Application of Bayes Theorem to Accident Analysis and Classification

Chengjiang Xu*

Department of Applied Mathematics, Stony Brook University, Henan, China

*Corresponding author: ChengjiangXu@StonyBrook.edu

Abstract. In today's social context, "data" is a very common term. Data Science, Data analysis and other applications on different levels of Data have become very popular topics and even jobs. The main topic to be analyzed in this paper is Bayesian theorem. Bayesian theorem is a basic theorem in statistics, but with sufficient data support, Bayesian theorem can be applied in many fields. In this paper, this paper will analyze the application of Bayes' theorem in these two aspects: Bayes Theorem to Accident Analysis and Bayes Theorem to Classification. In terms of Bayes Theorem to Accident Analysis, this paper will analyze three papers that use Bayesian theorem to analyze accidents. In these papers, the authors have used Bayesian theorem to analyze different accidents. These accidents are: The causes and analysis of engineering accidents in mining areas, the analysis of causes of roof accidents in coal mines, and the analysis of the causes of highway traffic accidents. way of accident. In terms of Bayes Theorem to Classification, most of the authors use naive bayes to classify different data sets. The benefits of this classification are that whenever new data enters, researchers can easily classify new data sets. Incoming data is categorized.

Keywords: Bayes, Accident, Classification.

1. Introduction

Bayes' theorem is a relatively basic theorem in statistics, but now Bayes' theorem has been widely used in many ways. In this review of the paper, the application of Bayesian theorem in different fields will be analyzed. A total of six papers will be analyzed in this review of the paper. These six papers will be divided into two main parts, also It is the analysis of Bayesian theorem on accidents and the application of Bayesian theorem in classification. Bayesian can play a great role in statistics when there is sufficient data. In the paper analyzed in this paper, The authors have used Bayes' theorem or Naive Bayes to analyze the data and provide corresponding information in the context of sufficient data support.

For the literature analyzed in this paper, in terms of the analysis of accidents by Bayesian theorem, the papers analyzed in this review of paper are the analysis of the causes of engineering accidents in mining areas based on Bayesian networks, and the analysis of accidents in mining areas based on Bayesian networks. Research on the causes of roof accidents in coal mines, analysis of behavioral causes of highway traffic accidents based on Bayesian networks. For the application of Bayesian theorem to classification, the papers to be analyzed are as follows: detection of the number of lanes based on naive Bayesian classification, classification of Chinese comments based on naive Bayesian algorithm, and classification of water quality data based on naive Bayesian Research. First, in the papers on the analysis of accidents by Bayesian theorem, the author's steps can be roughly divided into three steps:

1. The author first proposes the types of accidents to be analyzed and explains the importance of accident prediction
2. The author's general process description based on Bayes' theorem
3. The author uses Bayesian theorem step by step to analyze the different causes of accidents

For example, in the paper Analysis of Causes of Engineering Accidents in Mining Areas Based on Bayesian Networks, the author first explained that once an accident occurs in a mine, it will bring serious consequences such as casualties, property losses, etc., so avoid It is necessary to have an accident, which corresponds to the first step stated in the paper. Next, the author explained that

Bayesian theorem can be used to analyze the cause of the accident, and then the author explained how to use Bayesian Adams' theorem is used for analysis, and the data are summarized as diagrams.

In the application of Bayesian theorem for classification, the authors generally use the method: Naive Bayesian (this will be explained in more detail in the methodology section), the Naive Bayesian authors' analysis of the data Some features are used to classify and classify the data into different types by these data features, and these types are usually displayed at the end of the papers. For example, in the paper on the classification of water quality data based on Naive Bayesian, the author used Naive Bayesian to analyze the water quality in the data set. From the aspects of dissolved oxygen, permanganate index, ammonia nitrogen, and total phosphorus of water quality. The water quality was analyzed, and the water quality was divided into five categories through Naive Bayesian, and summarized with a chart. In this review of the paper, the analysis of Bayesian theorem in the analysis of accidents and the use of Bayesian theorem in classification problems will be carried out in two parts. In each part, three documents will be used for analysis. And explain the role played by Bayes' theorem.

2. Methodology

Bayesian theorem is an extension based on conditional probability. The role of Bayesian theorem is to speculate on the different combination probabilities of these events based on a set of events and the different probabilities of these events. Before using Bayes' theorem, the event set should satisfy the condition:

1. All events in event set A must be mutually exclusive
2. All events in event set A must be exhaustive
3. The probability of event set B is not 0

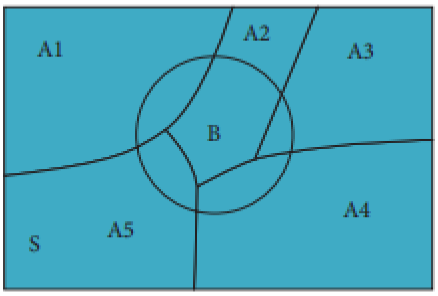

Fig. 1 Example of Bayes Theorem [1]

As shown in Fig.1, if the above conditions are met, then can use Bayes' theorem to find that when event B occurs, different events Ax (expressed here using A1, A2, A3, A4, A5) in the case that B has already occurred probability of occurrence.

In Bayesian theorem, P(A), P(B) is usually used to represent the probability of event A or event B occurring, as shown in Fig.1, if the requirement is to find out where B has occurred The probability of A happening, then the formula, then the Formula (1) is shown below [2]:

$$P(A \mid B)=\frac{P(B \mid A) * P(A)}{P(B)} \tag{1}$$

Among them, P(A | B) is mentioned in the requirement, when B is known to happen, the probability of A happening, and P(B | A) is the occurrence of event B when A is known to be correct The probability of these types of probabilities can be calculated using the method of calculating conditional probabilities, then the formula, then the Formula (2) is shown below [3]:

$$P(B \mid A)=\frac{P(A \cap B)}{P(A)} \tag{2}$$

It should be noted that when calculating P(B) here, it is necessary to list and calculate the probabilities of the simultaneous occurrence of A and B in all different situations. Assuming that the requirement is to find the probability of occurrence of P(Ak|B) (k is a specific number as shown in

Fig.1), then the probability can be calculated with the following formula, then the formula, then the Formula (3) is shown below [4]:

$$P(A\mathrm{k} \mid B) = \frac{P(Ak \cap B)}{P(B)} = \frac{P(Ak)P(B \mid Ak)}{\sum P(Ai)P(B \mid Ai)} \tag{3}$$

In the papers reviewed in this paper, especially in the application of Bayesian theorem in classification, the authors also use a branch of Bayesian theorem, called "naive bayes", which is a widely used Bayesian theorem in statistics, Naive Bayes has been simplified on the basis of Bayesian theorem, of course, there is also a key prerequisite for using Naive Bayesian, that is, all data sets The attributes and conditions of the data are independent of each other, which is consistent with the preconditions of Bayesian theorem, which also means that no attribute plays a decisive role in the classification of types in Naive Bayesian. Naive Bayesian Bayesian greatly simplifies Bayes' theorem and brings great help to classification.

3. The Role of Bayes Theorem in Accident Analysis

Bayesian theorem has a strong reasoning function in the case of sufficient information. In today's data age, the causes and frequencies of various accidents will be recorded by the government or relevant agencies, and this just satisfies the need to use Bayesian The precondition of Bayes' theorem is to have sufficient information, so in the field of analyzing the causes of accidents, Bayes' theorem has played a very good role.

In the paper "Cause Analysis of Engineering Accidents in Mining Areas Based on Bayesian Network", the author used Bayesian theorem to analyze the types and probabilities of different accidents caused by different factors in the mining area. Divided into three types and illustrated with the letter A plus different numbers, the main factors are:

1.A1: Management factors
2.A2: Human factors
3.A3: Environmental factors

The author classifies the different causes of accidents into these three factors, thus obtaining a tree diagram of accident causes

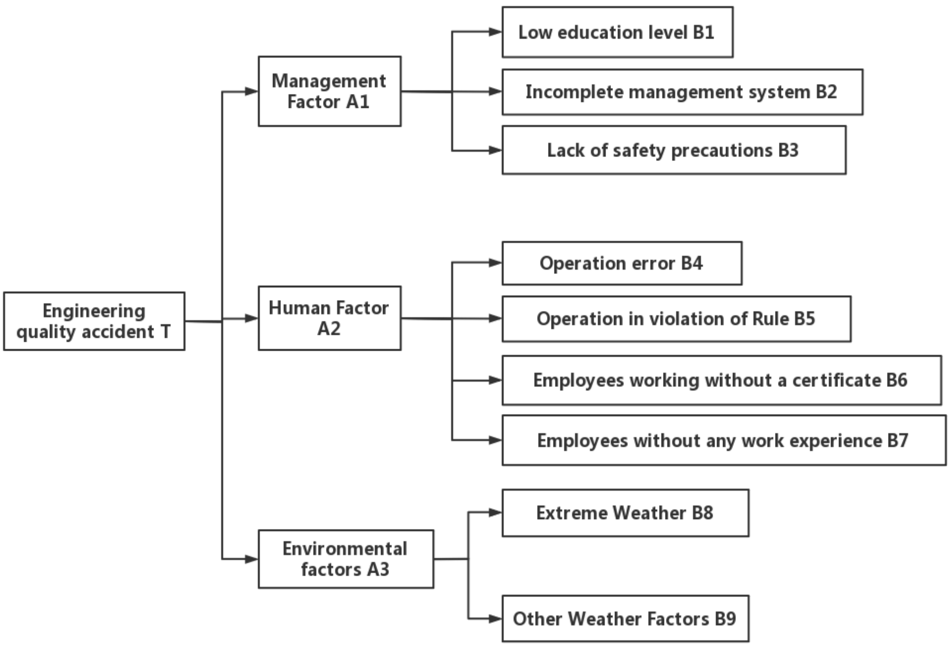

Fig. 2 A Bayesian tree diagram of the causes of accidents in mines

Photo credit: Original

As shown in Fig.2, it can be seen that each factor summarized by the author is followed by another letter B plus different numbers, and these letters B represent the specific cause of the accident, such as B1-Low Education Level, which is classified by the author as A1: Management factors, while B4-Opreation error is classified by the author as A2: Human Factor and other weather-like factors such as B8-Extreme Weather are classified by the author into A3: Environmental factor. Because there is sufficient data, the author can divide these different factors and find the corresponding probability after listing the tree-shaped diagram [5].

Table 1. Probability of Occurrence of Different Mining Accidents

Basic Event Content	Basic Event Probability/%
B1	15.71
B2	64.35
B3	5.41
B4	43.65
B5	34.42
B6	20.15
B7	10.49
B8	5.20
B9	10.84

As shown in Table 1, although with the advancement of science and technology, the probability of accidents in mines is gradually decreasing today, once an accident occurs in a mining area, the economic losses or casualties caused are immeasurable, but the causes of accidents in mining areas are very complicated, which is why The probabilities in the charts compiled by the author cannot add up to 100%, because there are some extremely accidental accidents that cannot be counted due to too few samples. But through the data compiled by the author, the managers of the mine can pay attention to what causes the accident in the mine, and use corresponding means to prevent the accident from happening. For example, the information that can be obtained from the chart provided by the author is A1: Management Factors and A2: Human factors are the most common cause of accidents in mines, so mine managers can mainly strengthen and improve these two factors, so as to avoid accidents, prevent casualties and economic losses.

The next article to be analyzed is about the analysis of coal mine roof accidents and their causes using Bayesian. First of all, the author explained the importance of this research. An accident on the roof of a coal mine will have a great negative impact. Once an accident occurs, coal mining enterprises will face great property losses and casualties. Being able to predict the cause of accidents on the roof of a coal mine in advance can help Greatly reduce the probability of accidents to avoid casualties and property losses. Here, the author will use Bayesian theorem to analyze the accidents with humans, equipment, environment, and management as the main factors. Before using Bayesian theorem, the author followed the basic rules of Bayesian theorem, which means having sufficient data. The data selected by the author here are sixty-four articles about coal mine roof safety accidents collected by the State Coal Mine Safety Supervision Bureau. Here, the author also replaces different factors with different letters, but the difference is that the author does not classify the main causes of the accident, and then classifies the small factors in the main causes with different letters. In this paper, The author directly divides the factors into letters, such as the letter "H" representing the thinking factor, the letter "D" representing the equipment factor, and the letter "M" representing the management factor. As shown in Table 2 below the author lists all the factors that lead to the accident directly in the table

Table 2. Causes of mine roof accidents [6]

Roof Accident	B	1(Y) / 2(N)
Execute Knockout	H_1	1(Y) / 2(N)
Into the Collapse zone	H_2	1(Y) / 2(N)
Fulfill Operating Produces	H_3	1(Y) / 2(N)
Safety Consciousness	H_4	1(Y) / 2(N)
Supervision and Inspection	H_5	1(Y) / 2(N)
Roof Collapse	D_1	1(Y) / 2(N)
Risk Identification Monitoring	D_2	1(Y) / 2(N)
Support Problem	D_3	1(Y) / 2(N)
Geological Condition	E_1	1(Y) / 2(N)
Roadway Deformation	E_2	1(Y) / 2(N)
Sound laws and Regulations	M_1	1(Y) / 2(N)
Response System	M_2	1(Y) / 2(N)
Technology Management	M_3	1(Y) / 2(N)
Production Management	M_4	1(Y) / 2(N)
Safety Management	M_5	1(Y) / 2(N)
Education and Training in place	M_6	1(Y) / 2(N)
Illegal Command	M_7	1(Y) / 2(N)
Illegal Work	M_8	1(Y) / 2(N)
Sufficient Talent Pool	M_9	1(Y) / 2(N)
System of Loading	M_{10}	1(Y) / 2(N)
Undocumented Employment	M_{11}	1(Y) / 2(N)

As shown in Table 2, after summarizing the causes of the accidents, the author's next step is to analyze the correlation between the causes of these accidents. The author uses the Bayesian model to screen out the correlations of different factors by reducing the dimension and simplifying the model, as shown in Table 2 shown.

Table 3. Relationship among causes leading to mine roof accidents [7]

B	H_3	-0.262
	M_5	-0.568
	M_8	0.201
	D_1	0.817
	D_3	-0.174
	M_3	0.115
H_2	H_3	-0.312
	H_5	0.286
	D_1	0.143
	D_3	0.302
	E_2	0.260
	M_8	0.102
H_5	D_3	0.213
	M_4	0.242
	M_9	0.221
D_3	M_3	0.134
	M_4	0.176
M_2	M_6	0.240
	M_9	0.252
	M_3	0.314
	M_5	0.260
	M_{10}	0.376
	M_{11}	-0.249

As shown in Table 3, the data shown here are variable symbols, that is, each individual letter that appears before all the letters. The appearance of variable symbols means that when this variable changes, it will also affect other variables. This is also in these individual letters. The meaning of those letters appearing after the letters, these letters are called related variables, and the meaning of the numbers behind these letters is the degree of change of the related variables when the value of the variable symbol changes, 0.01 is the smallest variable, and negative numbers are Values greater than 0.568. The complete data is not shown in Table 3, please refer to [7] for the complete data.

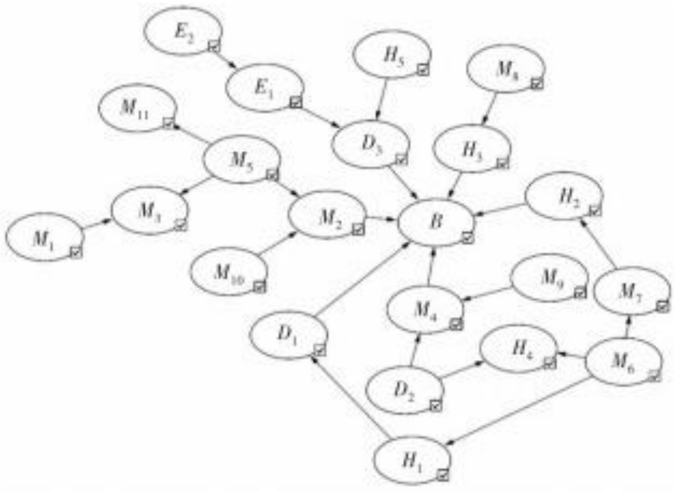

Fig. 3 Bayesian Networks of the Relationships among the Causes of Mine Roof Accidents [8]

As shown in Fig. 3, after obtaining the correlation of the variables, the author constructed the corresponding Bayesian network and imported the causes of all coal mine roofs. As shown in Fig. 3, it can be observed from this Bayesian network that most coal mine roof accidents are There are causal relationships. For example, E2 Roadway deformation is closely related to E1 geological conditions. This shows that when Roadway deformation occurs, the geological condition is very likely to occur. This can also warn coal mine managers that the most causally related Early warning of accident causes to avoid more accidents.

In the third article, the authors use Bayes' theorem to conduct a causal analysis of highway traffic accident behavior. The author first puts forward the importance of analyzing and telling the causes of highway traffic accidents. With the development of China's economy, the number of cars is also increasing with increase in economic development. Statistics on the causes of highway traffic accidents can greatly prevent accidents. occur. In this article, the author uses data to support a total of 77 motor vehicle traffic accidents on the Beijing-Harbin Expressway in 2019 according to the statistics of the Chinese government and uses Bayesian analysis of these data to obtain an analysis of the causes.

First of all, like the previous two articles, the author summarized the causes of accidents into different major factors. The first major factor is "unsafe actions". In the category of unsafe actions, all violations by drivers' Behaviors are categorized into this section. As shown in Table 4, the author summarizes the illegal operation of the driver itself, such as fatigue driving, speeding, and other factors, into the part of unsafe actions.

Table 4. Accident causes are classified under unsafe acts [9]

NUM		Frequency	Fre/%
1	Fatigue Driving	7	13.20
2	Over Speed	19	35.84
3	Sleazy Driving	17	32.07
4	The Driver Throws Illegally	2	3.77
5	Not Keeping a Safe Distance	8	15.09

As shown in Table 4, the information that can be obtained from the table listed by the author is that speed and sleazy driving occurred 19 times and 17 times respectively, which are the two accidents with the largest proportion of accidents caused by unsafe actions in highway traffic accidents, and then the author will make statistics on another main factor leading to accidents, which is the unsafe state of matter.

An unsafe state of matter is the cause of accidents not caused by improper operation of the driver. Here, the author analyzed 24 highway safety accidents caused by the unsafe state of matter. As shown in Table 5, from the table summarized by the author, the information can be obtained that other motor vehicle facility issues have caused the most highway safety accidents in the factor of unsafe state of matter.

Table 5. Causes of accidents classified as unsafe states of matter [10]

NUM		Frequency	Fre/%
1	Illegal Parking	15	28.30
2	Brakes Fail	9	16.98
3	Construction Accident	5	9.43
4	Other Motor Vehicle Facility Issues	24	45.28

As shown in Table 5, the data summarized by the author can make road managers and vehicle drivers understand how to avoid highway traffic accidents. For motor vehicle drivers, drivers can understand which factors are the most critical in unsafe behavior It may lead to traffic accidents. For example, after understanding the information summarized by the author, drivers can understand that they should not speed and sleazy driving because these two unsafe behaviors can easily lead to highway traffic accidents, and for roads from the information summarized by the author, managers can understand that all cars on the expressway should be inspected and vehicles that have not passed the safety test should be prohibited from driving on the expressway. Through such means, expressway traffic can be greatly reduced the occurrence of security incidents.

4. The Role of Bayes Theorem in Classification

In classification problems, Bayes' theorem can also solve problems very well. Here authors will use a branch of Bayes' theorem, which is the "naive Bayesian" mentioned in the article. These simple Bayesian network models can provide very accurate classification and can be used to solve real-world problems. The next paper Review uses Naive Bayesian to classify water quality. Like Bayesian theorem, using Naive Bayesian also requires sufficient data support. In this article, the author collected 500 pieces of water quality data. As a sample, the water quality was classified by four indicators including dissolved oxygen, permanganate index, ammonia nitrogen and total phosphorus. The author used A1, A2, A3, and A4 to refer to these four indicators respectively, and used the letter C to represent the five levels of water quality categories. After completing these, the author got the following formula, then the Formula (4) is shown below [11]:

$$\begin{aligned} &P(Ci \mid a1, a2, a3, a4) \\ &= \frac{P(a1, a2, a3, a4 \mid Ci)P(Ci)}{P(a1, a2, a3, a4)} \\ &= aP(a1, a2, a3, a4 \mid Ci)P(Ci) \end{aligned} \tag{4}$$

Naive Bayes assumes that all events are independent of each other, that is, defines "All events in event set A must be mutually exclusive", and obtains a new formula through probability multiplication, then the Formula (5) is shown below [12].

$$P(Ci \mid a1, a2, a3, a4) = aP(Ci)\prod_{j=1}^{N} P(aj \mid ci) \tag{5}$$

In this formula, the letter N represents the number of attributes. Naive Bayes calculates each variable Ci in this formula to find the category with the largest posterior probability and expresses this part of the formula with Cmap. Since all events are assumed to be independent, a will not affect the constant of Ci, which means that it has no effect on the settlement result. The formula of Cmap is[10], then the Formula (6) is shown below [13]:

$$Cmap = \arg\max P(Ci)\prod_{j=1}^{N} P(aj \mid ci) \tag{6}$$

After obtaining the prior probabilities P(Ci) and P(Aj|Ci), what the author did next was to perform a Naive Bayesian classification on all attributes. Before that, the author also made assumptions: the number of attributes is N, then the formula of the weight table W is, then the Formula (7) is shown below [14]:

$$N * P \backslash \sum_{k=1}^{N} Pk \tag{7}$$

Since the author defined the four factors that determine water quality as A1, A2, A3, and A4 before, and has obtained the formula of the W weight table, then the attribute A can be substituted into the weight Value table 2 (denoted by letters as W2) gets weight table 2, then the formula of the weight table W is, then the Formula (8) is shown below [15]:

$$aj \backslash \frac{1}{N}\sum_{k=1}^{N} ak \tag{8}$$

According to the definition of the final weight table, then the Formula (9) is shown below [16]:

$$wj = \frac{w1, j + w2, j}{2} \tag{9}$$

And after having all the information of weight table 1 and weight table 2, the author calculated the final weight table, which is weight table 3, and then the author called the formula of Cmap to get the classification results of different instances X.

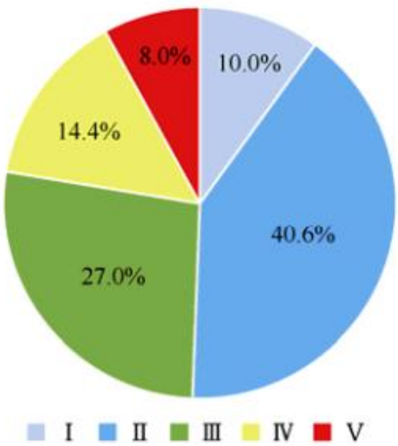

Fig. 4 Different types of water quality classification [17]

As shown in Fig.4, after a series of settlements of Naive Bayesian, the author obtained the final result of classifying water quality. The author integrated different indicators of different factors and divided the categories of water quality into five categories. The pie chart here is The proportions of the five water qualities in the 500 pieces of data.

In this paper, the author uses the Naive Bayesian algorithm to classify Chinese comments. The author mainly aims at the problem of comment classification. Most of the problems are solved by using the Naive Bayesian classifier. The author passed the Naive Bayesian algorithm. The Adams classifier screens the comments, and the author proposes that the benefits of doing so include managing users and managing the network environment. First of all, the author puts forward the principle of Chinese comment classification algorithm. First, it is necessary to determine the different attribute characteristics of different comments. After determining the characteristics, it is necessary to use the corresponding algorithm and bring it into the training data set for Naive Bayesian. Finally, it is necessary to perform classification tests and adjust weights, etc. factors to improve classification accuracy, as shown in Fig. 5.

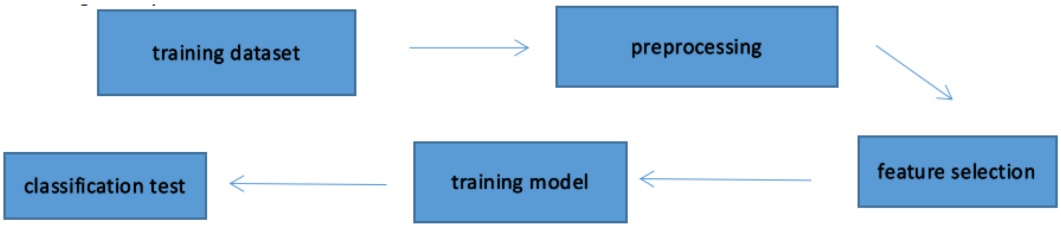

Fig. 5 Naive Bayesian classification process for Chinese reviews

Photo credit: Origin

Next, the author introduces the methods used in this paper in detail. First, the text classification process is introduced. The first step is text preprocessing, which mainly processes some common words in Chinese, such as The removal of word segmentation, the removal of stop words, etc., the author explained that the purpose of this step is to simplify the words to be analyzed, because a large number of repeated word segmentation and stop words will bias the data processing of Naive Bayes and lead to some expectations Errors other than that, after simplifying the Chinese words and words, the next step is to select further features. In this step, the author needs to identify what attributes can be used as the largest feature for distinguishing texts. The first feature is called document frequency, which is the frequency statistics of a certain attribute appearing in the target text. In this step, Naive Bayes will lower the weight of words that appear with high frequency, because it is always inevitable to use a certain attribute in the text. Some words, such as "the" and "that" in English, these similar words will also appear in Chinese texts, and avoid focusing on data processing in these words.

Table 6. The data obtained by the sample selected by the author through the Naive Bayesian classifier [18]

Type	NUM	PER/%
Positive	1123	51.51
Negative	1057	48.49
Total	2180	100

As shown in Table 6, by analyzing the 2180 samples taken this time, it can be found that there is not a big gap between negative and positive proportions, and through such classification sample data, network managers can have a general understanding of today's network environment and whenever new texts enter, managers can quickly classify.

In the third article, the author uses the naive Bayesian classification method to detect the number of cars in the lane. The author first puts forward the importance of this research. Guidance systems lead to more accurate analysis and software that allows assisted driving such as GPS to provide drivers with better predictions. Before conducting the research, the author puts forward two characteristics of the study, which are mass and real-time. Mass will lead to a large sample size, while real-time will lead to the number of lanes in a certain period. It can reflect the true characteristics of the number of lanes on the current road.

First of all, the author proposes two different road types, the first is the occluded class, and the author defines this road type as the letter a. The second type of road is unoccluded, and the author defines this type of road as the letter b. The difference between these two types of roads lies in type b. The road building density of the road with shelter is much higher than that of type a with shelter. The authors' data will also be different for these two different road types.

The data taken by the author in this article is supported by the floating car data collected by taxis in Wuhan, China as a database. Before processing these data, the author filters the data to filter irregular driving behaviors such as lane changes and stoppages. These irregular sample data will be filtered because these road data cannot make accurate predictions for the data. Next, the author used the naive Bayesian method to analyze the number of cars in different lanes. In the naive Bayesian method here, the author reassigned new letters to the two road types, that is, T1 with the occluded environment and T2 unoccluded environment, and introduces a new number of training samples N1 is the number of training samples in the unoccluded environment, N2 is the number of training samples in the occluded environment, where Hi represents the number of 4 lanes, and i is a different number (2, 3, 4, 5) represent one-way two-lane, one-way three-lane, one-way four-lane, one-way five-lane. After introducing these variables, the author brought these variables into the Naive Bayesian model and carried out analyze.

After substituting the data into the Naive Bayesian model, the author obtained the following data, as shown in Table 7:

Table 7. Number of vehicles in different lanes [19]

LANE	Data on Different Lanes /%				Correct rate %	Recall %
	2	3	4	5		
2	78.43	21.57	0.00	0.00	76.30	87.26
3	11.45	76.80	11.75	0.00		68.16
4	0.00	14.30	74.21	11.49		67.13
5	0.00	0.00	24.58	75.42		86.77
2	76.23	23.77	0.00	0.00	71.86	74.67
3	12.56	70.23	10.42	6.79		59.92
4	7.20	13.40	71.20	8.20		74.21
5	6.10	9.80	14.32	69.78		82.31

Through the data summarized by the author, it can be observed that under such complex road types and road conditions, the author can still achieve a correct rate of up to 70% using the naive Bayesian observation method, and the navigation can be greatly improved through these data correct rates.

5. Limitations & Future outlooks Suggestion

Bayes' theorem is widely used today. Mathematicians evaluate Naive Bayes as the simplest and most efficient classifier. This is also the case. When there is sufficient data, Naive Bayes can be very good. The data is classified according to certain characteristics, and when new data is added, Naive Bayes can quickly classify the new data by analyzing the characteristics of the data. The same Bayesian theorem is also very accurate in predicting the probability of the occurrence of a certain event in a specific situation. In the articles analyzed in this review of the paper, most of the results obtained by using Bayesian are accurate. The recall is higher than 90% or even higher than 95%, which fully proves the benefits of using Bayes' theorem.

Of course, Bayes' theorem is not perfect. Through this review of the paper, readers can observe that there is a precondition before analyzing each paper, that is, sufficient data must be available before using Bayes' theorem. However, in some cases, researchers do not have sufficien7t data, so Bayes' theorem cannot be used in these cases. And as mentioned in the Methodology part of this review of the paper, when the events in the event set are not mutually exclusive and not exhaustive, it is also impossible to use Bayes' theorem for analysis.

Of course, for the future of Bayesian theorem, through the review of so many papers, readers can understand that the future development of Bayesian theorem is very bright, especially in today's data age, the precondition of Bayesian theorem is sufficient The data is relatively easy to satisfy. With the improvement of data collection methods, Bayes' theorem will be applied to more and more fields. Nowadays, in computer languages such as R code and Python, Naive Bayesian is often used for classification. Many email companies use Naive Bayesian to distinguish spam from normal emails. Many pharmaceutical companies collect patient data on Whether the parents have had cancer and the patient's physical condition to analyze the probability of the patient getting cancer, these are new functions brought to Bayesian theorem with the development of science and technology, so Bayesian theorem will be accompanied by the advancement of science and technology progress.

6. Conclusion

By analyzing these papers, it can be found that Bayesian theorem has a wide range of functions in today's statistics. In the main analysis field of this paper: Bayesian theorem has brought great benefits to accident analysis and data classification. Benefits: In the accident analysis, Bayesian theorem analyzes the cause of the accident and summarizes the chart. Through the chart, the government or other departments can summarize the cause of the accident and find out the main cause of the accident to avoid the accident. occur to avoid personal injury or economic loss. In the application of Bayesian theorem to data classification, researchers can use the data classifier summarized by Naive Bayes to classify data and provide research assistance. This review of the paper selects these fields for analysis. Enables the reader to understand the application of Bayes' theorem in today's statistics.

References

[1] Cornfield, J. “Bayes Theorem.” Revue de l’Institut International de Statistique / Review of the International Statistical Institute, vol. 35, no. 1, 1967, pp. 34–49. JSTOR, https://doi.org/10.2307/1401634. Accessed 9 Dec. 2022.

[2] Efron, Bradley. “Bayes’ Theorem in the 21st Century.” Science, vol. 340, no. 6137, 2013, pp. 1177–78. JSTOR, http://www.jstor.org/stable/41985447. Accessed 9 Dec. 2022.

[3] PUZA, BOREK. “Bayesian Basics Part 1.” Bayesian Methods for Statistical Analysis, ANU Press, 2015, pp. 1–60. JSTOR, http://www.jstor.org/stable/j.ctt1bgzbn2.7. Accessed 9 Dec. 2022.

[4] Iversen, Gudmund R. “Statistics According to Bayes.” Sociological Methodology, vol. 2, 1970, pp. 185–99. JSTOR, https://doi.org/10.2307/270790. Accessed 9 Dec. 2022.

[5] Yao, Ding, and AnYa Qiang. “Cause Analysis of Engineering Accidents in Mining Area Based on Bayesian Network.” 5 Sept. 2022.

[6] Qing, Yan, and Wu Sheng. “Research on the Causes of Coal Mine Roof Accidents Based on Bayesian Network.” 25 June 2022.

[7] Qing, Yan, and Wu Sheng. “Research on the Causes of Coal Mine Roof Accidents Based on Bayesian Network.” 25 June 2022.

[8] Qing, Yan, and Wu Sheng. “Research on the Causes of Coal Mine Roof Accidents Based on Bayesian Network.” 25 June 2022.

[9] Luo, Xu, et al. “Behavior Analysis of Expressway Traffic Accidents Based on Bayesian Network.” 18 Apr. 2022.

[10] Luo, Xu, et al. “Behavior Analysis of Expressway Traffic Accidents Based on Bayesian Network.” 18 Apr. 2022.

[11] Fang, ZhiHao, et al. “Research on Water Quality Data Classification Based on Weighted Naive Bayes.” Mar. 2022.

[12] Fang, ZhiHao, et al. “Research on Water Quality Data Classification Based on Weighted Naive Bayes.” Mar. 2022.

[13] Fang, ZhiHao, et al. “Research on Water Quality Data Classification Based on Weighted Naive Bayes.” Mar. 2022.

[14] Fang, ZhiHao, et al. “Research on Water Quality Data Classification Based on Weighted Naive Bayes.” Mar. 2022.

[15] Fang, ZhiHao, et al. “Research on Water Quality Data Classification Based on Weighted Naive Bayes.” Mar. 2022.

[16] Fang, ZhiHao, et al. “Research on Water Quality Data Classification Based on Weighted Naive Bayes.” Mar. 2022.

[17] Fang, ZhiHao, et al. “Research on Water Quality Data Classification Based on Weighted Naive Bayes.” Mar. 2022.
[18] Ma, Wen, and Geng Chen. “Classification of Chinese Reviews Based on Naive Bayes Algorithm.” 31 Dec. 2021.
[19] Yang, Xue. “A Lane Quantity Detection Based on Naive Bayesian Classification.” 22 Mar. 2016.