An Efficient Naïve Bayes Classifier with Negation Handling for Seismic Hazard Prediction

Kalyan Netti
Senior Scientist
Knowledge Network Division
CSIR-National Geophysical Research Institute
Uppal Road, Hyderabad, India
netti_kalyan@ngri.res.in.in

Dr.Y Radhika
Associate Professor,
Dept. of Compute Science Engineering
GITAM University,
Visakhapatnam, India
radhika@gitam.edu

Abstract— **Classification is the one of the most important techniques in Datamining for data analysis. In Datamining, different Classification Techniques are available to predict outcome for a given dataset. There are many classification techniques for predicting and estimating accuracy, one such famous technique is Naïve Bayes Classifier. Naïve Bayes is very popular as it is easy to build, not so complex and when combined with smoothing techniques give better accuracy. In this paper Naïve Bayes Classifier for estimating Seismic Hazard activity is proposed. Hazard indicates a possible threat to life, health, property and environment. Mitigation of hazard when crossing stipulated level is very important, otherwise it may lead to an emergency. One of the most dangerous hazards in mining activities is Mining Hazard. Mineral, Diamonds/Gold and Coal exploration involves mining in a big way where hazard occurrence is quite common and addressing these mining hazards is a challenging task. An important threat of Mining Hazard is Seismic Hazard which is normal in underground mines. Thus Predicting Seismic Hazard is one of the most important aspect in countering Mining Hazards. In this paper the authors are proposing a new approach to improve accuracy of predicting seismic hazard by using Naive Bayes classifier with negation handling. This approach, outperforms the traditional Naïve Bayes Classifier in terms of accuracy.**

Keywords-Supervised Classification,Naïve Bayes, Prediction,Negation Handling, Smoothing, Accuracy, Gaussian Distribution,Seismic Hazard.

I. INTRODUCTION

Data Mining is a process of extracting useful and relevant information from data. There are many techniques in Data Mining to extract information from data. With different advanced technologies employed in the areas of engineering, finance, health etc., the data collected/accumulated, resulting in the exploration, is increasing exponentially. Now, with all the huge amounts of data available, the major task is to understand the data and extract knowledge from the data. In current scenario, extracting useful information from huge amounts of data is a complex task and need very efficient algorithms/techniques. This area is explored in a big way by employing new processes, methods along with statistical techniques. One such efficient technique in Data Mining is Classification. There are many Classification techniques available; like Bayesian Networks, Decision Trees, Nearest Neighbor, and Neural Networks. In general, classification is one of the analysis techniques, used to derive models by bringing in prior observations to predict the outcome. Naïve Bayes classifier (NBC) is a very popular and efficient technique in data classification. Naïve Bayes, a Supervised Classification Technique, is an efficient one because it is easy to build, computationally not complex and is capable of handling huge datasets. Moreover, Naïve Bayes Classifier performs well compared to other predictive models as it assumes conditional Independence among predictors.

This is the reason, for choosing NBC for accuracy estimation in seismic hazard event especially in Mining Activities. There is an urgent need to mitigate seismic hazard event and classification techniques may address this issue. One of the most dangerous hazards is Mining Hazard which are common in mining activities. An important threat of Mining Hazard is Seismic Hazards which is normal in underground mines. Thus Predicting Seismic Hazard is one of the most important aspect in countering Mining Hazards.

Also, in this paper an efficient Data Cleaning Technique is employed to further improve the accuracy in estimating Seismic Hazard.

Data cleaning is an important activity with techniques employed on datasets before using them for analysis by removing unwanted data or replacing missing values. Generally, smoothing techniques are used for data cleaning. There are different smoothing techniques like Dirichlet smoothing, Absolute Discounting smoothing to name a few. In this paper negation handling as a smoothing technique is chosen before doing classification. It is observed that the accuracy improved after applying smoothing to a greater extent.

The next Section i.e. Section-II, discusses the Naïve Bayes Classifier. Section-III describes Negation Handling and Section – IV describes the data, source and attributes. Section-V presents the implementation and results and the last section presents the conclusions.

II. NAÏVE BAYES CLASSIFIER

The Naïve Bayes Classifier is a datamining classification method which takes probabilities of attributes belonging to a class for prediction. NBC is a supervised classification approach which can be used effectively to model a predictive problem probabilistically.

Naïve Bayes classifier is based on Bayes' Theorem where predictors are treated as Independent. In Naïve Bayes method the overall probabilities of attributes belonging to a class are calculated by presuming that the probability of an attribute with respect to a given class value is not dependent on other attributes. This presumption leads NBC for better results and is called conditional independence.

Naïve Bayes theorem is described as follows,

$$\underbrace{P(c\mid x)}_{\text{Posterior Probability}}=\frac{\overbrace{P(x\mid c)}^{\text{Likelihood}}\overbrace{P(c)}^{\text{Class Prior Probability}}}{\underbrace{P(x)}_{\text{Predictor Prior Probability}}}$$

$$P(c\mid \mathrm{X}) = P(x_1\mid c)\times P(x_2\mid c)\times\cdots\times P(x_n\mid c)\times P(c)$$

The posterior probability, P(c/x) can be calculated, from P(C), Class Prior Probability, P(x) Predictor Prior Probability and P(x/c) Likelihood. The conditional Independence is explained in this scenario as, the predictor (x) value on a class (c) has no effect on the other predictor's values.

The Naïve Bayes Classifier in this paper takes numeric attributes as input and the values of each numeric attribute are Gaussian distributed. This was considered for robust results.

Gaussian Distribution

Gaussian distribution is defined by mean and standard deviation, which were defined as below

$$\mu=\frac{1}{n}\sum_{i=1}^{n}x_i \qquad \text{Mean}$$

$$\sigma=\left[\frac{1}{n-1}\sum_{i=1}^{n}(x_i-\mu)^2\right]^{0.5} \qquad \text{Standard Deviation}$$

$$f(x)=\frac{1}{\sqrt{2\pi}\sigma}e^{-\frac{(x-\mu)^2}{2\sigma^2}} \qquad \text{Gaussian distribution}$$

III. Negation Handling

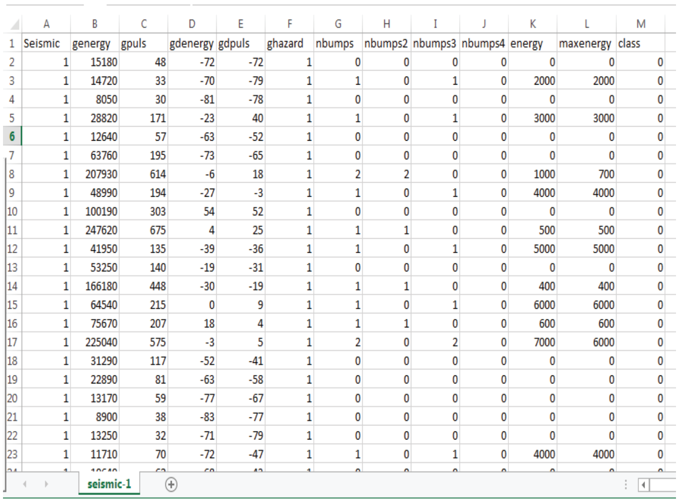
Negation Handling is a factor that plays a prominent role in estimating the accuracy in the classification. In this paper negation handling is carried out in a manner so that, wherever there is a sequence of values which are actually hazardous but not properly described in decision attribute are filled with appropriate value.

For example in the dataset which was considered here for accuracy estimation, there are instances where the reading of tremors with energy > 10^4j is recorded, decision attribute, which generally describes the event as hazard or non-hazard by 1 or 0, is either kept null or irrelevant. Thus, in this paper, the data was smoothed by using negation handling, before giving as an input to the Classifier so that the irrelevant/null field values are filled with appropriate values i.e 0 or 1 as per the energy recorded. This kind of smoothing technique leads Naïve Bayes Classifier to estimate exact accuracy by considering the entire data and gives best results when compared to NBC applied without smoothing.

IV. Data

Data in .csv format which was given as input to NBC for accuracy estimation was download from UCI Machine Learning Repository website which falls under Multivariate category. This data is a collection of forecasted seismic bumps in a coal mine, collected from two of long walls of a Polish coal mine [1].

Each row of the data describes the seismic activity in the rock mass within one shift (8 hours). If decision attribute has the value 1, then any seismic bump with energy > 10^4 J is registered in the next shift. This is the main attribute for accuracy estimation in this paper. A sample screen shot of data for accuracy estimation is shown in Fig-1.



	A	B	C	D	E	F	G	H	I	J	K	L	M
1	Seismic	genergy	gpuls	gdenergy	gdpuls	ghazard	nbumps	nbumps2	nbumps3	nbumps4	energy	maxenergy	class
2	1	15180	48	-72	-72	1	0	0	0	0	0	0	0
3	1	14720	33	-70	-79	1	1	0	1	0	2000	2000	0
4	1	8050	30	-81	-78	1	0	0	0	0	0	0	0
5	1	28820	171	-23	40	1	1	0	1	0	3000	3000	0
6	1	12640	57	-63	-52	1	0	0	0	0	0	0	0
7	1	63760	195	-73	-65	1	0	0	0	0	0	0	0
8	1	207930	614	-6	18	1	2	2	0	0	1000	700	0
9	1	48990	194	-27	-3	1	1	0	1	0	4000	4000	0
10	1	100190	303	54	52	1	0	0	0	0	0	0	0
11	1	247620	675	4	25	1	1	1	0	0	500	500	0
12	1	41950	135	-39	-36	1	1	0	1	0	5000	5000	0
13	1	53250	140	-19	-31	1	0	0	0	0	0	0	0
14	1	166180	448	-30	-19	1	1	1	0	0	400	400	0
15	1	64540	215	0	9	1	1	0	1	0	6000	6000	0
16	1	75670	207	18	4	1	1	1	0	0	600	600	0
17	1	225040	575	-3	5	1	2	0	2	0	7000	6000	0
18	1	31290	117	-52	-41	1	0	0	0	0	0	0	0
19	1	22890	81	-63	-58	1	0	0	0	0	0	0	0
20	1	13170	59	-77	-67	1	0	0	0	0	0	0	0
21	1	8900	38	-83	-77	1	0	0	0	0	0	0	0
22	1	13250	32	-71	-79	1	0	0	0	0	0	0	0
23	1	11710	70	-72	-47	1	1	0	1	0	4000	4000	0

Fig-1: Screenshot of Seismic Bumps Dataset

Value – '1' in the decision attribute i.e *Class* in the above dataset means a high energy seismic bump occurred in the next shift that means 'hazardous state', '0' indicates a not high energy seismic bumps occurred in the next shift that means a 'non-hazardous state' [1].

As mentioned in Section-III negation handling is used on the dataset for correcting fields where values are not mentioned even for higher energy Seismic Bumps.

V. IMPLEMENTATION

As mentioned in Section –III, NBC in this paper uses numerical attributes and implementation is divided into following steps,

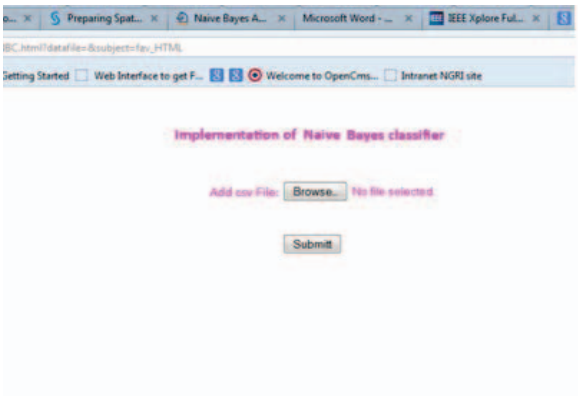
a. Negation Handling

Before giving as input to NBC the data which is in .csv format will be smoothened. The output data after negation has irrelevant/null field values filled with appropriate values i.e 0 or 1 as per the energy recorded.

b. *Naïve Bayes Classifier*

Deciding Training and Test data ratio

The dataset which is in .csv format after smoothing will be divided into training and test sets.
In this paper, 60% for training and 40% for test is considered, which is normal for testing algorithm on a dataset. The web application in this paper asks for the dataset, in .csv format to be uploaded for the prediction.



Implementation of Naive Bayes classifier

Add csv File: Browse... No file selected

Submitt

Fig-1

Calculation of probabilities and predictions

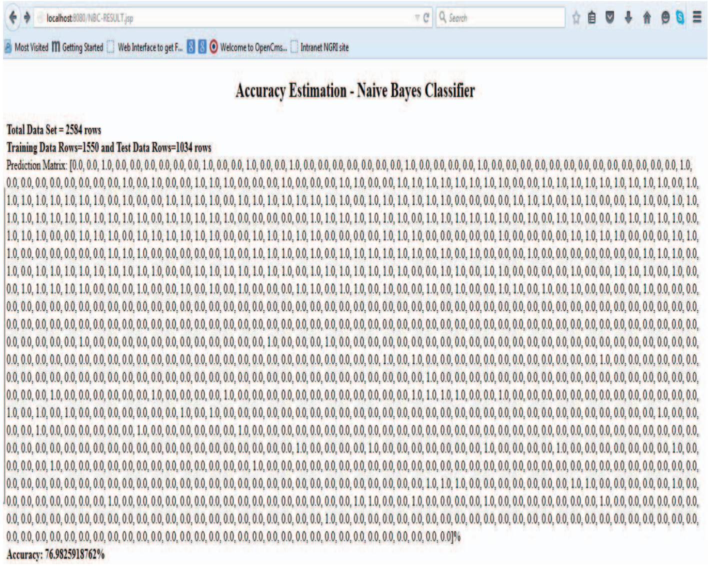
The Naïve Bayes model presented in this paper groups the properties of training dataset and this is used for calculating predictions. This part involves calculation of mean, standard deviation for each attribute belonging to a class. The mean and Standard deviation of each attribute is calculated for a Class. When calculating probabilities, this is used to define the spread of attributes belonging to a class in Gaussian distribution.

Prediction Calculation

After grouping data from the training set, NBC is ready to make prediction. With the probabilities belonging to each class that is available, the predication is calculated by picking the class with highest probability. As discussed in Section-II this can be carried out using calculation of Gaussian Function

Gaussian Probability Density Function

With the values of mean, standard deviation of attributes belonging to a class available from the training data, Gaussian function is used to calculate the probability of an attribute value. Knowing the likelihood of the attribute (here as per our dataset attribute 'class') value belonging to a class, the value of attribute, mean, standard deviation are given as input to Gaussian function. Now, with the probabilities of all the attribute values are combined, the probability of entire class can be calculated together by multiplying each other as mentioned in Section-II. The result is shown in Fig-2,



Accuracy Estimation - Naive Bayes Classifier

```
Total Data Set = 2584 rows
Training Data Rows=1550 and Test Data Rows=1034 rows
Prediction Matrix: [0.0, 0.0, 1.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 1.0, 0.0, 0.0, 1.0, 0.0, 0.0, 1.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 1.0, 0.0, 0.0, 0.0, 0.0, 1.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 1.0,
0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 1.0, 0.0, 1.0, 0.0, 0.0, 1.0, 1.0, 1.0, 0.0, 0.0, 0.0, 1.0, 0.0, 0.0, 0.0, 1.0, 1.0, 0.0, 0.0, 1.0, 1.0, 1.0, 1.0, 1.0, 1.0, 1.0, 1.0, 0.0, 0.0, 1.0, 1.0, 1.0, 1.0, 1.0, 1.0, 1.0, 1.0, 1.0, 0.0, 1.0,
1.0, 1.0, 1.0, 1.0, 1.0, 1.0, 1.0, 0.0, 1.0, 0.0, 0.0, 1.0, 1.0, 1.0, 1.0, 0.0, 1.0, 1.0, 1.0, 1.0, 1.0, 1.0, 1.0, 1.0, 0.0, 1.0, 1.0, 1.0, 1.0, 1.0, 1.0, 0.0, 0.0, 0.0, 0.0, 1.0, 1.0, 1.0, 0.0, 1.0, 1.0, 0.0, 0.0, 1.0, 1.0, 0.0, 1.0, 1.0,
1.0, 1.0, 1.0, 1.0, 1.0, 1.0, 1.0, 1.0, 1.0, 1.0, 0.0, 1.0, 1.0, 1.0, 1.0, 1.0, 0.0, 0.0, 0.0, 0.0, 0.0, 1.0, 1.0, 1.0, 1.0, 1.0, 1.0, 1.0, 0.0, 1.0, 1.0, 1.0, 1.0, 1.0, 1.0, 0.0, 1.0, 0.0, 1.0, 1.0, 0.0, 0.0, 1.0, 1.0, 1.0, 1.0, 1.0, 0.0,
1.0, 1.0, 1.0, 0.0, 0.0, 1.0, 1.0, 1.0, 0.0, 1.0, 1.0, 1.0, 1.0, 1.0, 1.0, 0.0, 0.0, 1.0, 1.0, 1.0, 1.0, 1.0, 0.0, 0.0, 0.0, 0.0, 1.0, 1.0, 1.0, 0.0, 0.0, 0.0, 0.0, 0.0, 1.0, 0.0, 0.0, 0.0, 0.0, 1.0, 1.0, 1.0, 1.0, 0.0, 0.0, 0.0, 1.0, 1.0,
1.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 1.0, 1.0, 1.0, 0.0, 0.0, 0.0, 0.0, 0.0, 1.0, 0.0, 1.0, 1.0, 1.0, 0.0, 1.0, 0.0, 1.0, 1.0, 1.0, 1.0, 1.0, 1.0, 0.0, 1.0, 0.0, 1.0, 0.0, 0.0, 0.0, 1.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 1.0, 1.0, 1.0, 0.0,
1.0, 0.0, 1.0, 1.0, 1.0, 1.0, 1.0, 1.0, 1.0, 1.0, 1.0, 0.0, 0.0, 1.0, 1.0, 1.0, 1.0, 1.0, 1.0, 1.0, 0.0, 1.0, 1.0, 1.0, 1.0, 1.0, 1.0, 1.0, 0.0, 0.0, 1.0, 1.0, 0.0, 1.0, 1.0, 0.0, 0.0, 0.0, 0.0, 1.0, 0.0, 0.0, 1.0, 1.0, 1.0, 0.0, 1.0, 0.0,
0.0, 1.0, 1.0, 1.0, 1.0, 1.0, 0.0, 0.0, 0.0, 1.0, 0.0, 0.0, 1.0, 0.0, 1.0, 0.0, 1.0, 0.0, 0.0, 0.0, 1.0, 1.0, 0.0, 1.0, 1.0, 0.0, 1.0, 0.0, 0.0, 1.0, 0.0, 1.0, 0.0, 0.0, 1.0, 1.0, 0.0, 1.0, 0.0, 1.0, 0.0, 0.0, 0.0, 0.0, 1.0, 0.0, 0.0, 0.0,
0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0,
0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0,
0.0, 0.0, 0.0, 0.0, 0.0, 1.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 1.0, 0.0, 0.0, 0.0, 1.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0,
0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 1.0, 0.0, 1.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 1.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0,
0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 1.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0,
0.0, 0.0, 0.0, 1.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 1.0, 0.0, 0.0, 0.0, 0.0, 1.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 1.0, 1.0, 1.0, 1.0, 0.0, 0.0, 1.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0,
1.0, 0.0, 1.0, 0.0, 1.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 1.0, 0.0, 1.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 1.0, 0.0, 0.0,
0.0, 0.0, 1.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 1.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 1.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0,
0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 1.0, 0.0, 0.0, 0.0, 0.0, 1.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 1.0, 0.0, 0.0, 0.0, 0.0, 0.0, 1.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 1.0, 0.0,
0.0, 0.0, 0.0, 1.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 1.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0,
0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 1.0, 1.0, 1.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 1.0, 1.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 1.0, 0.0,
0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 1.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 1.0, 1.0, 0.0, 0.0, 1.0, 0.0, 0.0, 0.0, 1.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 1.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0,
0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 1.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0,
0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0]%
Accuracy: 76.9825918762%
```

Fig- 2

Accuracy estimation

The accuracy is calculated as a value between 0% and 100% by comparing the resultant predictions with class values of the Test Data. This is called the accuracy ratio and as per the input dataset of seismic bumps the accuracy estimation with our proposed NBC in combination with Smoothing, is 76.983% as shown in Fig-2

When NBC without smoothing is employed on the same dataset with same Training and Test data ratio i.e. 60% and 40% the accuracy is around 64.5% which is significantly less when compared to the new approach proposed in this paper. The accuracy of different Naïve Bayes Classifier Algorithms when applied on the same dataset with same ratios of Training and testing are shown in Table-1.

Classifier	Training (60%)	Test (40%)	Accuracy
NBC with Negation Handling	1550	1034	76.983%
NBC without Negation Handling	1550	1034	64.5%
Matlab Native NBC Algorithm without Negation Handling	1550	1034	65.09%

Table-1: Accuracy comparison chart with different classifiers with same ratios of Training and Test datasets

As per the experimental results, the NBC classifier algorithm presented in this paper performs better when compared to the native MATLAB Naïve Bayes Function without Negation Handling, which gave an accuracy of around 65.09% when performed on the same dataset with same ratio of Training and Test data. Table-I illustrates that the proposed method has improved accuracy (+10%), due to preprocessing of input data using negation handling, thus reducing the impact of Class Conditional Independence, when compared to Matlab Native NBC (65.09%) and NBC (64.5%) without Negation Handling.

VI. Conclusions & Future Work

The proposed Naïve Bayes Classifier algorithm with Negation Handling used in this Prediction System for Seismic Hazard, with 60%-40% ratio of Training and Test Datasets, yielded a good accuracy of 76.983% which is better when compared to the native MATLAB Naïve Bayes Classifier without smoothing. Experimental results indicate that using NBC in combination with Negation Handling has significantly improved accuracy for better understanding, mitigating Seismic Hazards and Mining related dangers. Our future work will be exploring the feasibility of applying various distributions like Multinomial, Bernoulli etc. and using other smoothing techniques as mentioned in Section-I to further improve the performance of Naïve Bayes Classifier.

Acknowledgment

We sincerely acknowledge the creator and donors , Marek Sikora (marek.sikora@polsl.pl), Institute of Computer Science, Silesian University of Technology, 44-100 Gliwice, Poland and Lukasz Wrobel (lukasz.wrobel@polsl.pl), Institute of Innovative Technologies EMAG, 40-189 Katowice, Poland, which was download from UCI Machine Learning Repository.

References

[1] J. Han, M. Kamber, 'Data Mining Concept and Techniques', Morgan Kaufmann, 2001.

[2] Improving Naive Bayes Classifier Using Conditional Probabilities , Sona T, M Mammadov,A M. Bagirov, Proceedings of the 9-th Australasian Data Mining Conference (AusDM'11), Ballarat, Australia.

[3] Scaling Up the Accuracy of Naïve Bayes Classifers a Decision Tree Hybrid, Ron Kohavi

[4] Novel Frequent Sequential Patterns based Probabilistic Model for Effective Classification of Web Documents, H Haleem,P K Sharma,M M S Beg ,Proceeding in 2014 5th International Conference on Computer and Communication Technology

[5] An Improved Computation of the PageRank Algorithm, S J Kim, Sang H Lee, Springer-Verlag Berlin Heidelberg 2002.