International Journal of Research in Advent Technology, Vol.7, No.1, January 2019
E-ISSN: 2321-9637
Available online at www.ijrat.org

Empirical analysis of Birth Registration E-governance data using Naïve Bayes Data Mining

Dr Pushpal Desai[1]
Department of ICT[1], Veer Narmad South Gujarat University, Surat[1]
Email: desaipushpal@hotmail.com

Abstract-The Naïve Bayes algorithm is based on Thomas Baye's theorem. The Naïve Bayes algorithm is very simple to understand and Naïve Bayes algorithm can be effectively applied to solve verity of problems. In this paper, Naïve Bayes algorithm is applied on Birth Registration E-governance data. The implementation of Naïve Bayes algorithm finds novel relationship among various attributes related to Birth Registration e-governance data.

Index Terms-Naïve Bayes, Birth Registration, Data Mining.

1. INTRODUCTION

In Naïve Bayes data mining, all attributes are considered independent and all attributes are considered for equal importance. In the algorithm, the hypothesis is that the given data belong to a particular class [1]. We need to calculate probability to calculate that our hypothesis is true [1]. The major advantage of this algorithm is that we need to scan all data only once [1]. As per Bayes theorem,

$$(P(E|H) = (P(E|H)\ X\ P(H))/P(E)$$

Where H is a hypothesis, E is the evidence associated with the hypothesis. The evidence is determined by value of the input attributes. P(E|H) is the conditional probability that H is true given evidence E. P(H) is an a priori probability, which is probability of hypothesis based on evidence [2]. The Naïve Bayes algorithm is applied to solve different types of problems. Jasmine Norman et al, applied Naïve Bayes algorithm for sentiment analysis. Jasmine Norman et al. considered demonetization and budget as case study [3]. Their finding suggests that Naïve Bayes algorithm can be effectively utilized in sentiment analysis [3]. Pisote and Bhuyar reviewed use of Naïve Bayes algorithm for Opinion Mining [4]. Ankita R. Borkar and Prashant R. Deshmukh proposed use of Naïve Bayes algorithm for Swine Flu prediction [5]. Ishtiaq Ahmed et al used Naïve Bayes algorithm for SMS classification [6]. Mahajan Shubhrata D et al proposed block diagram based on Naïve Bayes algorithm for stock market prediction [7]. Babajide O. Afeni et al research paper proposed Naïve Bayes classifier for Hypertension Prediction. Their results indicate that Naïve Bayes classifier can be effectively used for prediction for diseases [8]. S Vijayarani and S Deepa proposed Naïve Bayes Classification for Predicting Diseases in Haemoglobin Protein Sequences. Their results suggest that Naïve Bayes Classification is useful in predicting diseases [9]. Considering various literature, in this paper, all attributes related to Birth Registration data are considered to predict Delivery Attention ID and Delivery Method ID. The finding of the implementation shows novel trends and interesting relationships among different attributes.

2. METHODOLOGY

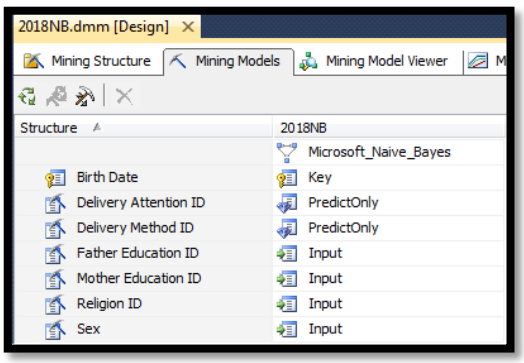
The Microsoft Analysis Service is used to implement Naïve Bayes algorithm. The algorithm implements Bayes' theorems and it can be used for data exploration and prediction [10]. This algorithm calculates conditional probability between input attributes and output attributes and it also consider all attributes independent. As all attributes are independent it is referred as Naïve [11]. The Microsoft Naïve Bayes algorithm was implemented considering parameters mentioned as per Table 1 [11].

Table 1. Naïve Bayes algorithm Parameters.

Name of parameter	Description
MAXIMUM_ INPUT_ ATTRIBUTES	To configure number of input attributes
MAXIMUM _OUTPUT _ATTRIBUTES	To configure number of output attributes
MINIMUM _STATES	To configure number of values to be considered for attributes

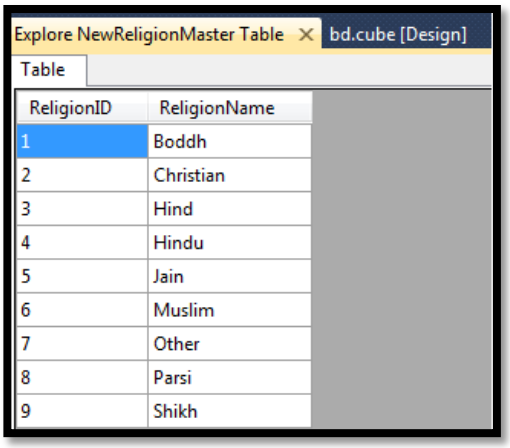
In the next phase, various input attributes and output attributes were selected to implement Naïve Bayes algorithm. The Birth Date attribute was selected as key value as it is required to identify each row uniquely from the data set. Considering analytical needs, various attributes were selected as per Figure 1.

Figure 1. Attribute selection



The Religion ID attribute contains values as per Figure 2.

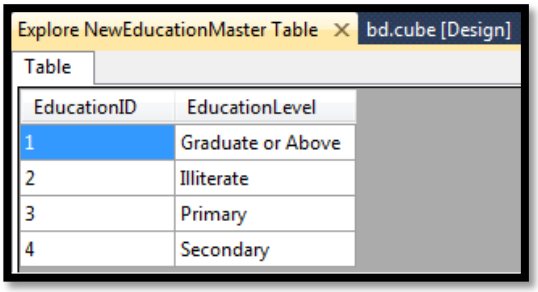
Figure 2. Religion ID Attribute values



ReligionID	ReligionName
1	Boddh
2	Christian
3	Hind
4	Hindu
5	Jain
6	Muslim
7	Other
8	Parsi
9	Shikh

The Sex attribute contain value Male or Female. The Father Education ID and Mother Education ID attributes contain values as per Figure 3.

Figure 3. Education ID Attribute values



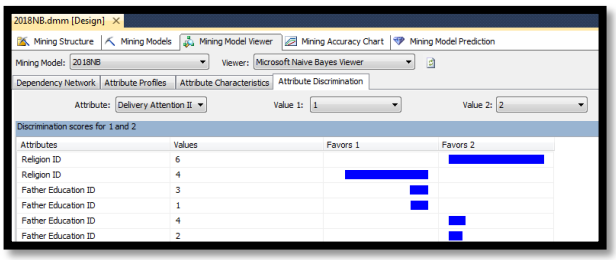
EducationID	EducationLevel
1	Graduate or Above
2	Illiterate
3	Primary
4	Secondary

Based on various parameters and attributes, Naïve Bayes data mining model was generated.

3. RESULTS

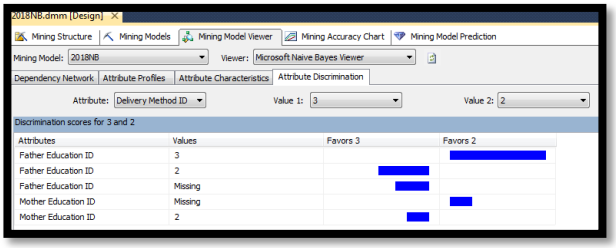
The Naïve Bayes data mining generated many interesting and novel trends from input attributes. For Delivery Attention ID predictable attribute, Naïve Bayes algorithm shows relationship among various input attributes. The Delivery Attention ID state 1= Doctor, Nurse or Trained Midwife and state 2= Institutional-Government is having novel relationship with Religion ID 6 and 4 which depicted in Figure 4.

Figure 4. Attribute Discriminations Delivery Attention ID



The result indicate that Religion ID = 6 (Muslim) favours Delivery Attention at Institutional-Government, whereas Religion ID = 4 (Hindu) favours Delivery Attention at Doctor, Nurse or Trained Midwife. Similarly, Delivery Method ID predictable attribute also shows interesting relationship among different input attributes. For example, Delivery Method ID=2 (Forceps/Vaccum) favours Father Education ID=3 which is Primary education, whereas Delivery Method ID=3 (Natural) favours Father Education ID=2 which is Illiterate. These results are shown in the Figure 5.

Figure 5. Attribute Discriminations: Delivery Method ID



4 CONCLUSION

The Naïve Bayes data mining model demonstrates that algorithm is very useful for finding novel relationship among various attributes and also for explanatory study of data. The study concludes that Religion ID attributes' value 6 and 4 are related with Delivery Attention ID attributes' value 2 and 1. Similarly, Delivery Method ID attributes' value 2 favours Father Education ID attributes' value 3 whereas Delivery Method ID attributes' value 3 favours Father Education ID attributes' value 2.

Acknowledgments

The Birth Registration data from year 2000 to year 2009 was provided by the municipal corporation for the research purpose only.

REFERENCES

[1] G. K. Gupta, Introduction Data Mining with Case Studies, PHI Publication..
[2] Richard Roiger, Michael Geatz, Data Mining A Tutorial Based Primer, Pearson Publication.
[3] Jasmine Norman et al, A NAIVE-BAYES STRATEGY FOR SENTIMENT ANALYSIS ON DEMONETIZATION AND INDIAN BUDGET 2017-CASE-STUDY, International Journal of Pure and Applied Mathematics, Volume 117 No. 17 2017, 23-31, ISSN: 1311 - 8080.
[4] PISOTE A. AND BHUYAR V., REVIEW ARTICLE ON OPINION MINING USING NAÏVE BAYES CLASSIFIER, Advances in Computational Research ISSN: 0975-3273 & E-ISSN: 0975-9085, Volume 7, Issue 1, 2015,p.-259-261.
[5] Ms. Ankita R. Borkar, Dr. Prashant R. Deshmukh, Naïve Bayes Classifier for Prediction of Swine Flu Disease, International Journal of Advanced Research in Computer Science and Software Engineering, Volume 5, Issue 4, April 2015, ISSN: 2277 128X.
[6] Ishtiaq Ahmed et al, SMS Classification Based on Naïve Bayes Classifier and Apriori Algorithm Frequent Itemset, International Journal of Machine Learning and Computing, Vol. 4, No. 2, April 2014.
[7] Mahajan Shubhrata D et al, Stock Market Prediction and Analysis Using Naïve Bayes, International Journal on Recent and Innovation Trends in Computing and Communication, Volume: 4, Issue: 11, pp- 124, ISSN: 2321-8169.
[8] Babajide O. Afeni1 et al, Hypertension Prediction System Using Naive Bayes Classifier, Journal of Advances in Mathematics and Computer Science, 24(2): 1-11, 2017; Article no.JAMCS.35610, ISSN: 2231-0851.
[9] S Vijayarani and S Deepa, Naïve Bayes Classification for Predicting Diseases in Haemoglobin Protein Sequences, International Journal of Computational Intelligence and Informatics, Vol. 3: No. 4, January - March 2014.
[10] https://docs.microsoft.com/en-us/sql/analysis-services/data-mining/microsoft-naive-bayes-algorithm-technical-reference?view=sql-server-2017, Last access date 29 Jan 2019
[11] https://docs.microsoft.com/en-us/sql/analysis-services/data-mining/microsoft-naive-bayes-algorithm-technical-reference?view=sql-server-2017, last access date 29 Jan 2019