



Probabilistic Determination of Student Performance using Naive Bayes Classification Algorithm

Madhavi Tripathi¹, Abhay Kumar Agarwal²
PG Scholar¹, Assistant Professor²
Department of CSE
KNIT, Sultan Pur, UP, India

Abstract:

This paper focus on identifying the slow learners and the performance of students and displaying it by using Naive Bayes Algorithm. Real World data set from a high school data set is taken and filtration of desired potential variables is done using Power BI Tool. The dataset of student academic records is tested and applied on Naive Bayes Classification Algorithm Using Power BI; Power BI is a visualization tool to visualize the probability of result. As a result, statistics are generated based on classification algorithms and presented the student performance on the basis of different attributes like mobile phones, computer at home, and net access, board of student etc and comparison of all the result in the form of chart. This Paper shows the importance of Naive Bayes Classification Algorithm.

I. INTRODUCTION:

This paper provides an overview regarding the work about the probalistic determination of student performance using Naive Bayes Classification Algorithm. This paper focus on identifying the slow learners and the performance of students and displaying it on the basis of different attributes like mobile phones, computer at home, net access, board of student etc and comparison of all the result in the form of chart.

II. RELATED WORK:

This section presents the related work to Naïve Bayes. Bandana Garg [1] In this paper designed and developed a naïve Bayes classifier for a better understanding of the algorithm. The classifier is tested on two different data sets from the University of California at Irvine machine learning repository. Different cross validation methods are used to calculate the accuracy of the developed classifier. S.L. Ting[2] the aim of this paper is to highlight the performance of employing, Naïve Bayes in document classification. Results show that Naïve Bayes is the best classifiers against several common classifiers.

Swati, Rejender Kaur[3] The high school students must be observed for their slow learning or quick learning abilities to provide them with the best education practices. Such analysis can be perfectly performed over the student performance data, using Naïve Bayes Classification Algorithm.

Mukesh Kumar,Shankar Shambhu and Poonam Aggarwal [4] In this paper the work has been done to find the slow learner in a High School class and then provide timely help to them for improving their overall result.

K. Prasada Rao, M.V.P Chandra Shekhara,B.Ramesh [5] This paper discusses about a learning model for predicting student performance using classification techniques. Also the paper shows the comparative performance analysis of J48, Naïve Bayesian classifier and Random forest algorithm.

P. Kavipriya [6] this paper reviews the previous research works done on students' performance prediction, analysis, early alert and evaluation by using different methods of data mining.

III. PROPOSED WORK:

This section presents about proposed work, Flow Chart and Algorithm.

A. Flow Chart:

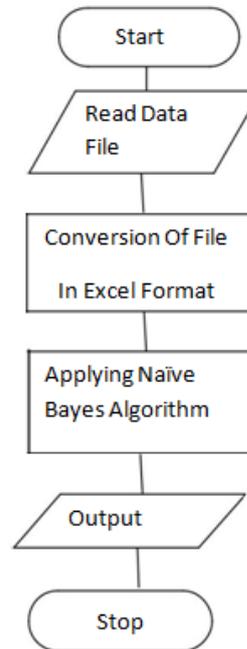


Figure.1. Flow Chart:

1. Description Of Flow Chart :

Firstly Read the data file for probalistic performance review of student, then convert the data file into excel then applying Naive Bayes formula for the probalistic calculation of student performance and shows the performance review in the form of output.

B. Algorithm:

Step 1: Begin

Step 2: Read Data File

- Step 3:** Conversion of data file into Excel
- Step 4:** Select attribute for calculation of profanity
- Step 5:** Apply Naive Bayes Formula
- Step 6:** Get result
- Step 7:** End

1. Description Of Algorithm: Firstly Read the data file for probabilistic perform ace review of student , then convert the data file into excel then select the attribute among n number of attribute for calculation, applying Naive Bayes formula for the probabilistic calculation of student performance and shows the performance review in the form of output.

C. Hardware Details:

D. For the implementation, hardware involves one computer system with the following specifications: processor Dual Core of 2.1 GHz, RAM of 2GB, Cache 2MB.

D. Software Details:

The existing and proposed algorithm has been implemented on operating system Windows 7. The tool used in implementing algorithms is Power BI. Microsoft Excel and used to draw graph for showing various results.

E. Power BI

Power BI is a collection of software services, apps, and connectors that work together to turn your unrelated sources of data into coherent, visually immersive, and interactive insights. Whether your data is a simple Excel spreadsheet, or a collection of cloud-based and on-premises hybrid data warehouses, Power BI lets you easily connect to your data sources, visualize (or discover) what’s important, and share that with anyone or everyone you want.

E. Data Set

sex	IIS-HIGH	TDR	MOI	TOS	PFUR	S-AREA	MOB	COM-IM	NETACS	SOLLNO	INT-GR	ATDN	CLASS(Response Variable)
M	PRIVATE	State Board	HINDI	CO-ED	YES	URBAN	YES	YES	YES	PR100	A	74% Q	
F	GOVERNMENT	State Board	HINDI	BOYS	NO	URBAN	YES	YES	NO	GR101	C	49% NQ	
F	PRIVATE	CBSE	ENGLISH	GIRLS	NO	RURAL	YES	YES	NO	PR100	A+	90% Q	
F	GOVERNMENT	CBSE	ENGLISH	CO-ED	YES	URBAN	YES	YES	YES	GR102	C	45% NQ	
M	GOVERNMENT	CBSE	ENGLISH	BOYS	NO	URBAN	YES	YES	NO	GR103	A	70% Q	
M	PRIVATE	State Board	HINDI	GIRLS	NO	RURAL	NO	NO	NO	PR100	A	75% Q	
M	PRIVATE	State Board	HINDI	BOYS	NO	URBAN	YES	YES	NO	PR100	A	76% Q	
M	PRIVATE	CBSE	ENGLISH	BOYS	YES	RURAL	NO	NO	YES	PR100	A	70% Q	
M	GOVERNMENT	CBSE	ENGLISH	CO-ED	NO	RURAL	YES	YES	NO	GR104	A+	90% Q	
F	GOVERNMENT	CBSE	ENGLISH	GIRLS	NO	URBAN	NO	NO	NO	GR105	B	50% Q	
M	PRIVATE	State Board	HINDI	GIRLS	NO	URBAN	YES	YES	NO	PR100	A	75% Q	
F	GOVERNMENT	CBSE	ENGLISH	CO-ED	YES	URBAN	YES	YES	YES	GR106	C	35% NQ	
M	PRIVATE	CBSE	ENGLISH	BOYS	YES	URBAN	YES	YES	YES	PR100	A	75% Q	
M	GOVERNMENT	CBSE	ENGLISH	BOYS	NO	URBAN	YES	YES	NO	GR107	B	62% Q	
M	GOVERNMENT	State Board	HINDI	CO-ED	YES	RURAL	NO	NO	YES	GR108	A+	90% Q	
M	PRIVATE	State Board	HINDI	BOYS	YES	URBAN	YES	YES	YES	PR100	C	38% NQ	
F	GOVERNMENT	State Board	HINDI	GIRLS	YES	RURAL	YES	YES	YES	GR109	A+	87% Q	
F	PRIVATE	State Board	HINDI	BOYS	NO	URBAN	NO	NO	NO	PR100	B	69% Q	
F	PRIVATE	State Board	HINDI	BOYS	NO	URBAN	YES	YES	NO	PR100	A+	80% Q	
M	PRIVATE	CBSE	ENGLISH	CO-ED	NO	URBAN	NO	NO	NO	PR101	C	42% NQ	
F	GOVERNMENT	CBSE	ENGLISH	GIRLS	YES	RURAL	YES	YES	YES	GR110	C	48% NQ	
M	GOVERNMENT	CBSE	ENGLISH	BOYS	NO	URBAN	YES	YES	NO	GR111	A	75% Q	
F	PRIVATE	State Board	HINDI	BOYS	NO	RURAL	NO	NO	NO	PR102	B	68% Q	
M	GOVERNMENT	State Board	HINDI	GIRLS	NO	RURAL	YES	YES	NO	GR112	A+	89% Q	
M	GOVERNMENT	CBSE	ENGLISH	GIRLS	NO	URBAN	NO	NO	NO	GR113	C	39% NQ	
M	GOVERNMENT	CBSE	ENGLISH	CO-ED	YES	URBAN	YES	YES	YES	GR114	A	76% Q	
M	GOVERNMENT	CBSE	ENGLISH	GIRLS	YES	URBAN	YES	YES	YES	GR115	A+	82% Q	

Figure.2. (a) student data used for calculation

M	GOVERNMENT	CBSE	ENGLISH	GIRLS	YES	URBAN	YES	YES	YES	GR115	A+	82% Q	
M	GOVERNMENT	State Board	HINDI	BOYS	YES	URBAN	YES	YES	YES	GR116	C	32% NQ	
F	GOVERNMENT	State Board	HINDI	GIRLS	YES	URBAN	YES	YES	YES	GR117	A+	89% Q	
F	GOVERNMENT	CBSE	ENGLISH	CO-ED	NO	RURAL	YES	YES	NO	GR118	C	25% NQ	
M	PRIVATE	CBSE	ENGLISH	GIRLS	NO	URBAN	YES	YES	NO	PR103	B	68% Q	
M	GOVERNMENT	State Board	HINDI	BOYS	YES	URBAN	NO	NO	YES	GR119	B	65% Q	
M	GOVERNMENT	CBSE	ENGLISH	BOYS	YES	URBAN	NO	NO	YES	GR120	A+	87% Q	
F	PRIVATE	CBSE	ENGLISH	GIRLS	NO	RURAL	YES	YES	NO	PR104	B	69% Q	
M	GOVERNMENT	State Board	HINDI	CO-ED	NO	URBAN	YES	YES	NO	GR121	C	33% NQ	
M	GOVERNMENT	State Board	HINDI	BOYS	YES	RURAL	YES	YES	YES	GR122	A	72% Q	
F	PRIVATE	CBSE	ENGLISH	BOYS	YES	URBAN	YES	YES	YES	PR105	A+	87% Q	
F	GOVERNMENT	CBSE	ENGLISH	GIRLS	NO	URBAN	YES	YES	NO	GR123	C	46% NQ	
M	GOVERNMENT	CBSE	ENGLISH	BOYS	NO	URBAN	YES	YES	NO	GR124	B	64% Q	
M	GOVERNMENT	State Board	HINDI	GIRLS	YES	RURAL	YES	YES	YES	GR125	A+	89% Q	
M	PRIVATE	State Board	HINDI	CO-ED	NO	URBAN	YES	YES	NO	PR106	C	47% NQ	
M	GOVERNMENT	CBSE	ENGLISH	GIRLS	NO	RURAL	NO	NO	NO	GR126	C	44% NQ	
F	PRIVATE	CBSE	ENGLISH	BOYS	YES	URBAN	YES	YES	YES	PR107	B	67% Q	
M	PRIVATE	CBSE	ENGLISH	BOYS	NO	RURAL	YES	YES	NO	PR108	A	75% Q	
M	GOVERNMENT	State Board	HINDI	BOYS	YES	URBAN	NO	NO	YES	GR127	A	74% Q	
M	GOVERNMENT	State Board	HINDI	GIRLS	NO	RURAL	YES	YES	NO	GR128	A	76% Q	
M	PRIVATE	CBSE	ENGLISH	CO-ED	NO	RURAL	NO	NO	NO	PR109	C	47% NQ	
F	GOVERNMENT	CBSE	ENGLISH	BOYS	NO	URBAN	YES	YES	NO	GR129	A+	80% Q	
M	PRIVATE	State Board	HINDI	BOYS	YES	URBAN	YES	YES	YES	PR110	C	43% NQ	
M	PRIVATE	State Board	HINDI	CO-ED	YES	URBAN	YES	YES	YES	PR100	A	70% Q	
F	GOVERNMENT	State Board	HINDI	BOYS	NO	URBAN	YES	YES	NO	GR101	C	41% NQ	
F	PRIVATE	CBSE	ENGLISH	GIRLS	NO	RURAL	YES	YES	NO	PR100	A+	90% Q	
F	GOVERNMENT	CBSE	ENGLISH	CO-ED	YES	URBAN	YES	YES	YES	GR102	C	40% NQ	

Figure.3. (b) student data used for calculation

IV. RESULT AND ANALYSIS:

This section presents the Results on the basis of different different attributes like Mobile, Net Access and Mode of Interaction.

A. Simulation Result student having Mobile Phone

The result obtained by implementing Naïve Bayes algorithm on the student data set on mobile number attribute how many student qualified and how many are not qualified using mobile

phone above is shown in the table 5.1. The results has been generated by,

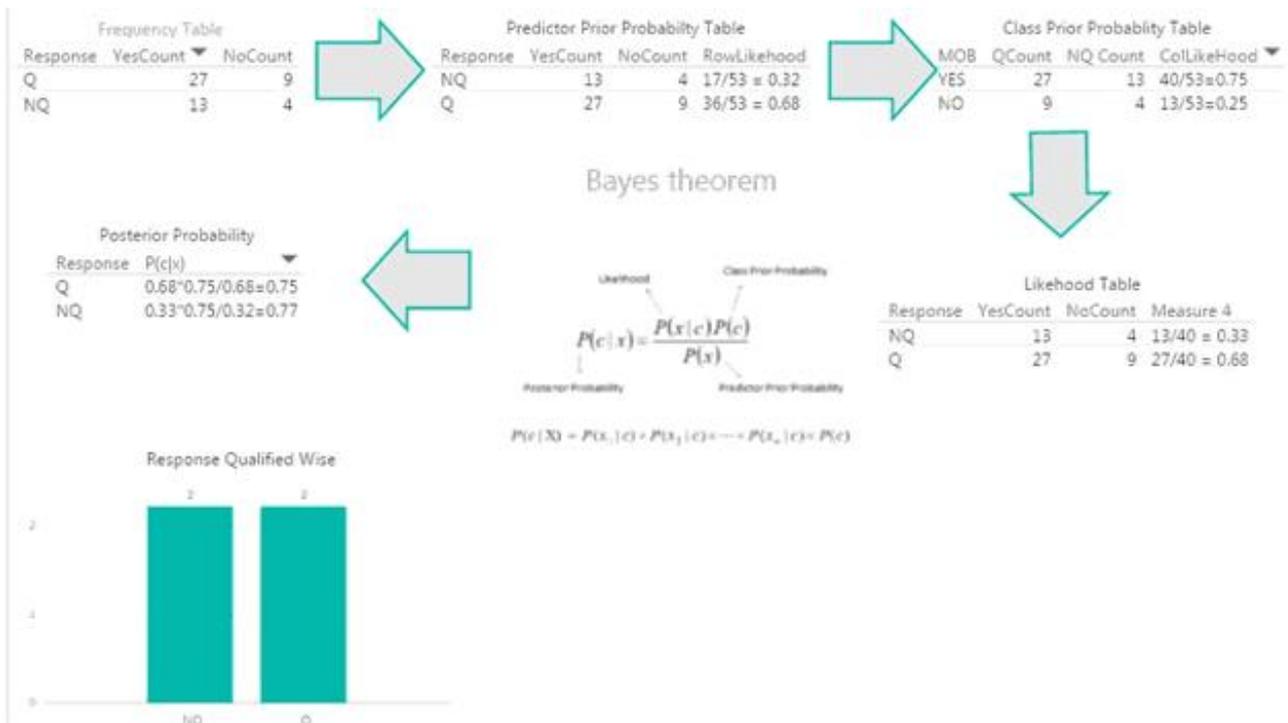


Figure.4. the results has been generated

We can say that if student using mobile phone then there is chances of not qualified student is .77 and qualified student is .75.

B. Simulation Result student having Net Access at Home:
The result obtained by implementing Naïve Bayes algorithm

on the student data set on Net Access in Home mobile number attribute how many student qualified and how many are not qualified using Net Access IN Home The results has been generated by Power BI



Figure.5. Simulation Result student having Net Access at Home

We can say that if student having Net Access at home then there is chances of not qualified student is 61% and qualified student is 63 %.

C. Simulation Result student having MOI Access IN Home using Naïve Bayes Algo: The result obtained by

implementing Naïve Bayes algorithm on the student data set on MOI attribute how many student qualified and how many are not qualified because of Mode Of Interaction , results has been generated by Power BI

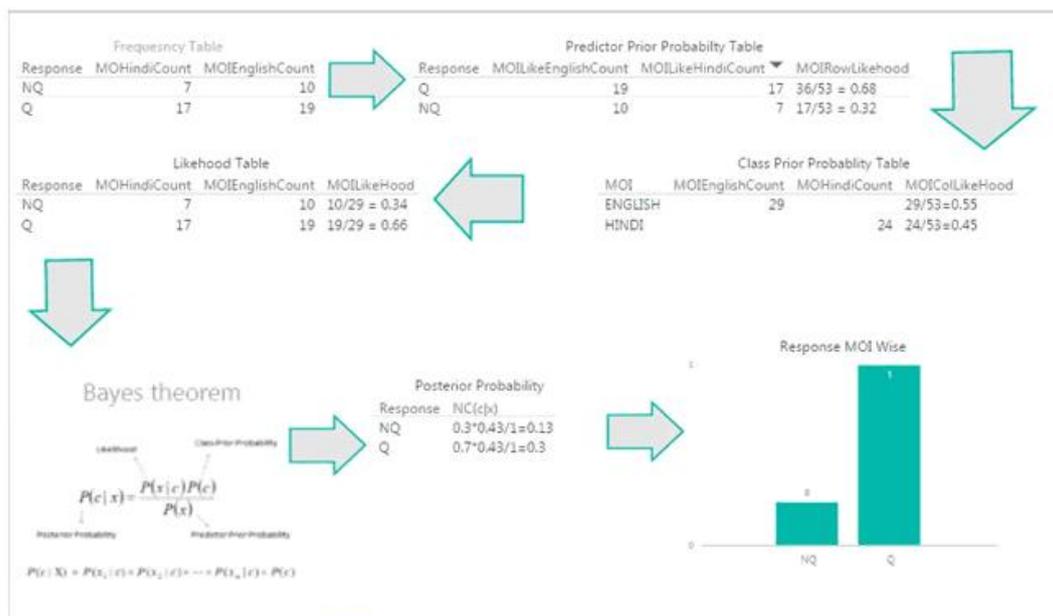


Figure.6. Generated by Power BI

If student using English Mode of Interaction then there is high chances of qualified and less chances of not qualified

4.4 Analysis: This section presents comparative analysis of three results on different attribute.

4.4.1 Comparative analysis of three charts

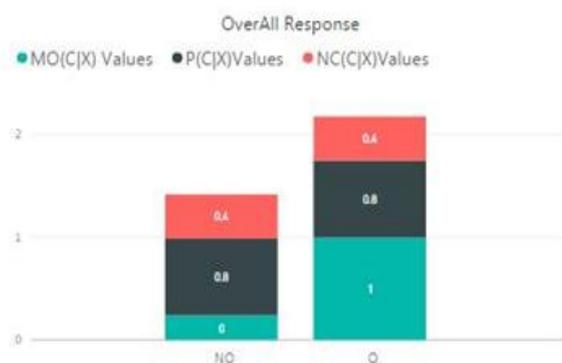


Figure.7. Summary Response Result

V. CONCLUSION

In this Paper we conclude how we can show the performance review of student using single single attribute like mobile phone, net access, computer at home and comparative analysis of all the result in the form of chart. By this it is very easy to analyse the performance on the basis of different different attributes using Naive Bayes Classification Algorithm.

VI. FUTURE SCOPE

The future scope of the work may be extended by presenting the same thing using different different attributes and give new performance chart using Naive Bayes algorithm. We can implement different classification algorithm for the same on Power BI.

VII. REFERENCES

[1]. Bandana Garg," DESIGN AND DEVELOPMENT OF NAIVE BAYES CLASSIFIER", North Dakota State University of Agriculture and Applied Science, pp 1-43, 2013

[2]. S.L. Ting, W.H. Ip., "Is Naive Bayes is good classifier for Document Classification," International Journal of Software Engineering and Its Applications, vol. 5, pp. 41-46, 2011

[3]. Swati,Rajendra Kaur," Multifactor Naive Bayes Classification for the slow learner prediction over multiclass student data set", International Journal on Computational Science & Applications , vol 6 ,pp 1-16,2016

[4]. Mukesh Kumar, Shankar Shambhu, Poonam Aggarwal "Recognition of slow learner using classification data mining techniques", Imperial Journal Of Interdisciplinary Research vol 2,pp 741-747,2016

[5]. K. Prasada Rao, M.V.P Chandra Shekhara,B.Ramesh "Predicting Learning Behavior of students using classification techniques", Intrnational Journal Of Computer Application, vol 139 ,pp 15-19,2016

[6]. P .Kavipriya" A review on predicting students academic performance Earlier, Using data mining techniques", Intrnational Journal of Advance Research in computer science and software Engineering, vol 6, pp 101-105, 2016

[7]. Shivani Rana,Roopali Garg " Slow Learner Prediction using Multi-Variate Naive Bayes classification algorithm", Intrnational Journal Of Engineering and technology Inovation, vol 7,pp 12-23,2017

[8]. Rish "An empirical study of the naive Bayes classifier", .J. Watson Research Center,pp 1-46

[9]. Parneet Kaura, Manpreet Singhb, Gurpreet Singh Josanc," Classification and prediction based data mining algorithms to predict slow learners", International Conference on Recent Trends in Computing, pp 501-508, 2015

[10]. I. Rish, "An empirical study of the naive Bayes classifier," IJCAI 2001 Workshop on Empirical Methods in Artificial Intelligence, vol. 22, pp. 41-46, 2001.