Task 2

APPLICATIONS OF MACHINE LEARNING TO IMPROVING THE WEBSITE(<https://isthisarealjob.com/> )

By @enfinity

With the spurge in demand for digital skills in the recent years, job applications have gone beyond manual submissions and personal interviews and the whole on-boarding process of most companies are now online. This leads to a high risk in fake companies posting fake adverts to extort or harm applicants. A lot of sites today have also eased the job searching and on-boarding process as anyone can post up a job with little or no verification done. This leaves intending applicants at the mercy of these fraudulent perpetrators. The website <https://isthisarealjob.com/> thus provides a solution to this menace by detecting fake jobs online. My approach to improving this website is outlined below:

**OPTIMIZED DATA COLLECTION AND UPDATE :** As the popular saying goes in the machine learning community, models are as good as the amount of data to train on. Oprtimizing our data collection route and regular updating would not only help us build a large dataset but an up-to-date one. Techniques such as web scrapping popular and authentic job sites using web-scrapping libraries like spacy,beautiful soup e.t.c on jobs daily updated on newspapers and other sources would help build a large dateset. The scrapping is both done to build a dataset of real and fake jobs to prevent the problem of false negatives due to an in-balanced dataset. The website could also have a page where users get to post fake jobs for a small incentive in order to drive users to improve data collection steps.Optimizing the collection pipeline is also important so that each and every day your dataset gets freshly updated with recent jobs in the market.

**DATA PREPROCESSING STEPS:** Usage of the python Pandas library, matplotlib, seabornn and other data analysis libraries as well as application of NLP techniques such as Names Entity Recognition,Tokenization e.t.c can be used to preprocess our data and parse out the important features we would need for our analysis which would be based on key words such company name, location, job title e.t.c. We could also check with the Corporate Affairs Commission to check if the company is registered and add that to a column in our dataset which would help in our model development.

**MACHINE LEARNING TECHNIQUES:**

Machine learning algorithms have been successful for decades fighting spam email, by [analyzing messages’ text](https://blog.malwarebytes.com/security-world/2017/02/explained-bayesian-spam-filtering/) and determining how likely it is that a particular message is a real communication from an actual person – or a mass-distributed [solicitation for pharmaceuticals](https://www.incapsula.com/blog/viagra-spam-botnet.html) or [claim of a long-lost fortune](https://www.fbi.gov/scams-and-safety/common-fraud-schemes/nigerian-letter-or-419-fraud). Building on this type of text analysis in spam-fighting, AI systems can evaluate how well a jobs is true or not.

However, those methods assume the people who spread fake news don’t change their approaches. They often shift tactics, [manipulating the content of fake posts](https://doi.org/10.1145/3159652.3159677) in efforts to make them look more authentic. Thus using sophisticated algorithms such as Neural Networks,Extra Trees,K-nearest neighbor classifier algorithm,SVM’s e.t.c we can train on our collected data to identify the pattern of recognizing fake and real jobs, build a model to predict fake news.

**ANALYSIS OF CUSTOMER SERVICE REPORTS:** As much as technology is wonderful and all, it is still an end user that may not necessarily understand the whole intricacies of the tech jargons that would use it. Thus, the customer service reports should be properly analyzed as the user are the ones that give actionable feedback. The issues raised would then be analysed and changes made to the either step of the process.

The above points are thus my suggestions on the application of Machine Learning to improving the website (<https://isthisarealjob.com/> ).