Naive Bayes Classification

Pritish Samant
Electronics Engineering
Hamm-Lippstadt University Of Applied Sciences
Lippstadt, Germany
pritish-sanjay.samant@stud.hshl.de

***Abstract*—In this world of Advanced Technology, it becomes necessary to run text or speech recognition to spam out text and speech from the data present. Naive Bayes Classifier is one of the many algorithm used in classification for solving real life problems. It includes probabilistic approach in classification. In this paper, we will see what is Naive Bayes and its mathematical foundation for the algorithm and also implementation of Naive Bayes Classifier algorithm into a machine learning model using a small data set.**

I. Introduction

Machine Learning can be divided into two parts, Supervised learning and Unsupervised learning. Further Supervised learning is categorized into two, Classification and regression. Further more, Regression and Classification are further categorized accordingly. Supervised learning is technique used to train model with known input and output for predictions and unsupervised learning is a technique which finds hidden patterns or features from input data. Our topic Naive bayes classifies as a classification technique under supervised learning. In this paper we will focus on Classification. In Classification, output of the model is discrete. The most common types of Classification are Logistic regression, Support vector Machine, Naive Bayes Classifier, Decision trees. Some of the algorithms are common both in Supervised Learning and Unsupervised learning. When the output is continuous, it is Unsupervised learning and when the output is Discrete, its Supervised learning.

Global optimization is an area of mathematics and computer science that develops techniques for locating global minima and maxima of continuous domain functions or a group of functions for a given dataset.

After the boom of the internet, lot of data is generated everyday. As the technology progresses everyday, data generation also increases. It becomes very difficult to sort and manage such huge amount of data. Such unmanaged data in form of text or speech has no value whatsoever. All this has to be processed. After the processing, data can be used in the correct form to understand the data and develop further. There has been a good variety of successful real-life applications that are based on Naive Bayes classifier, such as weather prediction services, customer credit evaluations, health condition categorizations and so on[1].

Identify applicable funding agency here. If none, delete this.

Naive Bayes Classification is based on highest probability value. It predicts the text or speech with the available data and predicts what the new data means. Naive Bayes model is extremely fast in predicting as there are less parameters. It works smoothly on simple data sets. This classification doesn't work well with complex data. With complex data, it might not be as accurate as with simple data sets.

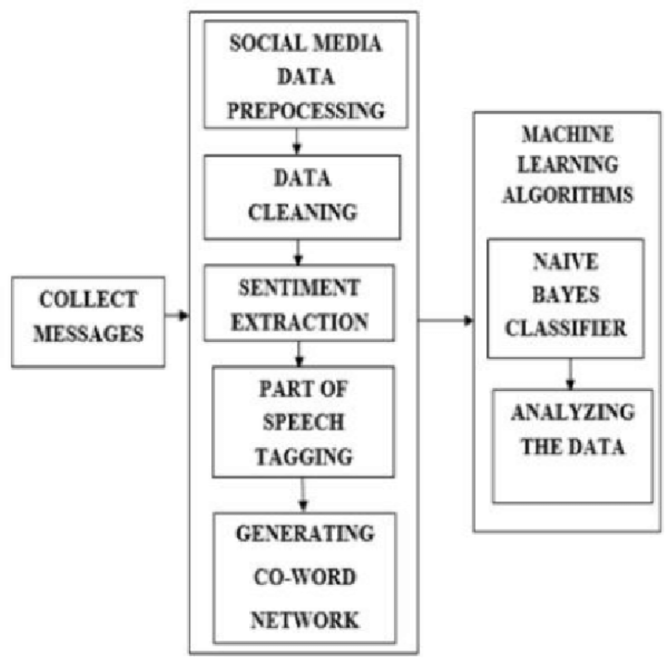

Fig. 1: Naive Bayes Classifier[2]

There are three types of Naive Bayes Classification, Multinomial Naive Bayes, Bernoulli Naive Bayes, and Gaussian Naive Bayes.

Multinomial Naive Bayes is used for sorting documents into different categories. It uses the predictors based on the frequency of words in a document which are pre-assigned to different categories. Bernoulli Naive Bayes uses Boolean variables and is extremely similar to Multinomial naive Bayes. Predictors in Gaussian Naive Bayes takes up a continuous value rather than discrete value.

II. Methodology

The methodology for the Naive Bayes starts with Bayes theorem. Before that we need to understand total probability theorem.let's assume the sets A and B.

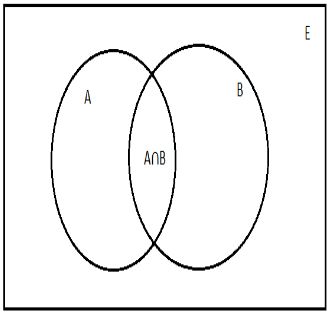

Fig. 2: Venn diagram for sets A and B

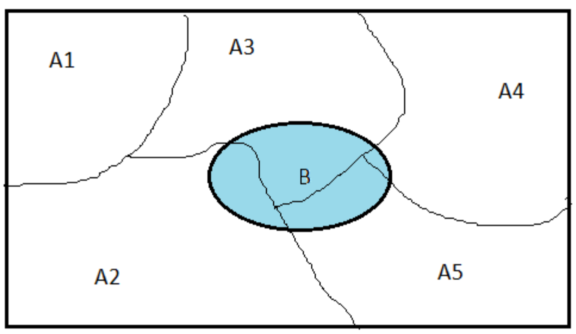

Fig. 3: Total Probability theorem

Before deriving Bayes' theorem, it is useful to consider the total probability theorem.

In probability theory, the elements $[\theta_i(0)]$ of the space $[\Theta]$ are experimental outcomes[2]. The subsets of $[\Theta]$ are called events and the event $[\theta_i]$ consisting of the single element $[\theta_i]$ is an elementary event[3]. The space $[\Theta]$ is called the sure event and the empty set $[\emptyset]$ is the impossible event[5]. We assign to each event A a number P(A) in [0, 1], called the probability of A, which satisfies the three Kolmogorov's conditions:
1) P($\emptyset$) $= 0$;
2)$P(\Theta)$=1; and
3) if A $\cap B = \{\emptyset\}, then[4], P(A \cup B) = P(A) + P(B)$. [5]

The fundamental Theorem of the probability theory is the Total Probability Theorem (TPT), also called a the law of total probability.[6]

Consider an event B and any partition1A$_1, A_2, \ldots, A_k of the space \theta.$
$Then$
$P(B) = P(B \cap A_1) + F(B \cap A_2) + \ldots + P(B \cap A_k)$.......[7]

Bayes Theorem can be stated as,

$$P(hypothesis|data) = [\frac{P(data|hypothesis)P(hypothesis)}{P(data)}]$$

[8]

"where P(data—hypothesis) is the likelihood of the data given the hypothesis ("if the hypothesis is true, then what is the probability of observing these data?"), P(hypothesis) is the prior probability of the hypothesis ("what is the a priori probability of the hypothesis?"), and P(data) is the probability of observing the data, irrespective of the specified hypothesis. The prior probability (short, prior) is also referred to as the (initial) degree of belief in the hypothesis. Bayes theorem states that , if we have to find probability of event A given Event B is true, we have to find probability of event B given event A is true multiplied by probability of event A and all divided by probability of event B. In other words, the prior quantifies the a priori plausibility of the hypothesis"[9]

This can also be written as,

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

III. IMPLEMENTATION

Data mining processes according to the cross industry-standard process data mining (CRISP-DM) has 6 phases, namely business understanding, data understanding, data preparation, modeling, evaluation, and deployment[10]. By considering the modularity and reusability, the entire scope of implementation is modeled into four class types, namely, the Priority Probability class type, the Conditional Probability class type, the Domain Knowledge class type, and the Classification class type[11].

Before training the model using naive bayes classification, we need to understand how the process works. Firstly, we need to collect the data then prepare the data in an organized way, do some exploratory data analysis on the data and then using python, we need to train the model using python libraries. Then the next step is to evaluate the model to find the conclusion or answers we are looking for. Once we have this, we must iterate the training session so as to find the best possible solution for our model.

To understand this better, we are going to use an example data set and work on its implementation. This data set is about the survival percentage on Titanic. This dataset was developed as a machine learning competition by kaggle. This data set contains data in form of test and train fulls. the data contains passenger information categorised based on gender, age, class, ticket(class), survival,etc. The data was developed for education purposes for machine learning training. In this report, I am using this data set to achieve the survival rate for people according to the mentioned categories. In order, for simplicity, we have restricted the number of entries(passenger id) to around 900.

For further implementation, we will see some entries from the data set so as we can know how the data set looks like. we will look at first 10 entries from the data set. this is the cleaned version of the data set[Fig.12].

After doing some analysis, we can say that majority of females survived even though there were more males than females. We can also see the count of males and females and total survival.

```
[ ] one['Survived'].value_counts()

0    549
1    342
Name: Survived, dtype: int64
```

Fig. 4: total survived

```
[ ] one['Sex'].value_counts()

male      577
female    314
Name: Sex, dtype: int64
```

Fig. 5: Male and Female

We will now start with the implementation of Bayes theorem using Machine learning.we are going to use data exploration firstly. For this, we are using google colab.

We have imported all the required libraries into google colab.For this, we are using pandas and Scikit learn libraries[fig.8].

Here, we have dropped most of the columns as they are of no use to us[fig.9].

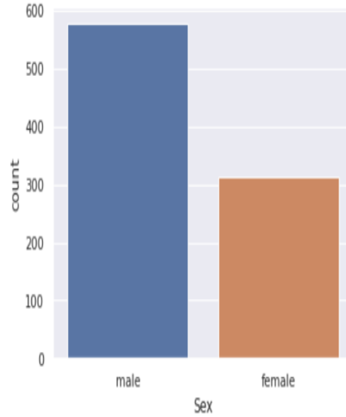

Fig. 6: Survived by sex

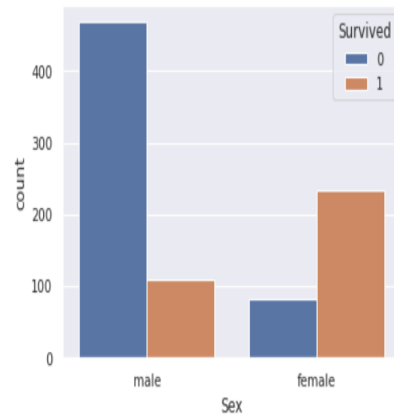

Fig. 7: Survival comparison by Gender

Now, we will use the dummies. dummies basically converts the texts into 1 and 0. this will help us later on. And then we use the concat function to use this stats in the main table[fig.10].

when we asked for the 10 values of age, we got NaN. the best way to deal with NaN value is to replace it with the

```
[114] import pandas as pd
      from sklearn.model_selection import train_test_split
      from sklearn.naive_bayes import GaussianNB
      from sklearn.model_selection import cross_val_score
```

Fig. 8

```
[117] one=df.drop(['PassengerId','Name','SibSp','Parch','Ticket','Cabin','Embarked'],axis='columns')

one.head()
```

	Pclass	Sex	Age	Fare	Survived
0	3	male	22.0	7.2500	0
1	1	female	38.0	71.2833	1
2	3	female	26.0	7.9250	1
3	1	female	35.0	53.1000	1
4	3	male	35.0	8.0500	0

Fig. 9

```
two = pd.concat([two,dummies],axis='columns')
two.head(5)
```

	Pclass	Sex	Age	Fare	female	male
0	3	male	22.0	7.2500	0	1
1	1	female	38.0	71.2833	1	0
2	3	female	26.0	7.9250	1	0
3	1	female	35.0	53.1000	1	0
4	3	male	35.0	8.0500	0	1

Fig. 10

```
[126] two.Age = two.Age.fillna(two.Age.mean())
      two.head(5)
```

Fig. 11

mean value of age[fig.11].

Now we will use the sklearn. This is how we train the split. I have divided my test and train sample to 80:20 ratio. As you can see, I have used size=0.2. To confirm the separation, you can also use 'len(X_train)'and 'len(X_test)' to check the ratio[fig.12].

Now we will use Naive Bayes model. For this data set we are using Gaussian Naive Bayes model.Here we created the model. Now we will train the model using (X_train) and (Y_train)[fig.13].

After training the model, we will check the score to measure accuracy. This score is going to vary every time you train the model[fig.14].

Then we use predict probability function to figure out survival in each class. The first part is the probability that the person didn't survive and the second part is the probability that the person survived. we can verify this with the predict model part to decide if the person survived or not[fig.15].

```
[127] X_train, X_test, y_train, y_test = train_test_split(two,surv,test_size=0.2)
```

Fig. 12

```
[128] titan = GaussianNB()

titan.fit(X_train,y_train)

GaussianNB()
```

Fig. 13

```
[130] titan.score(X_test,y_test)

0.7932960893854749
```

Fig. 14

```
[134] titan.predict_proba(X_test[:10])

array([[0.96292992, 0.03707008],
       [0.90336264, 0.09663736],
       [0.92386093, 0.07613907],
       [0.26979063, 0.73020937],
       [0.30227497, 0.69772503],
       [0.96042503, 0.03957497],
       [0.93811218, 0.06188782],
       [0.9630706 , 0.0369294 ],
       [0.92913774, 0.07086226],
       [0.96301718, 0.03698282]])

cross_val_score(GaussianNB(),X_train, y_train, cv=5)

array([0.74825175, 0.75524476, 0.8028169 , 0.80985915, 0.76056338])
```

Fig. 15

IV. CONCLUSION

In this paper, we have understood what is Naive Bayes algorithm. We have also tried to explain the mathematics behind the algorithm.And lastly, as an example, we used a dataset and tried to do some implementation using Naive Bayes Algorithm. As ther are three types of Naive Bayes, we had to use one of them and we have seen the logic behind choosing one of them. In some cases, Naive bayes reliability can be doubtful but from the implementation we can say that it is fast.

REFERENCES

[1] F. -J. Yang, "An Implementation of Naive Bayes Classifier," 2018 International Conference on Computational Science and Computational Intelligence (CSCI), 2018, pp. 301-306, doi: 10.1109/CSCI46756.2018.00065.

[2] P, Krishna Sankar Kumar, Sriram Sathyakala,. (2016). Predicting product features in large scale social networks using node classification.

[3] J. Dezert, A. Tchamova and D. Han, "Total Belief Theorem and Generalized Bayes' Theorem," 2018 21st International Conference on Information Fusion (FUSION), 2018, pp. 1040-1047, doi: 10.23919/ICIF.2018.8455351.

[4] J. Dezert, A. Tchamova and D. Han, "Total Belief Theorem and Generalized Bayes' Theorem," 2018 21st International Conference on Information Fusion (FUSION), 2018, pp. 1040-1047, doi: 10.23919/ICIF.2018.8455351.

[5] J. Dezert, A. Tchamova and D. Han, "Total Belief Theorem and Generalized Bayes' Theorem," 2018 21st International Conference on Information Fusion (FUSION), 2018, pp. 1040-1047, doi: 10.23919/ICIF.2018.8455351.

[6] J. Dezert, A. Tchamova and D. Han, "Total Belief Theorem and Generalized Bayes' Theorem," 2018 21st International Conference on Information Fusion (FUSION), 2018, pp. 1040-1047, doi: 10.23919/ICIF.2018.8455351.

[7] J. Dezert, A. Tchamova and D. Han, "Total Belief Theorem and Generalized Bayes' Theorem," 2018 21st International Conference on Information Fusion (FUSION), 2018, pp. 1040-1047, doi: 10.23919/ICIF.2018.8455351.

[8] Berrar, Daniel. (2018). Bayes' Theorem and Naive Bayes Classifier. 10.1016/B978-0-12-809633-8.20473-1.

[9] Berrar, Daniel. (2018). Bayes' Theorem and Naive Bayes Classifier. 10.1016/B978-0-12-809633-8.20473-1.

[10] E. Oktavianti, A. R. Yuly, F. Nugrahani and G. A. Siwabessy, "Implementation of Naïve Bayes Classification Algorithm on Infant and Toddler Nutritional Status," 2019 2nd International Conference of Computer and Informatics Engineering (IC2IE), 2019, pp. 170-174, doi: 10.1109/IC2IE47452.2019.8940894.
[11] F. -J. Yang, "An Implementation of Naive Bayes Classifier," 2018 International Conference on Computational Science and Computational Intelligence (CSCI), 2018, pp. 301-306, doi: 10.1109/CSCI46756.2018.00065.

PassengerId	Name	Pclass	Sex	Age	SibSp	Parch	Ticket	Fare	Cabin	Embarked	Survived
1	Braund, Mr. Owen Harris	3	male	22	1	0	A/5 21171	7.25		S	0
2	Cumings, Mrs. John Bradley (Florence Briggs Thayer)	1	female	38	1	0	PC 17599	71.2833	C85	C	1
3	Heikkinen, Miss. Laina	3	female	26	0	0	STON/O2. 3101282	7.925		S	1
4	Futrelle, Mrs. Jacques Heath (Lily May Peel)	1	female	35	1	0	113803	53.1	C123	S	1
5	Allen, Mr. William Henry	3	male	35	0	0	373450	8.05		S	0
6	Moran, Mr. James	3	male		0	0	330877	8.4583		Q	0
7	McCarthy, Mr. Timothy J	1	male	54	0	0	17463	51.8625	E46	S	0
8	Palsson, Master. Gosta Leonard	3	male	2	3	1	349909	21.075		S	0
9	Johnson, Mrs. Oscar W (Elisabeth Vilhelmina Berg)	3	female	27	0	2	347742	11.1333		S	1
10	Nasser, Mrs. Nicholas (Adele Achem)	2	female	14	1	0	237736	30.0708		C	1

Fig. 16: Dataset

V. Review

1. I find this paper very interesting.
2. It has lot of references.
3. Everything was explained in details.
4. Every paragraph there is a reference .
5. Everything figure is referenced name and explained.

VI. Declaration of Originality

I, ..., herewith declare that I have composed the present paper and work by myself and without the use of any other than the cited sources and aids. Sentences or parts of sentences quoted literally are marked as such; other references with regard to the statement and scope are indicated by full details of the publications concerned. The paper and work in the same or similar form have not been submitted to any examination body and have not been published. This paper was not yet, even in part, used in another examination or as a course performance. I agree that my work may be checked by a plagiarism checker.

18.06.2022&Lippstadt - Pritish Samant