

Detecting Fraud Calls vis-à-vis Natural Language Processing

Anurag Dutta^{1*} and Avdhesh Gupta¹

^{1*}Undergraduate, Computer Science and Engineering,
Government College of Engineering and Textile Technology, 12,
William Carey Road, Serampore, 712201, Calcutta, India.

²Professor, Information Technology, Ajay Kumar Garg
Engineering College, **27th** Km Milestone, Delhi - Meerut
Expressway, Ghaziabad, 201009, Uttar Pradesh, India.

*Corresponding author(s). E-mail(s):

anuragdutta.research@gmail.com;

Contributing authors: avvipersonal@gmail.com;

Abstract

Fraud is defined in law as the willful use of deception to obtain unfair or illegal gain or to deny a victim of a legitimate right. Fraud can be illegal under criminal law or civil law. Scam may not always result in a loss of money, property, or legal rights but still be a component of another civil or criminal wrong. A victim of fraud may sue the offender to stop the fraud or receive monetary compensation. The goal of fraud may be financial gain or other benefits, such as getting a passport, travel document, or driver's licence, or it may be mortgage fraud, when the offender makes false claims in an effort to qualify for a mortgage. The collection of enormous volumes of data combined with predictive analytics or forensic analytics, the use of electronic data to reconstruct or detect fraud, makes the detection of fraudulent acts on a wide scale conceivable. Particularly when using computer-based analytical techniques, errors, inconsistencies, inefficiencies, irregularities, and biases can be exposed, which frequently pertain to fraudsters favouring particular currency amounts in order to bypass internal control thresholds. Scam calls are false calls that persons or businesses make in an effort to deceive recipients into parting with their money or private information. Scammers frequently play down the situation by calling it a normal call or even by lying to the person they were trying to con. In this work, we will

try to detect fraud calls, by making use of Supervised Machine Learning Algorithms. For the same, we have collected a total of 5925 Data points. We have used the data to train the model. For obvious reasons, we have kept the source of the data as anonymous. This work would help in early detection of scam calls and would help a lot of innocent lives from getting tied by the knot of fraudulences. Numerous ML Algorithms have been used for the work, namely Support Vector Machine, k - Nearest Neighbours, Gaussian Naive Bayes, etc. Further, these Algorithms have been compared on the basis of their performance of prediction.

Keywords: Supervised Learning, Support Vector Machine, Gaussian Naive Bayes, k - Nearest Neighbours, Fraud Detection, Criminology, Natural Language Processing

JEL Classification: C810 , C820 , O390

1 Introduction

Fraud is a purposeful act of deception that leads to unethical or unfair behaviour, according to the law. Fraud is frequently categorised as a criminal offence. Fraudulent behaviour is motivated for a variety of causes. Others fabricate the assertions for their own personal gain, which cannot be done without committing a crime, while some people want to benefit from value. According to the most popular definition, fraud is "a deception with the goal to benefit financially or personally." Different legal systems have different definitions of what constitutes a valid explanation for fraud, which may be a civil wrong requiring proof and a rationale, or a criminal infraction if the intentional deception has been justified. Financial scams are common and cost the people who commit them a lot of money. For instance, a business or bank may trick a customer into paying for extra services. The Wells Fargo bank in the United States experienced one of the same situations. A bank is the target of a bank fraud, which is typically carried out by making false claims and utilising fictitious documentation. Everyone is concerned about bank fraud. It is a highly delicate subject since it impacts public trust, which is the foundation of the entire banking system. The term "bank fraud" refers to a variety of thefts, embezzlements, and falsifications of negotiable documents such checks, bank draughts, bills of exchange, statements of accounts, stocks, etc. The booming banking industry has led to an increase in bank fraud. Indian Penal Code addresses bank fraud. By using a fake signature on the check, frauds are frequently conducted in the domain of checks. The majority of bank frauds presumably occur in this manner. Hypothecation fraud is another type of fraud in which money is fraudulently obtained in exchange for a security of some sort. A fair financial system that enjoys the public's confidence is a prerequisite for an equitable economic system. To prevent it, a number of preventive steps can

be done, such as proper recruiting, constant attention, adherence to regulations, training programmes, etc. Telephonic Frauds are another strata of Fraudulences that affect many people round the world. People are constantly communicating. Sadly, they don't always communicate in the most polite manner. While some attempts at communication result in violent or unpleasant interactions, other attempts at communication result in frauds. Scams occur on a variety of platforms, including the internet and the phone. An individual is frequently deceived into divulging personal information unknowingly when a fraud occurs. Although there are many other ways available, intimidation tactics are frequently used. In particular, the general public is compelled to accept criminals when they employ tech assistance as a front, lest they leave their computers vulnerable to malware attacks. In essence, there are many other types of scams, and this is just one of the most recent ones that customers should be aware of. Scams, whether they are conducted over the phone or in another way, need a vulnerability to operate. There are many different kinds of breaches, ranging from data to financial. The 2014 public awareness campaign on the Telephone Technical Support Scam involved an unintentional disclosure. Technical support frightened people into providing personal information when they contacted, leading them to believe they were chatting with experts who would clean their machines of malware or viruses. They also kept sending money to a person they thought was a government official after learning that the technical support was a hoax. Despite the fact that they shared information, they had no intention of sharing it with a fraudster. Studies make sure that the Telephonic Phone is compromising the security and privacy of many individuals. In this work, we would build a model that can predict a call as being Fraud or Normal. From the problem statement, it's clear that it's a classification problem. So, for modelling the same, we would make use of Supervised Learning Algorithms, like Gaussian Naive Bayes, Support Vector Machine, k - Nearest Neighbours, etc.