BIG DATA AND BAYES THEOREM USED ANALYZE THE STUDENT'S PERFORMANCE IN EDUCATIONAL AND LEARNING FRAMEWORK

N.NASURUDEEN AHAMED[1], S.SARANYA DEVI[2], D. SAVITHA[3]

[1]*Assistant Professor, Dept of computer science engineering, Vivekanandha College of Engineering for Women, Tamilnadu, India*

[2,3]*U.G Scholar, Dept of computer science engineering, Vivekanandha College of engineering for women, Tamilnadu, India*

---***---

Abstract - *The paper presents a brief introduction to big data and its role in educational and learning applications. It is pragmatic that the use of big data architecture and techniques are continuously support in managing the speedy data growth in educational industry. Here, initially an experimental study is performed to analyze the role of big data in educational industry. It has been observed that important work has been done using big data in educational and learning. Nowadays, it is involved the based mathematics that has been increasingly used in artificial intelligence applications of computer science is "probability based decision making." The decision making (if, if-else, switch-case) based upon randomly generated numerical values. At the university and applied level we see decisions being made based upon more sophisticated criteria. A major part of this work is predicated upon knowledge of Bayes' Theorem. Here, a novel design of analyze the student's performance like to cross checking their assignments using the bayes theorem and its advanced mechanism to proposed to handle the big data of educational industry involving "probability based decision making."*

***Key Words*: Bayes Theorem, Decision Making, Classification, Supervised Learning, Regression.**

1. INTRODUCTION

Today, big data refers to the voluminous and complex amount of data collected from sources like web, enterprise applications, mobile devices and digital repositories which cannot be easily managed by using traditional tools. Big data is not only about the large data size; rather, it is an act of storing and managing data for eventual analysis. As the humans are getting digitized, therefore, the computing embroils data with greater variety, volume and velocity. The significance of 3V's is briefly mentioned below: - **Volume:** Big data analyze comparatively a huge quantity of data such as in terabytes. **Variety:** Big data incorporates data from distinct sources that appears in numerous formats such as structured, unstructured, multifactor, and probabilistic. **Velocity:** Big data handles the fast processing of data to promote the decision making process. In Computer Science Bayes Theorem is used in enhancing low resolution imaging and in filtering situations such as spam and noise filters...all situations involving "probability based decision making." In this research article, conceptual thoughts and theory of Bayes theorem are discussed in introductory part, then also discussed the application of Bayesian inference with illustration and discussed about Bayes in computational models, in sub sequential sections.

2. RELATED WORKS

Based on the concept of decision making system based on mathematics that has increasingly used in artificial intelligence application of computer science. Thomas Bayes develop a theorem to understand conditional probability. A theorem is a statement that can be proven true through the use of math. Bayes' theorem is written as follows:-

P (A | B)

This complex notation simply means, from the above notation is, the probability of event A given event B occurs. One key to understanding the essence of Bayes' theorem is to recognize that we are dealing with sequential events, whereby new additional information is obtained for a subsequent event, and that new information is used to revise the probability of the initial event. To use this concept to analyze the student's performance like how many students copy the assignment content of one student to other students. That means it's very helpful to rather than the cross check assignments in classroom. In Computer Science Bayes Theorem is used in enhancing low resolution imaging and in filtering situations such as spam and noise filters...all situations involving "probability based decision making."

3. PROPOSED SYSTEM

The main purpose of this method is used to analyze the student's performance in educational and learning framework like cross check the assignments. In, class room the teacher gives the common assignment topic to all students. After the submission the teachers to verify the cross checking assignment is complicated. So, analyze the student's performance using bayes theorem to decision making.

3.1 TECHNIQUES

3.1.1 BAYES THEOREM, AN ANTICIPATORY SET

In the educational world an "anticipatory set" is a preliminary discussion of a topic.

Problem: A given class room is 40% girls and 60% boys. 30% of the boy's carbon copy the assignment topic but only 10% of the girls carbon copy the assigment topic. A visual can make all this data immediately available to our brains to process.

Class Room P
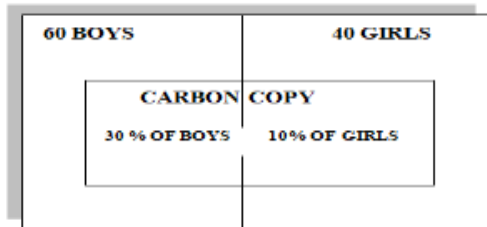
Probability a person chosen at random from set P is boys, **P (boys) = 60% of entire population**

Probability a person chosen at random from set P is girls, P **(girls) = 40% of entire population**

Probability of picking at random from set of boys a subject who likes to carbon copy:

P (carbon copy | boys) = 30% of boys (but 30% * 60% = 18% of entire population)

Probability of picking at random from the set of girls a subject who likes to carbon copy:

P (carbon copy | girls) = 10% of girls! (but 10% * 40% = 4% of entire population)

Notice the symbolism **P(carbon copy | boys) and P(carbon copy | girls).**

P (carbon copy | boys) is read as "the probability of choosing a person who likes to carbon copy from the set of boys."

P (carbon copy | girls) is read as "the probability of choosing a person who likes to carbon copy from the set of girls"

Probabilities of the form **P (set 1 | set 2) are called "conditional probabilities."**

> P (BOYS | CARBON COPY) = P (CARBON COPY | BOYS) * P (BOYS)
> $$\frac{}{P\text{ (CARBON COPY)}}$$
>
> ### BAYES THEOREM, A SNEAK PEAK

3.1.2 BAYES THEOREM, APPLIED...NO GRID

When doing an experiment there are four things that can happen. The experiment can:

1.) Correctly identify something as true (correct true test)

2.) Incorrectly identify something as true (false positive)

3.) Correctly identify something as false (correct false test)

4.) Incorrectly identify something as false (false negative)

For example: A staff declare that the student used for carbon copy there for assignment results in 95% true positives and 15% false positives.10% of the students did

carbon copy. What is the probability that your he/she uses carbon copy the assignment?

Let S represent those students on your team that use carbon copy. Hence S ' will be those students that do not use carbon copy. Let TP represent all the people who test positive for carbon copy...
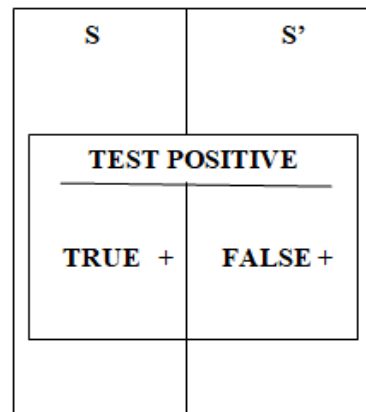
TP = true positive + false positive .Your friend just tested positive. What is the Probability that your friend is a user?

Given: P(S) = 10%

P (TP | S) = 95%

P (TP | S ') = 15%

Implied: P(S ') = 90%

The generic Bayes Theorems are:

$$P\ (A\ |\ B) = \frac{P\ (B\ |\ A) * P\ (A)}{P\ (B)}$$

$$P\ (A\ |\ B) = \frac{P\ (B\ |\ A) * P\ (A)}{P\ (B\ |A)*P\ (A) + P\ (B\ |A')*P\ (A)}$$

For this specific problem the formulas would be:

$$P(S\ |\ TP) = \frac{P\ (TP\ |\ S) * P(S)}{P\ (TP)}$$

$$P(S\ |\ TP) = \frac{P\ (TP\ |\ S) * P(S)}{P\ (TP\ |\ S)*P(S) + P\ (TP\ |\ S')*P(S')}$$

$$P(S\ |\ TP) = \frac{95\% * 10\%}{(95\% * 10\%) + (15\% * 90\%)}$$

$$= \frac{0.095}{0.23}$$

$$= 0.413$$

$$= 41\%$$

Probability, friends (41%) draw on carbon copy for another friend assignment topic in class room.

4. FUTURE SCOPE

The proposed methodology has describe predicts the class of data set is very fast. In spite of the great advances of the Machine Learning in the last years, it has proven to not only be simple but also fast, accurate and reliable. Naive Bayes is a family of probabilistic algorithms that take advantage of probability theory and Bayes' Theorem .In, future is easy to predict the carbon copy of assignments in educational and learning framework. In, Baye's Theorem is the statistical method for classification belongs to the supervised learning method and it's also easy to predicts the data set like classification and regression in the given data set. So, educational framework it's very fast and easily evaluates the student's performance. To, Baye's Theorem associated an underlying the probabilistic method. That is can solve the problems involving both categorical and continuous valued attributes.

5. CONCLUSION

The main aim of this work is to propose an Naive Bayes model is easy to build and particularly useful for very large data sets. Along with simplicity, Naive Bayes is known to outperform even highly sophisticated classification methods. It is easy and fast to predict class of test data set. It also performs well in multi class prediction. Naive Bayes is an eager learning classifier and it is sure fast. Thus, it could be used for making predictions in real time. This algorithm is also well known for multi class prediction feature. Here we can predict the probability of multiple classes of target variable. In, this method is going to be implemented in educational framework to analyze the student's performance like carbon copy the assignment topic among the students in class room very fast and reliable manner in statistical method of classification.

REFERENCES

[1]. Hitendra P. Dave, Krutarth H. Dave." Application of Bayesian Decision Theory In Management Research Problems" International Journal of Scientific Research Engineering & Technology (IJSRET) ISSN: 2278–0882, March-2015.

[2]. Thomas Z. Fahidy." Some Applications of Bayes' Rule in Probability Theory to Electrocatalytic Reaction Engineering" SAGE-Hindawi Access to Research International Journal of Electrochemistry Volume 2011, Article ID 404605, 5 pages doi:10.4061/2011/404605

[3]. Jasmine Norman, Mangayarkarasi R, Vanitha M, Praveen Kumar T, UmaMaheswari G." 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.

BIOGRAPHIES

N.Nasurudeen Ahamed received his Bachelor of Engineering degree in Computer science from Anna University Chennai in 2011, Master of Engineering degree in computer science from Anna University Chennai in 2013. He is currently working as an Assistant Professor in the Department of Computer science and Engineering at Vivekanandha College of Engineering for Women, Namakkal. His main research interests include IOT and Artificial Intelligence.