Fake Job Recruitment Detection Using Machine Learning Approach

Author: Shawni Dutta and Prof.
Samir Kumar Bandyopadhyay
Year: 2020

Journal Name: International Journal of Engineering Trends and Technology (IJETT), Volume 68, Issue 4, April 2020

Ethan Rangel-Torres

Dutta, Shawni, and Samir Kumar Bandyopadhyay. "Fake Job Recruitment Detection Using Machine Learning Approach." *International Journal of Engineering Trends and Technology (IJETT)*, vol. 68, no. 4, Apr. 2020.

Problem Statement

- The main objective of the project is to create a POC detection model to distinguish if a job recruitment is real or fake.

- Potential applications of this project is it can be further implemented into a software for applicants seeking jobs to use to weed out fake recruitments.

Dataset

- The dataset used for this project can be found in kaggle:https://www.kaggle.com/datasets/shivamb/real-or-fake-fake-jobposting-prediction

- It consists of the following:
  - Job post: 17,880
  - Attributes:18
    - Key features (Job title, company profile, description, requirements, benefits, employment type, required education, and industry)

Motivation

- The main motivation is to create a POC that can be further implemented by bigger development teams to create a tool for applicants to use. I have personally dealt with fake recruitments on LinkedIn and know its a common problem students face. The job market is currently difficult for new graduates and weeding out fake recruitments will help applicants spend more time on real opportunities.

Detection of Fake Job Postings by Utilizing Machine Learning and NLP

Authors: Aashir Amaar, Wajdi Aljedaani, Furqan Rustam, Saleem Ullah, Vaibhav Rupapara, Stephanie Ludi

- Applies six supervised machine learning models: RF, SVM, KNN, Logistic Regression, Extra Tree Classifier, and Multilayer Perceptron.

- This paper is important because it uses the same dataset as the primary paper but applies different feature extraction methods and class balancing techniques. Comparing the results with the original work helps evaluate the impact of these variations.

Amaar, Aashir, et al. "Detection of Fake Job Postings by Utilizing Machine Learning and NLP." *Neural Processing Letters (Springer)*, vol. 54, 2022. https://doi.org/10.1007/s11063-021-10727-z.

Online Recruitment Fraud (ORF) Detection Using Deep Learning Approaches

Authors: Natasha Akram, Rabia Irfan, Ahmad Sami Al-Shamayleh, Adila Kousar, Abdul Qaddos, Muhammad Imran, Adnan Akhunzada

- Focuses on using transformer-based models (BERT and RoBERTa) to classify job postings, extending the dataset with Pakistani and US job postings..

- This paper is valuable because it contrasts traditional ML approaches by exploring deep learning models, allowing us to examine the impact of advanced methods in detecting fake job postings.

Akram, Natasha, et al. "Online Recruitment Fraud (ORF) Detection Using Deep Learning Approaches." *IEEE Access*, vol. 12, 2024. https://doi.org/10.1109/ACCESS.2024.3435670.

Summary of the Method

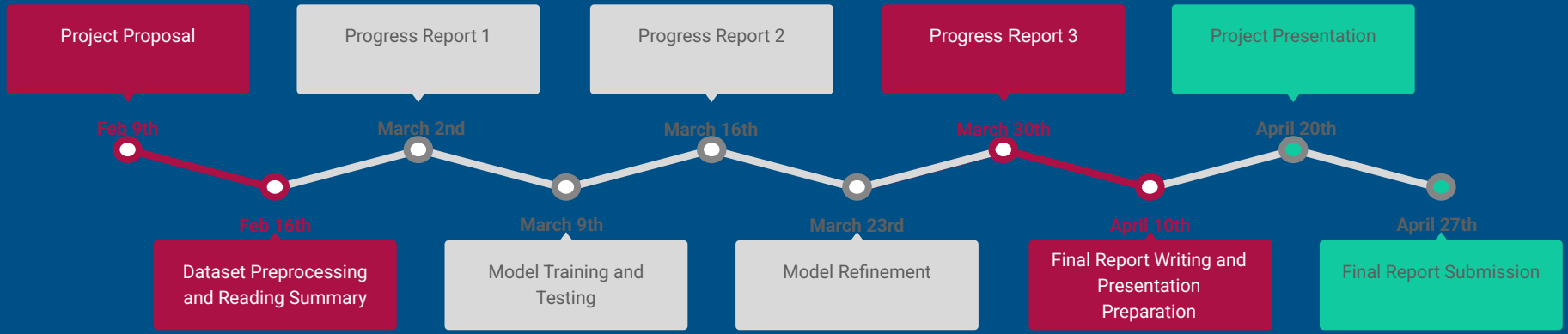
- **Paper**: Fake Job Recruitment Detection Using Machine Learning Approach by Shawni Dutta and Samir Kumar Bandyopadhyay.

- **Problem**: Identifying fraudulent job postings to protect job seekers from scams.

- **Method**: Multiple machine learning classifiers (Random Forest and Decision Tree) are trained on preprocessed job posting data, using structured features like job title and description.
  - Hyperparameters to modify and test:
    - n_estimaitors
    - class_weight
    - min_sample_leafs
    - leaf_nodes

- **Solution**: Build a classification model that distinguishes between legitimate and fake job postings using the dataset and evaluate performance with metrics like accuracy and F1-score. I will try to replicate and surpass the scores achieved by the paper using the same models and also a 3rd model K-Nearest Neighbors.

Timeline


Project Proposal
Progress Report 1
Progress Report 2
Progress Report 3
Project Presentation
Feb 9th
March 2nd
March 16th
March 30th
April 20th
Feb 16th
March 9th
March 23rd
April 10th
April 27th
Dataset Preprocessing and Reading Summary
Model Training and Testing
Model Refinement
Final Report Writing and Presentation Preparation
Final Report Submission

Sources

Primary Paper:

- Dutta, Shawni, and Samir Kumar Bandyopadhyay. "Fake Job Recruitment Detection Using Machine Learning Approach." *International Journal of Engineering Trends and Technology (IJETT)*, vol. 68, no. 4, Apr. 2020.
    a. https://ijettjournal.org/assets/Volume-68/Issue-4/IJETT-V68I4P209S.pdf

Additional Papers:

- Amaar, Aashir, et al. "Detection of Fake Job Postings by Utilizing Machine Learning and NLP." *Neural Processing Letters (Springer)*, vol. 54, 2022.
    a. https://doi.org/10.1007/s11063-021-10727-z
- Akram, Natasha, et al. "Online Recruitment Fraud (ORF) Detection Using Deep Learning Approaches." *IEEE Access*, vol. 12, 2024.
    a. https://doi.org/10.1109/ACCESS.2024.3435670