Enhancing Prescription Fraud and Error Detection in NHS Prescriptions Through Anomaly Detection

Abstract

This research explores the critical issue of fraud and errors in NHS prescriptions through data-driven methods. It investigates the landscape of prescription fraud within the healthcare system, delves into existing fraud control mechanisms, and scrutinizes fraud detection methods encompassing manual inspection, rule-based techniques, and advanced machine learning algorithms. The research adopts a structured approach, comprising data understanding, preparation, and modeling phases. Performance metrics and machine learning algorithms, including Neural Networks, Decision Trees, and Regression models and traditional outlier detection methods are employed to develop a robust fraud detection model. The findings of this pioneering study hold promise for revolutionizing fraud prevention and detection within the NHS, ultimately leading to improved patient care and cost savings. With a steadfast commitment to realizing its objectives, this study also extends an invitation to explore potential avenues for future exploration, emphasizing the importance data literacy and expert insights for refining fraud detection strategies in the healthcare sector.

Keywords: Prescription Fraud; Prescription Error; Machine Learning; NHS Prescription; CRISP-DM

1.0 Chapter One

1.1 Introduction, Background and Rationale

The British National Health Service (NHS), which came to life on 5th July 1948 has since been internationally admired and emulated (Sheard 2011). It is part of the investments made by countries into healthcare. This form of expenditure has been rising rapidly with reports indicating up to 10% of the gross domestic product (GDP) of many countries being spent on health (Joudaki et al. 2016).

According to Gee and Button (2015), healthcare organizations have traditionally been diligent in assessing and managing various costs such as staffing, accommodation, utilities, and procurement. Over the years, they have implemented measures to proactively address these costs and enhance operational efficiency. However, the same level of attention and focus has rarely been given to the issue of fraud and errors. In fact, many organizations have either denied the existence of fraud or planned to react only after fraud has occurred and as a result, fraud has become one of the significant ongoing expenses in the healthcare sector, largely unaddressed and unreduced (Gee and Button 2015). According to Johnson et al. (2021), up to 10% of global health care expenditure is lost due to fraud and abuse.

In a report as far back as 2011, the number of fraud cases in the NHS in the United Kingdom had risen by 37% in the three previous years with the value of fraud and unlawful action identified in 2009/10 being some £10,951,069 (Griffith and Tengnah 2011). In 2019 however, the costs of fraudulent activities had increased to £1.27 billion according to the NHS Counter Fraud Authority (NHS CFA) estimates and this was shown to be increasing by 17% annually (Griffith 2019).

The modus operandi of fraudsters in the NHS is already well known as there are tried and tested ways of siphoning off money, one of which is overprescribing by professionals due to fraud and error (Manning Julia 2011). Prescription is, by definition, “a request for the dispensing of one or more items or service to a patient”, according to NHS England (2023). The substantial cost associated with prescription items dispensed in England alone, amounting to £9.69 billion and over 1.14 billion items between 2021 and 2022 (NHS Business Services Authority, 2022) highlights the pressing need for a proactive approach to tackle fraud and errors.

According to Rashidian, Joudaki and Vian (2012), the interventions to combat this health care fraud can be categorized into three which are prevention of fraud, detection of fraud and response to fraud with artificial intelligence and data mining being able to help with detection of fraud. Furthermore, according to Johnson et al. (2021), a report by the National Academy of Medicine highlights the unprecedented opportunities for artificial intelligence (AI) to enhance healthcare by supporting specialists and addressing inherent human limitations such as fatigue and inattention which lead to errors being made. By leveraging AI technology, healthcare practitioners can overcome human limitations and reduce the risks associated with human error, ultimately improving the quality of patient care.

Given the increasing prevalence and financial impact of fraud and errors in healthcare, machine learning models can therefore be leveraged to analyse vast amounts of prescription data in order to identify patterns and anomalies that may indicate fraudulent or erroneous practices which can ultimately improve patient care.

1.2 Aim

This study aims to develop an advanced machine learning model employing anomaly detection techniques to effectively identify instances of fraud and errors within NHS prescriptions in England. The research adopts the comprehensive CRISP-DM methodology while focusing on the utilization and evaluation of specific methods and models. This study seeks to contribute to enhanced fraud detection and error prevention within the NHS prescription system.

1.3 Objectives

The specific objectives are:

1. Identify potential key patterns and indicators of fraudulent prescriptions within healthcare.
2. Explore and select suitable machine learning algorithms for anomaly detection in healthcare data.
3. Develop a predictive model using the selected algorithm(s) to detect fraudulent activities.
4. Evaluate the performance of the developed model and compare it with existing methods.
5. Provide recommendations for the implementation of the model in the NHS to enhance prescription fraud detection and prevention.

1.4 Research Outline

This research comprises five key chapters that systematically address the objectives set forth to enhance prescription fraud and error detection within the National Health Service (NHS) using machine learning techniques. Chapter One, "Introduction, Background, and Rationale," provides an overview of the study, including its motivation, objectives, and significance.

Chapter Two, "Literature Review," delves into the background, offering insights into the NHS, the prevalence of fraud and error in prescriptions, existing fraud control mechanisms, and various fraud detection methods, encompassing manual inspection, rule-based techniques, supervised and unsupervised machine learning methods, and hybrid approaches.

Chapter Three, "Methodology," presents the research's methodological framework, covering business understanding, data understanding, data preparation, modeling, and evaluation. It explores the process of data preprocessing, the selection of appropriate machine learning algorithms, and performance evaluation metrics.

Chapter Four, "Results, Discussion, and Findings," showcases the outcomes of the machine learning models, including their accuracy, precision, recall, and F1 score. The chapter facilitates a comprehensive discussion of the results and their implications for fraud and error detection in NHS prescriptions.

Chapter Five, "Conclusion, Limitations, and Future Work," concludes the research, summarizing the realization of objectives, acknowledging limitations, and outlining potential areas for future research in this pioneering exploration of machine learning in the realm of NHS prescription fraud detection.

2.0 Chapter Two – Literature Review

2.1 Introduction

This literature review explores fraud and errors in NHS prescriptions, aiming to establish clear definitions of these terms and provide an overview of the NHS structure. The review delves into the prevalence and consequences of fraud and errors, along with the NHS's current fraud control strategies. Additionally, it examines various fraud detection methods, including manual inspection, rule-based, supervised (neural networks, decision trees, support vector), unsupervised (clustering, outlier/anomaly), and hybrid approaches, offering examples from existing literature. The review sets the stage for building an efficient model to detect fraud and errors in NHS prescriptions, ultimately contributing to patient safety and healthcare fraud prevention.

There are many definitions of fraud in literature. According to the Association of Certified Fraud Examiners (2023), fraud is the intentional deception or concealment of material facts to induce another person to act in a way that results in their harm or loss. Levi and Burrows (2008 p. 299) define fraud as “the obtaining of financial advantage or causing of loss by implicit or explicit deception; it is the mechanism through which the fraudster gains an unlawful advantage or causes unlawful loss.” On a basic level, the NHS Counter Fraud Authority (2023a) defines fraud as “deception carried out for personal gain, usually for money” although, it could also involve the abuse of a position of trust. A common theme for the above can be seen as fraud being the intentional use of deception to gain a financial advantage or cause harm or loss to others, in this case, the NHS. In line with prescriptions, Aral et al. (2012 p. 37) defines prescription fraud as “the illegal acquisition of prescription drugs for personal use or profit, and could be observed in numerous ways”.

Similarly, Lisby et al. (2010) rightly note that no single definition is currently being used to determine medication/prescription errors as different organizations differ on the description of what constitutes an error. Aronson (2009 p. 599) notes that “a prescription error is a failure in the prescription writing process that results in a wrong instruction about one or more of the normal features of a prescription”. According to Dean, Barber and Schachter (2000 p. 233), a “clinically meaningful prescribing error occurs when, as a result of a prescribing decision or prescription writing process, there is an unintentional significant (1) reduction in the probability of treatment being timely and effective or (2) increase in the risk of harm when compared with generally accepted practice”. Various types of errors which could occur are: irrational prescribing, inappropriate prescribing, underprescribing, overprescribing and ineffective prescribing (Aronson 2009).

The coming together of fraud and errors is in the fact that with statistical analysis, both accidental errors and deliberately falsified data can be fished out through further investigations of anomalous observations (Bolton and Hand 2002). According to Thaifur et al. (2021), most of the perpetuators of fraud in healthcare field are medical personnel physicians. This further calls for the need of audits and checks at the physician level of prescriptions as fraud is characterized on the intention to cause harm or loss, without the requirement of demonstrating actual gain or loss suffered by the victim. (Griffith 2019 p. 1269).

In general, three possible ways fraud can be committed are set out in the 2006 Fraud Act and according to Griffith and Tengnah (2011), these are:

- False representation (Section 2)
- Failing to disclose information (Section 3)
- Abuse of position (Section 4)

Many examples of healthcare fraud can be found in literature. The United States Department of Justice in November 2022 (Anon.2022) released a report of a California man convicted for his role in an approximately $723,000 health care fraud and prescription drug diversion scheme involving two Southern California pharmacies. According to court documents and trial evidence, Shahriar "Michael" Kalantari, a resident of Beverly Hills, was involved in a health care fraud and unlicensed wholesale distribution scheme between 2016 and 2017. Kalantari and his co-conspirators obtained beneficiary information, which he then used to generate false and fraudulent prescriptions for costly medications, including those used for HIV treatment. The co-conspirators submitted claims to Medicare and Medicaid in California through their pharmacies, falsely claiming the drugs were dispensed to beneficiaries, when in reality, they were provided to other co-conspirators for illegal resale. Kalantari was found guilty of conspiracy to commit health care fraud, health care fraud, and conspiracy to engage in unlicensed wholesale distribution of prescription drugs.

In the NHS, an example was cited by Manning Julia (2011) in which a dentist from Liverpool was sentenced to two years in prison for fraudulently claiming over £300,000 from the NHS. John Hudson, who worked at HMP Altcourse and ran a dental practice in Rochdale, exploited an administrative error at another prison, leading to him wrongfully receiving NHS payments in addition to his earnings from the privately run prison. The court deemed his actions as calculated and persistent dishonesty, despite his standing in the local community. Hudson failed to disclose his existing payment from HMP Altcourse when the NHS changed prison dental contracts in 2006, resulting in him fraudulently obtaining £307,000. He used the money for personal expenses, including his children's school fees and vacations.

2.2 NHS Overview

The National Health Service (NHS) is the publicly funded healthcare system in the United Kingdom, providing comprehensive healthcare services to all residents. Established in 1948, the NHS is one of the largest and oldest single-payer healthcare systems globally (Sheard 2011).

The NHS operates under the principles of universality, equity, and access to healthcare services based on need rather than ability to pay (United Kingdom Government 2021). It is funded through general taxation, with the government allocating a significant portion of the national budget to support its operations. The NHS encompasses a wide range of services, including primary care, hospitals, mental health services, dentistry, and community healthcare (NHS 2023c).

According to the NHS website (2023b), primary care forms the foundation of the NHS, with general practitioners (GPs) serving as the initial point of contact for patients. GPs provide a range of medical services, including diagnosis, treatment, preventive care, and referrals to specialized services. Hospital services are also an integral part of the NHS, offering specialized care, emergency services, surgeries, and treatments. Hospitals provide a wide range of medical and surgical specialties, along with diagnostic and imaging services. The NHS also emphasizes mental health services, aiming to provide accessible and comprehensive support for individuals facing mental health challenges.

Prescriptions also make up a core service within the NHS. Medicines are usually prescribed by a doctor, with other healthcare professionals such as nurses, pharmacists, dentists, and physiotherapists also having the authority to prescribe medicines (NHS Inform 2023). In the current NHS prescription process, the majority of prescriptions for medicines and essential supplies are now managed electronically (NHS 2023a).

According to the NHS Website (2023a), individuals have two options for how this system operates. Firstly, they can select a specific pharmacy or dispenser to handle all their prescriptions. In this case, when a prescription is issued, it will be electronically transmitted to the chosen dispenser, eliminating the need for a paper prescription. Patients can then conveniently collect their prescribed medicines or essential supplies without the requirement of submitting a physical document. Alternatively, individuals have the flexibility to decide on a case-by-case basis where they would like their prescription to be dispensed. Upon receiving a prescription, a paper copy will be provided, containing a unique barcode. This barcode can be scanned by any pharmacy or dispenser in England to access the prescription securely from the NHS database. Although paper prescriptions will still be available in specific circumstances, the electronic processing of prescriptions is the primary method employed.

According to Borland (2011), all prescriptions were free when the NHS was established in 1948 and charging was introduced three years later to pay for defence spending. Prescription charges were abolished in Wales on April 1st 2007. In Northern Ireland, all prescriptions dispensed were made free of charge in April 2010. In Scotland, prescription charges were gradually reduced from 2007 and abolished altogether on April 1st 2011, all according to The Drugs and Therapeutics Bulletin (2015). This leaves England as the only country in the United Kingdom that still pays for prescriptions, much to the dismay of the citizens (Borland 2011). However, in England, there are various circumstances in which individuals are eligible for free NHS prescriptions.

2.3 Fraud and Error in Prescriptions

As with any human led process, errors and fraud are prone to exist in this prescription procedure. In a study carried out in Qatar, the causes of errors in prescriptions can be attributed to various factors as highlighted by Stewart et al. (2018) which are explained below. One significant factor according to the paper is stress and high-pressure situations experienced by healthcare professionals, which can contribute to medication errors. Both workload issues and patients themselves exerting pressure can contribute to stress among healthcare professionals, increasing the likelihood of errors. Heavy workloads were particularly highlighted by doctors as a reason for errors to occur. Furthermore, a critical lack of staff during key times, such as weekends and evenings, was identified as another issue compromising patient safety. Another notable finding was doctors relying on pharmacists to correct their prescribing errors, leading to complacency in the prescribing process. Additionally, doctors displayed a reluctance to alter prescriptions written by their peers, especially if they were from different specialties, considering it the responsibility of the original prescriber even in cases of prescribing errors. These identified causes shed light on the multifaceted nature of prescription errors, emphasizing the importance detecting and preventing the errors.

According to Sanghera, Franklin, and Dhillon (2007), pharmacists and nurses play crucial roles in acting as the primary line of defence against errors in the prescription process. They are responsible for verifying prescriptions and ensuring the accuracy and appropriateness of medications. However, it is important to acknowledge that these healthcare professionals are also susceptible to the same causes of errors mentioned earlier. The stress and high-pressure situations they encounter, stemming from heavy workloads and patient-related pressures, can affect their performance and increase the risk of errors.

In terms of fraud, government programs are particularly vulnerable as it is harder to exclude problematic providers than in privately managed networks, according to Thornton et al. (2014). Thornton et al. (2014) further explain that the application of data analysis methods in healthcare fraud detection is not as prevalent as in other sectors and this is primarily due to several factors, including the reluctance to acknowledge the existence of fraud in healthcare, the intricacies of claim systems, the vast amount of distributed claim data, and the limited funding allocated to fraud detection initiatives. Consequently, the development of electronic fraud detection systems in the healthcare industry lags behind industries like banking and telecommunications, which have made significant strides in this area (Thornton et al. 2014). This is particularly true of the NHS where very little literature exists into the investigations of fraud and the use of data analytical methods to combat fraud and errors, especially in prescriptions.

2.4 Fraud Control in NHS

According to Griffith and Tengnah (2011), measures to counter NHS fraud were established in 1998 through the establishment of the NHS Counter Fraud Service. In 2003, the NHS Counter Fraud and Security Management Service (NHS CFSMS) was formed as a special health authority under the Department of Health. Its purpose was to safeguard the staff, assets, and resources of the NHS in England and Wales (NHS Counter Fraud Authority 2023b). In 2005, the NHS CFSMS became part of the NHS Business Services Authority (BSA), consolidating various functions under a single special health authority.

In 2011, the NHS CFSMS division within the NHS BSA was rebranded as NHS Protect, aligning its counter fraud role with the development and implementation of the Department of Health and Social Care's anti-fraud strategy. Notably, in 2014, NHS Protect achieved a significant milestone by recovering its first million-pound sum. This recovery involved Dentist Joyce Trail, who was required to pay back £1.4 million she had defrauded from the NHS (NHS Counter Fraud Authority 2023b).

In 2017, the NHS Counter Fraud Authority (NHSCFA) was established as a new special health authority. Its primary focus shifted towards identifying, investigating, and preventing fraud and other economic crimes within the NHS and the wider health group (NHS Counter Fraud Authority 2023b). It is independent from other NHS bodies and directly accountable to the Department of Health and Social Care (DHSC).

The available literature on the activities of the NHS Counter Fraud Authority (NHSCFA) is limited, offering little or no detailed information regarding their methods, operations, successes, and failures. The NHSCFA primarily maintains a portal that serves as a platform for reporting cases of fraud. However, comprehensive documentation regarding their specific activities and outcomes remains scarce or undisclosed in existing literature.

In terms of prescriptions however, and according to the NHS Business Services Authority (2023), regular checks are conducted on prescription forms and dental treatment claim forms as part of fraud and error prevention measures in the prescription process. The NHS Business Services Authority carries out monthly randomized checks on these forms, which are submitted by pharmacies and dental practices at the end of each month. These checks aim to identify any potential instances of fraud or errors in the prescription and dental treatment claims.

These steps are in line with manual inspection method of fraud detection, which is one of the two general methods of fraud detection, according to Haddad Soleymani et al. (2018), with the second being statistics-based methods. Detecting fraud and errors through manual inspection is known to be highly accurate. However, when applied to a large volume of data, this method becomes costly and time-consuming as experts are required to meticulously review numerous documents within a limited timeframe, adhering to specific criteria to identify fraudulent claims (Soleymani et al., 2018). This process holds true not only for fraud detection but also for identifying errors within the NHS. According to the Association of Certified Fraud Examiners (2022), more than 50% of organizations currently use anomaly detection, as well as automated monitoring of red flags and business analysis as part of their anti-fraud programs with the use of artificial intelligence and machine learning in anti-fraud programs expected to more than double by 2024. This shows a strong move from manual inspection methods to statistics-based methods, which will be discussed in more detail in the following section.

2.5 Fraud Detection Methods

As earlier mentioned, there are two general methods for fraud detection which are manual inspection and statistics-based methods with the manual inspection method being very accurate but costly and time-consuming (Haddad Soleymani et al. 2018). On the other hand, statistics-based methods, such as data mining, offer an alternative approach to detecting fraud by leveraging advanced analytical tools and processing large volumes of data (Haddad Soleymani et al. 2018); Han, Kamber and Pei 2011; IBM 2023). Data mining is a relatively new concept that emerged in the mid-1990s and has since been utilized for data analysis and knowledge discovery (Haddad Soleymani et al. 2018; Han, Kamber and Pei 2011). It involves uncovering patterns, models, and valuable information from extensive datasets (Han, Kamber and Pei 2011; IBM 2023). By combining statistics, computer science, machine learning, and database technology, modern data mining has revolutionized fraud detection, enabling the analysis of larger datasets and the discovery of complex patterns (Hand 2007). This integration of various analytical technologies distinguishes data mining from traditional statistics, which was developed based on relatively small datasets (Hand 2007). Therefore, statistics-based methods, particularly data mining, provide valuable tools for efficiently analysing extensive datasets and uncovering fraudulent patterns in a more scalable and automated manner (Leanndra Copeland et al. 2012; Haddad Soleymani et al. 2018; Hand 2007). These methods can be divided into supervised or unsupervised, according to Bolton and Hand (2002). Other approaches are also discussed below:

2.5.1 Manual Inspection/Audit

Manual inspection or audit is a commonly employed method in healthcare fraud detection, where according to Leanndra Copeland et al. (2012), medical and claims experts review transactional claims on a case-by-case basis to identify anomalies. This approach relies on the expertise and knowledge of the reviewers to detect suspicious patterns or inconsistencies in the claims (Leanndra Copeland et al. 2012). Auditing strategies often utilize random stratification sampling methods to obtain samples from different claim types, aiming to capture a representative subset of the claims for review. However, manual inspection has limitations in terms of scalability and the ability to pinpoint all fraudulent claims among a large volume of transactions (Leanndra Copeland et al. 2012). In the case of healthcare prescription system, such as the Social Security Administration (SSA) in Turkey, manual detection of fraudulent prescriptions is conducted by a committee of medical doctors. A human expert reviews a relatively small sample of prescriptions associated with a hospital, and if fraudulent or abusive claims are found, the hospital is charged based on the percentage of fraudulent claims detected in the sample and the total cost of the prescriptions issued by the hospital during the inspection period (Aral et al. 2012).

2.5.2 Rule-Based Methods

Rule-based methods in fraud detection are closely related to manual inspection, relying on expert knowledge and domain expertise to identify anomalies in billing practices (Kumaraswamy et al., 2022). They utilize the experience and expertise of fraud experts to establish criteria for detecting fraudulent activities (Travaille et al., 2011). These methods involve developing simple to medium-complex rules based on common fraud schemes and patterns observed in the healthcare domain (Kumaraswamy et al., 2022). By utilizing such rules, they can effectively identify billing errors, duplicate claims, and fraudulent practices, such as DRG creep (manipulating diagnostic and procedural codes) and up-coding (billing for a higher level of service than provided) (Kumaraswamy et al., 2022).

Rule-based systems have been widely used in various domains such as money laundering and telecommunications for detecting suspicious activities (Bolton and Hand 2002). In these systems, rules are developed based on specific patterns or behaviours indicative of fraudulent or suspicious transactions, such as flagging transactions from certain countries or identifying high-value and long-duration calls (Bolton and Hand 2002).

While rule-based methods offer a straightforward and effective approach, they have inherent limitations. Once fraudsters become aware of the rules, they can modify their fraudulent patterns to evade detection (Kumaraswamy et al. 2022). Additionally, building and maintaining rule-based systems can be expensive and challenging, requiring constant input from fraud experts and difficulty in keeping the system up to date with the evolving healthcare landscape (Kumaraswamy et al. 2022). Nonetheless, rule-based methods provide a valuable contribution to fraud detection by providing a structured and predefined approach to identify known fraud patterns and schemes.

2.5.3 Supervised Methods

Supervised learning methods in fraud detection involve the utilization of labelled data, where the response variables are mapped to the corresponding inputs based on known fraud or non-fraud instances (Kumaraswamy et al. 2022). These methods heavily rely on human-labelled data to train sophisticated computer data mining algorithms and construct models that can distinguish between fraudulent and non-fraudulent records (Leanndra Copeland et al. 2012). The models are built using a range of classification techniques such as support vector machines, logistic regression, decision trees, and neural networks (Han, Pei and Tong 2023).

The fundamental principle of supervised learning is to create models from past fraudulent and non-fraudulent examples to classify new observations into one of these predefined classes (Bolton and Hand 2002). By leveraging the labelled training data, supervised models learn the distinct characteristics and patterns associated with each class, enabling the identification of similar instances in the future (Leanndra Copeland et al. 2012). However, it is important to note that supervised models are limited to detecting fraud patterns that have been observed in the past (Bolton and Hand 2002). They are not capable of identifying new or emerging types of fraud without continuous updates to capture evolving fraudulent behaviours (Joudaki et al. 2016).

Supervised methods face specific challenges in fraud detection. One issue is the requirement for accurately labelled data, where human experts categorize the records as fraudulent or non-fraudulent (Bolton and Hand 2002). Class imbalance, with a significantly larger number of legitimate transactions compared to fraudulent ones, can also impact the performance of supervised models (Leanndra Copeland et al. 2012). Additionally, the costs associated with misclassifications and investigations, as well as the uncertainty in class membership, must be carefully considered during model development (Bolton and Hand 2002).

Examples of the use of supervised learning algorithms in literature include neural networks, decision trees and support vector machines.

2.5.3.1 Neural Networks

Neural networks, a prominent supervised learning method, have gained significant traction in the realm of fraud detection due to their ability to handle complex data structures and capture non-linear relationships (Li et al. 2008). Inspired by the functioning of the human brain, a neural network comprises interconnected nodes that work collectively to process and analyse data (Kou et al. 2004). This modelling approach has been widely utilized in detecting healthcare fraud, where intricate patterns and relationships exist within the data.

According to Li et al. (2008), one of the notable challenges encountered when employing neural networks is the issue of overfitting, which occurs when the model performs well on the training dataset but fails to generalize effectively to new data. This concern is particularly relevant in the context of skewed datasets, such as healthcare claims, where the number of legitimate cases far exceeds the number of fraudulent cases (Li et al. 2008). Consequently, mitigating overfitting and improving the generalization capability of neural networks are crucial considerations in the development of effective fraud detection models.

The potential of neural networks in combating fraud has been demonstrated in various studies. For instance, a study focusing on Medicare fraud detection utilized deep learning methods to address the class imbalance problem present in the dataset (Johnson and Khoshgoftaar 2019). By implementing techniques like random over-sampling (duplicates examples in the minority class in the training dataset to rebalance the class distribution for an imbalaced dataset (Brownlee Jason 2021) and hybrid sampling (a combination of multiple resampling methods (Jiang et al. 2020), the study achieved improved performance, with notable gains in training time efficiency. Additionally, a case study conducted by a private health insurance company in Chile employed multilayer perceptron neural networks to detect fraudulent and abusive medical claims (Ortega, Figueroa and Ruz 2006). The results demonstrated the effectiveness of the neural network-based system in detecting such cases in a timely manner, contributing to the fight against fraudulent behaviours.

2.5.3.2 Decision Trees

Decision trees, a widely used supervised learning method, have gained popularity in the field of fraud detection due to their interpretability and ability to handle missing values (Li et al. 2008; Sarkar and Natarajan 2019). These non-parametric algorithms, such as ID3, C4.5, CART, and C5.0, facilitate predictive modelling by recursively partitioning the data based on certain conditions or rules (Sarkar and Natarajan 2019; Singh et al. 2019). The decision tree model estimates class conditional probabilities and partitions the data based on testing conditions to separate the classified variable into different classes (Singh et al. 2019).

The strengths of decision trees lie in their interpretability, as they can generate rules that provide insight into the decision-making process (Li et al. 2008). Additionally, decision trees are capable of handling missing values and can adapt to both categorical and continuous attributes (Sarkar and Natarajan 2019). However, decision trees may generate a large number of rules, which can decrease interpretability, particularly in high-dimensional datasets and they also have few adjustable parameters available, which can limit their flexibility (Li et al. 2008).

According to Bhattacharyya et al. (2011), ensemble methods such as random forests have been developed to address the limitations of single decision tree models. Random forests combine the random subspace method with bagging to create an ensemble of decision trees. This approach enhances the generalization performance of decision trees by introducing variation among individual trees and mitigating issues of instability and reliability (Bhattacharyya et al. 2011).

In a study by Bhattacharyya et al. (2011) evaluating the performance of different data mining techniques for credit card fraud detection, random forests exhibited superior performance across various measures. The study utilized a real-life dataset of credit card transactions and found that random forests outperformed other methods, particularly at higher depths, effectively capturing more fraud cases while minimizing false positives.

2.5.3.3 Support Vector Machines

Support Vector Machines (SVMs) stand as a prominent supervised machine learning algorithm with versatile applicability in both classification and regression tasks (Patel, Chatterjee and Gorai 2019). Over the last decade, SVMs have emerged as a topic of significant development and research in the realm of machine learning (Martín Muñoz 2002). Their success can be attributed to their strong theoretical foundations in generalisation and convergence, alongside their exceptional performance in challenging problems (Martín Muñoz 2002).

While SVMs are renowned for their prowess in classification tasks, their utility extends to approximating functions as well, known as SVM regression (Mario, 2002). However, one limitation in wider application arises from the fact that SVMs are not readily applicable in online learning scenarios, such as sequential data acquisition, where data arrives sequentially, and learning must be initiated anew with each data point (Martín Muñoz 2002). This challenges their practicality in scenarios like on-line prediction of temporal series (Martín Muñoz 2002).

Nonetheless, SVMs have made substantial contributions in a variety of data mining and machine learning applications, notably in classification, clustering, and regression (Murty and Raghava 2016).

The popularity of SVMs in pattern classification is particularly noteworthy, as SVMs excel in assigning class labels to unlabelled patterns based on a set of labelled patterns (Murty and Raghava 2016).

In a study by Kumar, Ghani and Mei (2010), a data-driven approach using Support Vector Machines (SVMs) was proposed for healthcare error detection. The research aimed to address escalating administrative costs and errors in the U.S. healthcare system, impacting insurance providers and consumers. The study focused on predicting claims likely to require rework before payment through a binary classification task using SVMs. Extensive features were extracted from claims data, including member and provider information, claim headers, and line details. SVMs were chosen for their robustness with large feature sets.

The system met industry experts' requirements, offering prepayment prediction, generalization, high accuracy, explanations for auditors, and adaptability to changes. Labelled data was collected from multiple sources, such as Quality Control Audit and Provider Dispute systems. The system implementation included data collection, feature construction, model learning, feature selection, and scoring. Model learning experiments optimized classifier parameters and feature sets. User feedback and explanations improved adaptability.

Kumar, Ghani and Mei's (2010), study achieved promising results, accurately identifying potential rework claims, prioritizing manual examination, and reducing administrative costs. It showcased SVMs' effectiveness in healthcare fraud and error detection, promising more accurate claims processing.

2.5.4 Unsupervised Methods

Unsupervised methods of fraud detection play a vital role in identifying fraudulent activities in healthcare data when labelled instances of fraud are not available. These techniques are particularly useful in capturing anomalies and patterns based on the distributions of billing behaviour without the need for prior knowledge of fraud (Kumaraswamy et al. 2022). Unlike supervised methods that rely on labelled data, unsupervised learning approaches focus on identifying hidden structures and patterns in unlabelled data (Kumaraswamy et al. 2022; Ekin et al. 2018). By leveraging descriptive statistics and data mining techniques, unsupervised methods can effectively detect potential deviations and outliers that may indicate fraudulent behaviour (Ekin et al. 2018; Thornton et al. 2014).

According to Ekin et al. (2018), one significant advantage of unsupervised methods is their ability to serve as initial filters for identifying potentially fraudulent claims before conducting thorough investigations, which can reduce personnel costs and improve efficiency. These approaches are independent of specific classified data sets and can adapt to changing fraud patterns over time and the versatility of unsupervised learning makes it a valuable tool in the detection of fraud, even when dealing with unlabelled medical data that requires additional assessment from subject matter experts (Ekin et al. 2018).

In the context of fraud detection, two commonly used unsupervised methods are clustering and outlier detection (Ekin et al. 2018; Bolton and Hand 2002). These methods act as preliminary steps to flag potentially fraudulent activities and guide further investigations by domain experts (Ekin et al. 2018).

2.5.1 Cluster

Clustering techniques, as described by Chakraborty, Islam and Samanta (2022), involve the process of categorizing data into distinct classes or clusters, where the elements within each cluster exhibit similarities but maintain substantial differences from those in other clusters. As asserted by Ekin et al. (2018), clustering techniques facilitate the grouping of analogous claims based on their distinguishing attributes, allowing for the identification of underlying structures and correlations within the dataset.

In a study by Joudaki et al. (2016), clustering techniques were used to detect indicators of healthcare fraud and abuse among general physicians' outpatient claims. Data from insured patients' visits and drug prescriptions were obtained from the Social Security Organization (SSO) in Iran. Indicators representing fraud and abuse symptoms were created and validated through expert interviews and standardization using Z-scores was applied. Hierarchical clustering identified two clusters of physicians, labelled as healthy or suspect based on their characteristics. Discriminant analysis was then performed on data from the 12th month to assess indicator effectiveness. The results demonstrated successful identification of suspect physician groups involved in fraudulent activities.

In another study by Lin et al. (2008), the researchers aimed to explore general practitioners' (GPs') practice patterns using knowledge discovery in database (KDD) techniques and leverage expert knowledge to support management decisions. They utilized a two-stage unsupervised learning tandem approach, integrating self-organization maps (SOM) and principal component analysis (PCA) for clustering GPs' profiles based on health expenditure data.

The data pre-processing involved selecting GPs' profiles with specific reimbursement criteria and normalizing the raw data using the Min-Max method. The dataset covered a rolling 1-year period for 1210 GPs in the southern region of the Bureau of National Health Insurance (BNHI) in Taiwan. The SOM neural network and PCA were employed for clustering GPs' practice patterns. The silhouette coefficient which “is a metric to evaluate the quality of clustering” (Song, Wang and Pan 2023, p. 5), was used to determine the optimal number of clusters. The study identified five critical clusters, representing different dimensions of GPs' practice patterns related to quantity, price, and quality. The attributes' maximum values in each cluster were matched with principal components obtained from PCA, resulting in the critical clusters.

The managerial priority of these critical clusters was determined using the analytic hierarchy process (AHP) based on expert assessments. AHP is a decision-making approach that integrates rational and intuitive elements and serves as a theory and methodology for measuring alternatives relative to multiple criteria (Brunelli 2014; Saaty and Vargas 2012). The results of Lin et al. (2008) highlighted the importance ranking of the critical clusters for health expenditures, aiding in the prevention of health fraud and the improvement of prescription quality.

2.5.2 Outlier Detection

Outlier detection, as articulated by Ranga Suri, Murty and Athithan (2019), can be broadly defined as the process of identifying a specific subset of data objects within a larger dataset that exhibit significant dissimilarity, exceptional characteristics, and inconsistency when compared to the remaining data. This concept holds particular relevance in the domain of healthcare fraud detection, where these outliers often serve as crucial indicators of potentially fraudulent activities.

Unsupervised outlier detection methods, as explained by Boukerche, Zheng and Alfandi (2020), utilise unlabelled data to either construct models for outlier score calculation, as exemplified by techniques like the Isolation Forest, or directly compute outlier scores for input data without the need to build models, as seen in methods such as LOF (Local Outlier Factor).

According to Liu, Kai and Zhi-Hua Zhou (2008) Isolation Forest, often referred to as iForest, constructs an ensemble of iTrees when provided with a dataset. Anomalies within this dataset are identified as instances characterised by short average path lengths when traversing the iTrees. Liu, Kai and Zhi-Hua Zhou (2008) conclude in their paper that Isolation Forest emerges as a dependable and efficient anomaly detection tool, particularly well-suited for extensive databases as its remarkable capability to handle large-volume databases renders it exceptionally desirable for practical, real-life applications.

The Local Outlier Factor (LOF), as introduced by Breuniq et al. in 2000, represents a novel approach in the realm of outlier detection. LOF assigns each object within a dataset a specific degree of being an outlier, capturing the notion of how isolated an object is within its immediate neighbourhood (Breuniq et al. 2000). Unlike traditional outlier detection methods, as explained further by Breuniq et al. (2000), LOF stands out as it quantifies the extent to which an object deviates from the norm without relying on predefined clusters or explicit cluster notions. It provides a localised perspective, offering a more distinct understanding of outliers within a dataset (Breuniq et al. 2000).

In the context of healthcare fraud detection, outlier detection methods, including these unsupervised approaches, play a pivotal role in identifying unusual or anomalous observations that deviate markedly from the anticipated patterns (Ekin et al., 2018; Bolton and Hand, 2002). As emphasised by Ranga Suri, Murty and Athithan (2019), these outliers take on added significance, as their presence may signal the presence of fraudulent activities within the healthcare system.

In the study conducted by Thornton et al. (2014), a data-driven method for healthcare fraud detection was proposed using outlier detection as the primary method. The research aimed to address the significant issue of fraud, waste, and abuse in the U.S. healthcare system, estimated at $700 billion annually. The paper introduced a multi-dimensional data model for Medicaid claim data and identified specific metrics for dental providers to detect fraudulent activity.

Outlier detection, an unsupervised data mining technique, was employed to identify suspicious behaviour among dental providers. The metrics used for outlier detection were derived from various sources, including comparative research, fraud cases, and existing literature on healthcare fraud. The study analysed 11 months of Medicaid dental claim data and evaluated 14 different metrics using the R language and statistical packages. By applying the proposed outlier detection approach, the researchers successfully identified 35 dental providers with two or more potential predictive flags for fraud, and 17 providers with three or more flags. Upon review by qualified healthcare fraud subject matter experts, it was found that at least 12 of these 17 providers (71%) warranted immediate referral for audit and potential legal investigation.

In a separate study by Yamanishi et al. (2004), outlier detection was explored as the primary method for healthcare fraud detection, with clustering also playing a role. The study proposed SmartSifter, an outlier detection engine based on statistical learning theory, to detect outliers in an online process through unsupervised learning of a probabilistic model and was used in fraud detection, network intrusion detection and network monitoring, among others.

SmartSifter's unique features included its adaptability to non-stationary data sources, clear statistical/information-theoretic meaning for outlier scores, computational efficiency, and its ability to handle both categorical and continuous variables. The approach employed a hierarchical probabilistic model, with histogram density used for categorical variables and finite mixture models for continuous variables. The on-line learning algorithm developed, Sequentially Discounting Laplace Estimation (SDLE) for categorical domain and Sequentially Discounting Expectation and Maximizing (SDEM) for continuous domain, continuously updated the model and gradually discounted the effect of past examples.

The application of SmartSifter was demonstrated through various experiments, including network intrusion detection using the KDD Cup 1999 dataset and healthcare fraud detection using Australia's Health Insurance Commission's pathology data. In the latter case, SmartSifter identified several pathology providers as significant rare cases, which were later confirmed by human experts' manual review.

2.5.5 Hybrid Methods

Hybrid methods in fraud detection refer to an amalgamation of both supervised and unsupervised techniques, as highlighted by Kumaraswamy et al. (2022). According to Carcillo et al. (2021), hybrid learning can be leveraged to enhance the accuracy of fraud detection.

An illustrative example of hybrid methodology is demonstrated by Shin et al. (2012), as cited by Kumaraswamy et al. (2022). The study proposed a scoring model designed to identify outpatient clinics exhibiting abusive utilisation patterns based on profiling information extracted from electronic insurance claims. The model consisted of two core components: scoring and segmentation. The scoring phase quantified the degree of abusiveness, while the segmentation phase categorised problematic providers with similar utilisation patterns. This approach allowed for the creation of a comprehensive model that captured both the severity of abuse and the categorisation of providers based on their behaviour.

For their study, Shin examined 3,705 Korean internal medicine clinics using practitioner claims submitted to the National Health Insurance Corporation. The model's validity was assessed using data from the Health Insurance Review and Assessment Services, comparing the proposed scoring system against manual intervention decisions. The composite degree of anomaly (CDA) score, aggregating 38 indicators of abusiveness, was formulated and applied for detection. The model effectively grouped clinics based on CDAs, using decision trees for further segmentation into homogenous clusters according to their utilisation patterns.

This hybrid model demonstrated consistent results with manual detection techniques and offered automation capabilities, ensuring flexibility and ease of updating. This approach represents the synergy of supervised and unsupervised techniques, allowing for precise detection of abusive utilisation patterns in healthcare claims data.

2.6 Summary

In summarising the concepts, prescription fraud can be defined as the "illegal acquisition of prescription drugs for personal use or profit and may manifest in various forms" (Aral et al., 2012, p. 37). Conversely, a prescription error is characterised as a "failure in the prescription writing process that leads to incorrect instructions regarding one or more normal prescription attributes" (Aronson, 2009, p. 599). The methodologies employed to identify these issues are outlined in Table 1 below.

Table 1: Summary of Fraud Detection Methods (Source: Author 2023)

Method	Characteristics and Implications
Manual Inspection/Audit	- Involves medical and claims experts reviewing claims individually for anomalies.
	- Requires expert knowledge to detect suspicious patterns or inconsistencies.
	- Limited scalability and efficiency due to manual nature.
Rule-Based Methods	- Expert knowledge used to establish criteria for detecting fraud.
	- Simple to medium-complex rules based on common fraud patterns.
	- Effectively identifies billing errors, duplicate claims, and specific fraudulent practices.
	- Vulnerable to fraudsters adapting to rules and expensive to maintain and update.
Supervised Methods	- Utilizes labeled data to train models to distinguish between fraudulent and non-fraudulent instances.
	- Models learn characteristics of each class from past examples.
	- Limited to detecting known patterns and requires labeled data.
	- Challenges include accurate labeling, class imbalance, and cost considerations.
Unsupervised Methods	- Captures anomalies and patterns based on billing behavior distributions without prior fraud knowledge.
	- Useful for initial filtering and adaptability to changing fraud patterns.
	- Requires expert interpretation.
	- Includes clustering and outlier detection as preliminary steps for identifying potential fraud.

Hybrid Methods	- Amalgamation of both supervised and unsupervised techniques.
	- Enhances accuracy by leveraging strengths of both approaches.

Chapter Three – Methodology

3.1 Introduction

This chapter provides an overview of the model building process and the important considerations that guided it. To ensure the reliability and effectiveness of the analysis, the Cross-Industry Standard Process for Data Mining (CRISP-DM) framework was adopted. CRISP-DM is a well-established and widely accepted methodology in the field of data analytics and machine learning (Tripathi et al. 2021; Hemanth and Kose 2020).

As noted by Tripathi et al. (2021), CRISP-DM was first introduced in 1996 and has proven to be versatile across various domains, including healthcare. It offers a systematic and organized structure that breaks down the complex data mining process into distinct phases, enabling the practical use of data-driven knowledge discovery techniques. According to Plotnikova, Dumas, and Milani (2020), CRISP-DM was developed with input from industry experts, making it widely adopted by both industry and research communities and is considered the standard for data mining methodologies and serves as a benchmark for other frameworks.

Plotnikova, Dumas, and Milani (2020) outline the six phases of CRISP-DM as: 1) Business Understanding, 2) Data Understanding, 3) Data Preparation, 4) Modeling, 5) Evaluation, and 6) Deployment, as seen in Figure 1 below. Each of these phases will be discussed in detail in the following sections. By adhering to the CRISP-DM framework, this methodology chapter aims to bring structure and rigor to the model development process.

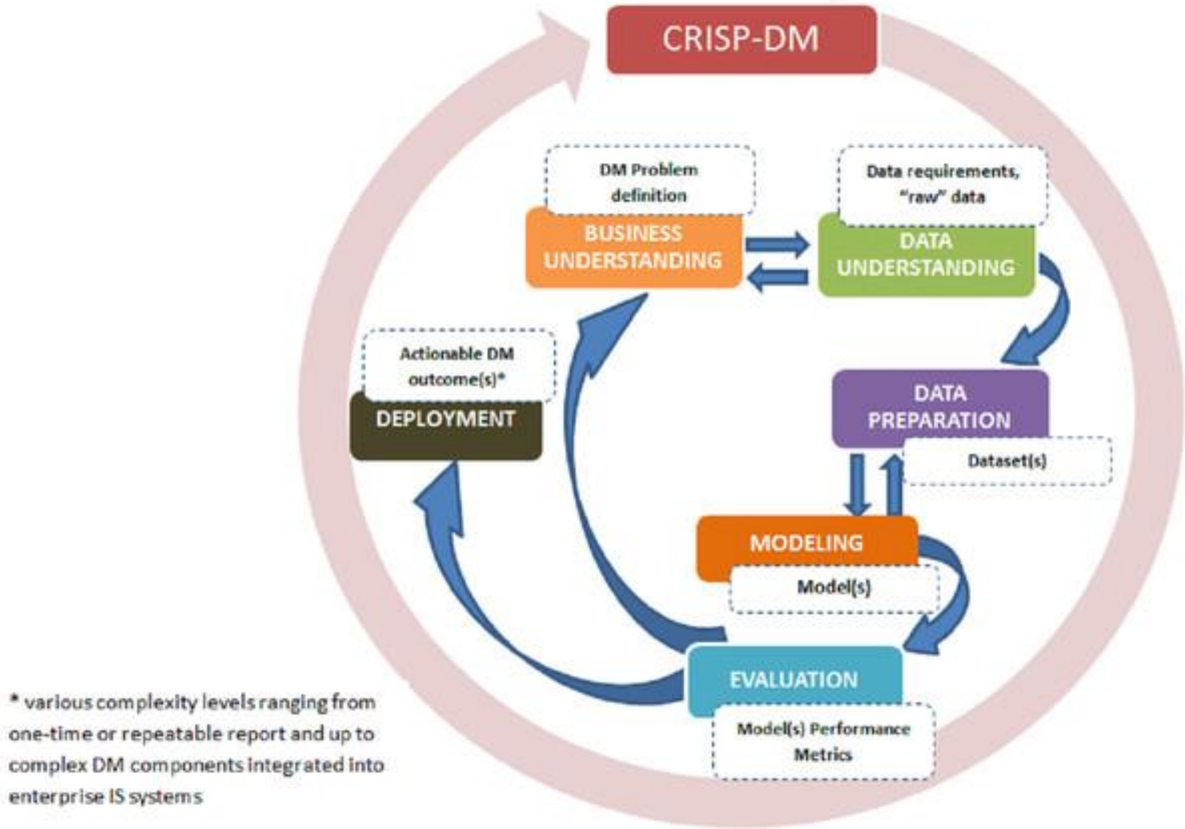

Figure 1: CRISP-DM Framework (Source: Plotnikova, Dumas, and Milani 2020)

3.2 Business Understanding

In accordance with Plotnikova, Dumas, and Milani (2020), the initial phase, referred to as "Business Understanding", entails a comprehensive grasp of the project's business objectives and requirements. This phase serves as the cornerstone upon which the entire data mining endeavor is constructed.

Building upon the insights of Iyengar, Hermiz and Natarajan (2014) in their study on fraud and abuse in prescription claims data, this approach was customized to pinpoint practices associated with abnormal and excessive prescriptions within a specific medication class. This targeted identification is of paramount importance for effective auditing and intervention. Similarly, Yamanishi et al. (2004) underscored the necessity of evaluating the degree of deviation of incoming data in relation to a normal pattern.

Hence, this project involved harnessing the available data to construct a model that captures the baseline behavior of prescription patterns within practices in England. Following a similar methodology to Yamanishi et al. (2004), the model was devised to flag outliers by assessing the extent of deviation from established norms. Furthermore, scores were assigned to the detected outliers in the input data to measure the change in the model post-learning, drawing inspiration from both

Yamanishi et al. (2004) and Iyengar, Hermiz and Natarajan (2014). A high score indicates a high likelihood that the data point is an outlier.

3.3 Data Understanding

The phase of Data Understanding, as elucidated by Plotnikova, Dumas, and Milani (2020), initiates with the acquisition of initial data, followed by a series of steps aimed at fostering familiarity with the data landscape. The data collection process involved a comprehensive review of data released by the NHS, specifically focusing on prescription data for practices in England. This approach aligns with the observations of Kumaraswamy et al. (2022), who outlined that publicly available data is predominantly utilised due to stringent restrictions imposed by data privacy and legal concerns when it comes to utilising private data.

For sourcing data, the English Prescribing Dataset available on the NHS Open Data Portal was utilised. As detailed in the guidance document provided by the NHSBSA (2023), this dataset encompasses information regarding the number and quantity of prescription items dispensed monthly, along with cost-related details for General Practitioner Practices and Cost Centres in England. The dataset includes items prescribed in England that were dispensed both in England and other parts of the United Kingdom. It does not include items whose prescription originated outside England even though they were dispensed in England.

Consistent with the approaches of Joudaki et al. (2016) and Lin et al. (2008), the analysis cantered on prescription data from a single year, in this instance, the year 2022. A comprehensive data dictionary is presented in Table 2 below.

Data Dictionary

Table 2: Data Dictionary (Source: NHSBSA 2023)

Column	Title	Type	Description
YEAR_MONTH	Year and Month as YYYYMM	number	Example: 201401
REGIONAL_OFFICE_NAME	Regional Office Name	string	The name given to a geographical region by NHS England. Each region supports local

Column	Title	Type	Description
			systems to provide more joined up and care for patients.
REGIONAL_OFFICE_CODE	Regional Office Code	string	The unique code used to refer to a Regional Office.
STP_NAME	STP Name	string	The name given to a geographical area by NHS England that is a smaller division of a Region. STP stands for Sustainability and Transformation Partnership
STP_CODE	STP Code	string	The unique code used to refer to an STP.
PCO_NAME	Primary Care Organisation Name	string	An NHS organisation that commissions or provides care services involving prescriptions that are dispensed in the community. For example: a Clinical Commissioning Group (CCG), an NHS Trust.
PCO_CODE	Primary Care Organisation Code	string	The unique code used to refer to a Primary Care Organisation.
PRACTICE_NAME	Practice Name	string	The name of an organisation that employs one or more prescribers who issue prescriptions that may be dispensed in the community. For example: a GP Practice, an Out-of-Hours service, a hospital department within an NHS Trust.
PRACTICE_CODE	Practice Code	string	The unique code used to refer to a Practice.

Column	Title	Type	Description
ADDRESS_1	Address Field 1	string	The Address used by a Practice. This data is supplied by Primary Care Support England (PSCE), NHS England Area Teams or the Clinical Commissioning Group, whenever a new practice is opened or if a change of details is required.
ADDRESS_2	Address Field 2	string	Same as above
ADDRESS_3	Address Field 3	string	Same as above.
ADDRESS_4	Address Field 4	string	Same as above.
POSTCODE	Post Code	string	Same as above.
BNF_CHEMICAL_SUBSTANCE	British National Formulary (BNF) Chemical Substance Code	string	A unique code used to refer to a BNF Chemical Substance. For example, 0501013B0
CHEMICAL_SUBSTANCE_BNF_DESCR	British National Formulary (BNF) Chemical Substance Description	string	The name of the main active ingredient in a drug or the type of an appliance. Determined by the British National Formulatory (BNF) for drugs, or the NHS BSA for appliances. For example, Amoxicillin
BNF_CODE	British National Formulary (BNF) Code	string	The unique code used to refer to a BNF Presentation. For example, 0501013B0AAABAB
BNF_DESCRIPTION	British National Formulary (BNF) Description	string	The name given to the specific type, strength, and formulation of a drug; or, the

Column	Title	Type	Description
			specific type of an appliance. For example, Amoxicillin 500mg capsules
BNF_CHAPTER_PLUS_CODE	British National Formulary (BNF) Description	string	The name given to a British National Formulatory (BNF) Chapter that includes the prescribed product. Includes the numerical code used to refer to the chapter. For example, 05: Infections
QUANTITY	Quantity	number	The quantity of a medicine, dressing or appliance for which an individual item was prescribed and dispensed, for each BNF Presentation. This represents a pseudo pack size, to illustrate the typical range of prescribed quantities of a given presentation. For example, a quantity of 28 for Amoxicillin 500mg capsules means that the pack size dispensed was 28 capsules.
ITEMS	Items	number	The number of times a product appears on a prescription form. Prescription forms include both paper prescriptions and electronic messages.
TOTAL_QUANTITY	Total Quantity	number	The total quantity of a drug or appliance that was prescribed. This is calculated by multiplying Quantity by Items. For example, if 2 items of Amoxicillin 500mg capsules with a quantity of 28 were prescribed, total quantity will be 56.
ADQUSAGE	Average Daily Quantity (ADQ)	number	Average Daily Quantity (ADQ) is the typical daily dose of a medication, prescribed to

Column	Title	Type	Description
			adult patients by GP Practices. This field shows the quantity prescribed multiplied by the strength, which is then divided by the Average Daily Quantity value.
NIC	Net Ingredient Cost (NIC)	number	In GBP. The amount that would be paid using the basic price of the prescribed drug or appliance and the quantity prescribed. Sometimes called the "Net Ingredient Cost" (NIC). The basic price is given either in the Drug Tariff or is determined from prices published by manufacturers, wholesalers or suppliers. Basic price is set out in Parts 8 and 9 of the Drug Tariff. For any drugs or appliances not in Part 8, the price is usually taken from the manufacturer, wholesaler or supplier of the product.
ACTUAL_COST	Actual Cost	number	In GBP. The basic cost after adjustment for the national average discount and some payments to the dispenser. The calculation is: Net Ingredient Cost - National Average Discount Percentage + (payment for consumables + out of pocket expenses + payment for containers)
UNIDENTIFIED	Unidentified	string	This field shows data from prescription forms that could not be allocated to a Practice.

3.4 Data Preparation

As outlined by Plotnikova, Dumas, and Milani (2020), the third phase encompasses tasks necessary to construct the definitive dataset from the initial raw data, with these operations conducted iteratively. Given the substantial size of the datasets—averaging at 6.5GB and containing approximately one million seven hundred thousand records per month—the initial step was to filter a specific category of medication based on the British National Formulary (BNF) Chemical Substance Description.

Iyengar, Hermiz, and Natarajan (2014) point out that prescription drug fraud and abuse are predominantly linked to certain drugs, which fall into two primary categories. The first category encompasses high-volume drugs that can be resold to pharmacies and potentially billed twice to health plans.

The second category includes drugs with a high street value owing to their association with non-medical and recreational misuse. For the scope of this project, emphasis was placed on the latter group of drugs, and further elaboration on these drugs is provided in Table 3.

The dataset was free from of missing values. However, since the project aimed to identify practices exhibiting unusual prescription behavior, all prescriptions associated with unidentified practices were excluded from the dataset. Additionally, irrelevant columns were eliminated, and an overview of the utilized columns along with their respective data types is illustrated in Figure 2.

The ultimate dataset encompassed twelve months of data, seven regions, eight thousand five hundred fifty-four practices, and six hundred eighty-eight drugs—all specific to England.

Table 3: Drugs of Focus (Source: Author 2023)

Class	Explanation	Generic Name
Narcotic Analgesics	This class contains two of the most widely abused prescription medications, oxycodone and hydrocodone, and also contains a variety of combination drugs which are often abused because they may have less stringent controls	Tramadol Oxycodone Fentanyl Methadone Dextromethorphan Meperidine

	on dispensing and distribution Iyengar, Hermiz and Natarajan (2014).	Codeine Buprenorphine Hydromorphone Tapentadol Morphine Hydrocodone Oxymorphone Butorphanol Nalbuphine Opium Propoxyphene Pentazocine Levorphanol Remifentanil Sufentanil Oliceridine Source: Cohen Brandon, Leigh J. Ruth and Preuss V. Charles (2023) and Pope Carmen (2023)
Ataractics-Tranquilizers	This class includes medications with benzodiazepines that are prescribed as antianxiety drugs but are also susceptible to addiction and abuse Iyengar, Hermiz and Natarajan (2014).	Benzos Diazepam Valium Source: Public Health Scotland (2023)
CNS Stimulants	This class includes medications like the generic methylphenidate that are prescribed for attention deficit hyperactivity disorder (ADHD), but are also abused due to their euphoria-inducing effects Iyengar, Hermiz and Natarajan (2014).	Methylphenidate Lisdexamfetamine Source: NHS Highlands (2023)
Amphetamine Preparations	These drugs are often abused for their performance-enhancing benefits and	Amphetamine Methylamphetamine Ecstasy-type drugs

	euphoria-inducing effects Iyengar, Hermiz and Natarajan (2014).	Source: Scottish Police Authority (2023)

```
YEAR_MONTH                          datetime64[ns]
REGIONAL_OFFICE_CODE                        object
REGIONAL_OFFICE_NAME                        object
PRACTICE_CODE                               object
PRACTICE_NAME                               object
CHEMICAL_SUBSTANCE_BNF_DESCR                object
BNF_DESCRIPTION                             object
QUANTITY                                     int64
ITEMS                                        int64
TOTAL_QUANTITY                               int64
```

Figure 2: Dataset Utilised Columns (Source: Author 2023)

Further exploration of the dataset was conducted following the data preparation phase to uncover its inherent characteristics. Understanding the nature of the dataset is crucial as the modeling and evaluation results are intricately linked to its attributes.

Regarding the distribution of practices across regions, the Midlands region stood out with the highest number, comprising 1,681 practices (see Figure 3). On average, each region contained approximately 1,233 practices, indicating a relatively consistent distribution.

Furthermore, an examination of unique medications within the categories of interest revealed a consistent pattern across each month of the year, with an average count ranging from 46 to 48 (see Figure 4). This consistency suggests a stable environment with no unusual spikes in the introduction or discontinuation of medications during the observation period.

The average monthly occurrences of each medication within practices also followed a similar uniform trend, albeit on a smaller scale, averaging around 1.4 (see Figure 5).

Shifting the focus to the volume of individual prescriptions, specifically Quantity and Items, the average quantity per prescription remained relatively stable each month, with an average of 96.5 (see Figure 6). However, this metric exhibited significant variability, ranging from a minimum value of 1.0 to a maximum of 12,000 (see Figure 9). Conversely, for Items, the average was 3.2, with a similar distribution of values, ranging from 1.0 to 972 (see Figure 7 and Figure 9).

In summary, the dataset showcases a considerable number of practices and medications, each with a relatively low frequency of occurrences within individual practices. Moreover, there exists a substantial variation in both Quantity and Items values.

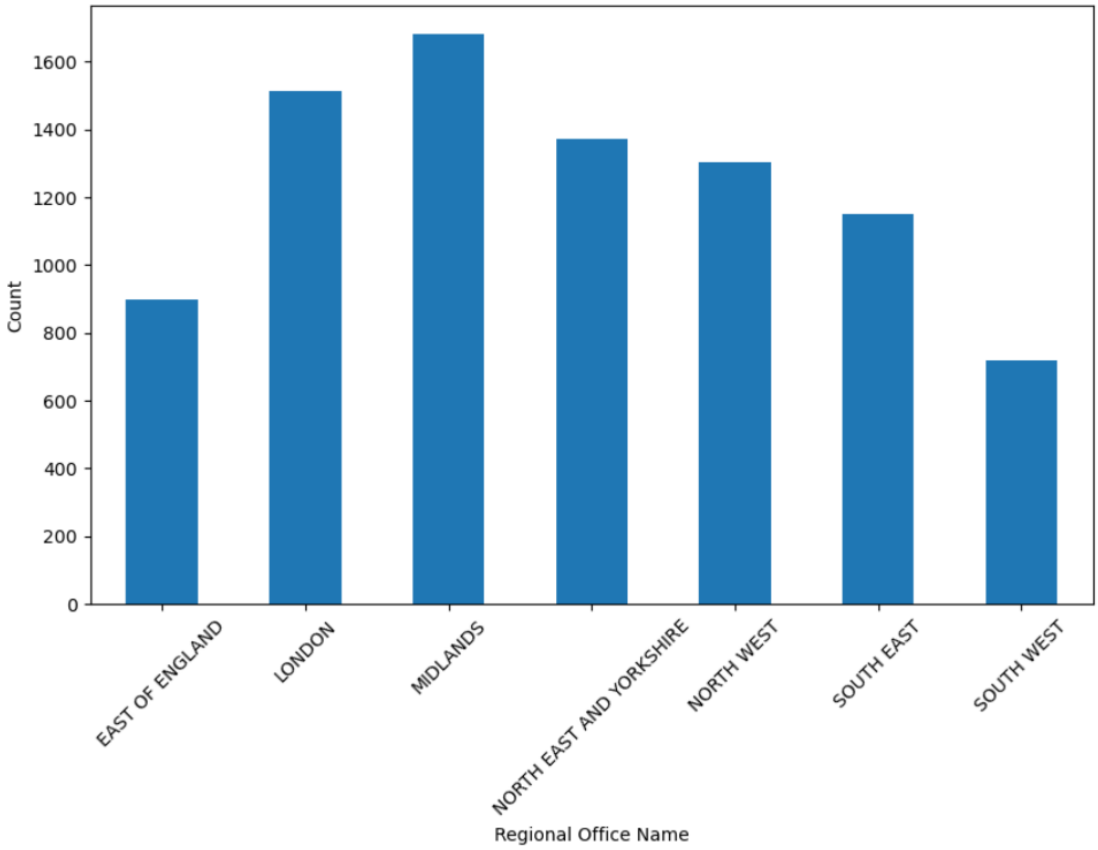

Figure 3: Number of Unique Practices in Each Region (Source: Author 2023)

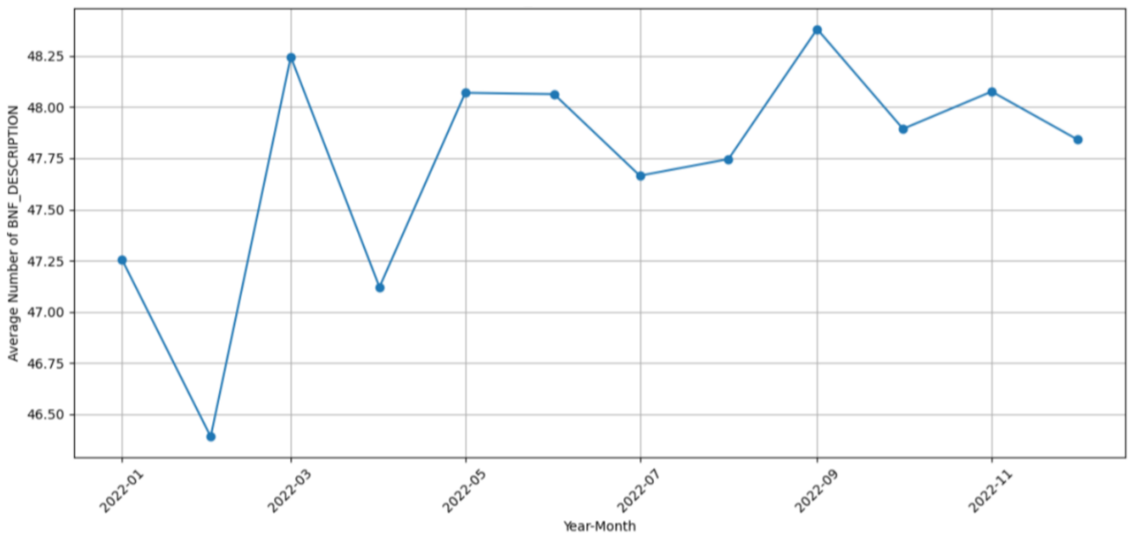

Figure 3: Average Number of BNF Description per Hospital for Each Month (Source: Author 2023)

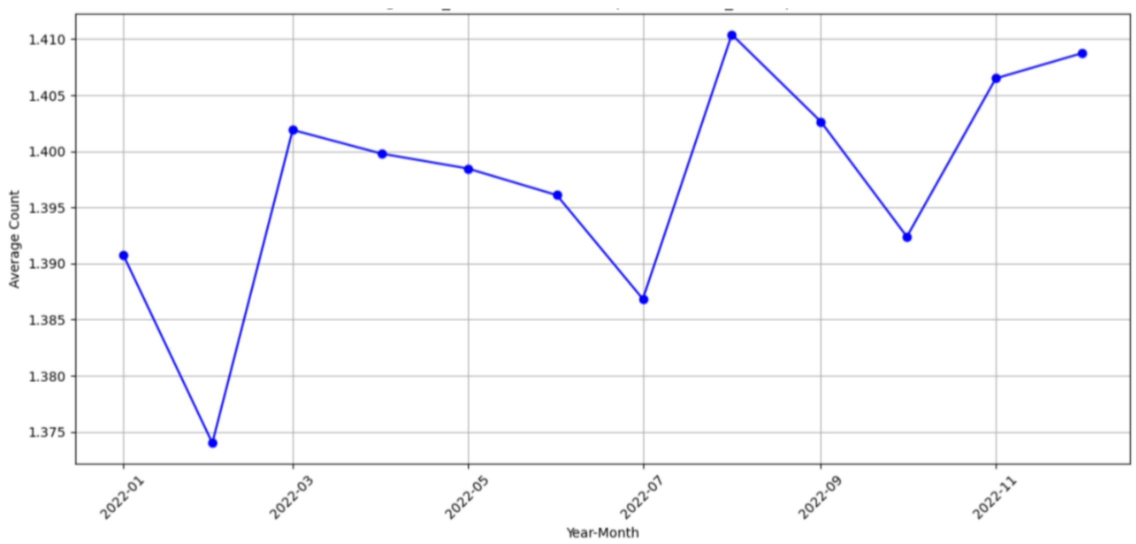

Figure 4: Average BNF Description Count per Practice Name per Month (Source: Author 2023)

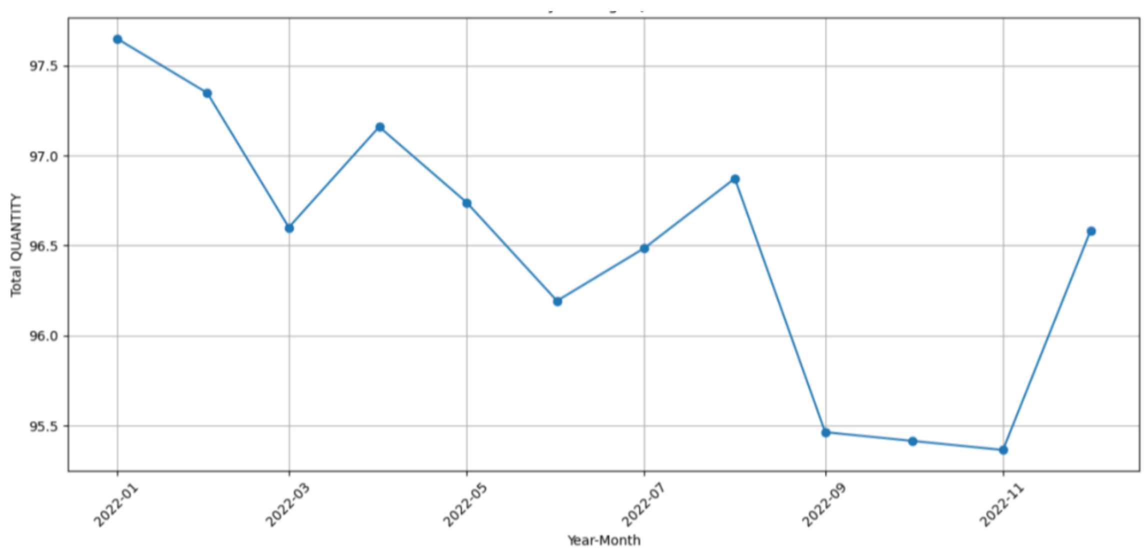

Figure 5: Monthly Average Quantity (Source: Author 2023)

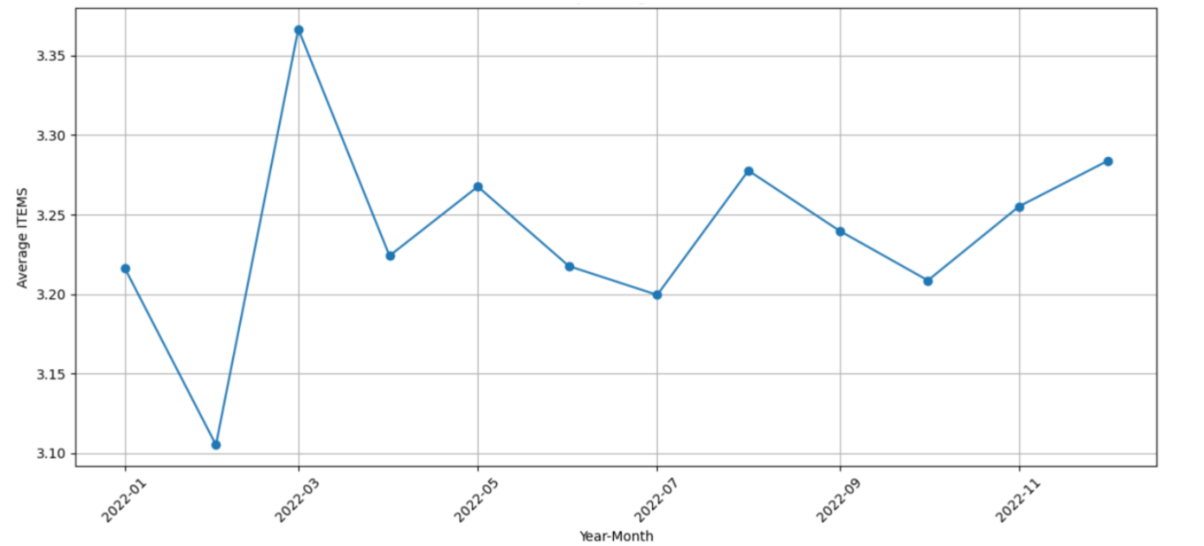

Figure 6: Monthly Average Items (Source: Author 2023)

Descriptive Statistics for Numeric Features:

	QUANTITY	ITEMS
count	9.314930e+06	9.314930e+06
mean	9.648494e+01	3.239038e+00
std	1.736448e+02	6.063373e+00
min	1.000000e+00	1.000000e+00
25%	2.000000e+01	1.000000e+00
50%	5.600000e+01	2.000000e+00
75%	1.000000e+02	3.000000e+00
max	1.200000e+04	9.720000e+02

Figure 7: Descriptive Statistics for Numeric Features (Source: Author 2023)

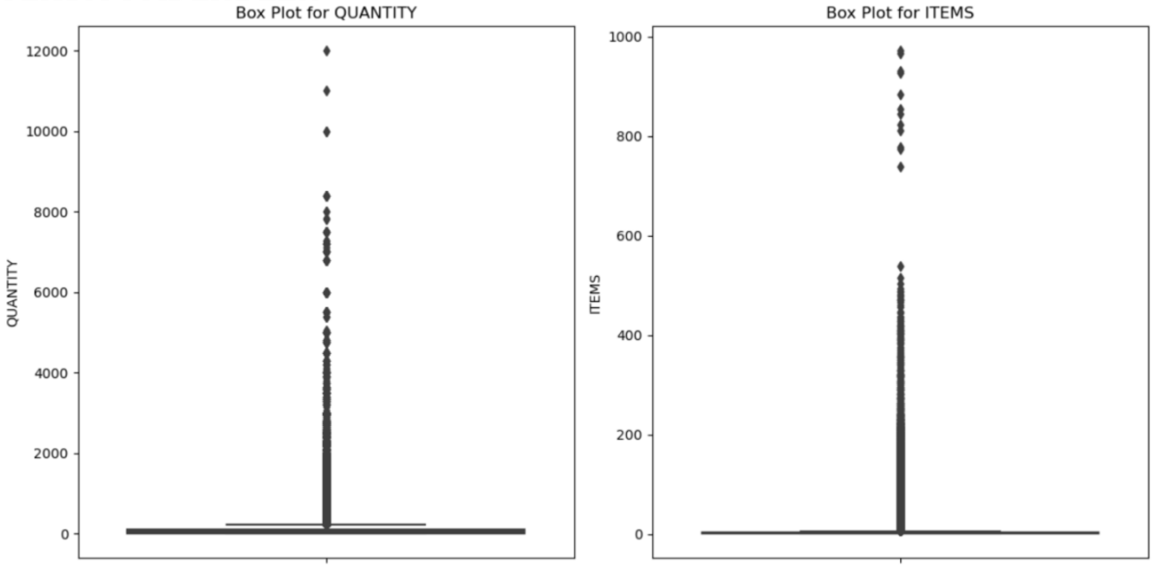

Figure 8: Box Plot for Quantity and Items (Source: Author 2023)

3.5 Modelling

According to Plotnikova, Dumas and Milani (2020), the phase of Modelling involves the development and refinement of predictive models using various machine learning techniques. Python, which is a high level, interpreted and open source programming language (Rajagopalan 2020), was chosen for this phase as it is one of the most sought after and rapidly growing programming languages in the world today and is the language of preference for data science and machine learning, according to Rajagopalan (2020).

3.5.1 Feature Selection and Encoding

In line with the nature of the dataset, the features for prediction used were the regional office, the practice and the BNF Description. In line with the widely accepted practice, as noted by Rodríguez et al. (2018), the one-hot encoding technique was employed to transform categorical features into a suitable numerical format for machine learning algorithms. This technique, still considered the most prevalent, involves converting categorical variables into binary vectors, where each category is represented by a binary value (0 or 1) as explained by Al-shehari and Alsowail (2021). this process not only retains the categorical information but also ensures compatibility with predictive algorithms. The encoding was conducted in Python utilising the pandas library, and the get_dummies function.

3.5.2 Data Preparation and Splitting

The above-mentioned data preparation steps were carried out, with January 2022 to November 2022 data used as the training dataset and December 2022 data used for testing. However, the dataset lacked explicit labels indicating fraudulent or erroneous transactions, rendering it unsupervised in nature. To address this challenge and evaluate the model effectively, pseudo data, which represents artificial data points simulating abnormal and excessive prescriptions, was introduced into the testing dataset. This pseudo data was generated by multiplying some existing values in the dataset by a factor of 10, effectively creating anomalies. The primary objective was to assess the model's capability to identify these 'pseudo anomalies' as potential instances of fraudulent or erroneous prescriptions. This approach allowed for the evaluation of the model's performance in the absence of labelled data, drawing inspiration from similar techniques applied in studies like Yamanishi et al. (2004).

3.5.3 Model Selection and Implementation

Multiple predictive algorithms were explored to harness the full spectrum of modelling capabilities. These algorithms spanned from traditional techniques such as Random Forest Regression, Linear Regression, Gradient Boosting Regression, Bayesian Regression, to advanced methods like Neural Networks, and Outlier Detection Methods such as Isolation Forest and Local Outlier Factor. Please refer to Appendix B for these code listings.

The primary objective across all models was to instil an understanding of the baseline behaviour for each medication category (BNF_DESCRIPTION) within individual practices, aligning with the methodology put forth by Iyengar et al. (2014). For the traditional modelling techniques (Random Forest Regression, Linear Regression, Gradient Boosting Regression, Bayesian Regression, and Neural Networks), the models embarked on calculating deviations from this baseline behaviour in new data. This calculation was based on the extent of deviation observed between the actual values of Quantity and Items in the prescriptions and the corresponding values predicted by the established baseline behaviour. Conversely, the Outlier Detection Methods, Isolation Forest, and Local Outlier Factor, assumed the role of flagging outliers within the incoming data. These were all carried out with the scikit-learn library, which is a Python module that integrates state-of-the-art machine learning algorithms for medium-scale supervised and unsupervised tasks (Pedregosa et al. 2011).

3.5.4 Normalization and Enhancement

As inspired by the work of Yamanishi et al. (2004) and Iyengar et al. (2014), a normalisation strategy was employed to transform deviation scores into a standardised range between 1 and 10 using the following min-max equation:

$$Normalized\ Score = 1 + \left(\frac{Deviation\ Score - Min\ Score}{Max\ Score - Min\ Score}\right) \times 9$$

Notably, this normalisation process was tailored to individual BNF descriptions for each practice, ensuring a contextually relevant assessment of deviations. In the context of this study, anomalies were defined as prescriptions with a normalised deviation score greater than 5 out of 10 in either the quantity or items columns. This criterion was applied to several machine learning models, including Neural Networks, Bayesian Ridge, Linear Regression, Gradient Boost Regression, and Random Forests. Please refer to Appendix C for these code listings.

For the Isolation Forest algorithm, anomalies were identified as prescriptions with a negative deviation score in either the quantity or items columns, aligning with the algorithm's principles. Conversely, when dealing with the Local Outlier Factor, the top 1% quantile was considered as anomalies. Please refer to Appendix D for these code listings.

3.6 Evaluation

In accordance with the CRISP-DM framework, Phase 5 initiates with a focus on the quality perspective, ascertaining the alignment of the model(s) with the predefined business objectives (Plotnikova, Dumas and Milani 2020). In this context, the model was evaluated on its ability to flag the introduced pseudo data as outliers, representing illegitimate transactions, as inspired by Yamanishi et al. (2004).

3.6.1 Confusion Matrix

The evaluation process hinges notably on the application of the confusion matrix and related accuracy metrics, drawing inspiration from various scholarly sources (Hasnain et al. 2020; Lovell et al. 2021). The confusion matrix serves as a fundamental tool in the evaluation of classification models, particularly for supervised learning tasks (Hasnain et al. 2020). It provides a structured breakdown of predictions, effectively categorising instances into four key groups, as explained by (Chicco, Tötsch and Jurman 2021):

True Positives (TP): These represent cases where the model correctly identifies positive instances, among the positive data instances. These are also called “hits”, according to Larner (2021). In the context of this study, these are the cases where the model is able to correctly identify the pseudo data.

True Negatives (TN): On the other side, the negative elements that are correctly labelled negative are called true negatives. According to Larner (2021), these are also called “non-events” or “correct rejections”. These signify the model’s proficiency in correctly recognising legitimate prescriptions without the pseudo data (anomalies).

False Positives (FP): Those which are wrongly predicted as positives are called false positives (FP). This category accounts for instances where the model incorrectly classifies genuine prescriptions as fraudulent or erroneous. They are often referred to as "false alarms" or “false hits” by Larner (2021).

False Negatives (FN): Those wrongly classified as negative are labelled false negatives (FN). FN instances signify situations in which the model inaccurately classifies fraudulent or erroneous prescriptions as legitimate, signifying "missed" detections.

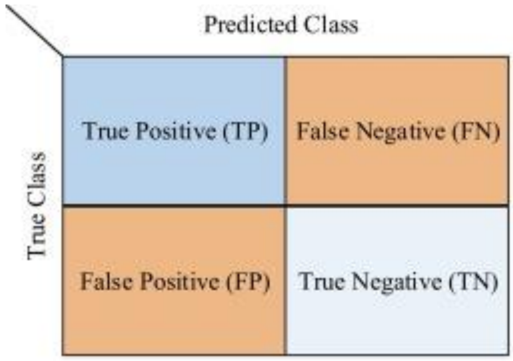

Figure 9: Confusion Matrix (Source: Demir 2022)

The aforementioned components of the confusion matrix form the basis for evaluating the model's performance. Various performance metrics, including accuracy, precision, recall, and the F1 score, are derived from these matrix values (Lovell et al. 2021). However, for the purpose of this thesis, the primary focus will be on accuracy.

3.6.2 Performance Metrics

Accuracy, as a metric for evaluating classification models, gauges the overall correctness of predictions by considering the ratio of correctly classified instances (TP and TN) to the total number of predictions made (TP, TN, FP, and FN) (Hasnain et al. 2020). Mathematically, it is represented as:

$$Accuracy = \frac{Total\ no.of\ corrected\ preditions}{Total\ no.of\ predictions}$$

OR:

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN}$$

Source: Hasnain et al. (2020)

Precision, according to Goutte and Gaussier (2005), can be defined as the probability that an object is relevant given that it is returned by the model. Tatbul et al. (2018) further explain precision as the fraction of all detected anomalies that are real anomalies. Mathematically, it is represented as:

$$Precision = \frac{TP}{TP + FP}$$

Source: Goutte and Gaussier (2005)

According to Tatbul et al. (2018), recall can be defined as the fraction of all real anomalies that are successfully detected. It is the probability that a relevant object is returned by the model (Goutte and Gaussier 2005). Mathematically, it is represented as:

$$Recall = \frac{TP}{TP + FN}$$

Source: Goutte and Gaussier (2005)

Larner (2021) defines F1 score as the harmonic mean of precision and recall. Mathematically, it is represented as:

$$F1\ Score = 2 * \frac{precision * recall}{precision + recall}$$

Source: Goutte and Gaussier (2005)

In the subsequent chapter dedicated to the results, discussion, and findings, these metrics will be meticulously computed and scrutinised, shedding light on the models' performance in the task of anomaly detection within the prescription data. The application of the confusion matrix and performance metrics, in accordance with the CRISP-DM framework, serves as a robust methodology for systematically assessing the model's alignment with the research objectives and its effectiveness in addressing the challenges of fraud and error detection in healthcare prescriptions. As the main objective is to assist with the audit process of prescriptions, the recall metric, which is a measure of all successfully detected anomalies, will be of paramount importance to auditors. However, all performance metrics will be outlined.

Chapter Four – Results, Discussion and Findings

4.1 Introduction

This section presents the outcomes of the model evaluation, along with an extensive discussion and key findings. In alignment with the evaluation methodology detailed in Section 3.5.2, synthetic data, referred to as "pseudo data," was incorporated into the testing dataset to assess the models' ability to detect anomalies. To introduce pseudo data into the testing phase, Python's sample function was employed. This approach randomly selected sixty-nine distinct medical practices from the dataset ensuring each of the nine regions was represented (please refer to Appendix A for this code listing). Approximately one hundred and twenty-three pseudo data entries were created. These pseudo data entries accounted for roughly 1% of the entire testing dataset, inspired by the methodology employed by Yamanishi et al. (2004), which identified a similar low proportion of suspect records in a study done in Australia. Also, according to Bolton and Hand (2002), there are typically many legitimate records for each fraudulent one which also justified the low percentage of introduced anomalies.

4.2 Results

The confusion matrixes of the machine learning models employed for fraud and error detection within the NHS prescription dataset are presented in Figure 11 to Figure 17. Each model's performance is evaluated based on the various metrics outlined in the evaluation subsection, including True Positives, False Positives, True Negatives, False Negatives, Accuracy, Precision, Recall, and F1 Score, which is presented in Table 4.

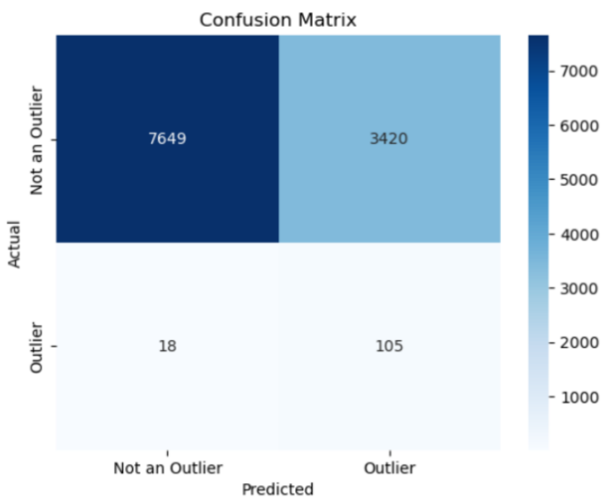

Figure 10: Random Forest Confusion Matrix (Source: Author 2023)

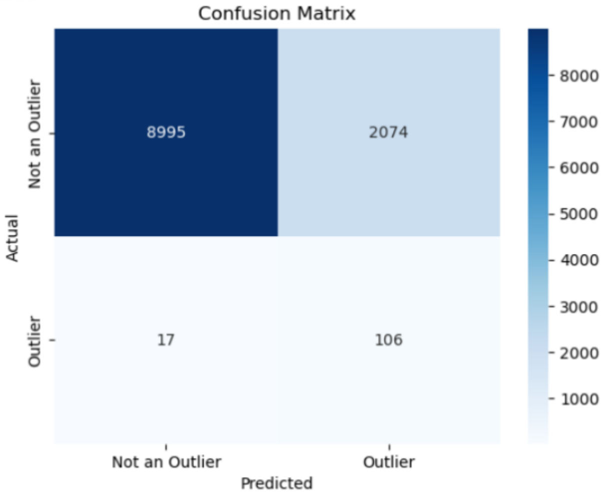

Figure 11: Bayesian Ridge Confusion Matrix (Source: Author 2023)

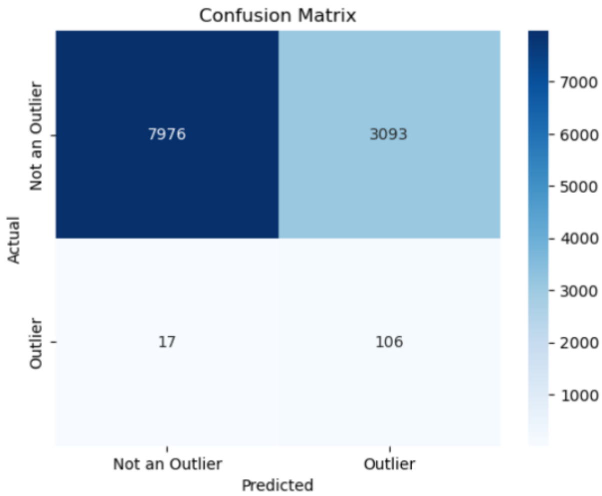

Figure 12: Gradient Boosting Regression Confusion Matrix (Source: Author 2023)

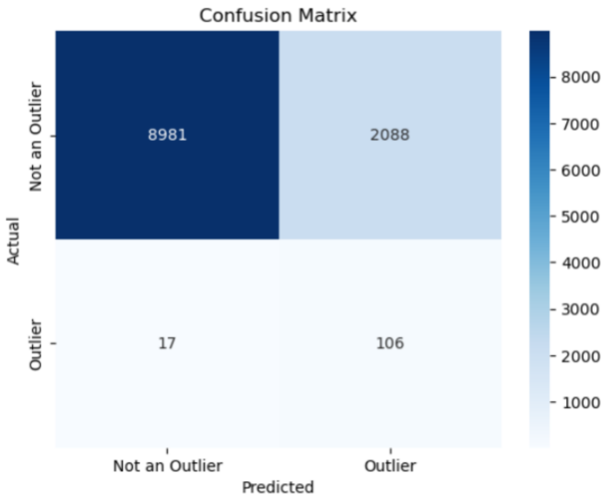

Figure 13: Linear Regression Confusion Matrix (Source: Author 2023)

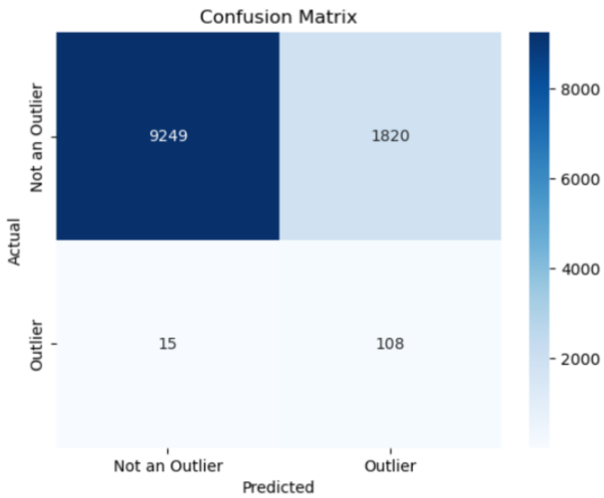

Figure 14: Neural Network Confusion Matrix (Source: Author 2023)

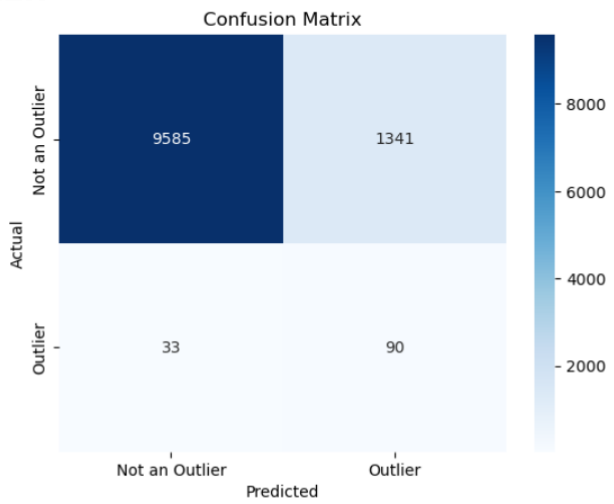

Figure 15: Isolation Forest Confusion Matrix (Source: Author 2023)

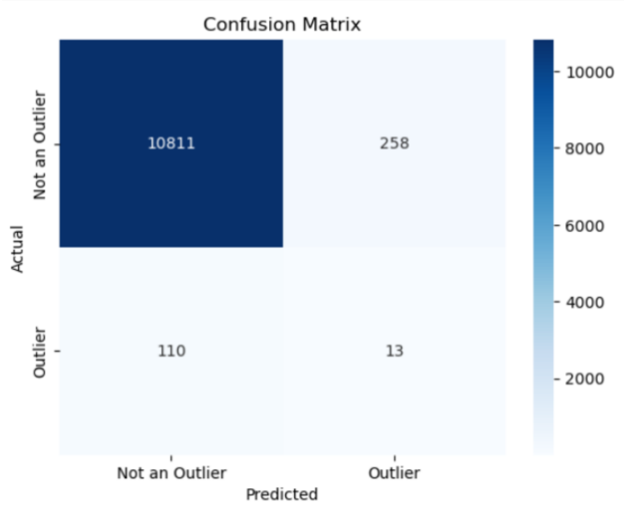

Figure 16: Local Outlier Factor Confusion Matrix (Source: Author 2023)

Table 4: Model Results (Source: Author 2023)

Model	True Positives	False Positives	True Negatives	False Negatives	Accuracy	Precision	Recall	F1 Score
Random Forest	105	3420	7649	18	0.69	0.03	0.85	0.06
Bayesian Ridge	106	2074	8995	17	0.81	0.05	0.86	0.09

Gradient Boosting Regression	106	3093	7976	17	0.72	0.03	0.86	0.06
Linear Regression	106	2081	8988	17	0.81	0.05	0.86	0.09
Neural Network	108	1820	9249	15	0.84	0.06	0.88	0.11
Isolation Forest	90	1341	9585	33	0.88	0.06	0.73	0.12
Local Outlier Factor	13	258	10811	110	0.97	0.05	0.11	0.07

4.3 Discussion and Findings

This section presents a comprehensive discussion of the results obtained from the various models employed in the analysis. The performance of individual models is compared and the effectiveness of outlier detection methods, specifically Isolation Forest and Local Outlier Factor are evaluated, in contrast to other traditional techniques. It is worth noting that this discussion is in line with the use of these models with the NHS dataset and is therefore in line with the nature of the particular dataset as described Section 3.4.

Accuracy: As previously defined, accuracy measures the proportion of correctly classified instances. The models achieved a range of accuracy scores, with the outlier detection methods, specifically Local Outlier Factor and Isolation Forest, displaying the highest accuracy rates (0.97 and 0.88, respectively). Among the supervised methods, Neural Network exhibited the highest accuracy (0.84), closely followed by Bayesian Ridge (0.81) and Linear Regression (0.81). These models demonstrated a robust ability to correctly classify data instances. In contrast, Gradient Boosting Regression and Random Forests exhibited comparatively lower accuracy rates, at 0.72 and 0.69, respectively.

Given the absence of existing literature on fraud or error prevention in the NHS using machine learning at the time of writing this report, benchmarking these results against similar work presented a challenge. However, when compared to the study conducted by Liou, Tang and Chen (2008), which investigated hospital fraud and claim abuse through diabetic outpatient services in Taiwan and used accuracy as a measure of success, these models achieved relatively lower accuracy rates. It's essential to note that the order of accuracy performance also varied significantly between the two studies. In Liou, Tang and Chen (2008), the classification tree had the highest accuracy of 99.37%, followed by the neural network with 95.73%, and regression with 92.18%. In contrast, the order of model performance in terms of accuracy in this study was as follows: Neural Network (84%), followed by Regression (81%), and Random Forests (69%).

Furthermore, it's worth noting that the study by Liou, Tang and Chen (2008) involved a more extensive set of variables for investigation and utilised a labelled dataset, whereas this study incorporated pseudo data for evaluation. These distinctions could contribute to variations in the observed accuracy levels and the order of model performance.

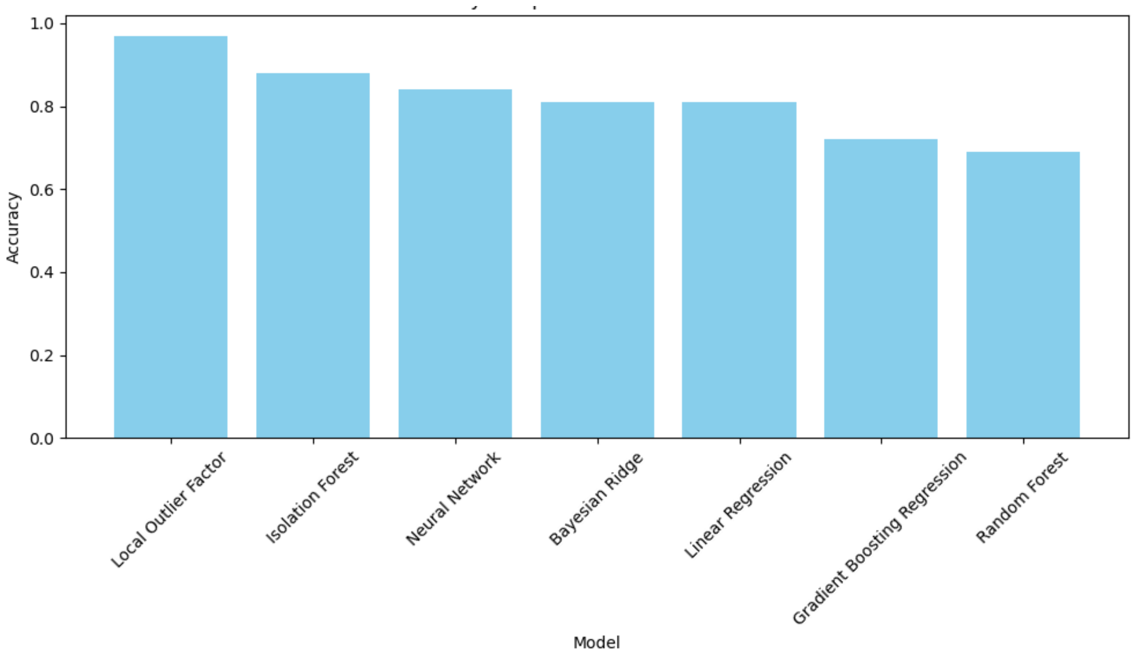

Figure 17: Accuracy Comparison of Different Models (Source: Author 2023)

Precision: Precision measures the ratio of true positives to the total instances predicted as positive, offering insights into a model's ability to minimize false positives. Neural Network displayed the highest precision score (0.06), closely followed by Linear Regression and Bayesian Ridge. Notably, the precision scores observed here are relatively low, as they should ideally approach 1 to be considered indicative of strong performance, as illustrated in the work of Mohammed, Rawashdeh, and Abdullah (2020).

This issue of low precision values aligns with the explanation provided by Lovell et al. (2021), which highlights that performance metrics reliant on class distribution, including precision, can pose challenges when dealing with imbalanced or skewed datasets. In the dataset used, where pseudo data was introduced for evaluation purposes, comprising approximately 1% of the total data, this issue becomes particularly pronounced. Consequently, the models, except for Local Outlier Factor, exhibited a relatively high number of false positives, ranging from 1,341 to 3,420. In practice, this may lead to the detection of more potential fraudulent or erroneous transactions than actually exist.

It is essential to note that the dataset's outliers were not removed during the model development process, and it's plausible that additional outliers, distinct from those intentionally introduced, will exist. As one of the recommendations for future research, outlined in the following chapter, it is advised that all flagged anomalies be reviewed by fraud experts to determine the likelihood of being fraudulent or erroneous. Such a review process could potentially lead to a reassessment of the models' precision scores, providing a more accurate reflection of their performance.

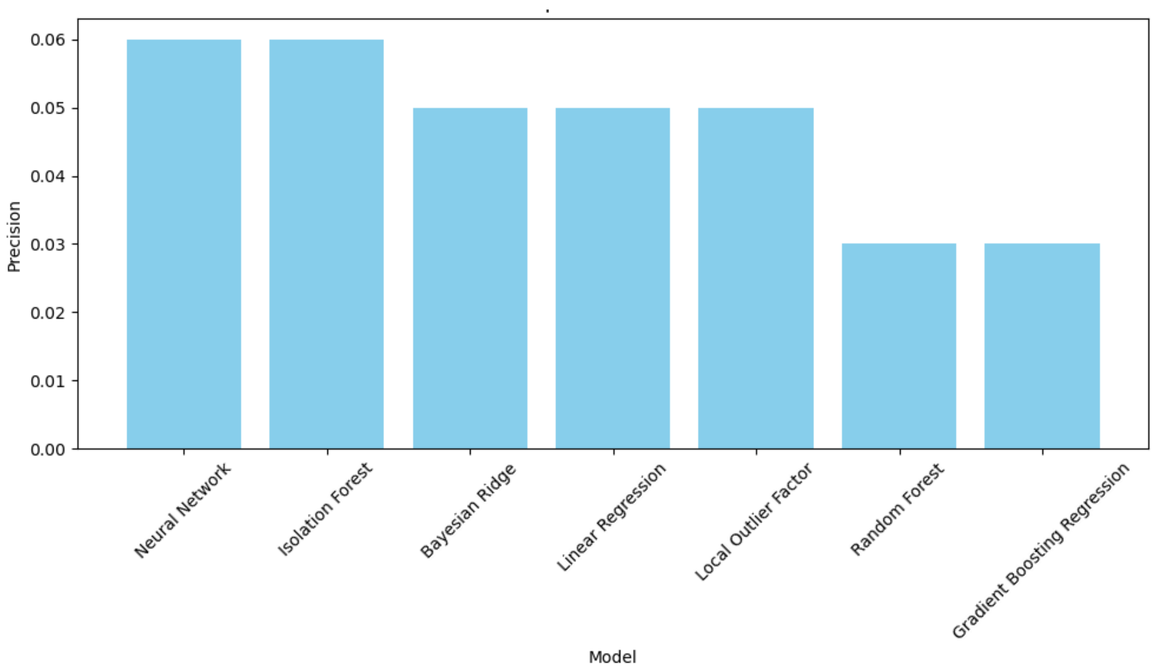

Figure 18: Precision Comparison of Different Models (Source: Author 2023)

Recall: Recall, also known as sensitivity or the true positive rate, assesses a model's proficiency in identifying genuine anomalies among all actual anomalies. Neural Network demonstrated superior recall performance, achieving a score of 0.88, underscoring its effectiveness in detecting a substantial portion of anomalies. Bayesian Ridge, Linear Regression, and Gradient Boosting Regression also exhibited robust recall scores, each registering 0.86, signifying their ability to identify anomalies effectively.

It's worth noting that these recall scores significantly surpassed those of traditional outlier detection methods, such as Isolation Forest and Local Outlier Factor, which yielded scores of 0.73 and 0.11, respectively. In essence, the recall metric measures the models' capacity to pinpoint genuine anomalous transactions, whether they are fraudulent or erroneous. However, it's essential to highlight that, in the literature reviewed, the emphasis has primarily favored accuracy as the preferred metric for evaluating model performance over recall.

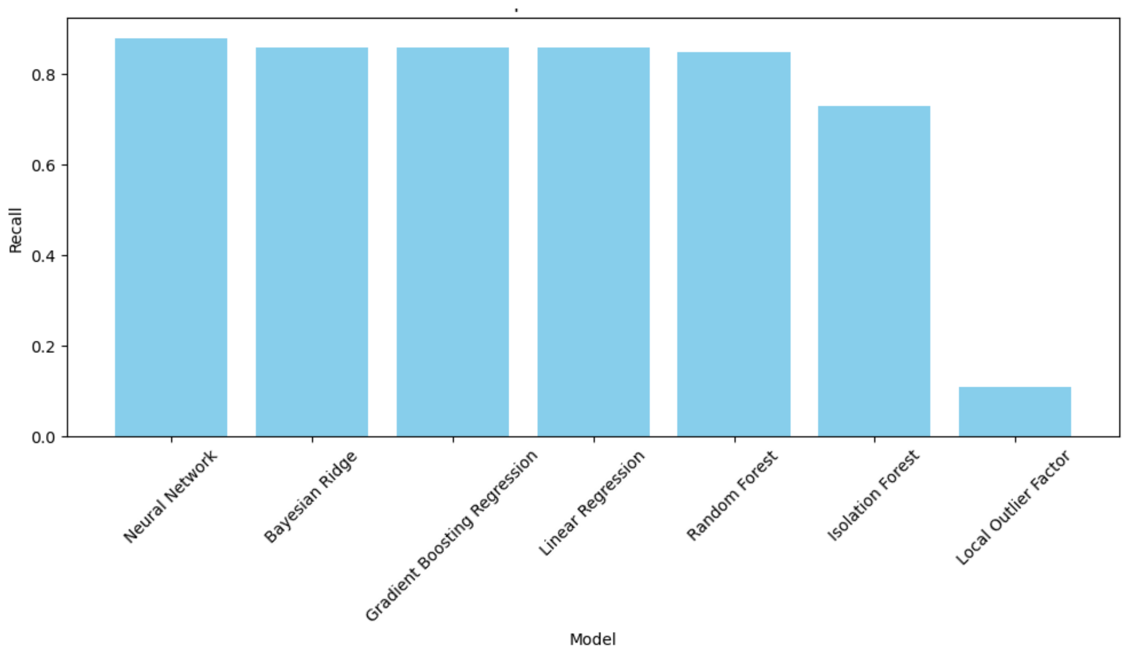

Figure 19: Recall Comparison of Different Models (Source: Author 2023)

F1 Score: The F1 Score is the harmonic mean of precision and recall. Isolation Forest exhibited the highest F1 Score (0.12), closely followed by Neural Network (0.11). However, these values are all relatively low as precision, which the F1 Score is highly dependent on, also had low values.

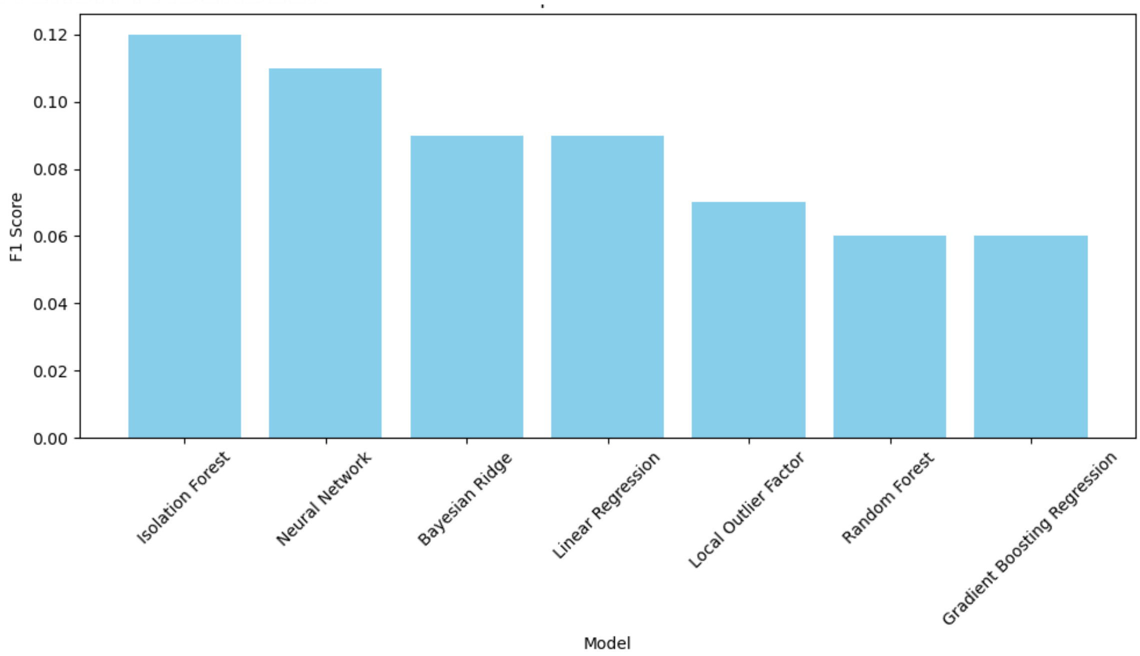

Figure 20: F1 Score Comparison of Different Models (Source: Author 2023)

Model Summaries:

Random Forest: Despite achieving a moderate accuracy of 0.69, the Random Forest model suffered from a low precision score of 0.03, indicating a high false positive rate, with 3,420 false positives. However, its recall (0.85) suggests it effectively identified most true anomalies, with 105 out of 123 identified. This aligns with the strength of decision trees in handling missing values, as discussed by Li et al. (2008) which may have led to its relatively lower performance compared to other models, as there were no missing records in the dataset.

Bayesian Ridge: Demonstrated a slightly better precision (0.05), although also low as with all models with a better accuracy of 0.81 while also maintaining a balanced recall (0.86), making it a reliable choice for anomaly detection. To the best of available knowledge, this use of Bayesian Ridge was not discovered in the literature for prescription fraud detection.

Gradient Boosting Regression: Achieved a much lower accuracy (0.72) relative to Linear Regression and only better than Random Forest. With a recall of 0.86 however, this was tied in second best together with Linear Regression and Bayesian Ridge. With a high false positive rate in a region similar to Random Forest, the ability of its use in fraud detection in this context is questionable.

Linear Regression: With a high accuracy of 0.81 and the same recall as Gradient Boosting Regression (0.86), this makes Linear Regression a better choice at detecting fraud and errors in the dataset. This is particularly true as it had a much lower false positive of 2,081 compared to 3,093 of the Gradient Boosting Regression.

Neural Network: With the exception of the traditional outlier detection methods (LOF and Isolation Forest), Neural Network stood out with the highest accuracy (0.84), overall highest precision (0.06), highest recall (0.88), and highest true positives (108), showcasing significant potential for identifying anomalies effectively. As discussed in the literature review by Li et al. (2008), neural networks are renowned for their ability to handle complex data structures and capture non-linear relationships, which correlates to the case here where it can be argued that the relationships between the features and variables were not necessarily linear.

Isolation Forest: Achieved a respectable accuracy (0.88) but with a relatively low recall of 0.73. When compared with Local Outlier Factor, it tallies with what was discovered in literature, as mentioned by Liu, Kai and Zhi-Hua Zhou (2008), as Isolation Forest being particularly well-suited for extensive databases with its remarkable capability to handle large-volume databases.

Local Outlier Factor: Excelled in accuracy (0.97), but had a much lower recall (0.11), suggesting it minimised false positives (the best at 258) but equally missed many true anomalies (the worst at 13). The high score in accuracy was as a result of the model's high score in the true negatives (10,811, the highest). However, in the context of prescription fraud and error detection, what Larner (2021) considers as the "hits" (true positives) are of paramount importance as these signify the actual prescriptions that should be audited, as opposed an excel in the "correct rejections" at the expense of true positives.

Recommended Model – Neural Network

The choice of the Neural Network model as the most promising candidate for detecting anomalies in the dataset stems from its exceptional performance across multiple metrics. Neural Networks have gained prominence in the realm of fraud detection due to their ability to handle complex data structures and capture non-linear relationships, as noted by Li et al. (2008). This inherent capacity aligns well with the nature of healthcare prescription data, which often exhibits intricate interdependencies between various variables as seen in this dataset.

The Neural Network model exhibited a remarkable accuracy score of 0.84, surpassing other algorithms in correctly classifying instances. Furthermore, it achieved a relatively better precision of 0.06. However, what truly sets the Neural Network apart is its outstanding recall score of 0.88. This metric underscores the model's ability to detect a high proportion of actual anomalies within the dataset. In practical terms, this means that the Neural Network excels in identifying potentially fraudulent or erroneous prescriptions, minimising the risk of missing critical cases.

Considering the delicate nature of healthcare fraud and error detection, where identifying true anomalies is of paramount importance, the Neural Network's superior recall rate is a decisive factor in its selection. It signifies that the model is adept at recognising instances that genuinely require further investigation, reducing the burden on auditors and increasing the efficiency of the detection process.

Subsequent to the evaluation with pseudo data, the final phase involved employing the Neural Network algorithm to analyse the primary dataset without the incorporation of pseudo data. Following the model's training on prescription data spanning from January to November 2022, the prescriptions from December 2022 were utilised as the testing dataset, mirroring the previous methodology.

It is worth noting that the anomaly detection threshold greatly impacts the percentage of anomalies identified. Furthermore, users have the flexibility to focus on specific anomaly types (such as anomalies in Items and/or Quantity and/or values which exhibited a more significant increase). For example, setting the deviation score threshold to a score of 5 in either Quantity or Items columns within a randomly sampled subset of ten practices comprising 1,566 records, resulted in 669 anomalies, constituting approximately 43% of the total records. However, when considering only records with actual values higher than what the model predicts as baseline behaviour, the anomaly rate drops to approximately 18% with a deviation score of 5 and further reduces to around 10% with a deviation score of 10 in either Quantity or Items columns. When a deviation score of 10 is considered in both Quantity and Items columns, the anomaly rate drops further to 0.25%.

To visualise the patterns with a deviation score of 10, Figure 22 and Figure 23 were generated below. These represent the behaviour of these medications (BNF Descriptions) within the practices where the anomalies were detected. Figure 22 reveals a fluctuating trend in the average quantity with December witnessing a notable upswing, marking the highest quantity since March. This trend is further highlighted in Figure 23, which focuses on Items values. Here, a consistent decline is observed in item values from March onwards. Strikingly, December saw an abrupt surge in Items values to the highest in all twelve months. This observation suggests that the detected anomalies had a stronger presence in the Items column than in the Quantity column.

As demonstrated above, the model's performance is highly adaptable to specific needs, offering the flexibility to fine-tune parameters such as the threshold and anomaly type criteria. This adaptability empowers domain experts to customise the model according to their precise requirements, thus optimising its effectiveness in identifying fraudulent or erroneous prescriptions. The varying percentages of identified anomalies underscore the model's versatility, as it accommodates different user-defined thresholds. This emphasises the crucial role of expert guidance in configuring anomaly detection parameters for practical deployment. It is essential to note that not all identified anomalies necessarily indicate erroneous or fraudulent prescriptions. Instead, they signify deviations from established dataset patterns. Therefore, further analysis and expert evaluation are vital to discern the precise nature of these anomalies and determine whether they warrant in-depth investigation.

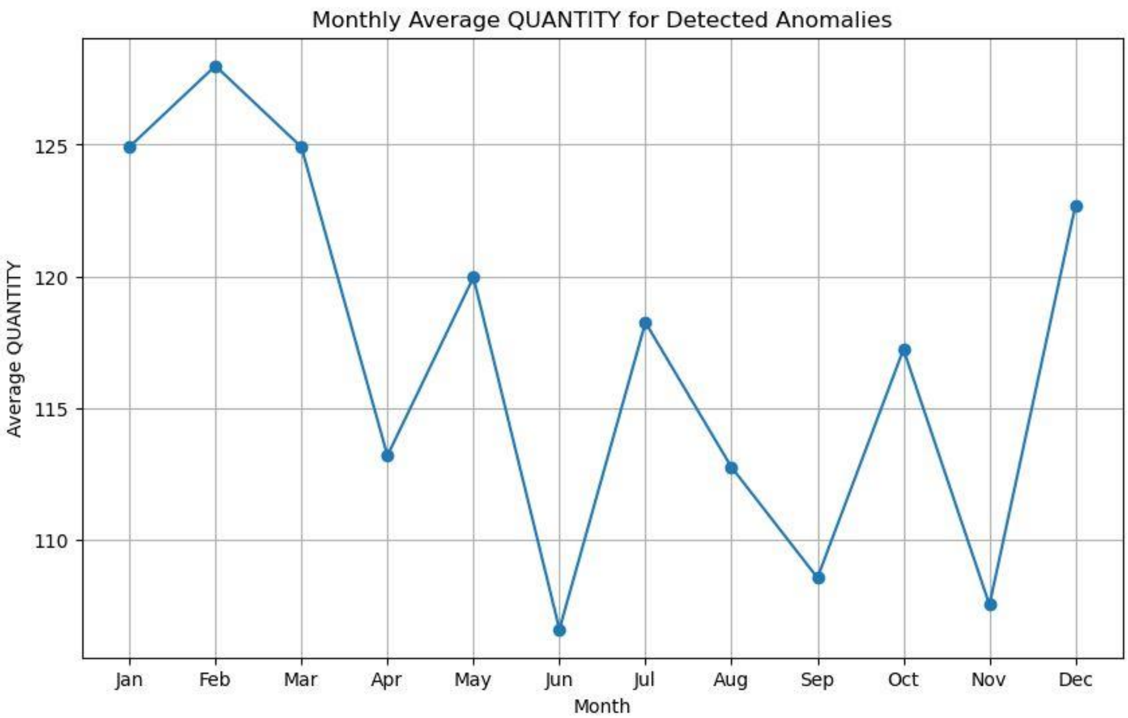

Figure 21: Monthly Average Quantity for Detected Anomalies (Source: Author 2023)

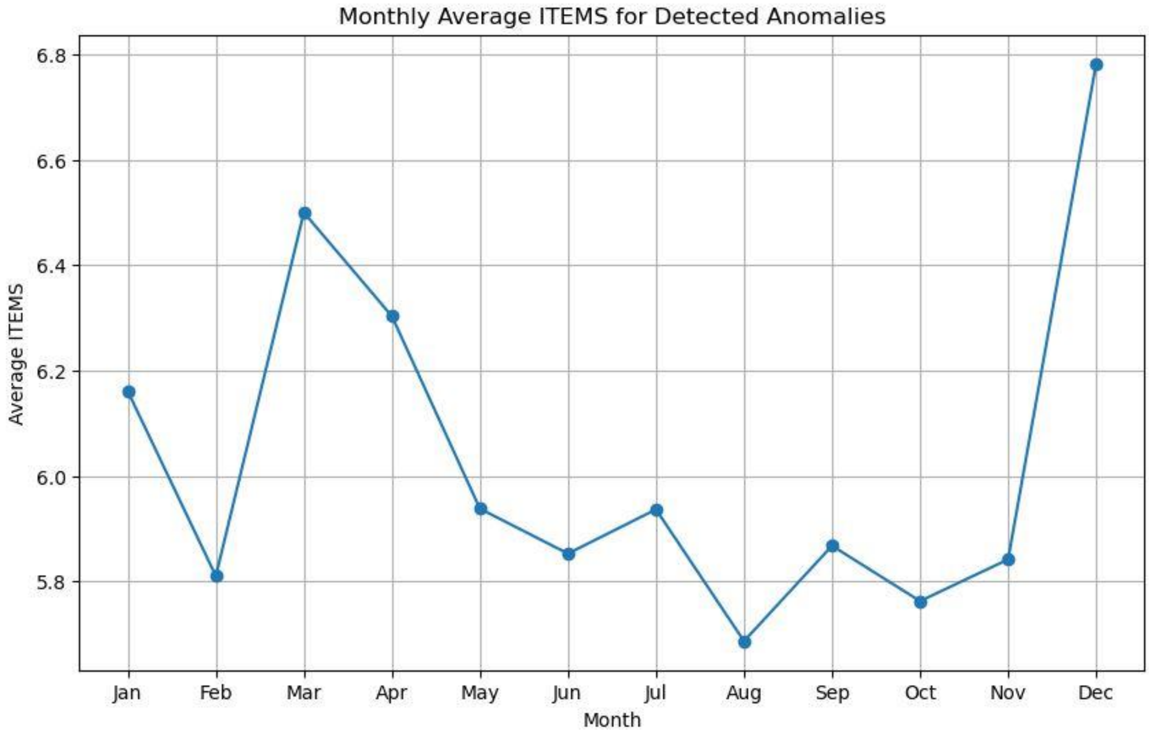

Figure 22: Monthly Average Items for Detected Anomalies (Source: Author 2023)

Chapter Five – Conclusion, Limitations and Future Work

In this study, a comprehensive exploration and development of a machine learning model for the detection of anomalies within NHS prescriptions were undertaken. The study aligned with the rigorous CRISP-DM methodology, focusing on specific methods and models to contribute to enhanced fraud detection and error prevention within the NHS prescription system.

5.1 Realization of Objectives

Objective 1: Identifying Key Patterns and Indicators

The initial objective entailed the identification of essential patterns and indicators related to fraudulent prescriptions within the healthcare domain. This investigation, driven by extensive research, unveiled that prescription drug fraud and abuse frequently cantered around specific drug categories, as outlined by Iyengar, Hermiz, and Natarajan (2014). These categories consisted of high-volume drugs, which could be resold to pharmacies and potentially billed twice to health plans, and drugs with a high street value due to non-medical and recreational misuse. The project focused notably on the latter group of drugs, renowned for their susceptibility to fraudulent activities, and these details were comprehensively presented in Table 3.

Objective 2 & 3: Selecting Suitable Machine Learning Algorithms and Developing a Predictive Model

Objectives 2 and 3 concentrated on the exploration and selection of appropriate machine learning algorithms for anomaly detection within healthcare data, specifically in the NHS context. Extensive evaluation led to the determination that Neural Networks proved to be the most effective algorithm for detecting anomalies within the available prescription dataset. This phase encompassed meticulous data processing, encoding, normalisation, and comprehensive evaluation, as thoroughly detailed in Chapters 3 and 4, respectively, under the headings "Methodology" and "Results, Discussion, and Findings."

Objective 4: Evaluating Model Performance

Objective 4 was dedicated to the rigorous assessment of the model's performance. This evaluation process was explained in the fourth chapter, where the metrics of accuracy, precision, recall, and the F1 score for all algorithms were outlined and discussed. In the context of the NHS prescription dataset, being the pioneering study to employ machine learning for the detection of fraudulent and erroneous prescriptions, the findings were compared with similar studies conducted in other countries. This initiative not only added depth to the research but also laid the groundwork for prospective investigations within the realm of NHS prescription data analysis.

Objective 5: Providing Recommendations for Implementation/Deployment

In adherence to the CRISP-DM process, the fifth objective corresponds to the deployment phase, entailing the practical application of the developed models to support decision-making processes. This step aligns with the insights of Plotnikova, Dumas, and Milani (2020), as it aims to make these models accessible to end-users for informed decision-making within various healthcare processes.

Considering the specific context of fraud and errors in NHS prescriptions, it is proposed that the model be employed as a fraud detection system. This is due to the necessity of monthly data aggregation before subsequent analysis and evaluation can occur. As a prospective avenue of research, the model's potential application in fraud prevention can also be explored, aligning with the two critical stages outlined by Bolton and Hand (2002) - fraud prevention and fraud detection. Fraud prevention involves measures to halt fraud before it transpires, while fraud detection identifies fraud swiftly after its occurrence.

5.2 Limitations and Future Work

This study encountered certain limitations that bear consideration. Firstly, one notable constraint was the restricted access to comprehensive data for analysis and modelling. As elucidated by Kumaraswamy et al. (2022), stringent data privacy and legal regulations pose significant challenges when working with private healthcare data. Consequently, this research primarily relied on the volume of prescription items as the primary metric for fraud and error detection within prescriptions. This approach aligns with the rationale outlined by Iyengar, Hermiz, and Natarajan (2014). Nevertheless, it's imperative to acknowledge that other studies have explored alternative metrics such as visit length, patient retention rates, visit frequency, percentages of reduplicative patients, patients-pharmacy ratios, the proportion of claims referred to high-cost pharmacies, and patient travel distances (examples are Thornton et al., 2014; Joudaki et al., 2016; Musal, 2010). These dimensions offer promising avenues for future research, contingent upon overcoming the data access barriers.

Another promising avenue for future inquiry lies in conducting interviews with fraud experts and medical professionals to gain deeper insights into all the anomalies detected by the models. This approach aligns with the methodologies employed by Thornton et al. (2014) and Aral et al. (2012). It also addresses a limitation of this study, where potential erroneous and fraudulent transactions were not explicitly categorised. Such interviews have the potential to provide a more nuanced understanding of the flagged prescriptions. By engaging with experts in the field, it may be possible to distinguish between legitimate errors and deliberate fraudulent activities more effectively.

In conclusion, this research has pioneered the integration of data mining techniques for the detection of fraudulent and erroneous prescription transactions within the NHS. As a pioneering study, it serves as a foundational stepping stone for future investigations in this critical domain. The collective insights gathered from this research, along with subsequent studies, hold the promise of yielding tangible cost savings within the NHS and enhancing the overall standard of patient care.

Acknowledgements

I am profoundly grateful to all those whose unwavering support, both personal and academic, has illuminated my path throughout this academic journey. Your boundless encouragement, tireless assistance, and unwavering understanding have been the driving force behind the successful completion of this dissertation. My heartfelt thank you for being the guiding stars in this voyage of knowledge and self-discovery.

References

NHS prescription charges. 2015. Drug and therapeutics bulletin, 53(4), pp. 45-48.

Man Convicted for Health Care Fraud and Prescription Drug Diversion Scheme. 2022. Washington: Federal Information & News Dispatch, LLC.

AL-SHEHARI, T. and ALSOWAIL, R.A., 2021. An insider data leakage detection using one-hot encoding, synthetic minority oversampling and machine learning techniques. Entropy (Basel, Switzerland), 23(10), pp. 1258.

ARAL, K.D. et al., 2012. A prescription fraud detection model. Computer methods and programs in biomedicine; Comput Methods Programs Biomed, 106(1), pp. 37-46.

ARONSON, J.K., 2009. Medication errors: definitions and classification. British journal of clinical pharmacology, 67(6), pp. 599-604.

ASSOCIATION OF CERTIFIED FRAUD EXAMINERS, 2022. Anti-Fraud Technology Benchmarking Report. [online]. Available from: https://www.acfe.com/fraud-resources/anti-fraud-technology-benchmarking-report [Accessed June 25, 2023].

ASSOCIATION OF CERTIFIED FRAUD EXAMINERS, 2023. Fraud 101: What Is Fraud? [online]. Available from: https://www.acfe.com/fraud-resources/fraud-101-what-is-fraud [Accessed June 17, 2023].

BECKER, R.A., VOLINSKY, C. and WILKS, A.R., 2010. Fraud Detection in Telecommunications: History and Lessons Learned. Technometrics, 52(1), pp. 20-33.

BHATTACHARYYA, S. et al., 2011. Data mining for credit card fraud: A comparative study. Decision Support Systems, 50(3), pp. 602-613.

BOLTON, R.J. and HAND, D.J., 2002. Statistical Fraud Detection: A Review. Statistical science, 17(3), pp. 235-249.

BORLAND SOPHIE, 2011. From today, only the English have to pay for ... prescriptions; (and no, that's not an April Fool's joke!). Daily mail (London, England), Apr 1, pp. 33.

BOUKERCHE, A., ZHENG, L. and ALFANDI, O., 2020. Outlier Detection: Methods, Models, and Classification. ACM computing surveys, 53(3), pp. 1-37.

BREUNIQ, M.M. et al., 2000. LOF: Identifying density-based local outliers. SIGMOD record, 29(2), pp. 93-104.

BROWNLEE JASON, 2021. Random Oversampling and Undersampling for Imbalanced Classification. [online]. Available from: https://machinelearningmastery.com/random-oversampling-and-undersampling-for-imbalanced-classification/ [Accessed August 5, 2023].

BRUNELLI, M., 2014. Introduction to the analytic hierarchy process. Cham: Springer.

CAMBRIDGE DICTIONARY, 2023. Meaning of error – Learner's Dictionary. [online]. Available from: https://dictionary.cambridge.org/dictionary/learner-english/error [Accessed June 18, 2023].

CARCILLO, F. et al., 2021. Combining unsupervised and supervised learning in credit card fraud detection. Information Sciences, 557, pp. 317-331.

CHAKRABORTY, S., ISLAM, S.H. and SAMANTA, D., 2022. Data classification and incremental clustering in data mining and machine learning. Cham, Switzerland: Springer.

CHICCO, D., TÖTSCH, N. and JURMAN, G., 2021. The matthews correlation coefficient (Mcc) is more reliable than balanced accuracy, bookmaker informedness, and markedness in two-class confusion matrix evaluation. BioData mining; BioData Min, 14(1), pp. 1-22.

COHEN BRANDON, LEIGH J. RUTH and PREUSS V. CHARLES, 2023. Opioid Analgesics. [online]. Available from: https://www.ncbi.nlm.nih.gov/books/NBK459161/#_ncbi_dlg_citbx_NBK459161 [Accessed July 20, 2023].

DEAN, B., BARBER, N. and SCHACHTER, M., 2000. What is a prescribing error? Quality in Health Care; Qual Health Care, 9(4), pp. 232-237.

DEMIR, F., 2022. 14 - Deep autoencoder-based automated brain tumor detection from MRI data. In: V. BAJAJ and G.R. SINHA, eds. Artificial Intelligence-Based Brain-Computer Interface. Academic Press. pp. 317-351.

EKIN, T. et al., 2018. Statistical Medical Fraud Assessment: Exposition to an Emerging Field. International statistical review, 86(3), pp. 379-402.

GEE, J. and BUTTON, M., The Financial Cost of Healthcare Fraud 2015:What Data from Around the World Shows. PKF Littlejohn.

GOUTTE, C. and GAUSSIER, E., 2005. A probabilistic interpretation of precision, recall and F-score, with implication for evaluation. Proceedings of the LECT NOTES COMPUT SC2005. Berlin, Heidelberg: Springer Berlin Heidelberg, pp. 345-359.

GRIFFITH, R., 2019. Fraud in the NHS. British journal of nursing (Mark Allen Publishing); Br J Nurs, 28(19), pp. 1268-1269.

GRIFFITH, R. and TENGNAH, C., 2011. Looking at the extent of fraud within the NHS. British journal of community nursing; Br J Community Nurs, 16(10), pp. 507-510.

HADDAD SOLEYMANI, M. et al., 2018. Detecting medical prescriptions suspected of fraud using an unsupervised data mining algorithm. Daru; Daru, 26(2), pp. 209-214.

HAN, J., KAMBER, M. and PEI, J., 2011. Data mining: Concepts and techniques. 3rd ed. Burlington: Elsevier Science.

HAN, J., PEI, J. and TONG, H., Data Mining Concepts and Techniques (4th Edition). Elsevier.

HAND, D.J., 2007. Principles of Data Mining. Drug safety; Drug Saf, 30(7), pp. 621-622.

HASNAIN, M. et al., 2020. Evaluating Trust Prediction and Confusion Matrix Measures for Web Services Ranking. IEEE access, 8, pp. 90847-90861.

HEMANTH, D.J. and KOSE, U., 2020. Analyzing and Processing of Supplier Database Based on the Cross-Industry Standard Process for Data Mining (CRISP-DM) Algorithm. Switzerland: Springer International Publishing AG. pp. 544-558.

IBM, 2023. What is data mining? [online]. Available from: https://www.ibm.com/topics/data-mining#:~:text=Data%20mining%2C%20also%20known%20as,information%20from%20large%20data%20sets. [Accessed July 6, 2023].

IYENGAR, V.S., HERMIZ, K.B. and NATARAJAN, R., 2014. Computer-aided auditing of prescription drug claims. Health care management science; Health Care Manag Sci, 17(3), pp. 203-214.

JIANG, K. et al., 2020. Network Intrusion Detection Combined Hybrid Sampling With Deep Hierarchical Network. IEEE access, 8, pp. 32464-32476.

JOHNSON, J.M. and KHOSHGOFTAAR, T.M., 2019. Medicare fraud detection using neural networks. Journal of Big Data, 6(1), pp. 1-35.

JOHNSON, K.B. et al., 2021. Precision Medicine, AI, and the Future of Personalized Health Care. Clinical and translational science; Clin Transl Sci, 14(1), pp. 86-93.

JOUDAKI, H. et al., 2016. Improving fraud and abuse detection in general physician claims: A data mining study. International journal of health policy and management; Int J Health Policy Manag, 5(3), pp. 165-172.

KOU, Y. et al., 2004. Survey of fraud detection techniques. Proceedings of the ICNSC2004. IEEE, pp. 749-754 Vol.2.

KUMAR, M., GHANI, R. and MEI, Z., 2010. Data mining to predict and prevent errors in health insurance claims processing. Proceedings of the International Conference on Knowledge Discovery and Data Mining2010. ACM, pp. 65-74.

KUMARASWAMY, N. et al., 2022. Healthcare Fraud Data Mining Methods: A Look Back and Look Ahead. Perspectives in health information management; Perspect Health Inf Manag, 19(1), pp. 1-18.

LARNER, A.J., 2021. The 2x2 matrix: Contingency, confusion and the metrics of binary classification. Cham: Springer.

LEANNDRA COPELAND et al., 2012. Applying Business Intelligence Concepts to Medicaid Claim Fraud Detection. Journal of Information Systems Applied Research, 5(1), pp. 51-61.

LEVI, M. and BURROWS, J., 2008. MEASURING THE IMPACT OF FRAUD IN THE UK: A Conceptual and Empirical Journey. British Journal of Criminology, 48(3), pp. 293-318.

LI, J. et al., 2008. A survey on statistical methods for health care fraud detection. Health care management science; Health Care Manag Sci, 11(3), pp. 275-287.

LIN, C. et al., 2008. Intelligent physician segmentation and management based on KDD approach. Expert Systems with Applications, 34(3), pp. 1963-1973.

LIOU, F., TANG, Y. and CHEN, J., 2008. Detecting hospital fraud and claim abuse through diabetic outpatient services. Health care management science; Health Care Manag Sci, 11(4), pp. 353-358.

LISBY, M. et al., 2010. How are medication errors defined? A systematic literature review of definitions and characteristics. International journal for quality in health care, 22(6), pp. 507-518.

LIU, F.T., KAI, M.T. and ZHI-HUA ZHOU, 2008. Isolation forest. Proceedings of the ICDM2008. IEEE, pp. 413-422.

LOVELL, D. et al., 2021. Taking the confusion out of multinomial confusion matrices and imbalanced classes. 2021. Singapore: Springer Singapore, pp. 16-30.

MANNING JULIA, 2011. Stealing the NHS?: how careless is the NHS being with our tax money? London: 2020health.

MARTÍN MUÑOZ, M., On-line support vector machines for function approximation.

MOHAMMED, R., RAWASHDEH, J. and ABDULLAH, M., 2020. Machine learning with oversampling and undersampling techniques: Overview study and experimental results. Proceedings of the ICICS2020. IEEE, pp. 243-248.

MURTY, M.N. and RAGHAVA, R., 2016. Support vector machines and perceptrons : Learning, optimization, classification, and application to social networks. Switzerland: Springer.

MUSAL, R.M., 2010. Two models to investigate Medicare fraud within unsupervised databases. Expert systems with applications, 37(12), pp. 8628-8633.

NHS, 2023a. Electronic prescriptions. [online]. Available from: https://www.nhs.uk/nhs-services/prescriptions-and-pharmacies/electronic-prescriptions/ [Accessed June 20, 2023].

NHS, 2023b. How to register with a GP surgery. [online]. Available from: https://www.nhs.uk/nhs-services/gps/how-to-register-with-a-gp-surgery/ [Accessed June 20, 2023].

NHS, 2023c. NHS services. [online]. Available from: https://www.nhs.uk/nhs-services/ [Accessed June 20, 2023].

NHS BUSINESS SERVICES AUTHORITY, 2022. Prescription Cost Analysis – England – 2021/22. [online]. Available from: https://www.nhsbsa.nhs.uk/statistical-collections/prescription-cost-analysis-england/prescription-cost-analysis-england-202122 [Accessed June 1, 2023].

NHS BUSINESS SERVICES AUTHORITY, 2023. How we check claims. [online]. Available from: https://www.nhsbsa.nhs.uk/penalty-charges-explained/how-we-check-claims [Accessed June 25, 2023].

NHS COUNTER FRAUD AUTHORITY, 2023a. What is NHS fraud? [online]. Available from: https://cfa.nhs.uk/fraud-prevention/what-is-nhs-fraud [Accessed June 17, 2023].

NHS COUNTER FRAUD AUTHORITY, 2023b. Who we are. [online]. Available from: https://cfa.nhs.uk/about-nhscfa/who-we-are [Accessed June 24, 2023].

NHS ENGLAND, 2023. NHS Data Model and Dictionary. [online]. Available from: https://www.datadictionary.nhs.uk/classes/prescription.html#:~:text=A%20request%20for%20the%20dispensing,or%20SERVICE%20to%20a%20PATIENT. [Accessed June 1, 2023].

NHS HIGHLANDS, 2023. CNS stimulants and drugs used for attention deficit hyperactivity disorder. [online]. Available from: https://tam.nhsh.scot/formularies/highland-formulary/central-nervous-system/cns-stimulants-and-drugs-used-for-attention-deficit-hyperactivity-disorder/#:~:text=CNS%20stimulant%20medication%20(methylphenidate%2C%20lisdexamfetamine,provided%20by%20the%20relevant%20specialist. [Accessed July 20, 2023].

NHS INFORM, 2023. Prescriptions. [online]. Available from: https://www.nhsinform.scot/tests-and-treatments/medicines-and-medical-aids/prescriptions/#about-prescriptions [Accessed June 20, 2023].

ORTEGA, P.A., FIGUEROA, C.J. and RUZ, G.A., 2006. A Medical Claim Fraud/Abuse Detection System based on Data Mining: A Case Study in Chile. DMIN, 6, pp. 26-29.

PATEL, A.K., CHATTERJEE, S. and GORAI, A.K., 2019. Development of a machine vision system using the support vector machine regression (SVR) algorithm for the online prediction of iron ore grades. Earth science informatics, 12(2), pp. 197-210.

PEDREGOSA, F. et al., 2011. Scikit-learn: machine learning in Python. Journal of machine learning research,

PLOTNIKOVA, V., DUMAS, M. and MILANI, F., 2020. Adaptations of data mining methodologies: A systematic literature review. PeerJ.Computer science; PeerJ Comput Sci, 6, pp. 1-43.

POPE CARMEN, 2023. Opioids (narcotic analgesics). [online]. Available from: https://www.drugs.com/drug-class/narcotic-analgesics.html# [Accessed July 20, 2023].

PUBLIC HEALTH SCOTLAND, 2023. Benzodiazepines (benzos, diazepam, valium). [online]. Available from: https://www.nhsinform.scot/healthy-living/drugs-and-drug-use/common-drugs/benzodiazepines-benzos-diazepam-valium#:~:text=Benzodiazepines%20are%20a%20group%20of,anxiety%20or%20occasionally%20sleeping%20problems. [Accessed July 20, 2023].

RAJAGOPALAN, G., 2020. A python data analyst's toolkit: Learn python and python-based libraries with applications in data analysis and statistics. Berkeley, CA: Apress L. P.

RANGA SURI, N.N.R., MURTY, M.N. and ATHITHAN, G., 2019. Outlier detection--techniques and applications: A data mining perspective. Cham, Switzerland: Springer.

RASHIDIAN, A., JOUDAKI, H. and VIAN, T., 2012. No evidence of the effect of the interventions to combat health care fraud and abuse: A systematic review of literature. PloS one; PLoS One, 7(8), pp. e41988.

RODRÍGUEZ, P. et al., 2018. Beyond one-hot encoding: Lower dimensional target embedding. Image and Vision Computing, 75, pp. 21-31.

SAATY, T.L. and VARGAS, L.G.(.G., 2012. Models, methods, concepts & applications of the analytic hierarchy process. 2nd ed. New York: Springer.

SANGHERA, I.S., FRANKLIN, B.D. and DHILLON, S., 2007. The attitudes and beliefs of healthcare professionals on the causes and reporting of medication errors in a UK Intensive care unit. Anaesthesia; Anaesthesia, 62(1), pp. 53-61.

SARKAR, D. and NATARAJAN, V., 2019. Ensemble machine learning cookbook: Over 35 practical recipes to explore ensemble machine learning techniques using python. Place not identified: Packt Publishing.

SCOTTISH POLICE AUTHORITY, 2023. Types of Drugs. [online]. Available from: https://www.spa.police.uk/forensic-services/drug-analysis/types-of-drugs/ [Accessed July 20, 2023].

SHEARD, S., 2011. A creature of its time. Michael quarterly (trykt utg.), 8(4), pp. 428-441.

SHIN, H. et al., 2012. A scoring model to detect abusive billing patterns in health insurance claims. Expert Systems with Applications, 39(8), pp. 7441-7450.

SINGH, N. et al., 2019. Data-driven auditing: A predictive modeling approach to fraud detection and classification. The Journal of Corporate Accounting & Finance, 30(4), pp. 121.

SONG, W., WANG, Y. and PAN, Z., 2023. A novel cell partition method by introducing Silhouette Coefficient for fast approximate nearest neighbor search. Information Sciences, 642, pp. 119216.

STEWART, D. et al., 2018. Perspectives of healthcare professionals in Qatar on causes of medication errors: A mixed methods study of safety culture. PloS one; PLoS One, 13(9), pp. e0204801.

TATBUL, N. et al., 2018. Precision and Recall for Time Series.

THAIFUR, A.Y.B.R. et al., 2021. How to detect healthcare fraud? “A systematic review”. Gaceta sanitaria, 35, pp. S441-S449.

THORNTON, D., POEL, M., HILLEGERSBERG, J., MUELLER, R. and CAPELLEVEEN, G.C., Outlier-based Health Insurance Fraud Detection for U.S. Medicaid Data. SCITEPRESS.

TRAVAILLE, P. et al. , Aug 4, 2011. Electronic fraud detection in the U.S. medicaid healthcare program: Lessons learned from other industries. Proceedings of the Proceedings of the 17th Americas Conference on Information Systems (AMCIS)Aug 4, 2011. Association for Information Systems, pp. 1-11.

TRIPATHI, S. et al., 2021. Ensuring the Robustness and Reliability of Data-Driven Knowledge Discovery Models in Production and Manufacturing. Frontiers in artificial intelligence, 4, pp. 576892.

UNITED KINGDOM GOVERNMENT, 2021. The NHS Constitution for England. [online]. Available from: https://www.gov.uk/government/publications/the-nhs-constitution-for-england/the-nhs-constitution-for-england [Accessed June 20, 2023].

YAMANISHI, K. et al., 2004. On-line unsupervised outlier detection using finite mixtures with discounting learning algorithms. Data mining and knowledge discovery, 8(3), pp. 275-300.