IARJSET

ISSN (O) 2393-8021, ISSN (P) 2394-1588

International Advanced Research Journal in Science, Engineering and Technology

Vol. 8, Issue 8, August 2021

DOI: 10.17148/IARJSET.2021.8857

Identifying Real and Fake Job Posting-Machine Learning Approach

Devi.A P[1], Sandhiya.S[2], Gayathri.R[3]

[1] UG – Information Technology, Meenakshi Sundararajan Engineering College, Chennai, Tamilnadu

[2] UG - Information Technology, Meenakshi Sundararajan Engineering College, Chennai, Tamilnadu

[3] Assistant Professor - Information Technology, Meenakshi Sundararajan Engineering College, Chennai, Tamilnadu

ABSTRACT

The process of searching jobs is one of the most problematic issue freshers face, this process is used by various scamsters to lure freshers into scams and profit from the students. In order to avoid this, this paper proposes a system with deep learning and flask for front-end, that can identify fake jobs. The deep learning algorithm extracts specific features from the website's article and based on those features predicts if the job is genuine or not. The proposed system makes use of a deep learning based system and a web page to help non-technical users to analyze these fake scams and secure their jobs .

While browsing for jobs online we saw that many scamsters demanded money for booking slots to interviews that did not exist and also extort money from students with promise of giving them jobs in return, this served as motivation for this proposal.The objectives that are to be considered are: Prediction of real or fake job. And a front-end page to allow non-technical user to use the model

The proposed system is basically an ANN classification model based on Multinomial Naive Bayes algorithm to determine fake job posting or real one. The model is trained to be as efficient as possible by making the dataset to be a part of double-blind study and also considering the various formats of posting jobs in professional websites and other sites too. This therefore makes searching of jobs much more efficient and also allows the users to be worry free when they search for jobs online.

Keywords: Jobs, Deep learning, ANN

1.INTRODUCTION

In modern day, the major challenge faced by any graduate is searching for his or her dream job, the pity however is that they generally fall for fake job postings and end up losing money and time, the proposed system makes use of a deep learning based system and a web page to help non-technical users to analyze these fake scams and secure their jobs .

The rise of fake job now-a-days highlighting not only the dangers of the effects of fake job but also the challenges presented when attempting to separate fake job from real job. However, advances in technology and the spreading of news through different types of social media have increased the spreading of fake jobs today. As such, the effects of fake jobs have increased exponentially in the recent past and some steps must be taken to prevent this from continuing in the future. Therefore, our goal is to use machine learning to classify, at least as well as humans, more difficult between real and fake jobs.

IARJSET
ISSN (O) 2393-8021, ISSN (P) 2394-1588
International Advanced Research Journal in Science, Engineering and Technology
Vol. 8, Issue 8, August 2021
DOI: 10.17148/IARJSET.2021.8857

1. Proposed System

The proposed system is basically an ANN classification model based on Multinomial Naive Bayes algorithm to determine fake job posting or real one. The model is trained to be as efficient as possible by making the dataset to be a part of double-blind study and also considering the various formats of posting jobs in professional websites and other sites too. This therefore makes searching of jobs much more efficient and also allows the users to be worry free when they search for jobs online.

The dataset used is highly efficient as there is a clear analysis of the dataset. There is a front-end that could be used by a user to predict job descriptions.The proposed system uses flask and python to create a web interface for non-technical users to use the application with ease. Also, we intend to develop a highly accurate solution to determine the fake/real job posting. The dataset collected is scrapped of from various reliable sources and based of different perspectives, this gives data its integrity and helps to solve the problem of over fitting.

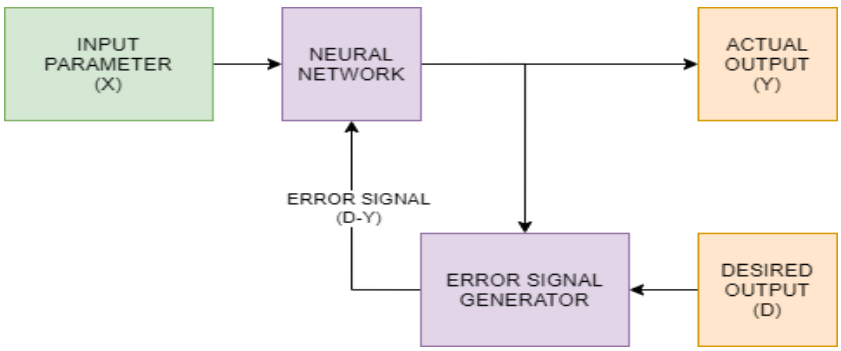
2.1 Dataset Analysis and Web scrapping

The proposed system has a UI front where the user enters the URL into the search box and presses enter, now the system in backend connected by Flask, scrapes the Web for data and the ANN model uses this algorithm to determine its credibility. Thus, the Web scrapping tool also serves as a dataset that can be allowed to train model in future works down the road. The major factor determining the performance of ANN model is the Dataset .

Data allows math to determine the exact outputs and requirements of a specific task, this project proposes a dataset that is huge, authentic and is based on a double-blind study on people who have applied for jobs. There are about 7796 entries across various countries, that are part of this study and about 17, most trusted websites are taken into account for referring formats. This allows the system to take in contexts also as in input while processing causing the whole system to very powerful and accurate..

2.2 Architecture of Naïve-Bayes Algorithm

Naive Bayes Algorithm is one of the popular classification machine learning algorithms that helps to classify the data based upon the conditional probability values computation. It implements the Bayes theorem for the computation and used class levels represented as feature values or vectors of predictors for classification. Naive Bayes Algorithm is a fast algorithm for classification problems. This algorithm is a good fit for real-time prediction, multi-class prediction, recommendation system, text classification, and sentiment analysis use cases. Naive Bayes Algorithm can be built using Gaussian, Multinomial and Bernoulli distribution. This algorithm is scalable and easy to implement for the large data set.



2.3 Organizing data between app and model

The data has to be accessed both to app as well as the model this makes it tough to wire a solution. However, an architecture called pickle helps in saving model weights thus providing a faster solution in training the dataset much better than that of training the dataset every time before we predict the model

IARJSET
ISSN (O) 2393-8021, ISSN (P) 2394-1588
International Advanced Research Journal in Science, Engineering and Technology
Vol. 8, Issue 8, August 2021
DOI: 10.17148/IARJSET.2021.8857

2.4 Modules

flask-cors:

Flask is a micro web framework written in Python. It is classified as a microframework because it does not require particular tools or libraries. It has no database abstraction layer, form validation, or any other components where pre-existing third-party libraries provide common functions.

Numpy:

NumPy is a library for the Python programming language, adding support for large, multi- dimensional arrays and matrices, along with a large collection of high-level mathematical functions to operate on these arrays.

Newspaper3k:

This is an on-demand module hosted by python foundation that allows web scrapping data from new websites. In our case jobsites along with format in which it was posted.

Sklearn:

Scikit-learn is a free software machine learning library for the Python programming language. It features various classification, regression and clustering algorithms including support vector machines

Pandas:

In computer programming, pandas is a software library written for the Python programming language for data manipulation and analysis. In particular, it offers data structures and operations for manipulating numerical tables and time series. It is free software released under the three- clause BSD license.

Nltk:

The Natural Language Toolkit, or more commonly NLTK, is a suite of libraries and programs for symbolic and statistical natural language processing for English written in the Python programming language. This is used in the proposed system to understand the context of jobs posted and its authenticity.

3.IMPLEMENTATION AND RESULT ANALYSIS

3.1Testing app features and connecting model predictions:

app.py contains python code to construct app with flask models and also communicate with the detection module.

3.2 fake_job_detection.py:

This module has the python code to run the classifier model based on Multinomial Naive Bayes algorithm, this module is the backbone of this project.

3.3 Classification

Naive Bayes Algorithm is one of the popular classification machine learning algorithms that helps to classify the data based upon the conditional probability values computation. It implements the Bayes theorem for the computation and used class levels represented as feature values or vectors of predictors for classification. Naive Bayes Algorithm is a fast algorithm for classification problems. This algorithm is a good fit for real-time prediction, multi-class prediction, recommendation system, text classification, and sentiment analysis use cases. Naive Bayes Algorithm can be built using Gaussian, Multinomial and Bernoulli distribution. This algorithm is scalable and easy to implement for the large data set.It helps to calculate the posterior probability P(c|x) using the prior probability of class P(c), the prior probability of predictor P(x) and the probability of predictor given class, also called as likelihood

IARJSET

ISSN (O) 2393-8021, ISSN (P) 2394-1588

International Advanced Research Journal in Science, Engineering and Technology

Vol. 8, Issue 8, August 2021

DOI: 10.17148/IARJSET.2021.8857

3.4 Overall Result

After training, using ANN, the result displays whether the corresponding jobs is authentic or fake.

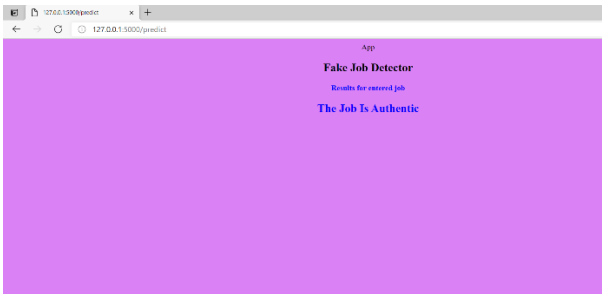


Fig. 5. Sample output generated

CONCLUSION AND FUTURE WORK

The main contribution of this project is to support for the idea that deep learning could be useful in a way for the task of classifying fake jobs. Our findings show that after much pre- processing of relatively small dataset, a simple ANN classification is able to pick up on a diverse set of potentially subtle language patterns that a human may (or may not) be able to detect. Many of these language patterns are intuitively useful in a humans manner of classifying fake jobs. Some such intuitive patterns that our model has found to indicate fake jobs include generalizations, colloquialisms and exaggerations.

The next steps involved in this project come in different aspects. Comparing the accuracies would be beneficial in deciding whether the dataset is representative of how difficult the task of separating fake from real Jobs is or not. If humans are more accurate than the model, it may mean that we need to choose more deceptive fake jobs examples

REFERENCES

In [1], Gulshan Shrivastava , Member, IEEE, Prabhat Kumar, Senior Member, IEEE, Rudra Pratap Ojha , Pramod Kumar Srivastava, Senthil Kumar Mohan "Defensive Modeling of Fake News Through Online Social Networks", — Online social networks (OSNs) *IEEE TRANSACTIONS ON COMPUTATIONAL SOCIAL SYSTEMS*

In[2], de Beer, Dylan & Matthee, Machdel," Approaches to Identify Fake News: A Systematic Literature Review." **:** *Integrated Science in Digital Age 2020.*

In[3], Bandar Alghamdi, Fahad Alharby," An Intelligent Model for Online Recruitment Fraud Detection". *Journal of Information Security. (2019)*

In[4], *Pham, Trung Tin" A Study on Deep Learning for Fake News Detection." Journal of Information Security. (2019)*

In[5], *Manoj Kumar Balwant." Bidirectional LSTM Based on POS tags and CNN Architecture for Fake News" 10th International Conference on Computing, Communication and Networking Technologies (ICCCNT). (2019)*

In[6], Amjad, Maaza , Sidorov, Grigoria, Zhila, Alisaa, Gómez-Adorno, Helenab, Voronkov, Iliac, Gelbukh, Alexander." Bend the truth" *Special section: Selected papers of LKE 2019*

In[7], Rami Mohawesh, Son Tran, Robert Ollington, Shuxiang Xu" Analysis of concept drift in fake reviews detection." *Expert Systems with Applications. (2021)*

In[8], Joma George; Shintu Mariam Skariah; T. Aleena Xavier." Role of Contextual Features in Fake News Detection: A Review" *International Conference on Innovative Trends in Information Technology (ICITIIT). (2021)*

In[9], Sultana Umme Habiba; Md. Khairul Islam; Farzana Tasnim." A Comparative Study on Fake Job Post Prediction Using Different Data mining Techniques." *2nd International Conference on Robotics, Electrical and Signal Processing Techniques (ICREST) (2021)*