UNIVERSITY OF NAIROBI

SCHOOL OF COMPUTING AND INFORMATICS

SUSTAINABILTY OF MACHINE LEARNING IN HEALTH CLAIMS AUTOMATION IN THE KENYAN INSURANCE INDUSTRY

BY

MANYWANDA JOSEPH MAENA

REG. NUMBER -P54/35949/2019

SUPERVISOR. PROF.DANIEL ORWA

A project submitted in partial fulfilment of the requirements for the award of the degree of Master of Science in Information Technology Management of the University of Nairobi, School of Computing & Informatics

MARCH 2021

DECLARATION

The research project below is my original work and has not been presented to any other examination body.

Signed: Joseph Manywanda Date: 8/20/2021

JOSEPH MANYWANDA

P54/35949/2019

This project report has been submitted in partial fulfilment of the requirements for the award of the degree of Master of Science in Information Technology Management of the University of Nairobi with my approval as the University supervisor

Signed Date 20th August 2021

PROF: DANIEL ORWA

ACKNOWLEDGMENTS

I would like to thank the Almighty God for the grace and endurance He gave me to undertake this project. I would like to acknowledge my parents Mr. and Mrs. Manywanda, my siblings Sheila Manywanda, Gloria Manywanda and Chris Hani for their enduring support while I was working on this project.

I want to specifically recognize my supervisor Prof Daniel Orwa for his guidance and contributions while working on this project.

Finally, I would like to appreciate Carol Wanjohi, David Njoroge and Purity Mwalozi on assisting me to debug my code and guiding me to the right direction.

ABSTRACT

Background

Hospitals in Kenya have embraced the use of hospital management systems to tracking patient's data. The automation of this data has improved the efficiency of the insurance system and ultimately the settlement of healthcare claims. Claims processing is the most important function for any insurance company. The speed and convenience with which the claims are settled has a bearing on the general reputation of the insurance company.

Problem

The insurance health industry in the country faces difficulties which include increasing fraud, rising costs, and a high number of claims ratios. There are approximately 30 million claims in the country every year that must be reviewed, approved, and paid every year. The major challenge affecting this is the process is not fully automated.

Purpose

This research analyzed the various machine learning models to determine the accuracy of health claims automation.

Method

This was achieved by using, the supervised machine learning model was used to determine the accuracy. The supervised machine learning models used were K- Nearest Neighbor, Naïve Bayes and LDA- Linear Discriminant Analysis.

Findings

Experiments were conducted and K -Nearest Neighbor was the best model to determine the accuracy of claims with a 99.9% accuracy and it produced the best results with smaller and larger datasets.

Conclusion

ML is a possible solution for the challenges that are facing insurers due to lack of automation. Applying ML not only in a health insurance industry but also other insurance industries. e.g., general insurance can be used to improve the business, increase income of the insurer, and reduce costs. The timely payment of a claim to a hospital will also improve the relationships between the insurer and the hospital.

Key words: Machine learning, Naïve Bayes, K-NN, claims automation, claims industry

Table of Contents

List of Figures

List of Tables

Acronyms

NHIF – NATIONAL HEALTH INSURANCE FUND

IRA – INSURANCE REGULATORY AUTHORITY

AI- AIRTIFICIAL INTELLIGENCE

ML – MACHINE LEARNING

EMR- ELECTRONIC MEDICAL RECORDS

UHC – UNIVERSAL HEALTH CARE

IP – INPATIENT

OP- OUTPATIENT

K-NN -K-NEAREST – NEIGHBOR

CHAPTER ONE

INTRODUCTION

The public and private health care sector has improved significantly in Kenya. Most hospitals have embraced the use of hospital management systems to track patient's data. The automation of this data has improved the efficiency of the insurance system and the settlement of healthcare claims (Thomson, Sagan and Mossialos, Private Health Insurance: History, Politics and Performance 2020). Claims processing is the most important function for any insurance company. The speed and convenience with which the claims are settled has a bearing on the general reputation of the insurance company (Luhach, et al. 2019). Most hospitals in the country have identified the need to embrace technology especially regarding patient's information. This data has enabled hospitals to capture crucial trends and patterns about patient's health, ailments, and medication. In recent years, the growth in Kenya's health insurance industry has been the fastest in comparison to all other general insurance sectors. The growth of this sector can be attributed to different factors (Barnes, et al. 2015).

The government has played a key role in promoting the health care system of the country. However, the insurance health industry in the country faces difficulties which include increasing fraud, rising costs, and a high number of claims ratios (Mulaki and Muchiri 2019). According to a 2012 report by the Insurance Regulatory Authority, health insurers experienced a collective claims ratio of 77 percent (Insurance Regulatory Authority 2019). Half of the insurers in the country were unable to generate an underwriting profit during this period (Insurance Regulatory Authority 2019). Ultimately, this resulted in the insurers raising their premium rates. The challenges of fraud, rising costs and dealing with high number of claims can be tackled by embracing technology (Mulaki and Muchiri 2019). Automation of the insurance system and alignment of the system with the service providers in the country is the solution for the challenges faced by the industry (Langlois 2016).

1.1 Background to the problem

The increase in number of health claims calls for improved management systems capable of storing all relevant patient information. Automation of health claims and adoption of an advanced claims management system is important in ensuring proper coordination (Thomson, Sagan and Mossialos, Private Health Insurance: History, Politics and Performance 2020). Most hospitals in Kenya have adopted the electronic hospital management system (Langlois 2016). This paper will focus on how the adoption of E-Systems in the healthcare system will streamline the claims process in both the public and private sector (Barnes, et al. 2015). There are approximately 30 million claims in the country every year that must be reviewed, approved, and paid every year. The major challenge affecting this is the system is the process is not fully automated. The fact that the system is done manually makes it expensive, time consuming and subject to many fraud-related cases (Deloitte 2020).

1.2 Statement of the problem

Health insurers in the country maintain a database that contains medical information used by claimants (Association of Kenya Insurers 2019). Proper implementation of this database will not only enable the industry to develop but also to process claims at a faster and more efficient rate. A medical insurance policy is designed to cater for the costs that result from injury or sickness. This cover caters for the expenses of the doctors, hospital bed, nursing services as well as the prescribed medication for the patient. For an individual that is covered, he or she can benefit either from direct reimbursement of the expenses that they have incurred provided they fall within the limit.

The Insurance Regulatory Authority plays a crucial role of supervising and ensuring the smooth operation of the insurance industry in the country (Association of Kenya Insurers 2019). In Kenya, we have the public medical insurance plan known as the National Hospital Insurance Fund. NHIF is governed by the NHIF Act 9 of the year 1998 and has been in existence since the year 1966 (Association of Kenya Insurers 2019). This system is open, and it covers individuals in both the informal and the formal sector of the economy. Individuals in the formal sector are compulsory members of NHIF, while membership in the informal sector is voluntary. NHIF covers a total of 3.3 million members while the number of beneficiaries is around 7.8 million (Association of Kenya Insurers 2019).

1.3 Research outcomes and significance to key audiences

Private medical insurance in the country has two major categories. The first category is the direct private medical insurance that covers insurance for individuals (Luhach, et al. 2019). There is also the employment-based insurance that is medical cover taken by employers on behalf of their employees as part of the employment package that is offered to them. The employer may have a healthcare facility at the work premise to cater for the employees or may contract external healthcare organizations to provide this service (Langlois 2016). The general cost of healthcare in the country has been on the rise. It is because of this reason that most Kenyans have embraced health insurance cover with the hope that this will provide access to quality healthcare service. The establishment of a medical framework that provides access to essential health care including maternal care and HIV related diseases is one of the major policies that the Kenyan government has adopted (Marquez 2020).

With many individuals in the country turning to insurance companies to provide health cover, certain challenges have been felt over the years. One of the main problems in Kenya is the long turnaround time that it takes for insurance claims to be processed. Most insurance providers provide the maximum number of days that it takes for claims to be processed and paid. However, in most cases the claims take longer to be processed leading to frustration on the side of the insured party. Hospitals on the other hand have also become disgruntled by the delays in disbursement of funds to them by insurance providers. This has created uncertainty in the healthcare industry. Insurance companies on the other hand blame the delays on the logistics involved in the processing huge volumes of claims from different hospitals. This is a problem that has been there for quite some time hence the need for a lasting solution (Langlois 2016).

1.4 Research questions and objectives

According to the Association of Kenyan Insurers, fraud and abuse are very prevalent and are certainly very costly to the Kenyan healthcare sector (Marquez 2020). Medical insurance fraud has resulted in an increase in the premium on medical business. According to the regulator, 143 cases of medical insurance fraud were reported in 2012 (Deloitte 2020). This resulted in a loss of 253.6 million Kenyan shillings. There has been a consistent loss ration from the year 2008 to 2014. The AKI contracted an external party known as MaxWorth Associates to conduct an extensive survey on the extent of fraud in the Kenyan health industry (Nath and Mandal 2018).

The study report indicated a general increase in cases of health insurance related fraud in the country. In the survey, 28 percent of the respondents had come across health insurance claims that were suspicious health insurance. 21 percent of the respondents had detected fraud in the claims presented to them in the past year. 48 percent of the respondents indicated that they had signed claim forms before obtaining health services (Association of Kenya Insurers 2019). The main reason for this rapid increase in fraud is the collusion between beneficiaries and health service providers. Beneficiaries have found a way of working together with certain healthcare providers to file false claims that are later approved by insurance providers (Luhach, et al. 2019). There is also a lack of a sophisticated software system capable of detecting fraudulent claims. Poorly trained claims processing staff and lack of proper internal control mechanisms are other factors that greatly contributed to increased fraud cases in Kenya. From this comprehensive survey, it is evident that there is need a to adopt certain measures that will improve the efficiency of the healthcare system in the country (Mulaki and Muchiri 2019).

1.4.1 Objectives of the study

Main Objective

The main objective of the study is to determine the accuracy of health claim automation through machine learning.

Specific Objectives

1. To determine the level of automation among health claims in the insurance industry
2. Challenges facing health insurance due to lack of automation.
3. To develop and test a prototype for claims automation in the health insurance claim industry.

1.5 Assumptions and limitations of the research

Adaptation of an automated system that tracks and covers all aspects of insurance health claims in Kenya will go a long way in navigating the challenges faced. Most healthcare providers have automated their systems but there is a need to streamline the systems that they have adopted with insurance providers to ensure a smooth processing of claims (Association of Kenya Insurers 2019). This will also reduce cases of fraud reported and the number of rejected claims. Intelligent automation of systems is a trend that has been adopted in different sectors of the economy and it has proved to be a game changer. The main reasons for automating the healthcare system are to lower the expenses incurred while improving the efficiency of the industry. Automation also ensures that operations are optimized while enhancing customer experience. This will result in improved confidence among the public on the ability of insurance providers to receive and process their claims in good time which ensures quality healthcare service (Barnes, et al. 2015).

An automated system accelerates the process as personnel are redirected to handle other critical tasks (Deloitte 2020). Cases of employees colluding with beneficiaries will be greatly reduced as most of the processes are automated. The long manual processes on the other hand will be replaced by more efficient machines with the ability to detect and provide immediate responses to claims. This eliminates the need for individuals to perform repetitive tasks such as confirming personal information and other data processes that waste valuable time (Thomson, Sagan and Mossialos, Private Health Insurance: History, Politics and Performance 2020).

CHAPTER TWO

LITERATURE REVIEW AND THEORY

2.1 Introduction

This chapter looks to explore theoretical and realistic objectives of what other scholars have researched. The chapter is structured as follows: Health Insurance Survey, Health Insurance Sector in Kenya, Automation of Health Insurance Claims, Clinicians Working with Artificial Intelligence

2.2 Health Insurance Survey

The Kenyan healthcare system has been marred by cases of fraud over the years. According to the Association of Kenya insurers (AKI), the lack of adequate information on medical insurance fraud has led to high financial losses. According to the survey there are various forms of fraud and the existing strategies to counter the vice are limited (Barnes, et al. 2015). The research established that 48% of the respondents indicated they sign claim forms before obtaining health services (Association of Kenya Insurers 2019). The research indicates that fraudulent claims increased from 22 in 2008 to 225 in 2012 (Insurance Regulatory Authority 2019). The survey further indicates that the total amount paid increased from 46,869,450 in 2008 to 497,047,607 in 2012. The total value of the fraudulent claims paid was 3.7% (Association of Kenya Insurers 2019). The report highlights that the major types of healthcare fraud include membership substitution, generic medicine provision over branded medication, over servicing, prior ailments' non-disclosure and claims falsification. The survey highlights the health service providers in the country as the major perpetrators of insurance fraud (Marquez 2020). Cases of collusion between health service providers and beneficiaries as well as poor software and control systems were identified as the main factors contributing to fraud in the country (Mulaki and Muchiri 2019). The results of the survey indicate that key stakeholders are aware of the need to implement policies that will curb the increased fraud (Nath and Mandal 2018). The business leaders paint a lack of comprehensive and integrated approach in tackling fraud risk management. 65.9 percent of the respondents reported health insurance fraud in their organizations. The organizations have health insurance fraud detection and handling mechanisms (Barnes, et al. 2015). The survey further indicates a 31.9 percent fraud awareness campaign among employees. The use of fraud detection software and maintaining a fraud policy and a code of conduct are some of the key policies in fighting fraud (Deloitte 2020).

Survey on the use of a dedicated forensic analysis unit to detect and curb fraud incidents among organizations highlighted marked improvement in the detection of fraud. 30.2% of organizations have a fraud detection unit with an average budget of 937,500.75 per year (Association of Kenya Insurers 2019). 63.1% of the in agreement that these units were able to successfully reduce the number of reported cases. On average, 9,209,092.36 Kenyan shillings were recovered within the last financial year (Insurance Regulatory Authority 2019). The survey indicates that to prevent fraud, implementing a comprehensive and integrated approach to fraud is important. Data analytics tools with the capability to identify red flags play a crucial role in monitoring and detecting fraud (Insurance Regulatory Authority 2019).

2.3 The Health Insurance Sector in Kenya

According to a research by Fatima Badat, growth in Kenya’s health insurance industry has been the fastest in comparison to all other general insurance sectors (Badat 2018). Kenya’s health insurance industry contribution to the overall gross premium accounted for a quarter of the total in 2013. The insurance industry is facing challenges that include a high claims ratio, rising costs, and an increase in fraud related cases (Langlois 2016). The Insurance Regulatory Authority (IRA) 2012 industry report indicates that the industry received an overall claims ratio of 77 percent (Langlois 2016). Half of the insurance firms in the country were unable to generate any underwriting profit during this period. According to the author, this resulted in insurers increasing their premium rates in response to a market that is highly sensitive to changes in price. According to Fatima, innovation in the sector is the only way this challenge can be tackled (Badat 2018). Innovation can be done through the development of new and innovative products that will be able to match the risk profile of policy holders (Langlois 2016).

The overall process of managing and handling claims should also be improved because increased efficiency results in a total reduction of the total healthcare costs (Association of Kenya Insurers 2019). The total healthcare costs can be reduced further through a combination of strategies that include providing incentives for appropriate use of healthcare services (Barnes, et al. 2015). This can be further strengthened through aligning reimbursement by service providers and the quality of care. Reduction in fraudulent activity and expenditure that is non-healthcare related will reduce total healthcare costs according to the author (Association of Kenya Insurers 2019).

Appropriate use of healthcare services can be incentivized by capping reimbursable types, healthcare visits and premiums on insurance covered procedures (Insurance Regulatory Authority 2019). Other strategies may include reducing drug costs through subsidies on generic drug use or alternatively using brand name drugs that are affordable. This can be done without compromising on the healthcare quality. The article further looks at the reimbursement method that will be used. The fee for service strategy is criticized as it encourages health services to be over serviced. This is because health care providers are paid for each service that they can offer

(Langlois 2016). This in return leads to the delegation of services to less qualified professionals as the service provider maximizes their income (Luhach, et al. 2019). The article suggests the adoption of an innovative method that combines the fee for service model with the allocated budget.

The establishment of independent divisions, with the capacity to prevent the prevalent fraud in the health insurance sector, within anti-fraud departments by insurers is vital (Luhach, et al. 2019). A CIC Insurance study on outpatient claims highlights that there were 30 to 40 percent of fraud cases committed by policy holders (Mulaki and Muchiri 2019). The idea of handling fraud cases manually has proved to be costly for insurance companies in the country. This is a result of growth in unstructured data that is in most cases not adequately analyzed (Thomson, Sagan and Mossialos, Private Health Insurance: History, Politics and Performance 2020). This unstructured data in most cases creates room for most of the fraud cases passing through the system undetected (Mulaki and Muchiri 2019). The use of superior analytical solutions that utilize algorithmic models will be able to detect, manage, predict, and report fraudulent activity (Deloitte 2020). Administering claims is a process that creates an additional challenge to insurers. Standards recognized worldwide suggest that the expenditure on non-healthcare services, such as administrative costs and profits, should only account for 15 to 20 percent of total healthcare expenditure at any time (Association of Kenya Insurers 2019). Improvement of any HMIS system can be achieved by implementing electronic systems that allow the exchange of information on eligibility claims and other administrative services between payers and providers. This allows payers to submit their benefits statements electronically which will phase out paper statements (Deloitte 2020). Administrative and clinical functions such as billing or prior authorization, should be integrated into electronic health records over the next five years

(Association of Kenya Insurers 2019). Most companies should reassess their entire business model with proper automation. This will ensure the effective growth and performance of health insurance industry in the country (Thomson, Sagan and Mossialos, Private Health Insurance: History, Politics and Performance 2020).

2.4 Automation of Health Insurance Claims

Data mining is a technique that has been employed extensively by many organizations around the world (Nath and Mandal 2018). The healthcare industry is one of the industries that have employed data mining techniques. Data mining comes in handy when healthcare insurers want to detect cases of fraud and abuse. Physicians on the other hand can employ data mining while identifying effective treatment and best practices. This has enabled patients to be able to receive better and more affordable healthcare services. Data mining is important because the data that is generated by the healthcare industry in most cases will be too complex and voluminous to be processed by individuals manually (Thomson, Sagan and Mossialos, Private Health Insurance: History, Politics and Performance 2020). Data mining provides the required technology and methodology that will ensure the complex data is converted into useful information for the purpose of decision making. According to Professor David Gichoya, from the school of IT in Moi University, healthcare providers and patients have benefitted greatly from the advancements in technology (Association of Kenya Insurers 2019). Electronic medical and health records (EMR/EHR) have allowed service providers to access patient information at the touch of a single button. Patients can always access their information without having to visit healthcare facilities. However, a 2017 article by Reuters states that the claims management system still has some disruptions (Luhach, et al. 2019). The article states that medical billing systems are slowed down greatly by manual systems that are included in the system. The only solution for this is the complete automation of the entire process of claim management. (Luhach, et al. 2019) . The article states that 31 percent of service providers rely heavily on a manual system that approves or rejects the claims that are made (Thomson, Sagan and Mossialos, Private Health Insurance: History, Politics and Performance 2020).

The dependence on the manual system has created a platform for fraud to thrive in the insurance sector (Mulaki and Muchiri 2019). The rise in payment of fraudulent claims is largely attributed to the dependence on individuals to manage and process the claims. A system with the ability to process multiple claims should be in place to streamline the process (Nath and Mandal 2018). The system should be able to automate functions that are repetitive especially those with a predetermined response. It should also be able to provide access to third parties for the purpose of audit or scrutiny. This will improve public trust as well as improve the efficiency and transparency of the system (Luhach, et al. 2019). The system should be able to store historical data while complying with the requirements of all the regulatory authorities. This data must be stored in all forms which include e-forms and images. The system implemented should ensure confidentiality of client information by ensuring that information is not accessed by unauthorized parties (Marquez 2020). There are various layers through which claims will be initiated and approved.

2.6 Challenges Facing Health Insurers Due to Lack of Automation

According to the Journal of Public Health in Africa, the health system in Cameroon faces many challenges due to lack of automation (Ngwakongnwi 2018). The article states that Ministry of Health in Cameroon implemented a bottom-up approach of manual collection and reporting of health data in its National Health Information system (NHIS). The information on how it was implemented and functioned is lacking but it is evident that the lack of proper automation in this health system has resulted in many challenges.

District health officers encountered problems when it came to retrieving data and patient information (Ngwakongnwi 2018). For this reason, it is difficult for them to determine the claims received and to determine the claims that are genuine and those that are false. The article further states that this has created a significant backlog in the system with claims taking years to be processed. This has created a lack of trust in the system. The labor-intensive nature of the system makes the system slow and tedious (Ngwakongnwi 2018). By implementing a manual system many mistakes are bound to happen which in turn causes more delays.

2.7 Clinicians working with Artificial Intelligence.

According to Infinite Healthcare opportunities, healthcare automation plays a critical role in hospitals (Nath and Mandal 2018). The article states that automation has become a part of our daily lives and an integral part of various industries. However, the author states that the healthcare sector is still lagging when it comes to healthcare. The article states that the USA will experience a nursing shortage of 260,000 nurses by 2025 due to an aging population (Nath and Mandal 2018). The article further states that the inclusion of machine learning in the healthcare system is bound to generate immense benefits. Healthcare automation saves on time labor and the overall cost (Nath and Mandal 2018). Manually repetitive tasks are bound to be eliminated hence saving on time and the compensation. The article further states that human error and issues related to quality and lack of consistency are bound to be eliminated with the adoption of machines. The article further states that machine learning makes population health more manageable and feasible (Nath and Mandal 2018). The author states that the process of patient discharge and necessarily follow up will greatly be improved by automation. The article concludes by claiming that there are individuals that are worried that automation will affect human resources. It further states that there are those that claim, automation cannot replace quality human care that is provided by healthcare professionals. However, the author argues that the goal for automation is to streamline healthcare delivery (Nath and Mandal 2018).

2.8 Machine Learning

Machine learning is a component of computer science research that is progressively being embraced in data analysis and it is becoming popular and has earned high demand in the last decade (Geron 2019). Logic and conditions make it possible for machines to advance their proficiency in data understanding and definition of models with the least possible involvement of humans (Goodfellow 2016). The core bit of machine learning methodologies is the comparison among the specific objective that is to be attained and the result is derived by the machine state. The various learning machines models aim to determine models and patterns in data and use them to forecast future results. (D'Angelo 2017)

2.8.1 Models of Machine Learning

a) Supervised learning

This involves the collection and labelling of sample data, creation of representative features for that data, and the training of a model with the data and features. The model’s performance is then inspected to determine how to proceed in the next iteration. (Amershi 2016). Examples of the next iteration can be collecting more data or research with another learning algorithm. In supervised learning model, the algorithm receives input data and splits it into two categories, that is training data and test dataset (Mohamed and Ming 2021).

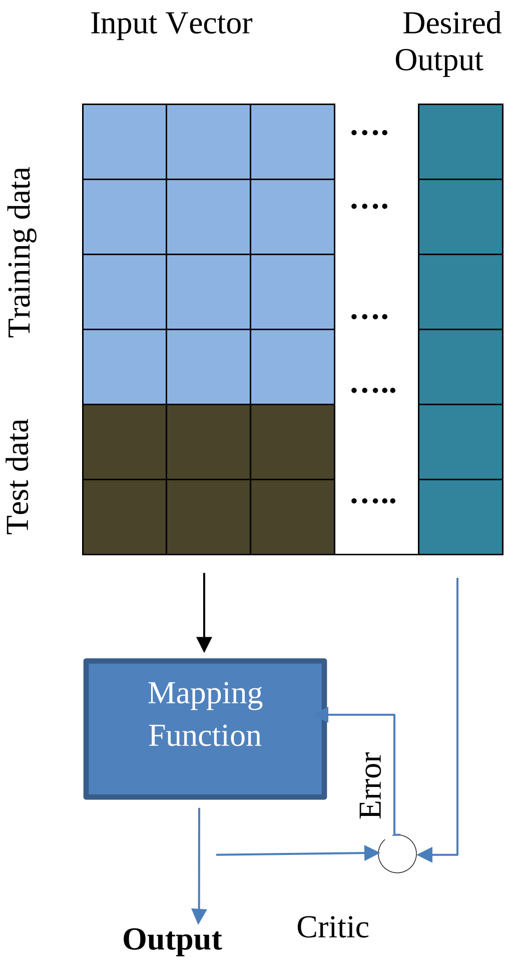


Figure 2 1: Supervised learning model

Several algorithms exist under supervised learning model, e.g., K Nearest Neighbor and naïve Bayes. In this paper we would look at the above examples.

a (i) K Nearest Neighbor

K-NN is a model that is centered around supervised learning and it is used for classification. K-NN takes a couple of labelled points, and it uses the labelled points to learn how to label other several other points. It is called Nearest neighbor because, when it labels a new point, it will look at any labelled point that is close to the new point which is its nearest neighbor. The “k” in K-NN is the total number of neighbors it checks. One of the advantages of K-NN algorithm is: for small training data sets it executes quite fast and it does not need any prior knowledge about the structure of data in the training set. One of the main disadvantages of K-NN is it might take a lot of space when the training set is large. (Priyadharsan Jeya 2019).

a (ii) naïve Bayes

Naïve Bayes is a machine learning algorithm that is based on the Bayes Theorem (used for calculating conditional probabilities). It has demonstrated that it cannot only be simple but also accurate and reliable. Naïve Bayes assumes the presence of a particular feature in a class is unrelated to the presence of any other feature. Naïve bayes shows that it can one of the best predictive performance compared to other classifiers e.g., decision trees (Freiedman 1997). Naïve Bayes is an efficient model for claims occurrence prediction (Jing 2017). One of the advantages of Naïve Bayes is, it simply needs a reduced quantity of the training data to compute the means and modifications of the parameters needed for classification.

b) Unsupervised learning

In unsupervised learning, no labels are given to the learning algorithm, therefore it must find its own structure on its input. Unsupervised has two phases, the first phase, the mapped function segments the data into classes and each input vector will be part of a class although the algorithm cannot apply labels to those classes. Unsupervised learning has no way to evaluate its performance. (Jones 2017). A well-known algorithm used in unsupervised learning is *k*- means clustering.

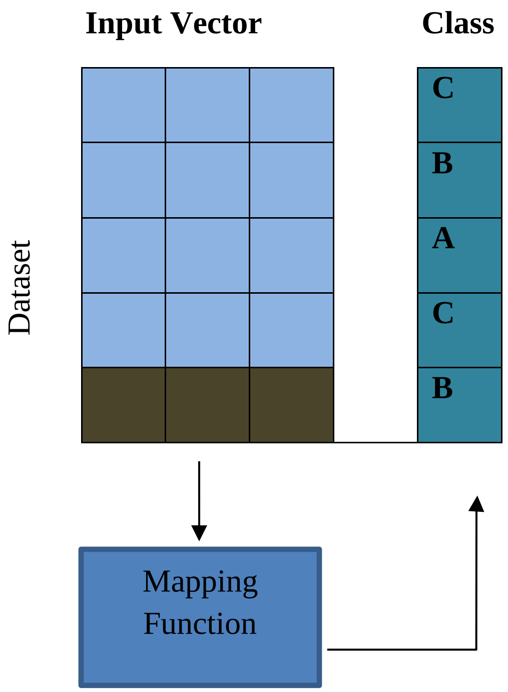


Figure 2 2: unsupervised learning model

2.9 High level Proposed Accuracy Model

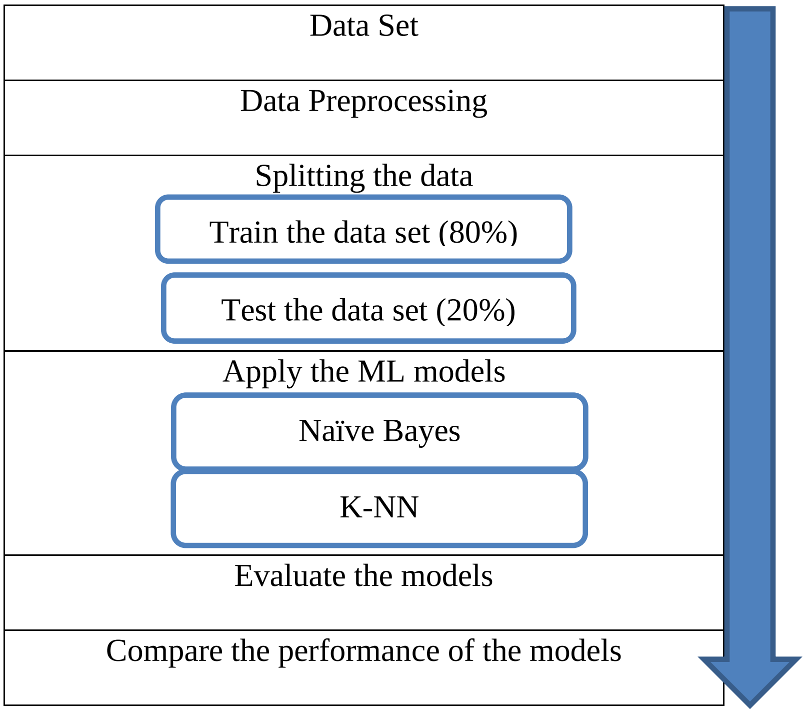


Figure 2 3: High Level Proposed Accuracy Model.

CHAPTER THREE

RESEARCH METHODOLOGY AND PROCESS

This chapter covers the research design, data collection and how it is relevant to the problem and analysis methods.

3.1 Research Design

Research design is evaluated as the design in which the research is contacted and comprises gathering, assessment which is also known as evaluation and analysis of data. (Akhtar, 2016) The system implemented will be accurate with the ability to capture data in a precise and accurate manner. The second step of this process will be to ensure that the claim goes through a set of predetermined steps before the settlement decision is made (Thomson, Sagan and Mossialos, Private Health Insurance: History, Politics and Performance 2020). Nearly all these predetermined steps are repetitive hence building a system that can analyze them is easy (Luhach, et al. 2019). According to the author, any irregularities that are flagged at this point are forwarded to investigators for further review and analysis. According to the author, this step should employ recent innovations and a robotic automation system (Luhach, et al. 2019).

3.2 Source of data

The data that will be used for this project will be obtained from a fintech company that deals with adjudication of claims. The data will contain private insurance claims. The data does meet the required standards for analysis. The data will contain OP claims. Personal names will not be shared because of General Data Protection Regulation (GDPR) purposes as GDPR does apply where personal data is processed exclusively or partially by automated means or even manual processing of personal data (Savic & Veinovic 2018)

3.3 Relevance of the Data to the Problem

This system will ensure that there is minimal disruption in the management of claims. The system ensures that all claims are processed in a timely manner based on the day the claims are made (Langlois 2016). The author states that the design employed will vary from one insurance provider to another based on the individual customer needs. The use of AI in analyzing insurance will enable service providers to make routine business decisions without having to employ human intelligence (Insurance Regulatory Authority 2019).

The benefits of automating the entire claims process include providing a single integrated environment that is more efficient than the manual systems that are employed in most cases (Association of Kenya Insurers 2019). According to the author, this system has the capacity to reduce cases of failed claims by 29 percent. This will also reduce the manual review wage bill by 200 hours hence saving the businesses money (Badat 2018). Automation of the system also ensures that all services are provided by the business without the need to contract external parties to conduct audits and to identify human errors. The system further ensures that third party claimants can submit their claims through a standard process. The article further states that big data analysis included in the analysis of business decisions ensures that the decisions that are made are more reliable and consistent (Insurance Regulatory Authority 2019). The automation of the claims process ensures that the workflow process is optimized, and administrators of the system can focus on the issue of irregular claims. Automation of the claims system also ensures that there are significant personnel cuts which reduce the overall business wage bill (Mulaki and Muchiri 2019). Decisions are made rapidly and efficiently hence claimants do not have to wait for long to get their responses (Association of Kenya Insurers 2019).

3.4 CRISP – DM Overview

ML includes a broad spectrum of methods, this research will focus on adapting CRISP-DM to the requirements of supervised learning, e.g., naïve Bayes and K-NN. CRISP-DM means Cross Industry Standard Process for Data mining. It is the model that was used for this research this is because it was initially built for data mining although it has widely been adopted for ML as well (Buczak.L 2016). CRISP-DM consists of six phases as shown in the figure below, they are: Business Understanding, Data Understanding, Data Preparation, Modelling, Evaluation and Deployment

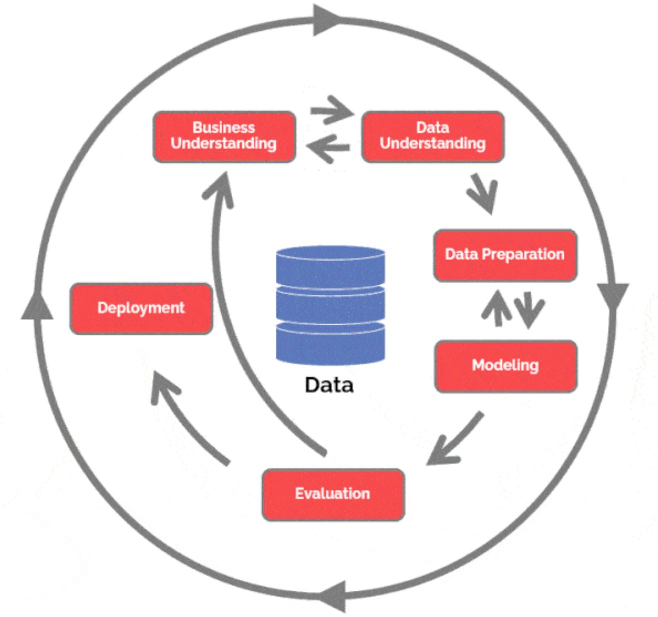


Figure 3. 1: Phases of CRISP-DM reference model. (Chapman .P 2000)

3.4.1 Business Understanding

For the specific goals for any ML project to be achieved the business objectives should be thoroughly assessed and understood. (Lukyanenko.R 2019). In this phase, the focus is on comprehending both the project objectives and its requirements from a commercial point of view. The project objectives were determined by the background of the problem which focuses on the health insurance industry which shades light on the claims module. This knowledge was later converted into a data mining problem definition as per the CRISP-DM model that focuses on ML and a plan designed to meet the objectives.

3.4.2 Data Understanding

In the second phase data understanding starts with collecting of the data then familiarizing with the data, identifying the quality of the data and if it is useful for the specific project area.

3.4.3 Data Preparation

In the third phase, the right datasets designed for this research was selected. The researcher answered the questions that arise in the phase, i.e., how do we organize the data for modelling. The phase covers all activities that are needed to construct the final dataset, one thing to note is that data preparation tasks may sometime be performed several times, some of the tasks that were repeated were cleaning of the data (e.g., removing duplicates), removing unnecessary labels that were not suitable to the problem (e.g. name of patient) and correcting noise (this is a value that is not correct and it is also known as corrupt data. Since the research focuses on Supervised learning, labels were added to the data. Finally, the data was split into training and testing set.

3.4.4 Modelling

In this phase, the methodology looks at what modelling techniques should be applied. The researcher selected 3 ML models to determine the accuracy of automation of claims. Naïve Bayes, LDA and K-NN were applied in this research.

3.4.5 Evaluation

In evaluation, the methodology identifies the best model that meets the business objective. The models that were applied from the modelling phase were evaluated if it met the original business objectives. The accuracy was achieved by using the training data set.

In this phase, the results that were captured by the models are interpreted by converting the structures generated into language that is understandable to business users. (Lukyanenko.R 2019).

3.4.6 Deployment

In this phase, the methodology identifies how the stakeholders will access the results. With the results from the model, a prototype will be developed to test claims automation. The prototype will be developed in Python. Once the deployment is done, enterprise users can refer to the original objectives and goals captured in the business needs.

3.4 System Design & Architecture

The starting point will be provided by the user, where they will input patient claim data through a simple user interface. The chosen algorithm will be implemented, and the output or result will be a notification whether the claim is approved or rejected.

CHAPTER FOUR

RESULTS AND DISCUSSION

4.1 Introduction

This chapter gives a detailed analysis of the research results and the discussions of the various Machine learning model and the prototype that was developed for this study.

4.2 Data Analysis, Preparation and Filtering

The claims dataset was categorized into private insurer claims and filtered into 2 statues that is Approved and Rejected claims, this is because the 2 statues are the last state of a claim and that is what is used to pay a provider or not.

The claims dataset had initially the following columns: Patient name, Hospital name, Age, Hospital name, gender, items, phone number, diagnosis, claim amount, currency, date the treatment was created, status of the claim. The dataset was then filtered to the following because the rest was considered as personal data and metadata claims status, claim amount, items and diagnosis, the following would be ideal to achieve the objectives of this research and to give an idea of when a claim should be Approved or Rejected.

The researcher had a dataset of 30000 claims, and this was sufficient to train the machine learning models that were stated in the objectives that is KNN, LDA and Naïve Bayes. The next steps of the process the researcher did was to install Python and SciPy (Scientific Python), SciPy providers several utility functions for optimization.

4.2.1 Load the dataset

The dataset used was claims dataset. The dataset contains 30000 claims, with four columns I.e., claim amount, diagnosis, item, and claim status. The fourth column is the status of the claim, which identifies the end state of the claim. The data was loaded using pandas. Pandas is a Python package built to provide expressive data structures that work with structured data in an easy, quick, and flexible way. Pandas will be used to explor**e** the data with descriptive statistics and data visualizations. Each column name needs to be specified when loading the data this helps in when the data is being explored.

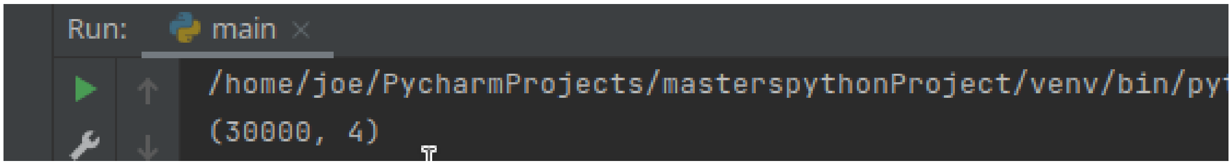
```
names = ['claim-amount', 'diagnosis', 'product', 'status']
```

Figure 1: Specification of the various columns

4.2.2 Summary /Dimension of the Dataset

The columns (attributes) and rows (instances) of the dataset needs to be confirmed.

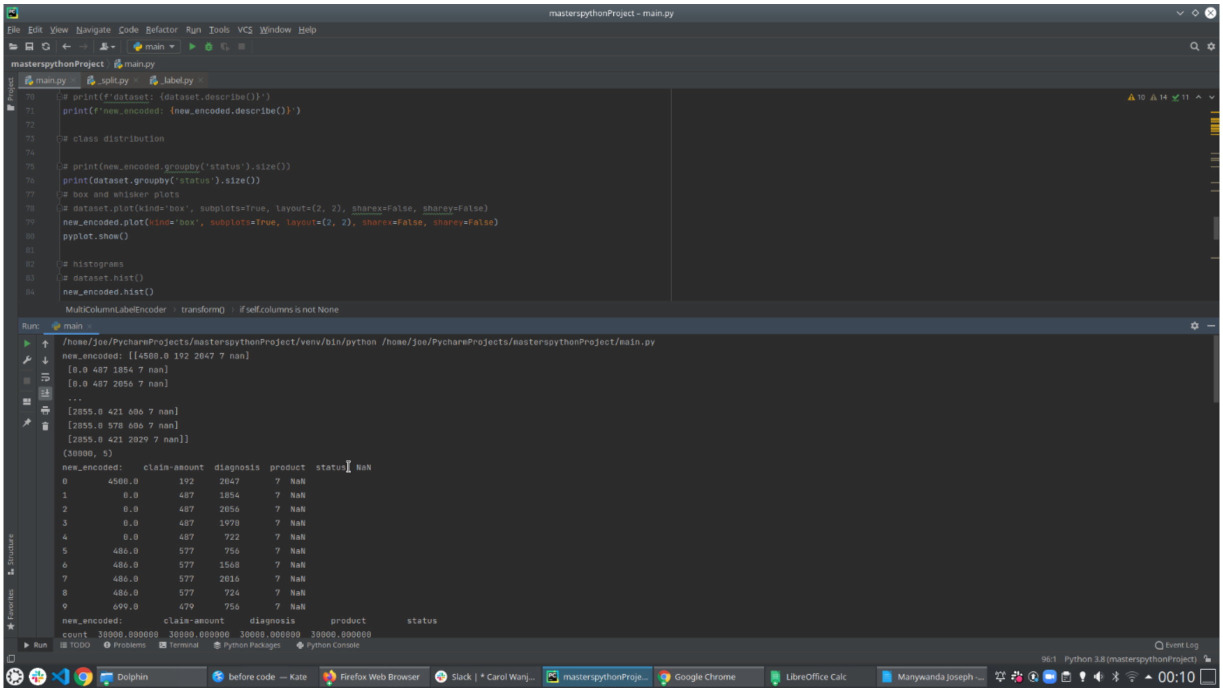
```
print (dataset. Shape)
```



```
/home/joe/PycharmProjects/masterspythonProject/venv/bin/py
(30000, 4)
```

Figure 4. 1: Confirming the dataset is correct

With the dataset in use, the string columns I.e., claim status, diagnosis, and item they had to be converted to unique integers.



```
# print(f'dataset: {dataset.describe()}')
print(f'new_encoded: {new_encoded.describe()}')

# class distribution

# print(new_encoded.groupby('status').size())
print(dataset.groupby('status').size())
# box and whisker plots
# dataset.plot(kind='box', subplots=True, layout=(2, 2), sharex=False, sharey=False)
new_encoded.plot(kind='box', subplots=True, layout=(2, 2), sharex=False, sharey=False)
pyplot.show()

# histograms
# dataset.hist()
new_encoded.hist()
```

```
/home/joe/PycharmProjects/masterspythonProject/venv/bin/python /home/joe/PycharmProjects/masterspythonProject/main.py
new_encoded: [[4500.0 192 2047 7 nan]
 [0.0 487 1854 7 nan]
 [0.0 487 2056 7 nan]
 ...
 [2855.0 421 606 7 nan]
 [2855.0 578 606 7 nan]
 [2855.0 421 2029 7 nan]]
(30000, 5)
new_encoded:    claim-amount  diagnosis  product  status  NaN
0        4500.0         192       2047        7   NaN
1           0.0         487       1854        7   NaN
2           0.0         487       2056        7   NaN
3           0.0         487       1970        7   NaN
4           0.0         487        722        7   NaN
5         486.0         577        756        7   NaN
6         486.0         577       1568        7   NaN
7         486.0         577       2016        7   NaN
8         486.0         577        724        7   NaN
9         699.0         479        756        7   NaN
new_encoded:        claim-amount     diagnosis       product        status
count  30000.000000  30000.000000  30000.000000  30000.000000
```

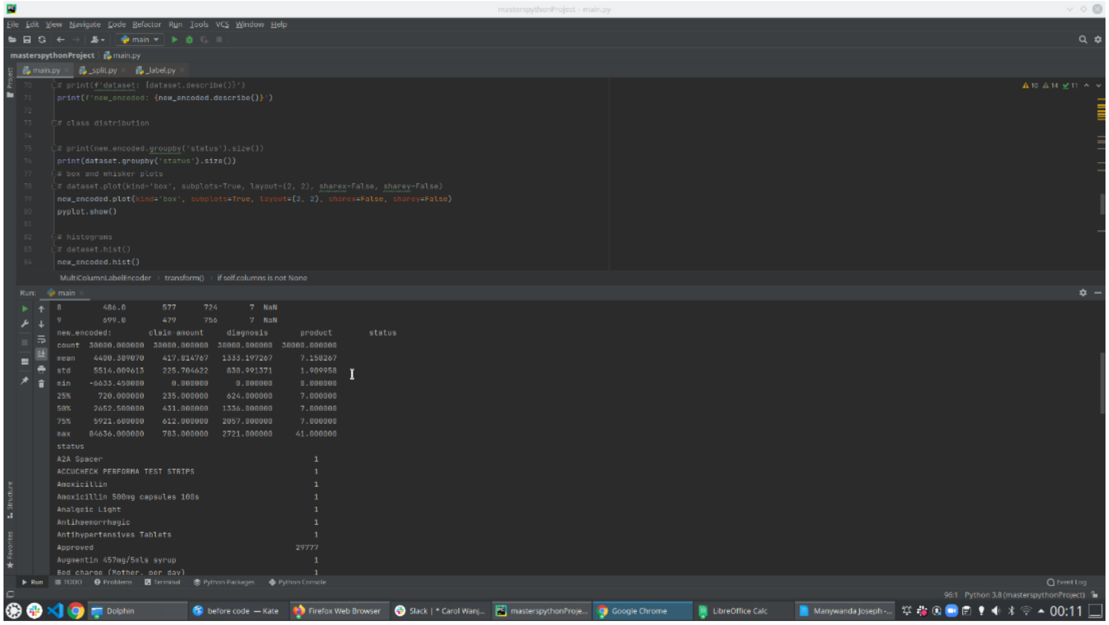
Figure 4. 2: Dataset has been converted to integers

4.2.3 Statistical Summary

In statistical summary, the researcher looks at a summary of each attribute. The summary includes the mean, count, the minimum and maximum values, and some of the percentiles. This was possible because of the conversion of the strings to integers.

From the results we can see the mean claim amount of a claim is Kshs. 4,400.30 which is typically a normal Outpatient claim (OP), the maximum value was Kshs. 84,636. The top item billed is Consultation and the common diagnosis is upper repository infection.

```
print (dataset.describe())
```


masterspythonProject - main.py
File Edit View Navigate Code Refactor Run Tools VCS Window Help
main
masterspythonProject main.py
main.py _split.py _label.py
MultiColumnLabelEncoder transform() if self.columns is not None
Run: main
Run TODO Problems Terminal Python Packages Python Console
Event Log
Python 3.8 (masterspythonProject)
Dolphin before code — Kate Firefox Web Browser Slack | * Carol Wanj... masterspythonProje... Google Chrome LibreOffice Calc Manywanda Joseph -...
00:11

```
# print(f'dataset: {dataset.describe()}')
print(f'new_encoded: {new_encoded.describe()}')

# class distribution

# print(new_encoded.groupby('status').size())
print(dataset.groupby('status').size())
# box and whisker plots
# dataset.plot(kind='box', subplots=True, layout=(2, 2), sharex=False, sharey=False)
new_encoded.plot(kind='box', subplots=True, layout=(2, 2), sharex=False, sharey=False)
pyplot.show()

# histograms
# dataset.hist()
new_encoded.hist()
```

```
8         486.0        577      724        7  NaN
9         699.0        479      756        7  NaN
new_encoded:        claim amount     diagnosis       product       status
count  30000.000000  30000.000000  30000.000000  30000.000000
mean    4400.389070    417.814767   1333.197267      7.158267
std     5514.009613    225.704622    830.991371      1.909958
min    -6633.450000      0.000000      0.000000      0.000000
25%      720.000000    235.000000    624.000000      7.000000
50%     2652.500000    431.000000   1336.000000      7.000000
75%     5921.600000    612.000000   2057.000000      7.000000
max    84636.000000    783.000000   2721.000000     41.000000
status
A2A Spacer                                        1
ACCUCHECK PERFORMA TEST STRIPS                    1
Amoxicillin                                       1
Amoxicillin 500mg capsules 100s                   1
Analgeic Light                                    1
Antihaemorrhagic                                  1
Antihypertensives Tablets                         1
Approved                                      29777
Augmentin 457mg/5mls syrup                        1
Bed charge (Mother, per day)                      1
```

Figure 4. 3: Statistical summary

4.3 Data Visualization

Data visualization is an important factor for understanding the data at hand. The two types of plots used was:

- Univariate plots this type of plot helps to comprehend each attribute.
- Multivariate plot helps to comprehend the relationship among attributes.

The function pyplot and matplotlib were used.

4.3.1 Univariate Plots

In univariate plots we will plot each individual variable. Since the input variables were converted from strings to integers and the variables are numeric, we created box and whisker plots of each.

The below figure shows a clearer distribution of the various input variables, the box and whisker plots are a better representation of the statistical summary.

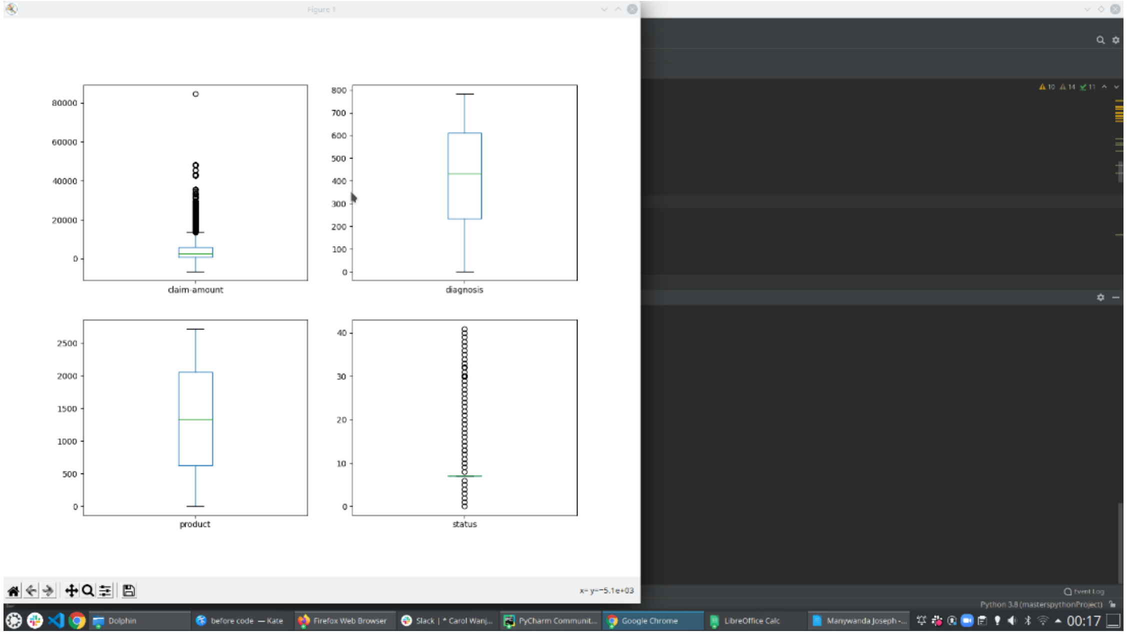


Figure 4. 4: A Box and whisker plot used for every input variable for the claim's dataset

A histogram was also created to get an idea of the distribution, the histogram below shows the correlation between box and whisker plots used for every input variable for the claim's dataset and the histogram.

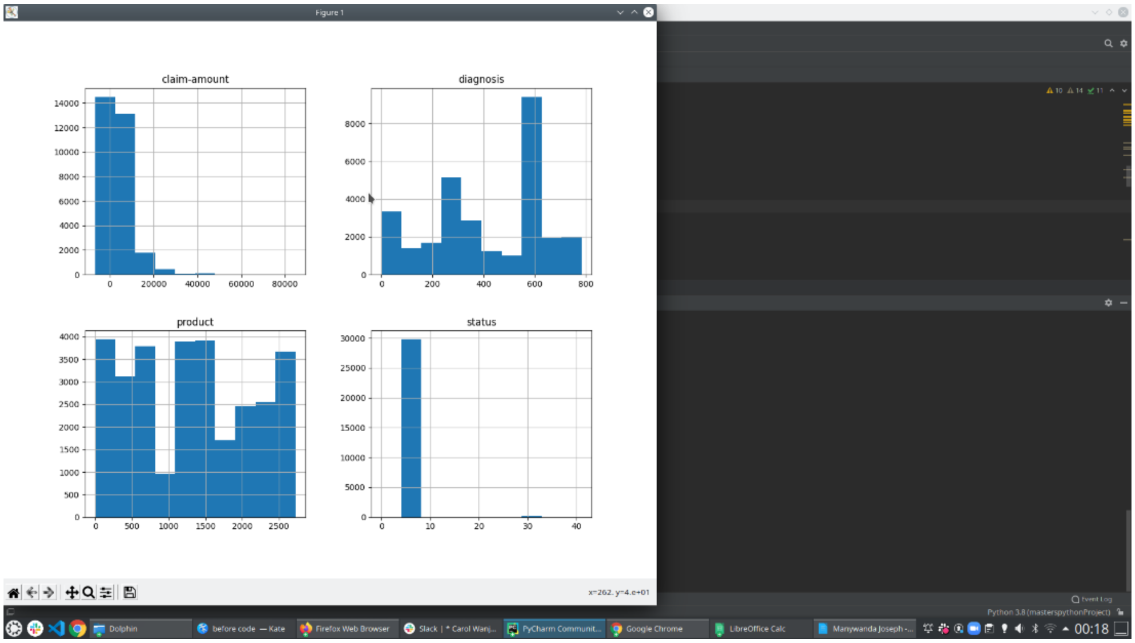


Figure 4. 5: Histogram plots used for every various input variables for the claim's dataset.

4.3.2 Multivariate Plots

The researcher considered to using Multivariate plots to look at the interactions between the variables. The below figure shows the scatterplots of all pairs of the attributes, this was valuable as it helps to point structured relationships between the various input variables. From the figure below, we can see the horizontal grouping of some pairs of attributes especially on the diagnosis and product, which suggests a very high correlation and a predictable relationship.

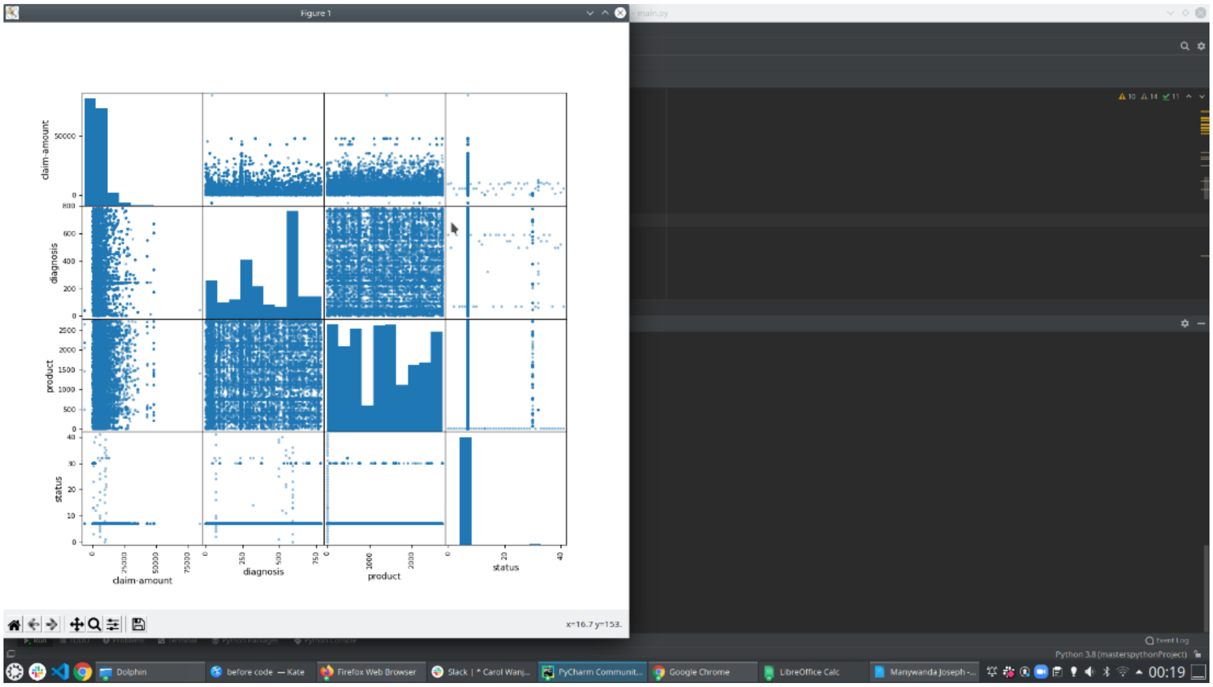


Figure 4. 6: Scatter matrix plot used for every various input variable for the claim's dataset

4.4 Evaluation of the Algorithms

One of the objectives of this research was to determine the level of accuracy of the models that were discussed in Chapter 2.8.1. The models of data will be used to approximate their accuracy of the unseen data. The data was trained using the following classification models: LDA, KNN and naïve Bayes.

4.4.1 Creation of a validation dataset

In this step, the validated dataset was separated; this was to get a concise assessment of the best model's accuracy on unseen data through its evaluation on authentic unseen data. This was achieved by holding back some data that unseen by the algorithms and was used to get a second and unbiased idea of the best model's accuracy. The dataset was then split into two, 80 percent (22,000 claims) was used to train and evaluate and 20 percent was held back as a validation dataset.

```
array = dataset.values
X = array[:,0:3]
y = array[:,3]
X_train, X_validation, Y_train, Y_validation = train_test_split(X, y, test_size=0.20, random_state=1)
```

From the above code snippet, the trained data in the X_train and Y_train which was used to prepare the models and X_validation and Y_validation sets that were later used to validate the data.

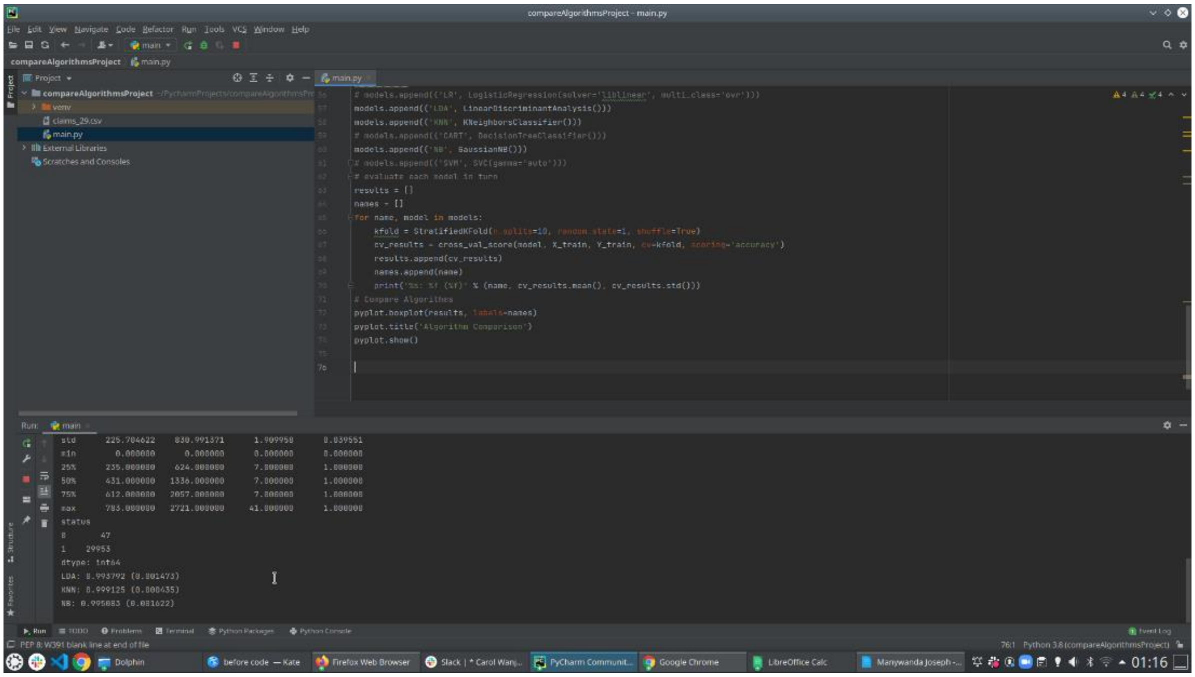
4.4.2 10-fold cross validation.

10-fold cross validation, split our dataset into 10parts, which in turn trains 9 and tested on 1 through the 30000 claims. The random seed through the random_state argument was set to a fixed number to ensure that each algorithm is assessed on the same splits of the trained dataset. In chapter 2.8.1 a(ii), which is K-NN, we used cross-validation on the training data set which determined the best estimate of the neighbors K for each data set.

The metric 'accuracy' was used to evaluate the models. This is the number of correct predictions instances divided by the total number of instances in the dataset multiplied by 100% to give a percentage e.g., 85.234%.

4.4.3 Build and select the best model.

The researcher used 3 models which were a good combination of simple linear (LDA- Linear Discriminant Analysis) and nonlinear (KNN, NB – naïve Bayes) the 3 models were compared, and the most accurate model was selected. The figure below shows the results of the model and the most accurate model. Hence, KNN has given us the best accuracy on the claims data, it was used to develop a simple prototype.



```python
# models.append(('LR', LogisticRegression(solver='liblinear', multi_class='ovr')))
models.append(('LDA', LinearDiscriminantAnalysis()))
models.append(('KNN', KNeighborsClassifier()))
# models.append(('CART', DecisionTreeClassifier()))
models.append(('NB', GaussianNB()))
# models.append(('SVM', SVC(gamma='auto')))
# evaluate each model in turn
results = []
names = []
for name, model in models:
    kfold = StratifiedKFold(n_splits=10, random_state=1, shuffle=True)
    cv_results = cross_val_score(model, X_train, Y_train, cv=kfold, scoring='accuracy')
    results.append(cv_results)
    names.append(name)
    print('%s: %f (%f)' % (name, cv_results.mean(), cv_results.std()))
# Compare Algorithms
pyplot.boxplot(results, labels=names)
pyplot.title('Algorithm Comparison')
pyplot.show()

```

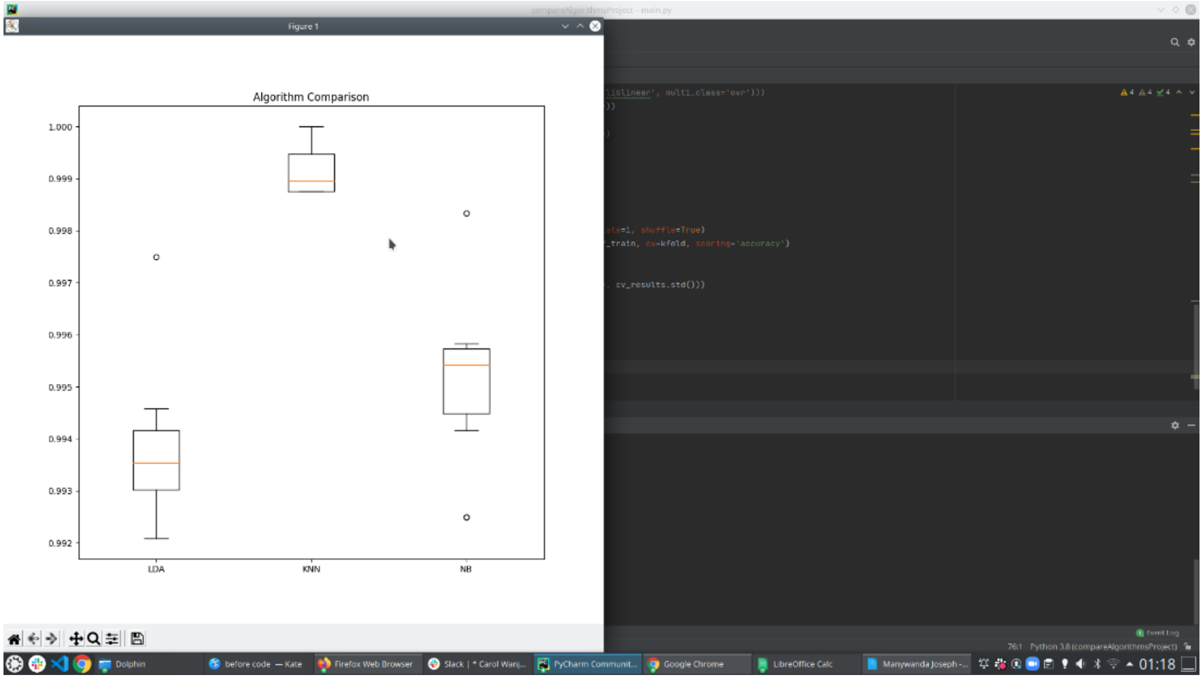
```
std     225.704622   830.991371    1.909958    0.039551
min       0.000000     0.000000    0.000000    0.000000
25%     235.000000   624.000000    7.000000    1.000000
50%     431.000000  1336.000000    7.000000    1.000000
75%     612.000000  2057.000000    7.000000    1.000000
max     783.000000  2721.000000   41.000000    1.000000
status
0      47
1   29953
dtype: int64
LDA: 0.993792 (0.001473)
KNN: 0.999125 (0.000435)
NB: 0.995083 (0.001622)
```

Figure 4. 7: Results of the evaluation of the algorithms

Model used	**Accuracy of the claims**
LDA	0.993792
KNN	0.999125
NB	0.995083

Table 4. 1: Results of the evaluation of the algorithms

From the observations above, the researcher found out that KNN had the largest estimated accuracy of around 0.9991 which is 99.9%, one effective way to determine comparison in the results of the various algorithms is by creating a box and whisker plot for every single distribution and evaluate them.



```
liblinear', multi_class='ovr')))
))

)

ate=1, shuffle=True)
_train, cv=kfold, scoring='accuracy')

, cv_results.std()))
```

Figure 4. 8: Algorithm comparison

The table below summarizes the frequency of the various models which ranked top 4 for all the datasets. KNN was constantly among the best performing model, and it was also observed it produces better results with large and smaller datasets.

Model used	**Number of times in Top 4**
KNN – (K- Nearest Neighbours)	4 Times
naïve Bayes	3 Times
LDA	1 Time

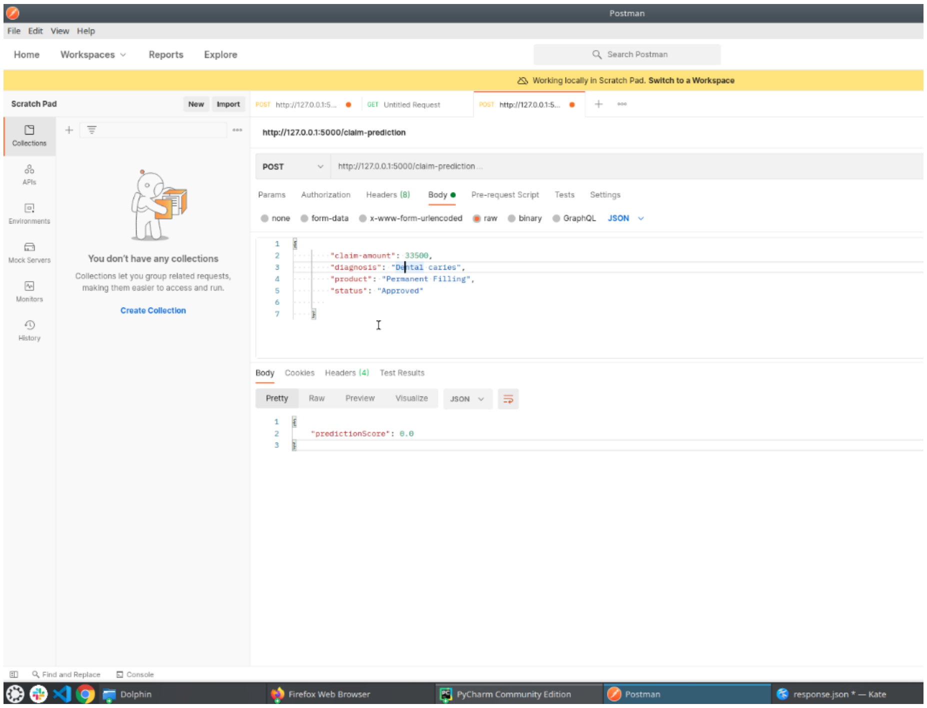
Table 4. 2: Frequency of various models

4.5 The prototype Development

The Machine learning model KNN was selected based on chapter 4.4.3, the prototype was developed using Python and Flask which was used to create a server to host the application in the researcher's local machine. The simple Python application (endpoint) gives the user 4 parameters to input that is the amount, diagnosis, item and claim status based, it then gives the user a prediction score and based on that prediction score, the claim can be relooked, or the claim paid by the insurer to the hospital.

This application could be integrated with an insurer claims module application to check the prediction score and based on the prediction score it can be parsed to the insurer finance module to pay the hospital in real time. The endpoint can be made available so that any insurer can use it and customize it to build a frontend to their liking.

During testing, the prototype was able to correctly identify approved and rejected claims and give an accuracy score. The researcher compared 250 manually adjudicated claims vs automated adjudicated claims, and the automated claims gave an 80% accurate score.


POST http://127.0.0.1:5000/claim-prediction

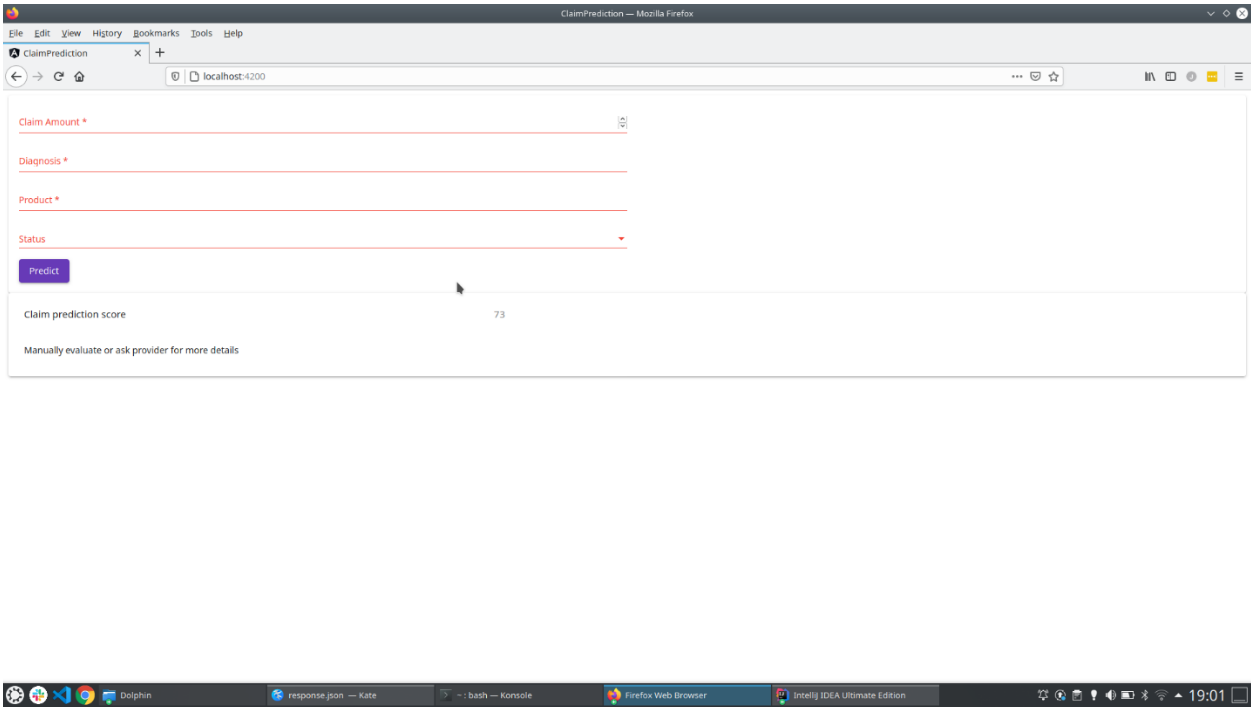
```
{
    "claim-amount": 33500,
    "diagnosis": "Dental caries",
    "product": "Permanent Filling",
    "status": "Approved"

}
```

```
{
  "predictionScore": 0.0
}
```

Postman
File Edit View Help
Home Workspaces Reports Explore
Search Postman
Working locally in Scratch Pad. Switch to a Workspace
Scratch Pad
New Import
Collections
APIs
Environments
Mock Servers
Monitors
History
You don't have any collections
Collections let you group related requests, making them easier to access and run.
Create Collection
Params Authorization Headers (8) Body Pre-request Script Tests Settings
none form-data x-www-form-urlencoded raw binary GraphQL JSON
Body Cookies Headers (4) Test Results
Pretty Raw Preview Visualize JSON
Find and Replace Console
Dolphin Firefox Web Browser PyCharm Community Edition Postman response.json * — Kate

Figure 4. 9: endpoint using the selected ML model


Claim Amount *

Diagnosis *

Product *

Status

Predict

Claim prediction score 73

Manually evaluate or ask provider for more details

ClaimPrediction — Mozilla Firefox
File Edit View History Bookmarks Tools Help
ClaimPrediction
localhost:4200
Dolphin response.json — Kate ~ : bash — Konsole Firefox Web Browser IntelliJ IDEA Ultimate Edition 19:01

Figure 4. 10: prototype basic User interface

CHAPTER FIVE

CONCLUSION AND RECOMMENDATIONS

5.1 Introduction

This chapter is the final stage of this study, and it presents a summary of the research findings, conclusion, and recommendations of the study.

5.2 Summary

Objective Number 1: To determine the accuracy of health claim automation through machine learning.

In this specific objective, the following machine learning models were used, LDA, KNN and naïve Bayes. These models were selected because they have a high predictive performance compared to the other models and the accuracy scores are quite high. From the evaluation of the algorithms, KNN had the highest accuracy compared to the other models.

Objective Number 2: Challenges facing health insurance due to lack of automation.

The research established that it is labor intensive for claims managers to look at every claim to adjudicate and in return quite intensive for the finance team to pay out claims in time. Insurers can take up to 3 months to pay out a claim to a specific hospital.

Objective Number 3: To develop and test a prototype for claims automation in the health insurance claim industry.

The prototype was developed based on the results in Chapter 4, the algorithm with the highest accuracy was selected (KNN), Python was used to create the backend and Flask was used to host the application.

5.3 Conclusion

There has been a substantial change in the health claims automation which has been played by the emergence of ML. In this study, the researcher showed that ML is a possible solution for the challenges that are facing insurers due to lack of automation. Applying ML not only in the health insurance industry but also other insurance industries. e.g., general insurance can be used to improve the business, increase income of the insurer, and reduce costs. The timely payment of a claim to a hospital will also improve the relationships between the insurer and the hospital.

The results of this study showed how several ML models can be used to accurately analyze health claims dataset and from this study KNN was the most accurate one. The prototype is built based on the accurate model where it gives a prediction score of the claim, insurers can parse their claim details in the prototype and get a prediction and pay hospitals immediately.

5.4 Recommendations

One of the recommendations is to carry out experiments based on the other ML models to know the accuracy of the other ML models and use large training sets, e.g., 500000 claims. It is recommended that the prototype be customized for other claims insurance industries.

REFERENCES

Akhatar. *Research Design.* 2016. *Research in Social Science: Interdisciplinary Perspectives.* 2016.

Amershi, Max Chickering, Steven M.Drucker. "Redesigning Performance Analysis Tools for Machine Learning." *Mircrosoft Research*, 2016.

Association of Kenya Insurers. *2019 Insurance Industry Report.* Nairobi: AKI, 2019.

Badat, Fatima. *Medicare Advantage.* Johannesburg: Dream Visual, 2018.

Barnes, Jeff, Barbara O'Hanlon, Frank Feeley III, Kimberly McKeon, Nelson Gitonga, and Caytie Decker. *Private Health Sector Assessment in Kenya.* Nairobi: World Bank Publications, 2015.

Buczak.L, Guven.E. "A Survey of Data Mining and Machine LearningMethods for Cyber Security Intrusion Detection." *IEEE COMMUNICATIONS SURVEYS & TUTORIALS*, 2016: 1153-1176.

Chapman .P, Clinton.J, Kerber.R. "CRISP-DM 1.0 Step-by-step data mining guide CRISP." *SPSS*, 2000.

D'Angelo, Gianni, Massimo Tipaldi, Luigi Glielmo. "Spacecraft Autonomy Modeled via Markov Deci-sion Process and Associative Rule-Based Machine Learning." *IEEE International Workshop on Metrol-ogy for Aerospace (MetroAeroSpace)*, 2017: 324-29.

Deloitte. *Insurance Outlook Report 2019/2020.* Nairobi: Deloitte, 2020.

Freiedman, Nir, Dan Geiger & Moises Goldszmidt. "Bayesian network classifiers." *Machine Learning*, 1997: 29: 131-163.

Geron, Aurelien. "Hands-on Machine Learning with Scikit-Learn, Keras, and TensorFlow." *Concepts, Tools, and Techniques to Build In-telligent System*, 2019.

Goodfellow, Ian, Yoshua Bengio. "Machine learning basics." *Deep learning*, 2016: 98-164.

Insurance Regulatory Authority. *Quarterly Reports.* Nairobi: Bima Bora, 2019.

Jing, Longhao, Wenjing Zhao, Karthik Sharma and Runhua Feng. "Research on Probability-based Learning Application on Car Insurance Data." *4th International Conference on Machinery, Materials and Computer (MACMC 2017).* Amsterdam: Atlantis Press., 2017.

Jones, Tim. "Models for machine learning." *Artificial Intelligence*, December 5, 2017.

Langlois, Adèle. *Negotiating Bioethics: The Governance of UNESCO's Bioethics Programme.* Chicago: Routledge, 2016.

Luhach, Ashish Kumar, Dharm Singh Jat, Kamarul Bin Ghazali Hawari, Xiao-Zhi Gao, and Pawan Lingras. *Advanced Informatics for Computing Research: Third International Conference, ICAICR 2019, Shimla, India, June 15–16, 2019, Revised Selected Papers, Part I.* Chicago: Springer Nature, 2019.

Lukyanenko.R, Castellanos.A, Parsons.J. "Using Conceptual Modeling to Support Machine Learning." *Information Systems Engineering in Responsible Information Systems*, May 2019.

Marquez, Lani Rice. *Improving Health Care in Low- and Middle-Income Countries: A Case Book.* Miami: Springer Nature, 2020.

Mohamed, Hanafy, and Ruixing Ming. "Machine Learinig Approches for Auto Insurance Big Data." *MDPI*, 2021.

Mulaki, Aaron, and Stephen Muchiri. *Kenya Health System Assessment.* Nairobi: Palladium, 2019.

Nath, Vijay, and Jyotsna Kumar Mandal. *Nanoelectronics, Circuits and Communication Systems: Proceeding of NCCS 2017.* Mumbai: Springer, 2018.

Ngwakongnwi, Dr. "Journal of Public Health in Africa." *Challenges to Implementing a National Health Information System in Cameroon: Perspectives of Stakeholders*, 2018: 322-325.

Priyadharsan Jeya, Sanjay Kabin,Kathiresan.S, Karthik Kiran, Prasath Siva. "Patient health monitoring using IoT with machine learning." *International Research Journal of Engineering and Technology (IRJET)*, 2019: 7518.

Savic, Veinovic. "Challenges of General Data Protection Regulation (GDPR)." *International Scientific Conference on Information Technology and Data Related Research*, 2018: 23-30.

Thomson, Sarah, Anna Sagan, and Elias Mossialos. *Private Health Insurance: History, Politics and Performance.* Chiago: Cambridge University Press, 2020.

—. *Private Health Insurance: History, Politics and Performance.* New York: Cambridge University Press, 2020.