and which indirectly affect so keeping both in mind we begin the further process to analyze the patterns more deeply. We have divided the whole data into 60:40 where 60% of data is taken as training set and the remaining 40% is used for testing.

### Machine Learning Algorithm

At this step we have applied all our mathematical, statistical and technological knowledge to apply best fit Machine Learning algorithm to get the predictive model that predicts the grade of the college. From qualitative analysis we got the quantitative data and applied the machine learning algorithm on the quantitative data. Here we used random forest algorithm to predict the analysis. As we are looking for particular grade random forest algorithm fits best for our prediction as it gives two possible outcomes that is A grade or B grade.

Collection of data is first step in random forest algorithm, we have loaded csv file which contains Data-set into program using pandas. To build prediction model to predict grade of the college

```
In [1]: import os
import numpy as np
import pandas as pd
import seaborn as sns
import matplotlib.pyplot as plt
from sklearn.model_selection import GridSearchCV
from sklearn.neighbors import KNeighborsRegressor
from sklearn.ensemble import VotingClassifier, RandomForestRegressor
from sklearn.linear_model import LogisticRegression
from sklearn.model_selection import train_test_split
```

**Figure 3:** import files

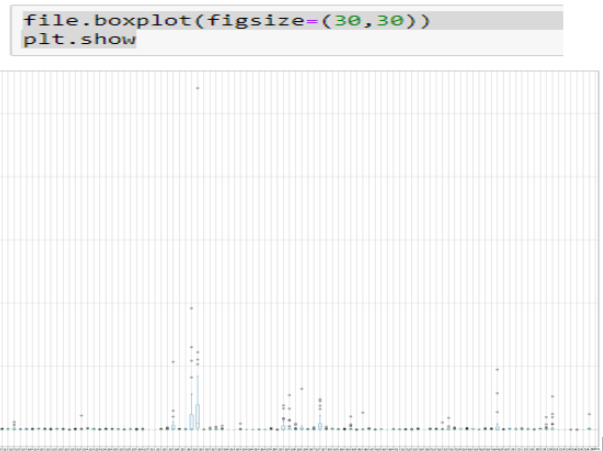
We have modified the data set according to target variable. We have eliminated all Null and String values from Data-set, by Using following code.

```
In [7]: print(file.info())
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 28 entries, 0 to 27
Columns: 581 entries, University to CGPA
dtypes: float64(56), int64(44), object(1)
memory usage: 22.2+ KB
None

In [8]: file.isnull()
Out[8]:
```

	University	1_1_2	1_1_3	1_2_1	1_2_2	1_3_2	1_3_3	1_3_4	1_4_1	1_4_2	...	7_1_9	7_1_10	7_1_11	7_1_12	7_1_13	7_1_14	7_1_15	7_1_16	7_1_17	
0	False	False	False	False	False	False	False	False	False	False	...	False	False	False	False	False	False	False	False	False	False
1	False	False	False	False	False	False	False	False	False	False	...	False	False	False	False	False	False	False	False	False	False
2	False	False	False	False	False	False	False	False	False	False	...	False	False	False	False	False	False	False	False	False	False
3	False	False	False	False	False	False	False	False	False	False	...	False	False	False	False	False	False	False	False	False	False
4	False	False	False	False	False	False	False	False	False	False	...	False	False	False	False	False	False	False	False	False	False
5	False	False	False	False	False	False	False	False	False	False	...	False	False	False	False	False	False	False	False	False	False
6	False	False	False	False	False	False	False	False	False	False	...	False	False	False	False	False	False	False	False	False	False
7	False	False	False	False	False	False	False	False	False	False	...	False	False	False	False	False	False	False	False	False	False
8	False	False	False	False	False	False	False	False	False	False	...	False	False	False	False	False	False	False	False	False	False

**Figure 4.** Elimination of null value



**Figure 5 :** Out layers

Once we check with irrelevant data such as null values and the value which is not useful for our model, as well as outlines we have further proceed with data splitting.

```
#Train & Test Splitting
from sklearn.model_selection import train_test_split
X_train,X_test,Y_train,Y_test=train_test_split(X,Y,test_size=0.2,random_state=0)
X_train.shape,X_test.shape

((22, 100), (6, 100))
```

**Figure 6 :** splitting of data

The data set is splits into training set and testing set with the ratio of 60:40 respectively. data is standardize and normalized using below code

```
#FeatureScaling
#Standardizing Data
from sklearn.preprocessing import StandardScaler
scaler=StandardScaler().fit(X)
rescaledX=scaler.transform(X)
rescaledX[0:5,:]

#FeatureScaling
#Normalizing Data
from sklearn.preprocessing import Normalizer
scaler=Normalizer().fit(X)
normalizedX=scaler.transform(X)
normalizedX[0:5,:]
```

Figure 7: Normalizing and standardizing

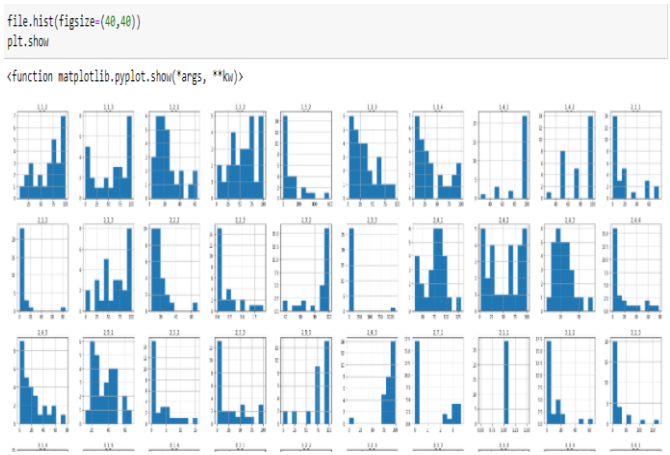


Figure 8 : Histogram

```
#plot co-relation heat map
def plot_correlation_map( COLLEGE_GRADING ):
    corr = file.corr()
    _, ax = plt.subplots( figsize = ( 20 , 20 ) )
    cmap = sns.diverging_palette( 220 , 10 , as_cmap = True )
    _ = sns.heatmap(
        corr,
        cmap = cmap,
        square=True,
        cbar_kws={ 'shrink' : .9 },
        ax=ax,
        annot = True,
        annot_kws = { 'fontsize' : 8 }
    )
    plot_correlation_map( file )
```

Figure 9 : correlation

Above code defines the co-relation between each criteria For model prediction Random Forest Regression function is implemented by importing below code

```
#create a new rf classifier
rf = RandomForestRegressor()

#create a dictionary of all values we want to test for n_estimators
params_rf = { 'n_estimators': [50000,100000,150000] }

#use gridsearch to test all values for n_estimators
rf_gs = GridSearchCV(rf, params_rf, cv=5)

#fit model to training data
rf_gs.fit(X_train, Y_train)
```

Figure 10 : RF algorithm

**Result Evaluation**

Once the machine is trained using 60% of data, then other40% of data is used for testing the model. Result are evaluated by compare the association data and model output.

Below screenshot explains the predicted Values.

```
Y_pred=rf_gs.predict(X_test)

results=pd.DataFrame({'Actual':Y_test,'Predicted':Y_pred})
results
```

	Actual	Predicted
0	3.22	3.064776
1	2.43	2.853589
2	2.56	2.595675
3	2.76	2.710657
4	3.44	2.899716
5	2.23	2.454871

```
#RMSE by Pandas
((results.Actual-results.Predicted)**2).mean()**.5
0.6787611263667555
```

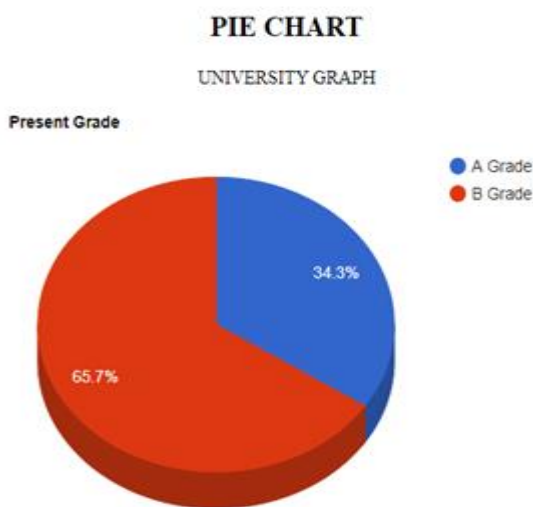
Figure 11 : Output

Criteria no	Response
1.1.2	
1.1.3	
1.2.1	
1.2.2	
1.3.2	
1.3.3	
1.3.4	
1.4.1	

**Figure 14 :** Screen Layout of User Response

Criteria no	Response	Grade	Suggestion
1.1.2	66	B Grade	Need to add 5.714285686% to be in A grade1B Grade
1.1.3	25.8	B Grade	Need to add 67.75% to be in A grade2B Grade
1.2.1	25.4	B Grade	Need to add 15.333333418% to be in A grade3B Grade
1.2.2	28.9	B Grade	Need to add 63.875% to be in A grade4B Grade
1.3.2	8	B Grade	Need to add 94.66666664% to be in A grade5B Grade
1.3.3	8	B Grade	Need to add 80% to be in A grade6B Grade
1.3.4	8	B Grade	Need to add 80% to be in A grade7B Grade
1.4.1	55	B Grade	Need to add 45% to be in A grade8B Grade
1.4.2	77	B Grade	Need to add 23% to be in A grade9B Grade
2.1.1	354	A Grade	No changes needed,you are in 'A' grade.10A Grade
2.1.2	6	A Grade	No changes needed,you are in 'A' grade.11A Grade
2.1.3	445	A Grade	No changes needed,you are in 'A' grade.12A Grade
2.2.2	89	B Grade	need to reduce -493.333333363% to be in A grade13B Grade
2.2.3	77	A Grade	No changes needed,you are in 'A' grade.14A Grade
2.3.2	45	B Grade	Need to add 50.00000005% to be in A grade15B Grade
2.3.3	32	B Grade	need to reduce -60% to be in A grade16B Grade
2.4.1	56	B Grade	Need to add 30% to be in A grade17B Grade

**Figure 15 :** Screen Layout of Grade and Suggestion



**Figure 16 :** Screen Layout of University Pie Chart

## CONCLUSION

Educational organization is always a concern about their grades and ranking of colleges, so by using Machine Learning and Data Mining techniques we have developed a model which can help them to improve their college grades or ranking.

The prediction models will not only give the prediction about the grades of the college, but will also provide real time advice to resolve their difficulties.

In proposed system have worked on the raw data, data has been collected from various college, from different portal such as internet sources like college websites and so on. In order to train our model we need huge amount of data. Once required amount of data has been collected processing has been done using data mining, had to manually verify the data in order to get a key attributes. Usually whatever data has been collected was complex data or data was not in the form the way needed for predicting. So cleaning and labeling of data has been done. We had to manually go through all the colleges to get related data which affects the grades of the college. From Data had to analysis which attribute will affect the grade and which are irrelevant. For proposed system have applied Random Forest Regression Algorithm to predict the Grades. Model also gives suggestion along with the grades.

## FUTURE ENHANCEMENT

Now model has the accuracy about 68%. As the size of the data set increases the model will be able to give better accuracy and suggestion to the user.

In future we can come across a better algorithm which can give the better accuracy rate.

## REFERENCES

- [1]. Harsh Valecha Department of Computer Science and Engineering, " Prediction of Consumer Behaviour using Random Forest Algorithm ", 2018 5th IEEE.
- [2]. Xu Zhang, Ruojuan Xue, Bin Liu, Wenpeng L, "Grade Prediction of Student Academic Performance with Multiple Classification Models",IEEE 4th, 2018.
- [3]. Jason Brownlee , "How to Save Your Machine Learning Model and Make Predictions in Weka", August 3, 2016.
- [4]. Neelam Naik & Seema Purohit, "Prediction of Final Result and Placement of Students using Classification Algorithm"International Journal of Computer Applications (0975 – 8887) Volume 56– No.12, October 2012
- [5]. Alaa M.El-Halees,Mohammed M. Abu Tair, "Mining Educational Data to Improve Students'Performance: A Case Study",International Journal of Information and Communication Technology Research, 2012.
- [6]. B.K. Bharadwaj and S. Pal,"Data Mining: A prediction for performance improvement using classification", International Journal of Computer Science and Information Security (IJCSIS), Vol. 9, No. 4, pp. 136-140, 2015.
- [7]. Suchita Borkar, K. Rajeswari, "Predicting Students Academic Performance Using Education Data Mining ", IJCSMC,Vol. 2, Issue. 7, July 2013, pg.273– 279.
- [8]. Randhir Singh, M.Tiwari, Neeraj Vimal,"An Empirical Study of Applications of Data Mining Techniques for Predicting Student Performance in Higher Education", 2013.

## Cite this article as :

Sushmita Gaonkar, "Grade Prediction of Educational Institute Using ML Algorithm ", *International Journal of Scientific Research in Computer Science, Engineering and Information Technology (IJSRCSEIT)*, ISSN : 2456-3307, Volume 6, Issue 3, pp.569-575, May-June-2020. Available at doi : <https://doi.org/10.32628/CSEIT2063131>  
Journal URL : <http://ijsrcseit.com/CSEIT2063131>