A NEW APPROACH IN WEBSITE FRAUD DEDUCTION BY CLASSIFYING FEATURE SELECTION METHOD

Mrs. S. DEEPIKA[1], M. PRATHIBA[2]

Mrs. S. DEEPIKA
ASSISTANT PROFESSOR, DEPARTMENT OF B.COM BUSINESS ANALYTICS
PSGR KRISHNAMMAL COLLEGE FOR WOMEN
deepika@psgrkcw.ac.in
M.PRATHIBA
UG SCHOLAR, B.COM (BUSINESS ANALYTICS),
PSGR KRISHNAMMAL COLLEGE FOR WOMEN
prathi172018@gmail.com

ABSTRACT:

This paper is used to detect the website fraud by using popup window. Because Website Site Fraud Detection is a crucial point that applies to numerous ventures including the monetary areas, banking, government organizations, protection, and law requirement, and other industries also. In spite of battles with respect to the grieved associations, countless amounts are squandered to misrepresentation every year. Pop-up windows are mostly generated by the websites. Web traffic is measured in visits, sometimes called "sessions," and is a common way to measure an online business effectiveness at attracting an audience. Website traffic refers to web users who visit a website. By using pop-up window and website traffic we are going to detect the fraud which are occurring in website.

Key words: NAIVE BAYES classifier, Machine learning.

I. INTRODUCTION

Website Site Fraud Detection, applies to numerous ventures including the monetary areas, banking, government organizations, protection, and law requirement, and other industries in crucial point. In spite of battles with respect to the grieved associations, every year countless amounts are squandered to misrepresentation. Almost a couple of tests affirm misrepresentation in a tremendous network, finding these can be more complex in nature. After a visitor arrives on your website pop-ups can show 60 seconds. It can be set to display on a specific landing page or across your entire site. Website traffic is a web users who visit a website. Web traffic is measured in visits called "sessions," and it is an online business attracting the audience effectively.

This paper shows the count of original websites and fake websites. Naïve Bayes classification algorithm which has been used to do simplest predictions. To make predictions machine learning creates a mathematical model using a sample data which can also be called as training data. Test data is also created to find the accuracy. With the help of training and test data Naive Bayes classifier is performed. Naive Bayes classifier helps us to find the accuracy of attributes.

II. RELATED WORKS

Phishing is a new word produced from 'fishing', it refers to the act that attacker allure users to visit a faked Website by sending them faked e-mails (or instant messages), and stealthily get victim's personal information such as user name, password, and national security ID, etc. Phishing is a form of identity theft that occurs when malicious Website impersonates a legitimate one in order to acquire sensitive information such as passwords, account details, or credit card numbers [1].
Though there are several anti -phishing software and techniques for detecting potential phishing attempts in emails and detecting phishing contents on websites, phishers come up with new and hybrid techniques to circumvent the available software and techniques. Pallavi D. Dudhe et al, International Journal of Computer Science and Mobile Computing, Vol.4 Issue.2, February- 2015 The internet is not only important for individual users but also for organizations doing business online [2]
Many of the organizations offer online trading and online sales of services and goods. Nevertheless, internet-users may be vulnerable to different types of online threats that may cause financial damages, identity theft, and loss of private information [3],[4].
Therefore, the internet suitability as channel for commercial exchanges comes into question. Phishing is considered a form of online threat that is defined as the art of impersonating website of an honest firm aiming to acquire users private information such as

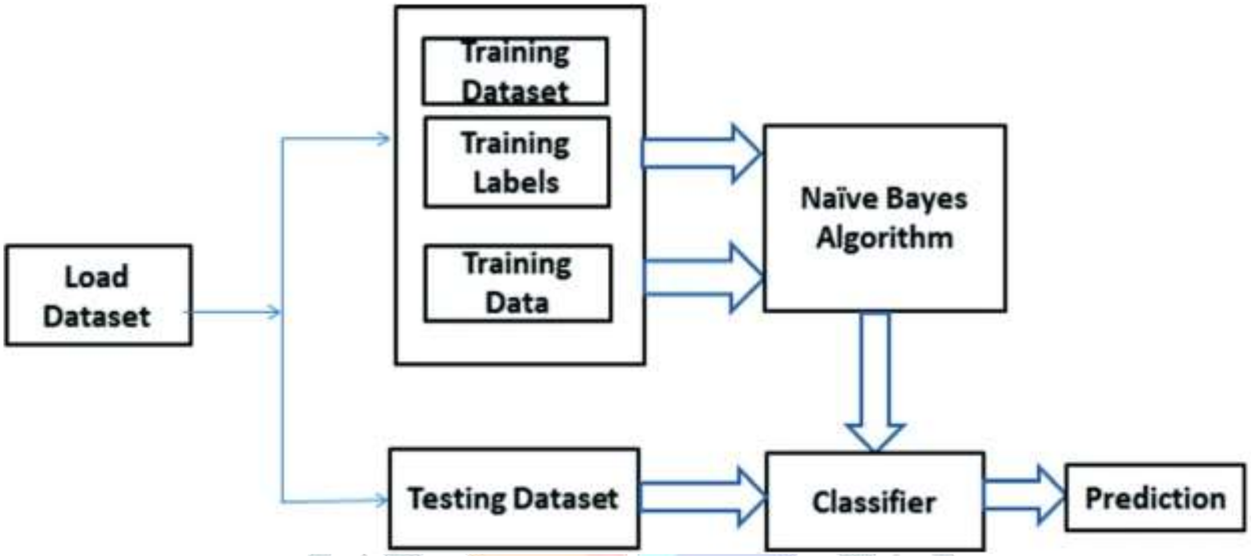
usernames, passwords and social security numbers. Phishing websites are created by dishonest individuals to imitate genuine websites [5].

III. METHODOLOGY

NAIVE BAYES CLASSIFIER ALGORITHM

Naive Bayes classifiers refers to "probabilistic classifiers" with strong independence assumptions They can achieve higher accuracy levels. The Bayesian network models is the simplest one, but it is coupled with kernel density estimation.

WORK FLOW MODEL



IV. RESULT

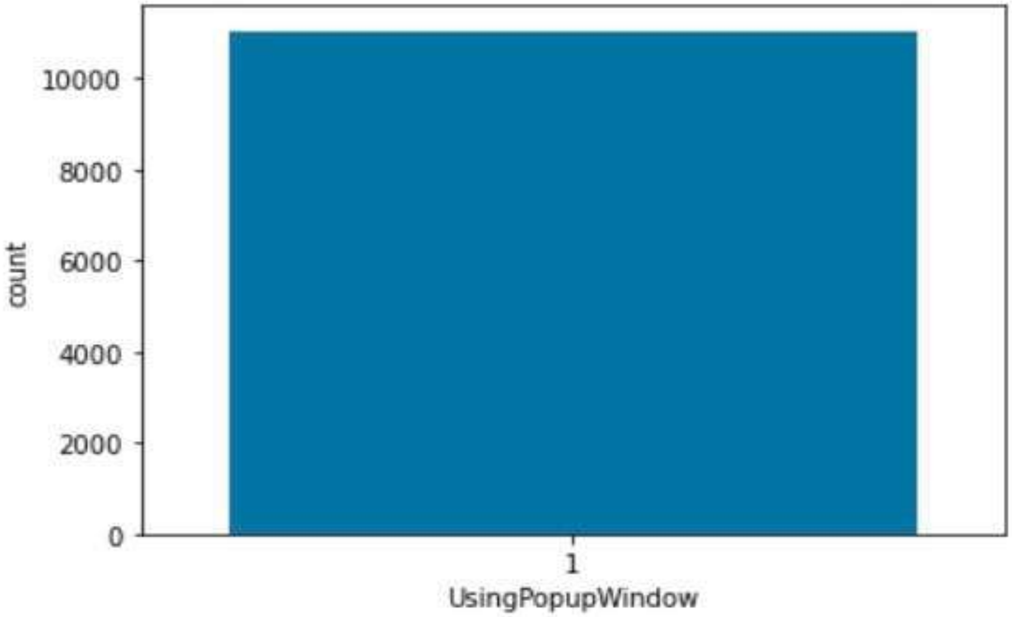


FIG:4.1

In the fig 4.1, it shows the count of original websites and fake websites in using popup window. It assures that the original website count is higher than the fake website count. In the attribute, the count of original website is more than 10000.

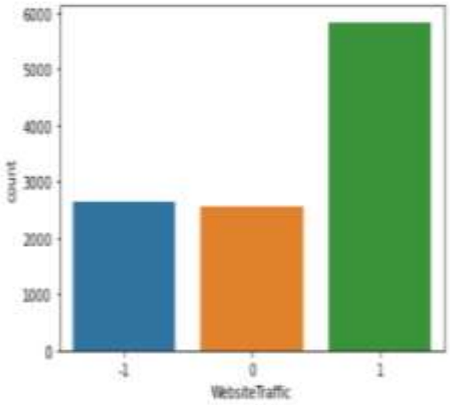


FIG:4.2

In the fig 4.2 it shows the count of original websites and fake websites in website traffic. It assures that the original website count is higher than the fake website count. In the attribute, the count of original website is more than 10000 and the count of fake website is less than 3000.

ACCURACY VALUE:

Accuracy=(y_test,y_pred)*100=95.085921

The accuracy score value is 95.086 that accuracy implies that there is 95% difference between the original and fake websites. In other words, 95% of original websites have found in accordance to the selective attributes.

V. CONCLUSION

In this paper JUYPTER is the tool which was used to analyze the Website fraud using NAÏVE BAYES classifier and the main objective is to find out the count and accuracy of original website by using various attributes. To Analyze the website fraud detection using popup window the small web browser and by website traffic where the data will send and received by visitors to a website and we have detect the fraudulent by popup window and website traffic using NAÏVE BAYES algorithm. Using the Naive Bayes Classifier Algorithm data is been predicted and model evaluation is done by fitting the model prediction into evaluation and the result of accuracy score in percentage.

FURTHER WORK:

It is suggested that comparing to categorical input variables and numeric variables. Naive Bayes is suitable for solving multi-class prediction problems. Naive Bayes is suitable for categorical input variables than numerical variables.

REFERENCES:

[1] Hicham Tout, William Hafner "Phishpin: An identity-based anti-phishing approach" in proceedings of international conference on computational science and engineering, Vancouver, BC, pages 347-352, 2009.
[2] J. Zdziarski, W Yang, and P. Judge, spam conference, Phishing activity trend report 1st half 2011
[3] Mather Aburrous, M.A. Hossain, Keshav Dahal, Fadi Thabtah "Prediction phishing websites using classification mining techniques with experimental case studies" in proceedings of Seventh International Conference on Information Technology, Las Vegas, NV, pages 176-181, 2010.
[4] Aaron, G., & Manning, R.APWG phishing reports 2012.
[5] Abdelhamid, N., Ayesh, A., & Thabtah, F. Associative classification mining for website phishing classification. In Proceedings of the ICAI(pp. 687–695), USA, 2013