International Journal of
Engineering in Computer Science

**E-ISSN:** 2663-3590
**P-ISSN:** 2663-3582
IJECS 2020; 2(2): 01-06
Received: 01-04-2019
Accepted: 02-06-2019

**Albin Thomas**
Department of Computer
Science, SAS SNDP Yogum
College, Konni, Kerala, India

**Correspondence**
**Albin Thomas**
Department of Computer
Science, SAS SNDP Yogum
College, Konni, Kerala, India

# A comparison study of SVM and NB machine learning algorithms performance on training dataset

**Albin Thomas**

**DOI:** https://doi.org/10.33545/26633582.2020.v2.i2a.31

**Abstract**
Support Vector Machine and Naive Bayes are popular classification algorithms in PDF malware detection, Spam filtering and scientific community training datasets. These algorithms incorporated classifications into the training datasets which they affected with the type of causative and evasion attack. The adversaries are insect the training dataset by injecting malicious sample data. This infected training datasets are used in the ML algorithms without knowing that they are infected for research purpose. Intelligent attackers mislead the SVM and NB learning algorithms functional task by modifying the training dataset. This may cause the security problems in the training dataset. To develop security mechanism, use to cope the attack on training dataset and avoid to decreases ML algorithms performance. This paper shows that the SVM and NB accuracy reduces dramatically when they used infected training dataset. The proposed defence method Rand Check used to prevent the trusted training dataset from causative and evasion attacks.

**Keywords:** causative, evasion, Rand Check, arm race

## 1. Introduction

Training dataset is the labeled data which used to train in ML algorithms for perform different actions like classification, prediction and accuracy. The adversaries manipulate the training dataset and achieve their motives. Adversary performs various attacks on training dataset such as causative attack and evasion attack.

In Causative attack, the adversaries analyze how Support Vector Machines and Naive Bayes algorithms to change decisions due to malicious data injected into training dataset. The adversaries able to manipulate training datasets with the malicious data or change training dataset labels for mislead the classification process in machine learning system [1].

Evasion attack is an adversarial attack on training dataset. The adversaries generated evasion attack by adding random malicious samples to training dataset and train the attacked dataset in the SVM and NB learning algorithms for make the algorithms decision to fool [2].

The training dataset collection reduced cost from our society will be increased machine learning dataset classification techniques access with large amount of data [3]. The users or researchers downloaded training dataset from any trusted data sources. They don’t know the training dataset attacked by some intelligent adversaries. The attacked training dataset train on SVM and NB learning algorithms, it will be changing the classification performance of the dataset. So, we must provide the security to the training dataset.

In this paper explains both SVM and NB algorithms are performed independent approach on mounting causative and evasion attacks in Electricity training datasets. Also demonstrate the achievements of evasion, poisoning attacks in Training datasets and secure training dataset using the proposed algorithm Rand Check against the attack by reduce SVMs and NBs machine learning models predictions and accuracy.

## 2. Background and related works

SVM is a supervised, unique way from all other learning algorithms for classification and regression. An attacker increasing crafted training datasets attack points that decrease classification accuracy of SVM [4]. NB algorithm is a decision making with all the attributes [5]. This NB algorithm is acceptable in the learning algorithm scenarios classification model in which all attributes conditionally are independent not necessarily. The reactive arms race method between ML classification designer and the adversary will be play attempts to reach his/her achievements.

The adversary purposely manipulated input training dataset to make false negative for classification production [15]. The false negative added to the training dataset increases the arm race of an adversary to make an attack.

### 2.1 SVM and NB in Causative attack

Causative attack leading new training of attack to compromised with malicious data to its training datasets [5]. SVM analyze training datasets and then it classifies the datasets into groups or classes [8]. SVM algorithms are divided training datasets into classes, the learners uses these classes for pointed pairs input samples [3]

$$T_{er} = \{(xi,yi) \mid x_i \in X \text{ and } y_i \in Y\}_i{}^n=1$$

Hyperlane of maximum marginal find for the separate class will be minimizing classification error of $T_{er.}$ The attacker aim is, to choose the pair value (xi, yi) from $T_{er}$ and modify that pair value for decrease classification accuracy of SVM.

Naive Bayes ML algorithms used in the situation of thousands of trainings datasets having few variables. This algorithm predicts the probability of various attributes on different classes of the training dataset. This algorithm result is a strong assumption but fast and successful method.

### 2.2 SVM and NB in Evasion attack

Evasion attack is known as exploratory attack, the attacker make arbitrary changes in training dataset features. An evasion attack directly change training samples in training datasets to lead misclassification and wrong decisions [6]. The evasion attack function in the training datasets X denoted as

$$x^* = \text{argmin}_{x'}\, c(x',x)$$

Where, x Є X that evades initial attack c(x′,x).

SVM train the evasion attacked training dataset and perform misclassification. Naive Bayes provides probability theorem as P(c|x) from probability class P(c), predicting attributes P(x) and probability of predictor P(x|c).

$$P(c|x) = \frac{P(x|c)P(x)}{P(x)}$$

The training input samples x Є X denoted as attributes and its class c labels with y Є{-1,+1}. The evasion attackers manipulate x as x′ which evade at the test time [7].

### 3. Datasets and Metrics

The training datasets collected from UCI repository. Tamil Nadu Electricity Board Hourly Readings collected the original training dataset and posted into the UCI repository. The dataset details of electricity reading values collected around from Thanajvur in Tamil Nadu. The dataset values retrieved based on Electricity Board produced the bill for hourly readings. The EB training dataset has five parameters forkva, forkw, type, sector, service with 45259 data.

### 3.1 Training Datasets Classification

Electricity utility power delivered to an organization charge as kW. The power factor supply between 0 and 1 depend the type of organization run.

The relationship forkW and for kVA attribute is

$$kW=\sqrt{3}\ (kVA)$$

To Measure power factor, divide working power (kW) by apparent power (kVA). In a linear system, the cosine Ø is referred to as result [16].

$$\text{Power factor}=kW/kVA=\text{cosine } Ø$$

The value of kVA decreases, the value of power factor increases. The measurement becomes low power factor means the organization not fully utilizing the electrical power, but they paying for it. To secure the attribute value of kVA in the EB training dataset, then we can get correct power factor value. The SVM and NB learning algorithms measurements are stated the EB training dataset infected or not. To train the EB training datasets in the machine learning algorithm, it classifies the dataset under the heading 'type' parameter. The classification algorithms before modify the training datasets is shown in figure1.

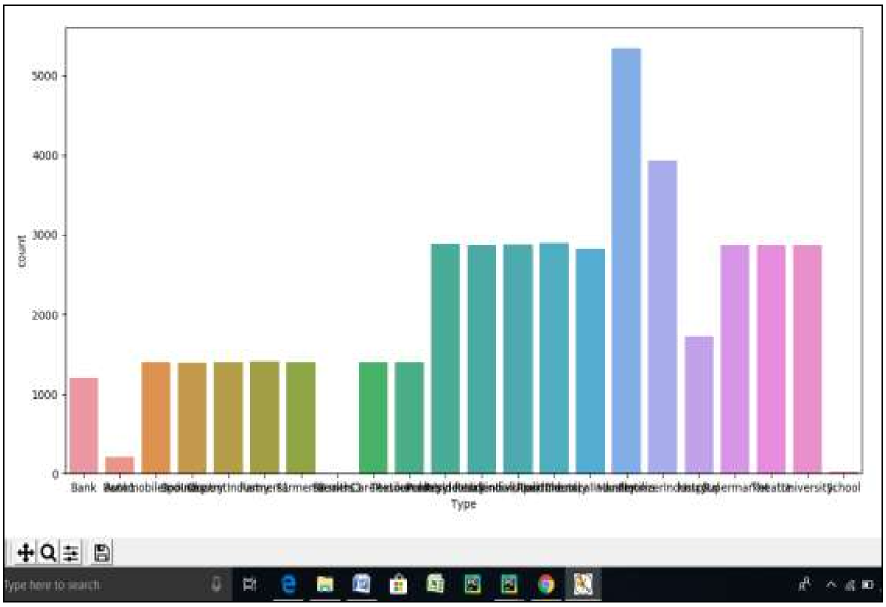
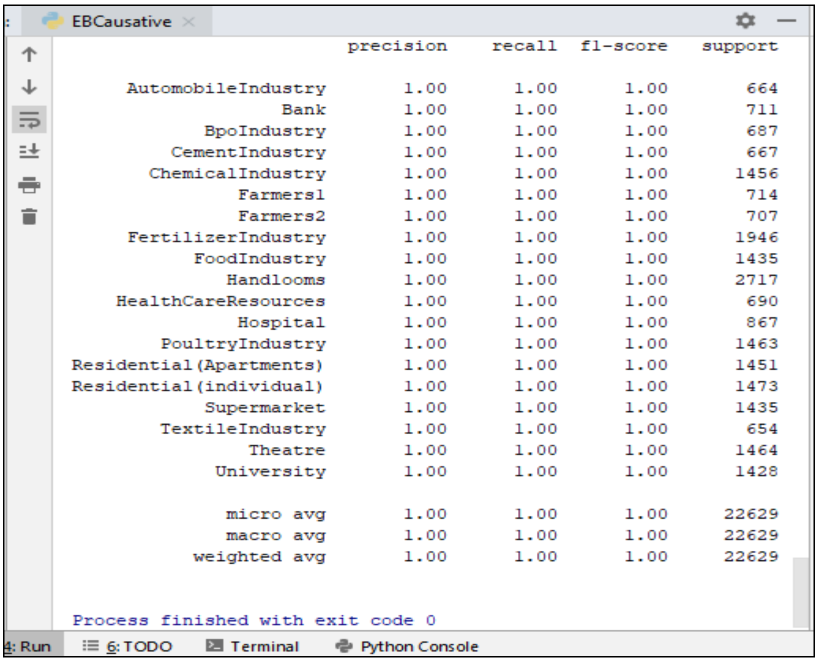


**Fig 1:** Classification of EB training datasets

### 3.2 Training datasets metrics

The metrics of precision value, recall, F1_score and support [10] are computed from the EB training datasets shown in Figure 2.



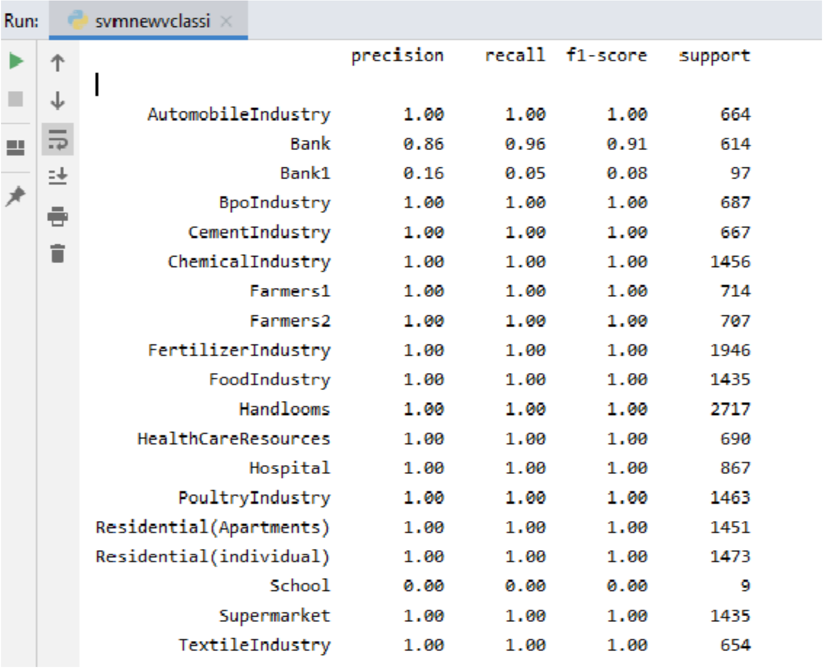
```
                       precision    recall  f1-score   support

    AutomobileIndustry       1.00      1.00      1.00       664
                  Bank       1.00      1.00      1.00       711
           BpoIndustry       1.00      1.00      1.00       687
        CementIndustry       1.00      1.00      1.00       667
      ChemicalIndustry       1.00      1.00      1.00      1456
              Farmers1       1.00      1.00      1.00       714
              Farmers2       1.00      1.00      1.00       707
    FertilizerIndustry       1.00      1.00      1.00      1946
          FoodIndustry       1.00      1.00      1.00      1435
             Handlooms       1.00      1.00      1.00      2717
   HealthCareResources       1.00      1.00      1.00       690
              Hospital       1.00      1.00      1.00       867
       PoultryIndustry       1.00      1.00      1.00      1463
Residential(Apartments)       1.00      1.00      1.00      1451
Residential(individual)       1.00      1.00      1.00      1473
           Supermarket       1.00      1.00      1.00      1435
       TextileIndustry       1.00      1.00      1.00       654
               Theatre       1.00      1.00      1.00      1464
            University       1.00      1.00      1.00      1428

             micro avg       1.00      1.00      1.00     22629
             macro avg       1.00      1.00      1.00     22629
          weighted avg       1.00      1.00      1.00     22629


Process finished with exit code 0
```

**Fig 2:** Metrics of EB Training Datasets

The precision computed as the ratio of true positive tp and false positive fp value is take as the form tp/(tp + fp). The cost of fp value is high the precision measure as good. The recall value computed as the ratio of true positive, false positive and false negative value is taking as the form tp/ (tp + fn). The cost of fn becomes to high, the recall select the best actual positive model. F1_score compute using the functions precision and recall.

$$F1 = 2 \text{ x } \frac{Precision * Recall}{Precision + Recall}$$

The adversary attack the original training datasets, the actual value of precision, recall and f1_score will be changed. The experiment of attacked training datasets result shown in figure 3.



```
                       precision    recall  f1-score   support

    AutomobileIndustry       1.00      1.00      1.00       664
                  Bank       0.86      0.96      0.91       614
                 Bank1       0.16      0.05      0.08        97
           BpoIndustry       1.00      1.00      1.00       687
        CementIndustry       1.00      1.00      1.00       667
      ChemicalIndustry       1.00      1.00      1.00      1456
              Farmers1       1.00      1.00      1.00       714
              Farmers2       1.00      1.00      1.00       707
    FertilizerIndustry       1.00      1.00      1.00      1946
          FoodIndustry       1.00      1.00      1.00      1435
             Handlooms       1.00      1.00      1.00      2717
   HealthCareResources       1.00      1.00      1.00       690
              Hospital       1.00      1.00      1.00       867
       PoultryIndustry       1.00      1.00      1.00      1463
Residential(Apartments)       1.00      1.00      1.00      1451
Residential(individual)       1.00      1.00      1.00      1473
                School       0.00      0.00      0.00         9
           Supermarket       1.00      1.00      1.00      1435
       TextileIndustry       1.00      1.00      1.00       654
```

**Fig 3:** Metrics after Attack of Training Datasets

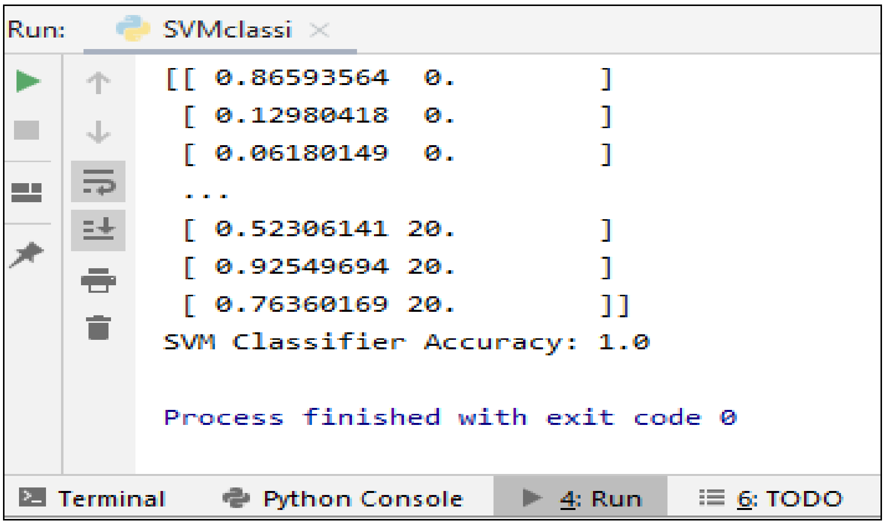
In the EB training dataset, the attacker makes an attack using the attribute type values. The attribute type has one of value organization names. The attacker change one of the organization name bank as bank1 in the real training dataset and then the attacked dataset train in ML algorithms. The result of precision value changed from 1 to 0.86, recall metric value change as 1 to 0.96, f1_score value change as 1 to 0.91. From the metric value difference we can identify the training datasets attacked by the adversary.

## 4. Experiments

The EB training dataset evaluation implemented in SVM and NB machine learning algorithms. The experiments stated original training dataset performance and attacked training dataset performance.

### 4.1 Causative attack against SVM

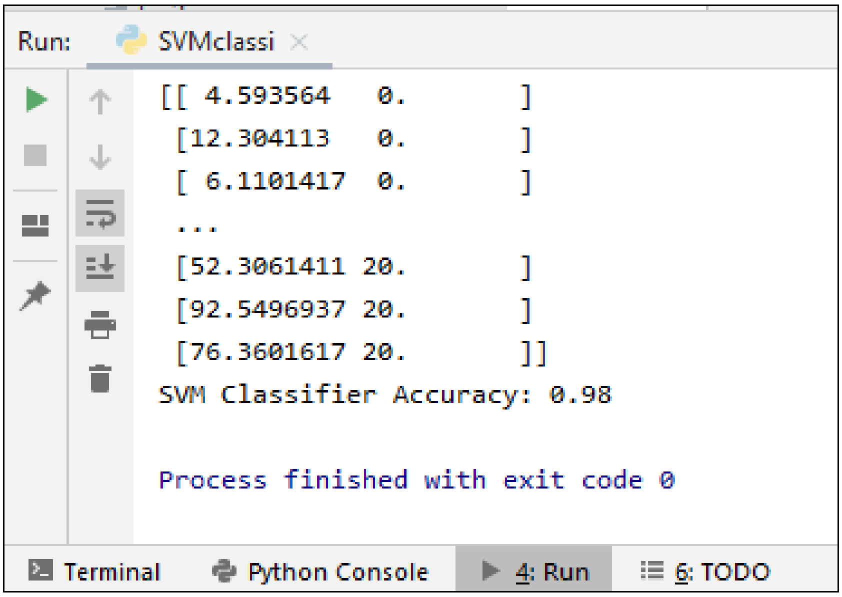
The SVM algorithm to misclassify causative attacked training datasets, by means of injecting malicious data samples into the training set. The EB training datasets trained in SVM learning algorithm and the accuracy of dataset computed. The accuracy of datasets before causative attack is shown in figure 4.



```
[[ 0.86593564  0.        ]
 [ 0.12980418  0.        ]
 [ 0.06180149  0.        ]
 ...
 [ 0.52306141 20.        ]
 [ 0.92549694 20.        ]
 [ 0.76360169 20.        ]]
SVM Classifier Accuracy: 1.0

Process finished with exit code 0
```

**Fig 4:** Accuracy of EB datasets using SVM

The malicious sample data injected to the original datasets, it will be changed as adversarial attacked training datasets. The attacked datasets applied in a linear SVM classification algorithm, the accuracy score reduced, it is stated in figure 5.



```
[[ 4.593564   0.        ]
 [12.304113   0.        ]
 [ 6.1101417  0.        ]
 ...
 [52.3061411 20.        ]
 [92.5496937 20.        ]
 [76.3601617 20.        ]]
SVM Classifier Accuracy: 0.98

Process finished with exit code 0
```

**Fig 5:** Accuracy of Attacked EB datasets using SVM

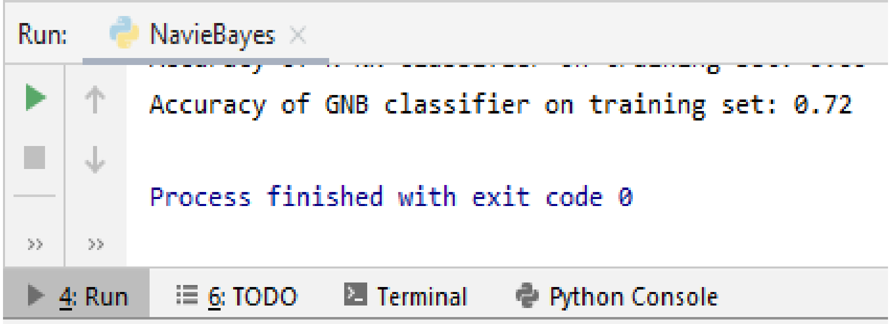
The accuracy score variation of before and after attack of training dataset analysis shows that the original datasets should be poisoned through adversaries. These results indeed to useful check whether the hand hold datasets are original training data or not.

### 4.2 Causative attack against Naive Bayes

Naive Bayes is a supervised classification machine learning algorithm using the concept of Basiyean theory [13]. The EB training dataset attribute fork VA follows Gaussian distribution and accept substitute probabilities to the normal distribution. The computation as

$$P(X|Y=c) = \frac{1}{\sqrt{2\pi\sigma^2 c}} *e^{\frac{-(x-\mu_c)^2}{2\sigma c^2}}$$

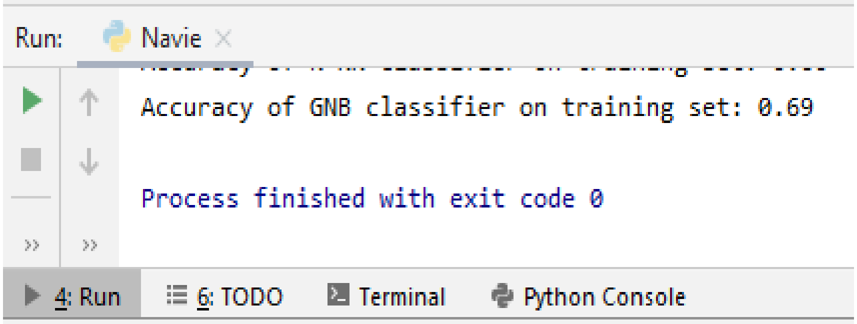
Where μ and σ are mean and variance computed for X for the given training dataset of Y. Accuracy measured from predictions of correct training datasets [14]. The figure 6 shows the accuracy of benign datasets applied in Gaussian Naive Bayes algorithm.



```
Accuracy of GNB classifier on training set: 0.72

Process finished with exit code 0
```

**Fig 6:** Accuracy of EB datasets using NB

The causative attack samples are changed the EB training set according to some adversarial knowledge [14]. The attacked data which did not helpful to improve the result accuracy value of the NB learning algorithm. However, the NB model evaluates the attacked training dataset and produces the reduced accuracy score described in figure7.



```
Accuracy of GNB classifier on training set: 0.69

Process finished with exit code 0
```

**Fig 7:** Accuracy of Attacked datasets using NB

## 5. Security using Rand Check Algorithm

The security algorithm Rand Check undertakes the adversarial attacked training dataset for evaluation. The signature based Training dataset privacy provide lot of security but sufficient level. However, the strong adversaries use the technical knowledge of attack on sensitive large training datasets [8]. The data holders to do review the training datasets and identifies lurking adversary at regular intervals. The review of training dataset done by both product managers and data scientists, they are taking this review as part of their job. This increasing dataset maintenance cost for data holders. The data holders prevent their data and should give confidential to the dataset users. The cost charging from adversary for injecting malicious data on training dataset should reduce the attack [12]. The simple queries to check through the data sample with the attacker's accessing dataset that is statistically not similar to the training datasets used to identify attack [11].

The research performs random testing on dataset and reviews the output. The researcher purchase small amount of original data from the particular region, then compare it to the training datasets which downloaded from trusted data source. The data from the purchased file randomly selected and it is checked with the training datasets. If the match is found anywhere inside the training dataset, the algorithm displays the message "Correct dataset" as output, otherwise it is assumed the training dataset contained malicious data injected. From the output message the researcher may noticed that the training datasets are affected by the adversaries. So, that they recollect training dataset from some other trusted source.

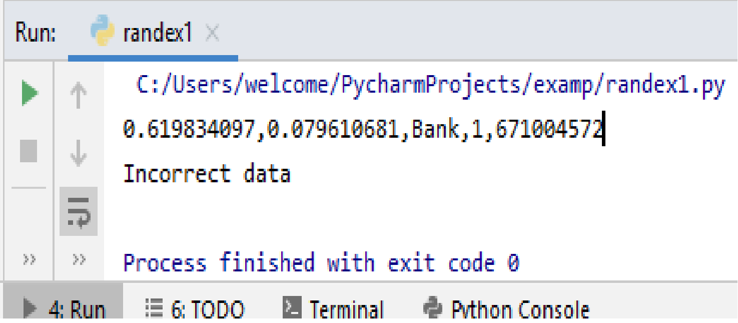
The data randomly selected from EB training dataset as $x_i$. The real time data directly collected from the region as $x_{1i}$. The algorithm compares both data and display the result will inform us the training dataset collected from trusted source is attack or not.

$t = x_i$

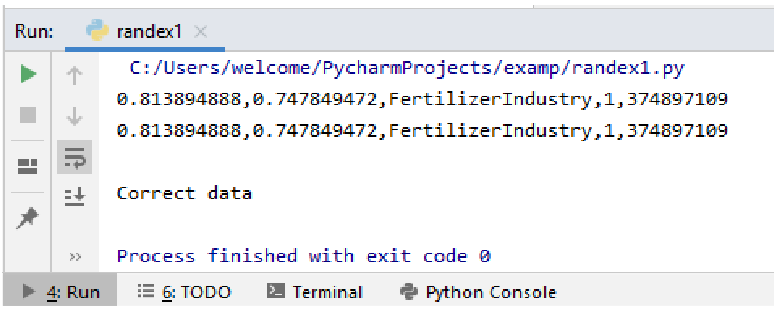
$t_1 = x_{1i}$

$t = t1 \rightarrow$ Training dataset not attacked

The algorithm implemented in python coding and executed with the EB training datasets. The Rand Check python program execution repeated number of times depends on the purchased amount of datasets. The outputs of Rand Check algorithm implementation is shown in figure 8.



```
C:/Users/welcome/PycharmProjects/examp/randex1.py
0.619834097,0.079610681,Bank,1,671004572
Incorrect data

Process finished with exit code 0
```



```
C:/Users/welcome/PycharmProjects/examp/randex1.py
0.813894888,0.747849472,FertilizerIndustry,1,374897109
0.813894888,0.747849472,FertilizerIndustry,1,374897109

Correct data

Process finished with exit code 0
```

**Fig 8:** Outputs of Rand Check

The attackers to modify the datasets with addition of new data or to change the datasets with new training sample. The python program checked both purchased dataset and downloaded dataset. The output of the program inform the researcher that his/her training dataset collection is real data or not through the message. The Training dataset has correct data then the researchers to keep the original training dataset themselves.

## 6. Conclusion and Future Work

The algorithms SVM and Naive Bayes accuracy rate proved that training dataset attacked or not. From the machine learning algorithm evaluation, we know that SVM is best suitable for EB training dataset. The SVM algorithm also identified the infected training dataset through decreasing accuracy. The adversary more knowledge about training dataset and machine learning algorithms, the detection of infected dataset becomes difficult. Because the adversary manipulating training dataset without affecting accuracy. In that situation the Rand Check algorithm helps to detect the infected training dataset or to save the original training dataset. The EB training dataset and purchased dataset are run on Rand Check algorithm. The algorithm produce the result EB training dataset is trustable or not. The small amount datasets used for checking with large datasets will be reduced purchasing training dataset amount cost.

In future, to extend the research to train single training dataset learning model to a troupe and each ML learner handles a different kind of attacked training dataset. The deep learning and clustering methods collaborated with the existing technique in order to produce better results.

## 7. References

1. Xiaojun Lin, Patrick PK Chan. “Causative Attack to Incremental Support Vector Machine”, International Conference on Machine Learning and Cybernetics, Lanzhou, 13-16 July, 2014 IEEE.
2. Zhiminhe, Junjiansu, Manzanhu, Gangrenwen, Shilinxu, Feizhang. “Robust Support Vector Machines against Evasion Attacks by Random Generated Malicious Samples”, International Conference on Wavelet Analysis and pattern recognition, china, 9-12 July 2017 IEEE.
3. Cody Burkard, Brent Lagesse. “Analysis of Causative Attacks against SVMs Learning from Data Streams”, IWSPA ’17, USA, March 24, 2017.
4. Battista Biggio, Blaine Nelson, Pavel Laskov. “Poisoning Attacks against Support Vector Machines”, Proceedings of the 29th International Conference on Machine Learning, Edinburgh, Scotland, UK, 2012.
5. Mehran Mozaffari-Kermani, Susmita Sur-Kolay, Anand Raghunathan, Niraj K Jha. “Systematic Poisoning Attacks on and Defenses for Machine Learning in Healthcare”, 2013, 2168-2194 (c) IEEE, DOI 10.1109/JBHI.2014.2344095.
6. Fei Zhang, Patrick PK Chan, Battista Biggio, Daniel S Yeung, Fabio Roli. “Adversarial Feature Selection against Evasion Attacks”, IEEE Transactions On Cybernetics, 2015, 2168-2267, IEEE.
7. Paolo Russu, AmbraDemonites, Battisa Biggio, Giorgio Fumera, Fabio Roli, “Secure Kernel Machines against Evasion Attacks”, AISec’16, October 28 2016, Vienna, Austria, DOI: http://dx.doi.org/10.1145/2996758.2996771.
8. Emmanual Gbenga Dada, Joseph Stephen Bassi, Haruna Chiroma, Shafii Muhammad Abdulamid, Adebayo Olusola Adetunmbi, Opeyemi EmmanuelAjibuwa, “Machine learning for email spam filtering: review, approaches and openresearch problems”, Published by Elsevier Ltd. This is an open access article under the CC BY-NC-ND license, Accepted 20 May 2019, https://doi.org/10.1016/j.heliyon.2019.e01802.
9. Battista Biggio, Igino Corona, Blaine Nelson, Benjamin I. P. Rubinstein, Davide Maiorca, Giorgio Fumera, Giorgio Giacinto, and Fabio Roli, “Security Evaluation of Support Vector Machines in Adversarial Environment”, arXiv: 1401.7727v1 [cs.LG] 30 Jan 2014.
10. sklearn.metrics.precision_recall_fscore_support
11. David J Miller, Zhen Xiang, George Kesidis. “Adversarial Learning in Statistical Classification: A Comprehensive Review of Defenses Against Attacks”, Pennsylvania State University, University Park, PA, 16803, arXiv: 1904.06292v2 [cs.LG] 13 May 2019.
12. Yan Zhou, Mura, Kantarcioglu, Bhavani Thuraisingham, Bowei Xi. “Adversarial Support Vector Machine Learning”, KDD’12, August 12–16, 2012, Beijing, China.
13. Pouria Kaviani, Mrs.Sunitha Dhatre. “Short Survey on Naive Bayes Algorithm” International Journal of Advance Engineering and Research Development, Volume 4, Issue 11, November-2017.
14. Junyan Peng, Patrick PK Chan. “Revised Naive Bayes Classifier For Combating The Focus Attack In Spam Filtering”, Proceedings of the 2013 International Conference on Machine Learning and Cybernetics, Tianjin, 14-17 July, 2013.
15. P Gnana Pavani, K Venkatesh, V Rajesh. “Security Evaluation of Pattern Classifiers under Attack”, IJDCST @. 2017; 5(5):SW-39.
16. EAN “Power factor correction: a guide for the plant engineer”, Technical Data SA02607001E Effective August, 2014.