Group 9:

Kuldeep Chawla

Akram Mohammed

Amarjeet

BAN 612 Team Project Proposal: Detecting Fake Job Postings on LinkedIn

Background: Fraudulent job postings have been on the rise on recruitment platforms because of the increase in remote work. Scammers are getting good at creating fake jobs, collecting personal data from the candidates, and then using that information for financial fraud or selling that data to achieve their monetary gains. The goal of this project is to utilize the skills that we have learned (Data Analytics and Machine Learning) in BAN612 to detect fake jobs postings, enhancing user awareness and reducing victimization. We will use different modeling techniques to differentiate between fake and legitimate job postings and develop a model to detect the patterns going forward automatically.

Problem Statement: Can technologies like machine learning and text analytics effectively be applied to identify fake job postings on professional networking platforms or job search sites? What common characteristics are found in those fraudulent listings, and how can those be utilized to construct a model?

Data: We will utilize the Kaggle dataset "Fake Job Postings," which comprises approximately 18,000 job records. This dataset consists of fields like job title, location, department, company profile, job description, requirements, and labels, which can be very instrumental in detecting a fraudulent listing. This dataset provides both structured and unstructured data, making it ideal for running different modeling techniques. We can get some additional job listing data from the parent companies to make a comparison between the fake and the legitimate listings.

Methods: We will begin our analysis by performing data cleaning and exploratory data analysis to identify the common patterns, handle or eliminate the key values, and define key variables. We will also utilize the NLP to prepare textual fields. We will then use multiple regression, logistic regression, random forest, SVM for data modeling. Techniques like SHAP values to interpret and highlight the predictors to detect fake job postings will be utilized.