

FAKE JOB RECRUITMENT DETECTION

R.Saiprasanna, M.Meghana, S.Vyshnavi, Dr.G.Thirupathi, Asso.Professor

SVS GROUP OF INSTITUTIONS, Hasanparthy, Bheemaram, Hanamkonda, Telangana 506015

ABSTRACT

The research proposes an automated solution based on machine learning-based classification approaches to prevent fraudulent job postings on the internet. Many organizations these days like to list their job openings online so that job seekers may access them quickly and simply. However, this could be a form of scam perpetrated by con artists who offer job seekers work in exchange for money. Many people are duped by this fraud and lose a lot of money as a result. We can determine which job postings are fraudulent and which are not by conducting an exploratory data analysis on the data and using the insights gained. In order to detect bogus posts, a machine learning approach is used, which employs numerous categorization algorithms. The system would train the model to classify jobs as authentic or false based on previous data of bogus and legitimate job postings. To start, supervised learning algorithms as classification techniques can be considered to handle the challenge of recognizing scammers on job postings. It will employ two or more machine learning algorithms, selecting the one that yields the highest accuracy score in the prediction of whether a job advertising headline is genuine.

1.INTRODUCTION

Employment scam is one of the serious issues in recent times addressed in the domain of Online Recruitment Frauds (ORF). In recent days, many companies prefer to post their vacancies online so that these can be accessed easily and timely by the job-seekers. However, this intention may be one type of scam by the fraud people because they offer employment to job-seekers in terms of taking money from them. Fraudulent job advertisements can be posted against a reputed company for violating their credibility. These fraudulent job post detection draws a good attention for obtaining an automated tool for identifying fake jobs and reporting them to people for avoiding application for such jobs. For this purpose, machine learning approach is applied which employs several classification algorithms for recognizing fake posts. In this case, a classification tool isolates fake job posts from a larger set of job advertisements and alerts the user. To address the problem of identifying scams on job posting, supervised learning algorithm as classification techniques are considered initially. A classifier maps input variable to target classes by considering training data. Classifiers addressed in the paper for identifying fake job posts from the others are described briefly.

These classifiers based prediction may be broadly categorized into -Single Classifier based Prediction and Ensemble Classifiers based Prediction.

2. LITERATURE SURVEY

2.1 Title: Fake Job Detection Using Machine Learning

Year : 2022

Author : Priya Khandagale , Akshata Utekar, Anushka Dhonde, Prof. S. S. Karve

Abstract : The research proposes an automated solution based on machine learning-based classification approaches to prevent fraudulent job postings on the internet. Many organizations these days like to list their job openings online so that job seekers may access them quickly and simply. However, this could be a form of scam perpetrated by con artists who offer job seekers work in exchange for money. We can determine which job postings are fraudulent and which are not by conducting an exploratory data analysis on the data. The system would train the model to classify jobs as authentic or false based on previous data of bogus and legitimate job postings. To start, supervised learning algorithms as classification techniques can be considered to handle the challenge of recognizing scammers on job postings. It will employ two or more machine learning algorithms, selecting the one that yields the highest accuracy score in the prediction of whether a job advertising headline is genuine or not.

2.2 Title: Fake Job Detection and Analysis Using Machine Learning

Year : 2021

Author : Anitha C S , P. Nagarajan , G. Adithya Sairam

Abstract : With the pandemic situation, there is a strong rise in the number of online jobs posted on the internet in various job portals. Some of the jobs being posted online are actually fake jobs which lead to a theft of personal information and vital information. Thus, these fake jobs can be detected and classified from a pool of job posts of both fake and real jobs by using machine learning classification algorithms. In this paper, machine learning and deep learning algorithms are used so as to detect fake jobs and to differentiate them from real jobs. The data analysis part and data cleaning part are also proposed, so that the classification algorithm applied is highly precise and accurate. A great importance is emphasized on data cleaning and pre-processing step in this paper. The classification and detection of fake jobs can be done with high accuracy and high precision. Hence the machine learning and deep learning algorithms have to be applied on cleaned and pre-processed data in order to achieve a better accuracy. Classification models are compared with each other to find the classification algorithm with highest accuracy and precision.

2.3 Title:Online Recruitment Fraud Detection

Year:2020

Author:Syed Mahbub

Abstract:Misinformation on the web has become a problem of significant impact in an information-driven society. Persistent and large volumes of fake content are being injected, and hence the content (news, articles, jobs, facts) available online is often questionable. This research reviews a range of machine learning algorithms to tackle a specific case of online recruitment fraud (ORF). A model with content features of job posting is tested with five supervised machine learning (ML) algorithms. It then investigates various crowdsourcing techniques that could enhance prediction accuracy and add human insights to machine learning automation. Each crowdsourcing method (explored as human signals online) was tested across the same ML algorithms to test its effectiveness in predicting fake job postings. The testing was conducted by comparing the hybrid models of machine learning and crowdsourced inputs. This study revealed that the best ML algorithm was different in the automated model compared to the hybrid model. Results also indicated that the net promoter type crowdsourced question resulted in the best accuracy in classifying fraudulent and legitimate jobs. The decision tree and generalized linear model demonstrated the highest accuracy among all the tested models.

3. PROBLEM STATEMENT

Review spam detection:Spammerscan manipulate reviews for gaining the profit and hence it is required to develop techniques that detects these spam reviews using NLP and Machine learning approach.

Email spam detection:Unwanted bulk mails, belonging to the category of spam emails, often arrive to user mailbox. To eradicate thisproblem, Gmail, Yahoo mail and Outlook serviceprovidersincorporate spam filters using Neural Networks with number of filters approach.

Fake news detection: The fundamental study of fake news detectionrelies on threeperspectiv-es how fake news is written, how fake news spreads, how a user is related to fake news. Features related to news contentand social context are extracted and machine learning models are imposed to recognize fake news.

3.1 LIMITATION OF SYSTEM

There were no such systems for fake job prediction. Less Accurate to be dump into real-time application

4. PROPOSED SYSTEM

For detecting bogus posts, a machine learning approach is used, which employs numerous classification algorithms. In this scenario, a classification tool detects and warns the user when it detects bogus job postings among a bigger set of job adverts. To begin addressing the challenge of spotting job posting scams, supervised learning algorithms as classification techniques are being examined. A classifier uses training data to map input variables to target classes. The classifiers covered in the research for distinguishing phony job postings from others are briefly outlined. These classifier-based predictions can be roughly classified into two types: single classifier-based predictions and ensemble classifier-based predictions.

Naive Bayes Classifier

Decision tree

K-nearest Neighbour Classifier

Random Tree classifier

4.1 Advantages

For detecting bogus posts, a machine learning approach is used, which employs numerous categorization algorithms. In this scenario, a classification tool detects and warns the user when it detects bogus job postings among a bigger set of job adverts. In this scenario, a classification tool detects and warns the user when it detects bogus job postings among a bigger set of job adverts. To begin addressing the challenge of spotting job posting scams, supervised learning algorithms as classification techniques are being examined. A classifier uses training data to link input variables to target classes. Classifiers were addressed.

4.Accuracy is improvised.

5. IMPLEMENTATION

5.1 DATA COLLECTION

Data collection is a process in which information is gathered from many sources which is later used to develop the machine learning models. The data should be stored in a way that makes sense for problem. In this step the data set is converted into the understandable format which can be fed into machine learning models.

5.2 DATA PRE-PROCESSING

Formatting

Cleaning

Sampling

5.3 FEATURE EXTRACTION

An attribute reduction process. Unlike feature selection, which ranks the existing attributes according to their predictive significance, feature extraction actually transforms the attributes. The transformed attributes, or features, are linear combinations of the original attributes. Finally, our models are trained using Classifier algorithm.

5.4 EVALUATION MODEL

Model Evaluation is an integral part of the model development process. It helps to find the best model that represents our data and how well the chosen model will work in the future.

6. ALGORITHM USED

6.1 Decision Tree:

Decision tree is the most powerful and popular tool for classification and prediction. A Decision tree is a flowchart like tree structure, where each internal node denotes a test on an attribute, each branch represents an outcome of the test, and each leaf node (terminal node) holds a class label.

6.2 Support Vector Machine

A Support Vector Machine (SVM) is a discriminative classifier formally defined by a separating hyperplane. In other words, given labeled training data (supervised learning), the algorithm outputs an optimal hyperplane which categorizes new examples. In two dimensional space this hyperplane is a line dividing a plane in two parts where in each class lay in either side.

6.3 naive Bayes Classifier

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

Example:

It tells us how often A happens given that B happens, written $P(A|B)$, when we know how often B happens given that A happens, written $P(B|A)$, and how likely A and B are on their own.

$\text{P}(A|B)$ is “Probability of A given B”, the probability of A given that B happens
 $\text{P}(A)$ is Probability of A
 $\text{P}(B|A)$ is “Probability of B given A”, the probability of B given that A happens
 $\text{P}(B)$ is Probability of B

When $\text{P}(\text{Fire})$ means how often there is fire, and $\text{P}(\text{Smoke})$ means how often we see smoke, then:

$\text{P}(\text{Fire}|\text{Smoke})$ means how often there is fire when we see smoke.

$\text{P}(\text{Smoke}|\text{Fire})$ means how often we see smoke when there is fire.

So the formula kind of tells us “forwards” when we know “backwards” (or vice versa) If Example: dangerous fires are rare (1%) but smoke is fairly common (10%) due to factories, and 90% of dangerous fires make smoke then:

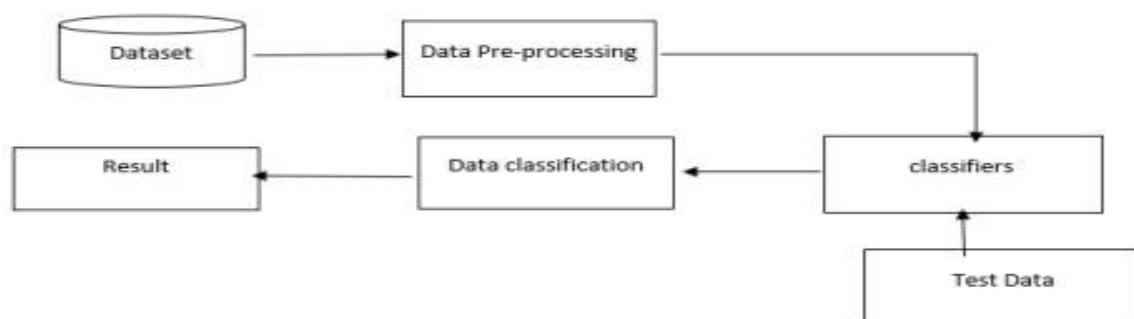
$\text{P}(\text{Fire}|\text{Smoke}) = \frac{\text{P}(\text{Fire}) \text{P}(\text{Smoke}|\text{Fire})}{\text{P}(\text{Smoke})} = \frac{1\% \times 90\%}{10\%} = 9\%$

 In this case 9% of the time expect smoke to mean a dangerous fire.

7. SYSTEM ARCHITECTURE

The purpose of the design phase is to arrange an answer of the matter such as by the necessity document. This part is that the opening moves in moving the matter domain to the answer domain. The design phase satisfies the requirements of the system. The design of a system is probably the foremost crucial issue warm heartedness the standard of the software package. It’s a serious impact on the later part, notably testing and maintenance.

System Design conjointly referred to as top-ranking style aims to spot the modules that ought to be within the system, the specifications of those modules, and the way they move with one another to supply the specified results.



System Architecture

8.EXPECTED OUTCOMES

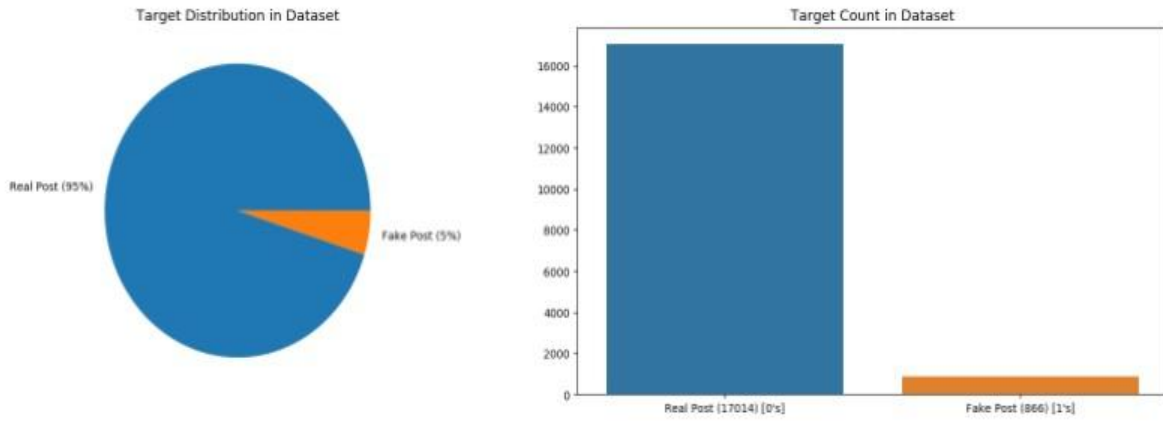
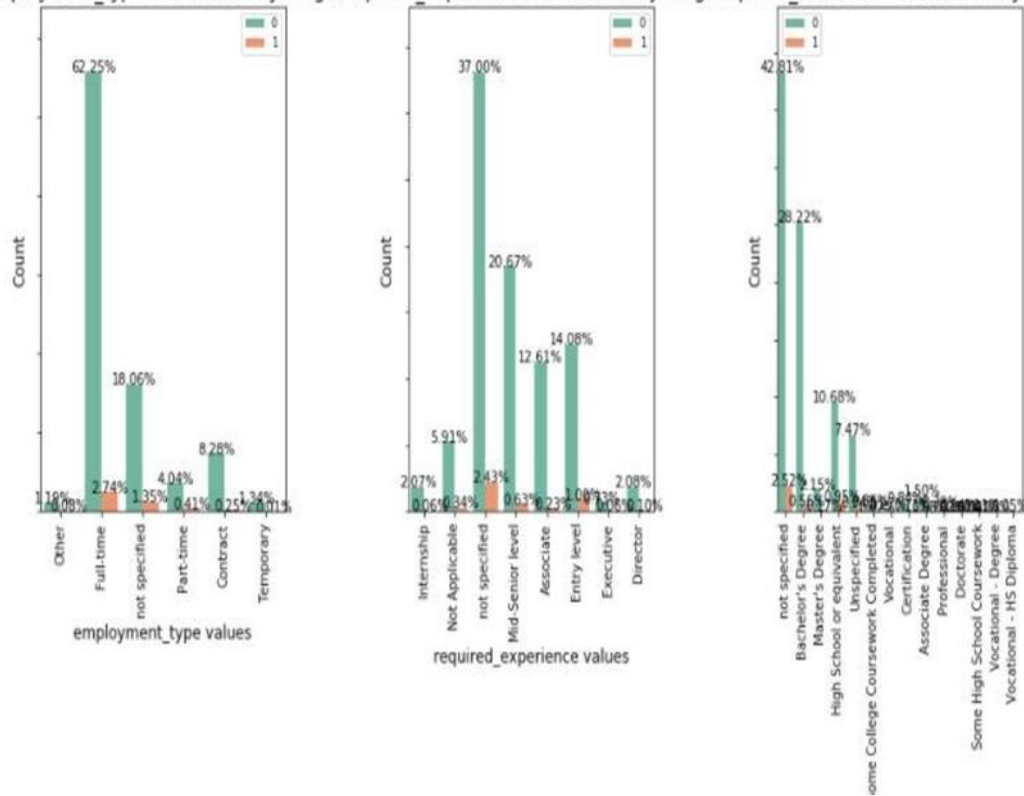


fig.8.1.1. Target Distribution in Dataset

employment_type Distribution by Targetrequired_experience Distribution by Targetrequired_education Distribution by Target



Employment Type Distribution

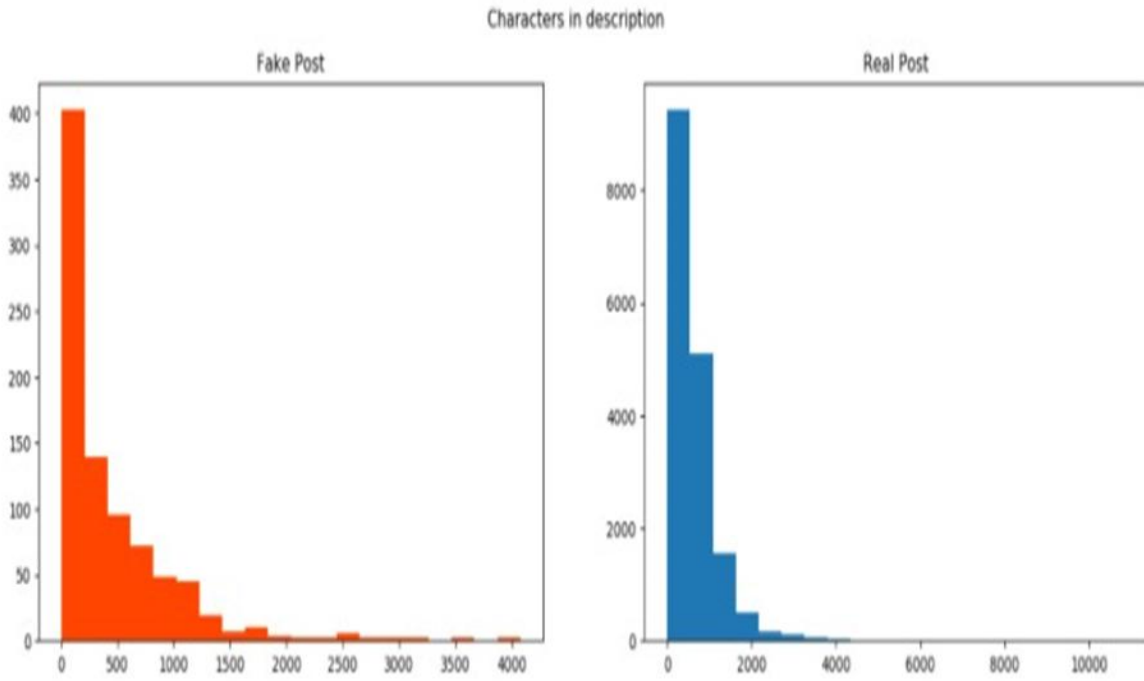


Fig.8.1.3.Bargraph based on characters in job Description

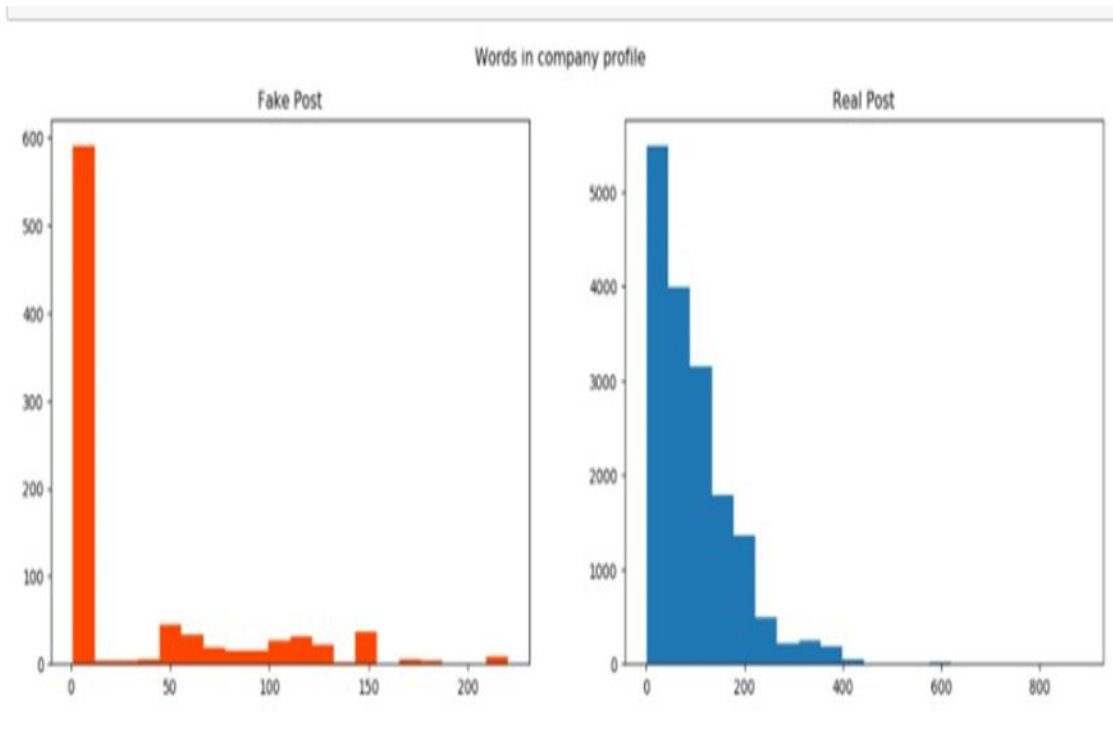


Fig.8.1.4.Bargraph based on Words in company profile

9. CONCLUSION

For Fake Job detection will guide job-seekers to get only legitimate offers from companies. tackling employment scam detection, several machine learning algorithms are proposed as countermeasures in this paper. Supervised mechanism is used to exemplify the use of several classifiers for Fraudulent Job detection. Experimental results indicate that Random Forest classifier outperforms over its peer classification tool. From the proposed approaches highest achieved accuracy is 98.76% which is much higher than the existing methods.

10. FUTHER ENCHANCEMENT

First and foremost you should lodge a complaint at your nearest police station. The police station has a special cyber cell group which is expert in dealing such crimes. You can either submit the complaint through email or letter. Start Social Media Campaign against Job consultancy. A recruiter can find a qualified candidate through a variety of websites. Fake recruiters will sometimes post a job on a job platform for the sole purpose of making money. Many job boards suffer from this issue. People later go to a new job portal in quest of legitimate employment, but phoney recruiters also migrate to this portal. As a result, it's critical to distinguish between legitimate and fictitious employment opportunities. Employment fraud is one of the most severe concerns that has been addressed in the arena of Online Recruitment Frauds in recent years (ORF). Many organizations these days like to list their job openings online so that job seekers may find them quickly and simply. This could, however, be one form of fraud perpetrated by the con artist. However, this could be a form of scam perpetrated by con artists who offer job seekers work in exchange for money.

11.1 REFERENCES

- [1] B. Alghamdi and F. Alharby, —An Intelligent Model for Online Recruitment Fraud Detection,” J. Inf. Secur., vol. 10, no. 03, pp. 155–176, 2019, doi: 10.4236/jis.2019.103009.
- [2] I. Rish, —An Empirical Study of the Naïve Bayes Classifier An empirical study of the naive Bayes classifier,|| no. January 2001, pp. 41–46, 2014.
- [3] D. E. Walters, —Bayes’s Theorem and the Analysis of Binomial Random Variables,|| Biometrical J., vol. 30, no. 7, pp. 817–825, 1988, doi: 10.1002/bimj.4710300710.
- [4] F. Murtagh, —Multilayer perceptrons for classification and regression,|| Neurocomputing, vol. 2, no. 5–6, pp. 183–197, 1991, doi: 10.1016/0925-2312(91)90023-5.

- [5] P. Cunningham and S. J. Delany, —K -Nearest Neighbour Classifiers,|| Mult. Classif. Syst., no. May, pp. 1–17, 2007, doi: 10.1016/S0031-3203(00)00099-6.
- [6] H. Sharma and S. Kumar, —A Survey on Decision Tree Algorithms of Classification in Data Mining,|| Int. J. Sci. Res., vol. 5, no. 4, pp. 2094–2097, 2016, doi: 10.21275/v5i4.nov162954.
- [7] E. G. Dada, J. S. Bassi, H. Chiroma, S. M. Abdulhamid, A. O. Adetunmbi, and O. E. Ajibuwa, “Machine learning for email spam filtering: review, approaches and open research problems,||Heliyon, vol. 5, no. 6, 2019, doi: 10.1016/j.heliyon.2019.e01802.
- [8] L. Breiman, —ST4_Method_Random_Forest,|| Mach. Learn., vol. 45, no. 1, pp. 5–32, 2001, doi: 10.1017/CBO9781107415324.004.
- [9] B. Biggio, I. Corona, G. Fumera, G. Giacinto, and F. Roli, —Bagging classifiers for fighting poisoning attacks in adversarial classification tasks,” Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics), vol. 6713 LNCS, pp. 350–359, 2011, doi: 10.1007/978-3-642-21557-5_37.
- [10] . Natekin and A. Knoll, —Gradient boosting machines, a tutorial,|| Front. Neurorobot., vol. 7, no. DEC, 2013, doi: 10.3389/fnbot.2013.00021

A

11.2 WEBSITES

<https://ijettjournal.org>

<https://ieeexplore.ieee.org>

<https://www.researchgate.net>

<https://www.kaggle.com>

<https://www.ijcaonline.org>