International Journal of Engineering and Advanced Technology (IJEAT)
ISSN: 2249-8958 (Online), Volume-9 Issue-3, February 2020

NBC Model for Early Prediction of At-Risk Students in Course

P. Sunanda, D. Kavitha

***Abstract**: Increase in computer usage for different purposes in different fields has made the computer important to learn things. Machine learning made systems to learn things and work accordingly on their own. Among the different fields that use machine learning, the education field is one. In the education field, machine learning has led to the advent of a digital-enabled classroom, speech recognition, adaptive learning techniques, and development of artificial instructor. Along with this, the prediction has its importance. In the education field, the main problem is students drop out. The machine learning predictive modeling approach can be used to identify the students who are at-risk and inform the instructor and students before reducing the dropouts. The main intention of this paper is to model a system that could be a solution to reduce the drop-outs and increase the education standards in students by early predicting their risk in a course.*

***Keywords**: Machine Learning, Prediction, at-risk, Naive-Bayes.*

I. INTRODUCTION

The world and the society around us say the importance of education through the inventions of new things each second. Education has become the basic need for a human to survive in this updated world. In this century, every child has many resources to acquire education. One such and most older resource is school or college which is an offline interactive mode of educating a child either using a student-centered approach or a teacher-centered approach. In either of these, a student can be guaranteed to gain knowledge. The problem here is since everyone is not equally intelligent, based on their understanding level, listening skills, attending the classes, interest in the course, there will be an effect in percentages or marks of the students. This results in dropouts in most schools and colleges.

This paper focuses mainly to predict different factors that affect student education. It predicts based on the marks, analyzing power of the student and then notifies him if he/she falls under risk or not before completing the course.

Revised Manuscript Received on February 05, 2020.

* Correspondence Author

P. Sunanda, Assistant Professor, Department of Computer Science & Engineering, G. Pulla Reddy Engineering College, Kurnool India.

Dr. D. Kavitha, Professor, Department of Computer Science & Engineering at G. Pulla Reddy Engineering College, Kurnool India.

II. MACHINE LEARNING

2.1 Introduction to Machine Learning

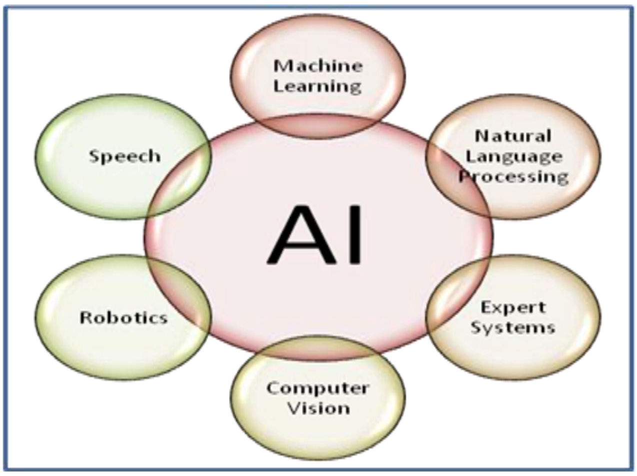

Figure1. Artificial Intelligence with its fields

Artificial Intelligence (AI) is the ability of a computer or a machine to think and learn. It has many fields like Natural Language processing, expert systems, robotics, etc. to make the computers smart. Among these, Machine Learning is one that allows computers or machines to automatically learn from past data, to enable data-driven design without any explicit programming.

In the year 1959, Arthur Samuel coined the name Machine Learning which explains the study and the construction of algorithms. These algorithms are also designed in such a way that, they can learn and improve themselves when exposed to the new data.

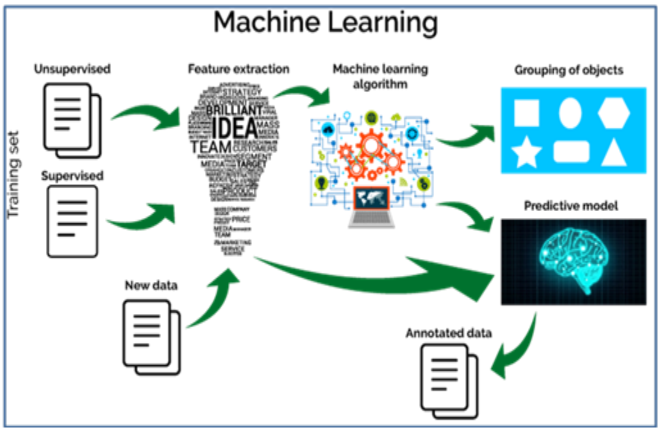

Figure2. Machine learning process

2.2 Classification of Machine Learning tasks

Machine learning tasks are classified typically into several broad categories as follows

- Supervised learning
- Unsupervised learning
- Semi-supervised learning
- Reinforcement learning

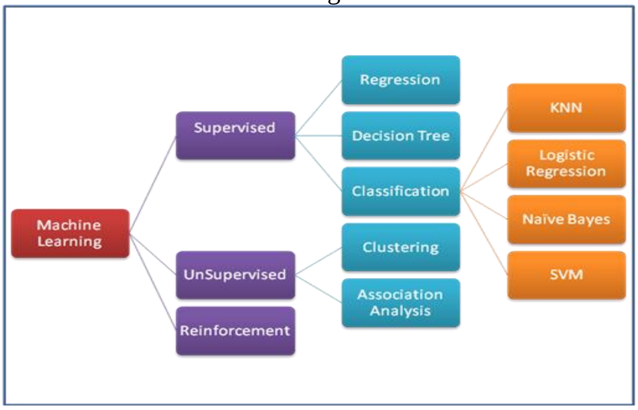

Figure3. Classification of Machine Learning

1) Supervised Learning

It is a task in which based on the example input-output pair called labelled dataset, the learning function maps input to an output. In this task, the model will get trained on the labelled dataset. The applications of supervised learning are

- Handwriting recognition
- Spam detection
- Pattern recognition

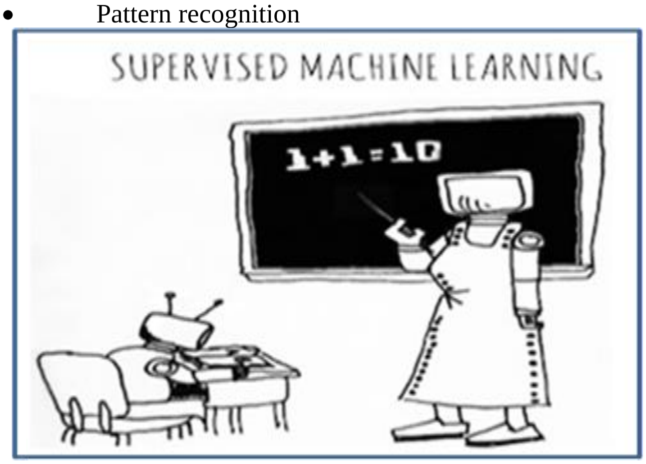

Figure4. Supervised Learning

2) Unsupervised Learning

This task helps in finding the unknown dataset patterns without pre-existing labels. It mainly deals with unlabelled data. The applications of unsupervised learning are

- Density estimation in statistics
- Clustering
- Dimensionality Reduction
- Visualization
- Anomaly Detection

3) Semi-Supervised Learning

This task combines a smaller volume of labelled data with larger volumes of unlabelled data. So, it falls in between supervised and unsupervised learning. The applications of semi-supervised learning are

- Speech analysis
- Web content classification
- Protein sequence classification

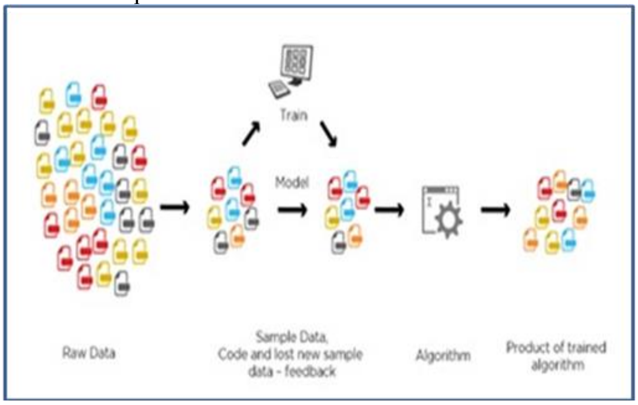

Figure5. Working model of Semi-Supervised Learning

4) Reinforcement Learning

It is a type of dynamic programming that trains algorithms to deal with taking suitable action to maximize the reward of a particular situation. The applications of Reinforcement Learning are

- Traffic Light Control
- Bidding and Advertising
- Web System Configuration

2.3 How does Machine Learning Works?

Machine learning has 7 major steps they are as follows

Step 1: Gathering Data - Predict model accuracy directly proportional to the quality of data.

Step 2: Data preparation - Prepare data as per our model and remove unnecessary data.

Step 3: Choosing a model - The most typical job is to select the most suitable model.

Step 4: Training - Train the model according to the dataset

Step 5: Evaluation - Evaluate Training dataset.

Step 6: Parameter Tuning - After evaluation, important parameters can be identified and added

Step 7: Prediction - And the final step is predicting using the model and dataset.

The working model of Machine Learning can be shown in the following figure.

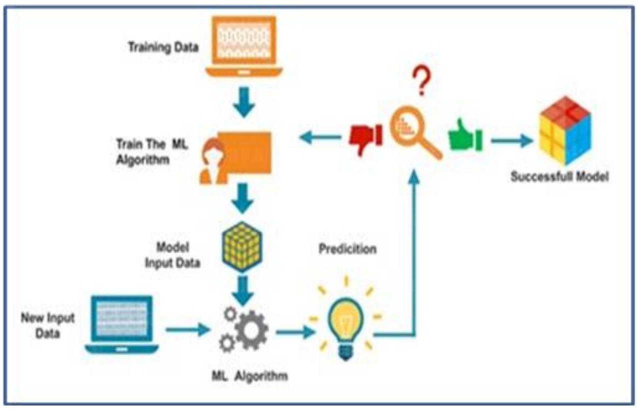

Figure6. Working Model of Machine Learning

2.4 Applications of Machine Learning

- Email Spam and Malware Filtering
- Search Engine Result Refining
- Playing Video Games Automatically
- Education
- Finance [1].

III. TECHNOLOGIES USED

To predict student's risk factors using Machine learning, a programming language with machine learning libraries is needed.

3.1 Available Languages

There are many programming languages available like C, Java, R, Python. Of these languages, every language has its' own specific role to be used to implement a particular technology. So, to implement machine learning algorithms and to work with a complex set of tasks, the preferred programming language is python.

3.2 Why Python?

According to the survey of Stacks Overflow in the year 2019, Python has raised in the ranks to the 4th position from among the most commonly used programming languages, from the 7th position in 2018. It is the fastest-growing major programming language today. It is the second most loved language.

There are several reasons like simple syntax, easy programming, many built-in modules, libraries, functions and methods that made to select python. Among built-in modules, the sklearn (scikit-learn) called "Machine Learning in Python" is used to work with machine learning algorithms to predict at-risk students.

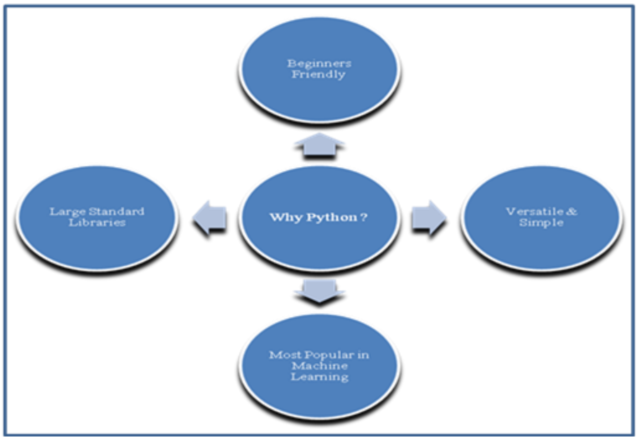

Figure7. Why Python?

3.3 Packages and Libraries in Python

The following table Tabe1 lists a set of built-in modules in python to work with machine learning algorithms

Table1. Python built-in modules

Module	Description
numpy	To work with N-dimensional array objects
pandas	To analyze the data through data frames
matplotlib	To create 2D graphs and plots
scikit-learn	To work with machine learning algorithms
seaborn	To visualize the data

IV. PROPOSED PREDICTION MODELS

The six prediction models are

- Logistic Regression (Log Reg)
- Support Vector Machine(SVM)
- Decision Tree (DT)
- Multi-Layer Perceptron (MLP)
- Naïve Bayes Classifier
- K- Nearest Neighbor (KNN)

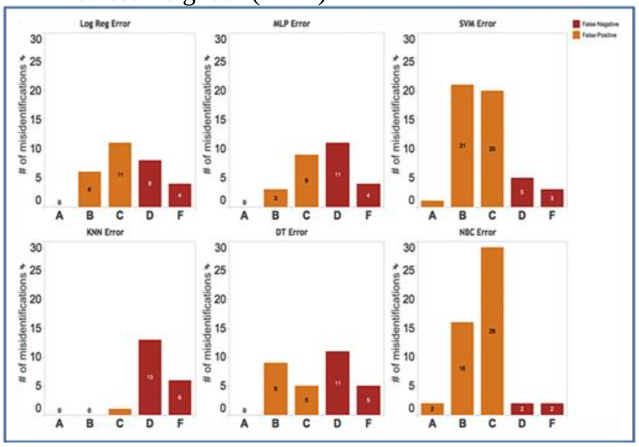

Figure8. Graphs showing the number of misidentifications

- **Logistic Regression** (Log Reg) in general is used to model the probability of certain existing events and classes such as pass/fail, win/lose, yes/no, more/less. In the same way, here it can be used to classify the students based on pass/fail, interested/uninterested, risk/no risk.
- **Support Vector Machine** (SVM) is a supervised learning model used for classification and regression associated algorithms. These associated algorithms are used to analyze the data. The analysis can be done through a kernel trick technique by transforming the data. And based on these data transformations, an optimal boundary between possible outputs will be found by it. Here the students can be classified based on their gender, test preparation course, parental education level.
- **Decision Tree** (DT) is a partitioning method model. Its structure looks like a tree with decisions and consequences. The consequences are selected based on the calculated entropy value of the decision. The fate of the consequence mainly depends on the entropy of the decision. Here the consequences like student interest in a course, predicting his risk factor mainly depends on the decisions of math score, reading and writing scores.
- **Multi-Layer Perceptron** (MLP) is an Artificial Neural Network (ANN) with one input, an arbitrary number of hidden layers and one output layer.

The input layer receives the input, the input is processed or computed by the hidden layers and decision or prediction is performed by the output layer. So, the input and output layers are fixed but the hidden layer count changes based on several computations to be performed. Here the student gender or parental education level can be taken as the input layer, the math, reading and writing scores are to be treated as hidden layers, and predicting the risk factor of a student is done by output layer.

- **Naive Bayes Classifier** (NBC) is a simple probabilistic classifier family member that follows the “Bayes theorem”. It follows Naïve independence assumptions between features. This is a family of algorithms that share “every classifier feature is equally important and independent of each other. Here the features such as gender, parental educational level, test preparation course, math score, reading score, and writing score are considered. The gender and parental educational level are independent of one another and race/ethnicity is independent of reading, writing and math scores. Even though these features are independent of one another, but they have their importance in predicting the student risk.
- **K-Nearest Neighbor** (KNN) is a simple, non-parametric algorithm used for both classification and regression. It stores all available cases and classifies new cases based on the similarity measures. It’s being in use for statistical estimation and pattern recognition since the 1970s. Here, based on student reading, writing, and math score, they can be classified as two sets one set with a set of students whose score is less than 50 and other greater than 50. The students in set less than 50 will be almost at risk. So, it becomes easy for prediction.

V. DATA PRE-PROCESSING AND EXPLORATORY ANALYSIS

Here all the steps needed for prediction like from data collection, data pre-processing, modelling are followed.

5.1 Data Source

Data is the primary thing needed to perform any operation. A data source is like a repository of data in different formats. It contains data as hard-coded data, datasets, spreadsheets, etc. Here the students performance dataset with 1000 records sis taken from 4Shared.com data source.

This dataset has many features like gender, race/ethnicity, and parental level of education, lunch, and test preparation course, math score, reading score and writing score. The Figure9. gives a snapshot of the dataset.

gender	race/ethnicity	parental level of education	lunch	test preparation course	math score	reading score	writing score
female	group B	bachelor's degree	standard	none	72	72	74
female	group C	some college	standard	completed	69	90	88
female	group B	master's degree	standard	none	90	95	93
male	group A	associate's degree	free/reduced	none	47	57	44
male	group C	some college	standard	none	76	78	75
female	group B	associate's degree	standard	none	71	83	78
female	group B	some college	standard	completed	88	95	92
male	group B	some college	free/reduced	none	40	43	39
male	group D	high school	free/reduced	completed	64	64	67
female	group B	high school	free/reduced	none	38	60	50
male	group C	associate's degree	standard	none	58	54	52
male	group D	associate's degree	standard	none	40	52	43
female	group B	high school	standard	none	65	81	73
male	group A	some college	standard	completed	78	72	70
female	group A	master's degree	standard	none	50	53	58
female	group C	some high school	standard	none	69	75	78
male	group C	high school	standard	none	88	89	86
female	group B	some high school	free/reduced	none	18	32	28
male	group C	master's degree	free/reduced	completed	46	42	46
female	group C	associate's degree	free/reduced	none	54	58	61
male	group D	high school	standard	none	66	69	63
female	group B	some college	free/reduced	completed	65	75	70
male	group D	some college	standard	none	44	54	53
female	group C	some high school	standard	none	69	73	73

Figure9. Snapshot of the dataset

5.2 Dataset Description

- **Gender:** In developing countries like India, gender inequality plays a main role in education. So, here this feature is considered to check the effect of it on student’s studies. Here Male and Female are two genders considered.
- **Race:** The race/ethnicity in the data set says to which group of race the student belongs. Based on the student race, it becomes easy to predict his risk factor.
- **Parental level Education:** The parental level education in the dataset about the degree of the parent.
- **Test preparation course:** The test preparation course in the dataset describes the course details of the student whether he has taken/completed the course or not.
- **Math score:** Math score in the dataset describes the score scored by the student.
- **Reading score:** The score that is scored by the student in the general daily basis reading tests.
- **Writing score:** The score that is score by the student in the general daily basis writing tests.

5.3 Data Pre-Processing

Before pre-processing the data it must be loaded into Data Frames.

```
#Reading data into python data frames
data=pd.read_csv("C:/input/studentsperformance/StudentsPerformance.csv")
data.head()
```

data - DataFrame

Index	gender	race/ethnicity	ntal level of educa	lunch	t preparation cou	math score	reading score	writing score
0	female	group B	bachelor's degree	standard	none	72	72	74
1	female	group C	some college	standard	completed	69	90	88
2	female	group B	master's degree	standard	none	90	95	93
3	male	group A	associate's degree	free/reduced	none	47	57	44
4	male	group C	some college	standard	none	76	78	75
5	female	group B	associate's degree	standard	none	71	83	78
6	female	group B	some college	standard	completed	88	95	92
7	male	group B	some college	free/reduced	none	40	43	39
8	male	group D	high school	free/reduced	completed	64	64	67
9	female	group B	high school	free/reduced	none	38	60	50
10	male	group C	associate's degree	standard	none	58	54	52

Format Resize Background color Column min/max OK Cancel

Figure10. Output after reading dataset into python

After successfully importing the dataset as shown in the Figure10, it is observed that the column names of the dataset seem to be confusing. So, they need to be renamed for better understanding and easy deploy and debug. The dataset with renamed columns is shown in Figure11.

1) Renaming columns:

```
#Renaming columns
data.columns=['gender','race','parentsdegree','lunch','course','mathscore','readings
core','writingscore']
data.head()
```

	gender	race	parentsdegree	lunch	course	mathscore	readingscore	writingscore
0	female	group B	bachelor's degree	standard	none	72	72	74
1	female	group C	some college	standard	completed	69	90	88
2	female	group B	master's degree	standard	none	90	95	93
3	male	group A	associate's degree	free/reduced	none	47	57	44
4	male	group C	some college	standard	none	76	78	75

Figure11. Data frame after renaming the columns

2) Checking for any missing values:

```
#Now check the missing values
miss=data.isnull().any()
miss# Hence no missing values
```

```
gender          False
race            False
parentsdegree   False
lunch           False
course          False
mathscore       False
readingscore    False
writingscore    False
dtype: bool
```

Figure12. Output after checking if any null values in the dataset

5.4 Implementation

For further implementation, an attribute called percentage is needed to represent the total scores of the students. For calculating the percentage, the mean of reading, writing and math score is considered. The mean is calculated as follows and the output is as in the Figure13.

```
#Lets check the percentage
data['Percentage']=(data['mathscore']+data['readingscore']+data['writingscore'])/3
data.head()
```

	gender	race	parentsdegree	lunch	course	mathscore	readingscore	writingscore	Percentage
0	female	group B	bachelor's degree	standard	none	72	72	74	72.666667
1	female	group C	some college	standard	completed	69	90	88	82.333333
2	female	group B	master's degree	standard	none	90	95	93	92.666667
3	male	group A	associate's degree	free/reduced	none	47	57	44	49.333333
4	male	group C	some college	standard	none	76	78	75	76.333333

Figure13. Dataset after calculating percentage

3) Dataframe.groupby()

```
DataFrame.groupby(by=None, axis=0, level=None, as_index=True, sort=True, group
_keys=True, squeeze=False, observed=False, **kwargs)
data.groupby(['race','parentsdegree']).mean()
```

		mathscore	readingscore	writingscore	Percentage
race	parentsdegree				
group A	associate's degree	61.000000	67.071429	63.571429	63.880952
	bachelor's degree	67.166667	68.083333	68.333333	67.861111
	high school	60.444444	62.888889	60.500000	61.277778
	master's degree	57.666667	64.666667	67.666667	63.333333
	some college	63.888889	65.777778	65.000000	64.888889
	some high school	58.916667	62.083333	58.583333	59.861111
group B	associate's degree	66.097561	69.585366	68.243902	67.975610
	bachelor's degree	69.300000	72.950000	71.650000	71.300000
	high school	59.791667	63.458333	61.250000	61.500000
	master's degree	67.166667	80.166667	77.166667	74.833333
	some college	63.189189	65.756757	64.189189	64.378378
	some high school	61.815789	66.447368	64.605263	64.289474
group C	associate's degree	66.730769	71.128205	70.269231	69.376068
	bachelor's degree	68.150000	75.675000	75.900000	73.241667
	high school	60.906250	64.421875	61.656250	62.328125
	master's degree	67.052632	70.526316	69.526316	69.035088
	some college	65.130435	69.420290	68.869565	67.806763
	some high school	60.551020	65.632653	63.285714	63.156463
group D	associate's degree	67.600000	70.540000	69.860000	69.333333
	bachelor's degree	67.571429	70.142857	71.892857	69.869048
	high school	62.863636	64.409091	63.159091	63.477273
	master's degree	72.521739	77.173913	79.739130	76.478261
	some college	68.731343	70.880597	71.701493	70.437811
	some high school	66.760000	69.980000	69.100000	68.613333
group E	associate's degree	74.897436	73.820513	73.205128	73.974359
	bachelor's degree	76.555556	74.833333	75.388889	75.592593
	high school	70.772727	70.318182	67.545455	69.545455
	master's degree	74.625000	82.125000	80.500000	79.083333
	some college	73.828571	72.628571	70.200000	72.219048
	some high school	72.111111	69.555556	66.555556	69.407407

Figure14. Group-wise analysis

By the analysis of the above output in the Figure14, it can be observed that as race (Group) increases, the score also increases. This helps further in the prediction of at-risk students from the dataset.

Now checking the scores of the students by gender-based

```
# Lets check the score according to gender
data.groupby('gender').mean()
```

	mathscore	readingscore	writingscore	Percentage
gender				
female	63.633205	72.608108	72.467181	69.569498
male	68.728216	65.473029	63.311203	65.837483

Figure15. Scores based on gender

The Figure15. clearly depicts that female percentage is greater than male percentage by some extent. While talking about marks, female math score is lower than the other two scores. The relation between the gender and the course completion is as shown in the Figure16.

```
# Lets check the relationship between genders, course and percentage
course_gender=data.groupby(['gender','course']).mean().reset_index()

sns.factorplot(x='gender', y='Percentage', hue='course', data=course_gender, kind='bar')
```

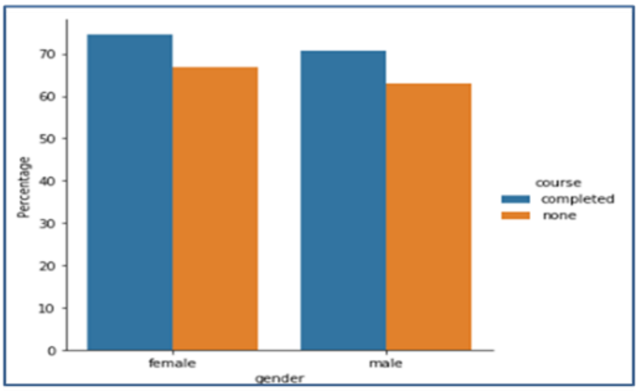

Figure16. Graph shows the relation between gender and course completion

```
# Now we can say that Parents Degree is also crucial in students score
course_gender=data.groupby(['gender','parentsdegree']).mean().reset_index()

sns.factorplot(x='gender',y='Percentage',hue='parentsdegree',data=course_gender, kind='bar')
```

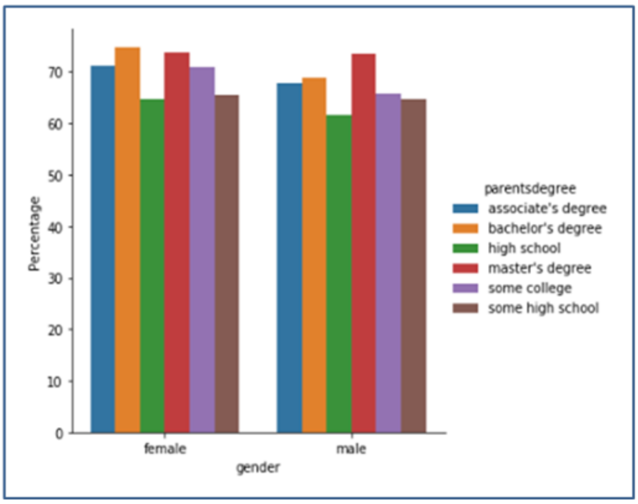

Figure17. Graph between gender and percentage with a parental degree as a factor

The students can be classified based on the parentsdegree.The "parentsdegree" column has 'has_degree' and 'no_degree' values. If a parentsdegree is either "high school" or "some high school" then, they comes under 'N0_Degree' category, else they come under 'has_Degree' category. The graph in the Figure17 shows the relation between gender and percentage with a parental degree as a factor.

```
#Finding out different types of degree available
data.parentsdegree.unique()
```

```
array(["bachelor's degree", 'some college', "master's degree",
       "associate's degree", 'high school', 'some high school'],
      dtype=object)
```

```
for i in range(len(data)):
        if data.iloc[i,2] in ['high school','some high school']:
                data.iloc[i,2]='No_Degree'
        else:
                data.iloc[i,2]='has_Degree'
data.head()
```

	gender	race	parentsdegree	lunch	course	mathscore	readingscore	writingscore	Percentage
0	female	group B	has_Degree	standard	none	72	72	74	72.666667
1	female	group C	has_Degree	standard	completed	69	90	88	82.333333
2	female	group B	has_Degree	standard	none	90	95	93	92.666667
3	male	group A	has_Degree	free/reduced	none	47	57	44	49.333333
4	male	group C	has_Degree	standard	none	76	78	75	76.333333

Figure18. Output after changing parentsdegree

Retrieval Number: C5440029320/2020©BEIESP
DOI: 10.35940/ijeat.C5440.029320
Journal Website: www.ijeat.org

Published By:
Blue Eyes Intelligence Engineering
& Sciences Publication

Now after changing the parentsdegree to either has_Degree or No_Degree, the students are again grouped.

```
final_data = data.groupby(['gender','parentsdegree','course','lunch','race']).mean().reset_inde
x()
after_sort = final_data.sort_values(by= ['Percentage'],ascending = False)
after_sort.drop(columns=['mathscore','readingscore','writingscore'],inplace = True)
after_sort
```

	gender	parentsdegree	course	lunch	race	Percentage
28	female	has_Degree	completed	standard	group E	84.023810
5	female	No_Degree	completed	standard	group A	82.666667
66	male	has_Degree	completed	standard	group C	79.980392
27	female	has_Degree	completed	standard	group D	79.738095
65	male	has_Degree	completed	standard	group B	79.100000
26	female	has_Degree	completed	standard	group C	78.342593
64	male	has_Degree	completed	standard	group A	77.761905
8	female	No_Degree	completed	standard	group D	76.566667
25	female	has_Degree	completed	standard	group B	76.555556
23	female	has_Degree	completed	free/reduced	group E	76.400000
48	male	No_Degree	completed	standard	group E	76.208333
68	male	has_Degree	completed	standard	group E	76.066667
38	female	has_Degree	none	standard	group E	75.840580
37	female	has_Degree	none	standard	group D	75.813725
9	female	No_Degree	completed	standard	group E	75.555556
19	female	No_Degree	none	standard	group E	75.222222
22	female	has_Degree	completed	free/reduced	group D	74.666667
24	female	has_Degree	completed	standard	group A	73.333333
67	male	has_Degree	completed	standard	group D	73.311111
7	female	No_Degree	completed	standard	group C	72.333333
63	male	has_Degree	completed	free/reduced	group E	71.333333
47	male	No_Degree	completed	standard	group D	71.300000
20	female	has_Degree	completed	free/reduced	group B	71.250000
36	female	has_Degree	none	standard	group C	70.809524
43	male	No_Degree	completed	free/reduced	group E	70.750000
4	female	No_Degree	completed	free/reduced	group E	70.500000
35	female	has_Degree	none	standard	group B	70.466667
62	male	has_Degree	completed	free/reduced	group D	70.000000
60	male	has_Degree	completed	free/reduced	group B	69.904762

Figure19. Output after finding mean

The Figure19. shows the generalization of all features.

5.5 Result Analysis

1. The result depicts that the Top students(mean) have completed their course, took standard Lunch, and they also had parent_Degree as a plus point.
2. Bottom students(mean) depicts that they didn't complete course, they didn't take good lunch, and their parent has no degree.
3. Out of Top 10(mean), 7 are female students
4. Interestingly, Out of Bottom 10(mean), 7 are male students

```
#see top performers
print("Top 10 Performer \n",after_sort[:10])
```

```
Top 10 Performer
     gender parentsdegree  ...     race Percentage
28  female    has_Degree  ...  group E  84.023810
5   female     No_Degree  ...  group A  82.666667
66    male    has_Degree  ...  group C  79.980392
27  female    has_Degree  ...  group D  79.738095
65    male    has_Degree  ...  group B  79.100000
26  female    has_Degree  ...  group C  78.342593
64    male    has_Degree  ...  group A  77.761905
8   female     No_Degree  ...  group D  76.566667
25  female    has_Degree  ...  group B  76.555556
23  female    has_Degree  ...  group E  76.400000

[10 rows x 6 columns]
```

Figure20. Result of the top performers

After analyzing the results, it can be concluded that if a student completes the course, has standard lunch, then he can score good grades. The Top 10 performers are as shown in the Figure20 with Bottom Performers in Figure21.

```
#see bottom performers
print("Bottom Performer \n",after_sort[-10:])
```

```
Bottom Performer
     gender parentsdegree  ...     race Percentage
54    male     No_Degree  ...  group A  56.000000
70    male    has_Degree  ...  group B  55.500000
49    male     No_Degree  ...  group A  54.208333
53    male     No_Degree  ...  group E  54.000000
11  female     No_Degree  ...  group B  53.066667
41    male     No_Degree  ...  group C  52.416667
10  female     No_Degree  ...  group A  52.066667
12  female     No_Degree  ...  group C  50.607843
40    male     No_Degree  ...  group B  50.333333
50    male     No_Degree  ...  group B  45.333333

[10 rows x 6 columns]
```

Figure21. Result of at-risk students

```
base = pd.get_dummies(final_data,columns=['gender','race','parentsdegree','course',
'lunch'],dtype =int)
base.sample()
base.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 79 entries, 0 to 78
Data columns (total 17 columns):
mathscore                   79 non-null float64
readingscore                79 non-null float64
writingscore                79 non-null float64
Percentage                  79 non-null float64
gender_female               79 non-null int64
gender_male                 79 non-null int64
race_group A                79 non-null int64
race_group B                79 non-null int64
race_group C                79 non-null int64
race_group D                79 non-null int64
race_group E                79 non-null int64
parentsdegree_No_Degree     79 non-null int64
parentsdegree_has_Degree    79 non-null int64
course_completed            79 non-null int64
course_none                 79 non-null int64
lunch_free/reduced          79 non-null int64
lunch_standard              79 non-null int64
dtypes: float64(4), int64(13)
memory usage: 10.6 KB
```

Figure22. Output of the base.info( ) command

VI. CONCLUSION AND FUTURE ENHANCEMENTS

The NBC model can be enhanced further by collecting more valid attributes in data set that helps in the prediction of the at-risk students. In this project, the most helping attributes are the parental education level and the student's reading score and writing score. More attributes like the student's percentage in the previous courses and present academic attendance help to increase the precision in the prediction. And the mean square error can be reduced by this.

REFERENCES

1. Machine learning for disease prediction by using neural networks, p. Sunanda, international journal of research and analytical reviews (ijrar), volume 6, Issue 2, Pages 712-719, Publisher, www.ijrar.org
2. http://www.ijrar.org/viewfull.php?&p_id=IJRAR19K1534
3. A survey on machine learning: concept,algorithms and applications - kajaree das1
4. bias-variance analysis of support vector machines for the development of svm-based ensemble methods - giorgio valentini
5. grade prediction with course and student specific models – agoritsa polyzou and george karypis
6. building course-specific regression-based models to identify at-risk students -mr. farshid marbouti
7. a deep feature selection approach - milad zafar nezhad
8. predicting at-risk students by analysing clicking behaviour in a virtual learning environment - annika wolff, zdenek zdrahal
9. survey on student performance prediction - mudit panwar
10. early detection of students at risk - predicting student dropouts using administrative student data and machine learning methods - johannes berens,
11. solutions to instability problems with sequential wrapper-based approaches to feature selection - kevin dunne
12. valentini, g., & dietterich, t. g. (2004). bias-variance analysis of support vector machines for the development of svm-based ensemble methods.
13. carberry, a. r., siniawski, m. t., & dionisio, j. d. n. (2012). standards-based grading:
14. links
15. https://en.wikipedia.org/wiki/machine_learning
16. https://www.tutorialspoint.com/
17. https://www.expertsystem.com/machine-learning-definition/
18. https://towardsdatascience.com/applications-of-reinforcement-learning-in-real-world-1a94955bcd12
19. https://resources.interactions.com/machine-learning-important/
20. https://towardsdatascience.com/machine-learning-classifiers-a5cc4e1b0623
21. https://www.edureka.co/blog/what-is-machine-learning/
22. https://bigdata-madesimple.com/8-ways-machine-learning-will-improve-education/
23. https://towardsdatascience.com/machine-learning-in-finance-why-what-how-d524a2357b56
24. https://theknowledgereview.com/machine-learning-impact-education-technology/
25. https://www.techopedia.com/how-machine-learning-can-improve-teaching-excellence/2/33323
26. http://psycnet.apa.org/fulltext/1992-40647-001.html
27. https://www.analyticsvidhya.com/blog/2017/09/understaing-support-vector-machine-example-code/
28. https://www.researchgate.net/publication/250145376_efficacy_of_participating_in_a_first-year_seminar_on_student_satisfaction_and_retention
29. https://www.mathworks.com/discovery/machine-learning.html
30. https://www.tandfonline.com/doi/abs/10.1080/07924360120043621
31. https://realpython.com/python-data-types/
32. 18.https://www.sciencedirect.com/science/article/pii/S0360131516301634

AUTHORS PROFILE

P. Sunanda, is Assistant Professor in the department of Computer Science & Engineering at G.Pulla Reddy Engineering College, Kurnool India. She obtained her M.Tech degree in Computer Science from Jawaharlal Nehru Technological University, India and B.Tech degree in Computer Science and Engineering from Sri Krishna Devaraya University, India. Her current research interests are in the area of Computer Vision, Big data and machine learning.

Dr. D. Kavitha, is a Professor in the department of Computer Science & Engineering at G.Pulla Reddy Engineering College, Kurnool India. She received her Ph.D from Sri Krishna Devaraya University, India in 2012. She obtained her M.Tech degree in Computer Science from Jawaharlal Nehru Technological University, India and B.Tech degree in Computer Science and Engineering from Sri Krishna Devaraya University, India. Her current research interests are in the area of data analytics and machine learning. She published more than 30 papers in various International conferences and journals.