SCHOOL OF GRADUATE STUDIES AND RESEARCH
FACULTY OF COMPUTING AND INFORMATION MANAGEMENT

PROJECT TITLE: APPLICATION OF MACHINE LEARNING FOR ESTIMATING MOTOR VEHICLE INSURANCE PREMIUM.

BY
STUDENT NAME: PAMBA FIDELIA.

REG. NO: 16/01039

SUPERVISOR: DR. LUCY MBURU WARUGURU

A RESEARCH PROJECT SUBMITTED IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE AWARD OF MSc. IN INFORMATION SYSTEMS MANAGEMENT AT THE FACULTY OF COMPUTING AND INFORMATION MANAGEMENT

DECLARATION

I declare that this research project "**Application of Machine Learning for Estimating Motor Vehicle Insurance Premium"** is my own work.

1. I therefore declare that I am the sole author of this dissertation.
2. I further authorize the KCA University to reproduce this thesis to other individuals for scholarly research.
3. I understand that my thesis may be made electronically available to the public.

STUDENT'S NAME : FIDELIA PAMBA

REG NO : 16/01039

SIGN…………………DATE……………………………………

SUPERVISOR'S NAME: DR LUCY MBURU WARUGURU

SIGN…………………DATE…………………………….

Abstract

Risk models need to be estimated by insurance companies so as to predict the magnitude of claim and determine the premiums charged to the insured. This is intended to prevent losses in the future. Motor vehicle damage insurance is the most common type of insurance in the world, forming the largest sector of the insurance industry. It is also the type of insurance that generates the largest amount of loss for most insurance companies. In Kenya especially, the challenge faced by insurers is to balance the growth of the motor vehicle insurance business by increasing the customer base while also maintaining the profitability of this sector. It is therefore important to identify the main causes of problems associated with motor vehicle damage insurance, its impact on the revenue of the insurers and factors that contribute to the high motor claims to enable more accurate estimates of risk versus premium paid. In recent years the interest has increased in the use of information technology (IT) and statistical machine learning methods. This is supported by increasing computing capabilities, data availability and the trend towards automation of cumbersome or repetitive tasks. Statistical regression models have numerous applications, where they have been used in many contexts. Using a linear or generalized linear regression model in predicting insurance premiums is an area in which only a few pioneer studies have been carried out with promising results.

This thesis explores applicability of new machine learning techniques such as tree-boosted models to optimize the proposed premium of prospective policy holders. It proposes two machine learning models for pricing motor vehicle damage insurance (decision trees and regression). The thesis is therefore aimed at identifying sources of risks in motor vehicles, identifying variable for motor vehicle premium determinants and then establishing a framework that will be used to regulate the premiums charged by motor vehicle insurers. Data from insurance companies has been used, which is made up of the premium rates and compensations, and other variables such as age, driver's experience, etc. Results of this thesis should be seen as successful for the use of generalized linear models in the making of car damage insurance premium rates. The established model will be used to advise the insurance companies on how to charge premiums dynamically.

Acknowledgement

I would like to thank my Almighty God for the many graces that He has continued to provide to me as work on the research.

I would also like to express my gratitude to my supervisor –Dr. Lucy Waruguru whose guidance and advice has greatly assisted me. I appreciate their invaluable and professional contribution this research. I would also want to thank my classmates for their support and encouragement to me.

I can't forget to thank my sick parents who in spite of their ailments their continued to encourage and pray for me.

Finally, I would like to thank my husband, Francis for his unending support for my studies. Thank you all for your support for without your support and assistance it would have been hard to work on this research.

Table of Contents

List of Figures

List of Abbreviations / Acronyms

IRA: Insurance Regulatory Authority.

AKI: Association of Kenya Insurers.

PSV: Public Service Vehicle.

SAS: Statistical Analysis System

PERT: Program Evaluation Review Technique

CRISP-DM: Cross Industry Standard Process for Data Mining

KSH: Kenya shilling

PMO: Project Management Office

CRB: Credit Reference Bureau

Definition of Insurance Terminology

IRA: As per guidelines stated in THE INSURANCE ACT laws of Kenya, The Insurance Regulatory Authority **IRA** (herein referred to as the Authority) has a mandate to formulate and enforce supervisory standards for the conduct of insurance business in Kenya as well as to protect the interests of policy holders and insurance beneficiaries in any insurance contract (CAP 487)

Unbiased: As per the English language the word is an ADJECTIVE showing no prejudice for or against something; impartial: 'his assessment of the benefits and drawbacks was unbiased' 'they give clear, unbiased advice' (oxford dictionary, 2012)

Premium: The cost of insurance in the monetary value for which two parties agree to exchange risk and certainty (Leaven & Governs, 2011).

Gross premium: The price of an insurance contract

Risk premium: Part of the premium corresponding to the insurance risk

Policyholder: The buyer of an insurance contract

Insurer: The issuer of an insurance contract, often an insurance company

Insurance risk: The probability that the insurer is obliged to pay the policyholder due to occurrence of insured events, defined by the insurance contract between the insurer and the policyholder

Claim Cost: The sum of payments to the policyholder from the insurer due to occurrence of insured events

Claim Frequency: The number of insurance claims for a specified time window

Claim severity: The cost that the insurer incurs for each claim

CHAPTER ONE

1.0 Introduction

The term "Insurance" can be defined differently from the point of view of different disciplines, but a more general definition is offered by the American Risk and Insurance Association as "the pooling of fortuitous losses by transfer of such risks to insurers, who agrees to indemnify insured people for such losses, to provide other pecuniary benefits on the occurrence, or to render services connected with the risk" (Redja, 2015). The key purpose of insurance companies is to protect their customers from economic stress in the case of unexpected events by selling insurance policies and charging premiums. An insurer assists the development of a country's economy in several ways. Primarily, they act as mobilizers of savings, financial intermediaries, and promoters of investment activity, stabilizer of the financial market, risk managers and agents to allocate capital resources efficiently (Desalegn, 2014).

All the risks that are quantifiable are potentially insurable. Certain types of risks that may give rise to claims are known as "perils". An insurance policy will set out in details which perils are covered by the policy and which one is not. The insurance market consists of two sectors; life- and non-life insurance sectors (Datamonitor, 2016). The amount and circumstances under which a customer who has paid premiums is to receive economic compensation is defined in the agreement between the insurer and the insured. Common types of insurance for private individuals are home and auto insurance.

Although the insurance industry has grown rapidly in the industrialized countries, its growth in developing countries like Kenya has neither been satisfactory nor in tandem with the growth of other sectors of the economy. A key issue is how to price an insurance contract, which is also known as rate making. As the Insurance industry continues to grow in Kenya, research by the Association of Kenya Insurers (AKI) has shown that customers often pick an insurance company as purely based on the premium to be paid (AKI Annual Report, 2015). If the price is too high, customers will turn to other insurance companies, and if the price is too low, the insurance company will not receive enough premiums to cover the insureds' claim costs. Premium determination in Kenya is currently dictated by IRA. In the recent past, there have been complaints that the IRA premium rate does not truly reflect the reality on the ground. Consequently, many motor insurance companies in Kenya have started discriminatory treatment of motor vehicles based on they own claim experience.

It seems reasonable to charge different premiums to different customers based on some well-chosen variables which are correlated to the insurance risk of the specific customer. In auto insurance, it could be that the risk is correlated to the brand of the car the insured drives or to how many years they have held the driving license. The part of the gross premium corresponding to the insurance risk is known as the risk premium, which is hence the expected claim cost. On average, an insurance company needs to charge more than the risk premium, since costs for administration need to be covered and a profit is often needed. The gross premium to be charged can also depend on price optimization strategies based on various factors such as price elasticity. Obviously, understanding the insurance risk of each contract is absolutely essential in an insurance business. If the risk is not understood, the profitability of the insurance company could decrease or the company might not even be able to meet its liabilities.

From a machine learning perspective, premium estimating represents a supervised classification problem that in the past has been solved in the actuarial science using regression models. Traditionally linear regression models have been used to model the risk premium. Recently, there has been a tendency to apply generalized linear models (GLM) such as logistic regression, since these types of models are known to be more suitable for rate-making than linear regression models. A key advantage offered by logistic regression is the easy interpretation of fitted parameters and the reasonable computational speed. Nevertheless, machine learning techniques such as Regression and Classification Trees, Random Forests, Gradient Boosted Machines, and Deep Learners (Kuhn& Johnson, 2013) have recently acquired increasing popularity in many business applications.

The interest of actuarial practitioners in machine learning models has grown in recent years, e.g. (Frees et al., 2014; Frees et al. 2016; Pozzolo, 2011) also used various machine learning algorithms to predict claim frequency in the Kaggle Allstate competition. Generally, machine learning techniques have been known to outperform logistic regression modeling in many applications, and particularly in matters involving classification (Pozzolo, 2011). However, two key issues prevent their widespread adoption in the insurance field. First, their parameters are often relatively more difficult to interpret (the “black box” issue). Second, the computational time required can be overwhelming compared to the time required to fit a GLM. Such issues have been addressed using random Forest models of machine learning (Breiman, 2001), but they have not been tested for predicting risk within the insurance industry.

1.1 Insurance Activities

The major activity of insurance companies is to receive premiums from customers and to pay out claims. However, Operating the insurance business involves several additional activities and is quite complex. Insurance companies will essentially conduct each of the following activities (Diacon, 2016):

Underwriting: here the insurer will evaluate the risk of a proposal before deciding whether to enter into a contract. He also must establish the terms of the contract.

Price decision: pricing is an important part of underwriting, and this process also called "premium rating". The price should as accurately as possible reflect the claim costs and expenses according to the contract. It must also allow for a reasonable margin of profit for the insurer. This allowance will depend on the level of surrounding competition from other insurers in the market.

Generation of new business: similar to all businesses insurers seek to increase the amount of business undertaken the right price. Insurance companies differ in their degree of reliance on a direct sales force rather than the use of brokers, and also in the amount they spend on advertisement.

Payment of insurance claims: this is the usual process that insurers use to pay claims.

Funds maintenance: insurers may not be able to pay all claims out of the revenue that they have received from premiums and investments. For this reason, they need to maintain a stable fund that can be used to pay the claims.

Investment of the fund for investment income: many classes of insurance will accumulate a good amount of funds. Careful investing of these funds allows insurers to earn investment income and make capital gains.

Purchase of reinsurance: the insurance company may be aware that some claim payments go beyond its financial resources. In that case, the insurer he will need to pass on part of liability for these claims to another insurance company by purchasing reinsurance.

Drawing up of accounts: like any other trading businesses, insurance companies remit money to creditors and receive the same from debtors. They must then balance their accounts for internal managements, shareholders, as well as for taxation and supervision authorities.

Payment of taxes: like other trading enterprises, insurance companies need to remit corporate tax, value added tax, and capital gained tax (Diacon, 2016).

Provision of extra services: together with other operations, insurance companies can provide additional services as well as advise their customers.

1.2 Problem statement

Motor vehicles in Kenya are the largest cause of accidents, damage and destruction of property (Peters et al., 2016). However, due to a low and uneven development of insurance in developing society companies rely on standardized measures for determining which premium to pay. The study carried out by IFAA in 2011 indicate, "out of the total 309,361 vehicles populated in Kenya only about 35% or 106,765 were insured voluntarily whereas the reaming 65% were dependent on their financial resources if liability arises".

The high hazard of motor risk makes insurance companies to adopt various strategies to minimize motor insurance adverse effect and maximize their profitability. Different premiums or tariffs are applied depending on degree of risk that is highlighted within the insurance portfolio price (Mahaela, 2015). The main problem is that the level of premiums charged to a customer does not depend on statistical data about the state of road, age of vehicle, age and experience of the driver etc (Dominique-Ferreira, 2017). For example, an 18-year-old man who drives a red sports car is more likely to pay a higher insurance premium than a 50-year-old man who drives a four-door sedan, (Yang et al., 2018). The underwriting process involves investigation into various factors and analysing them through statisticians called actuaries. In many cases the actuaries are not accurate.

Recently, more sophisticated machine learning models such as logistic regression models and tree-based decision models have been developed, but their usage in the insurance industry is not widespread (Frees et al., 2014; Frees et al. 2016; Pozzolo, 2011) Generally, decision tree techniques have been known to outperform logistic regression modelling in many applications, and particularly in matters involving classification (Pozzolo, 2011). To the researchers' knowledge, a comparison of machine learning techniques to model insurance premium prediction is still lacking in the literature and will hence be explored here.

1.3 Aim and objectives

1.3.1 Aim

The main aim of this study is to establish machine learning models for regulating motor vehicle premiums among insurance companies.

1.3.2 Specific objectives

The specific objectives of the study are:

1. Identify and classify sources of risk in motor vehicle.
2. Identify the variables for motor vehicle premium determination.
3. Establish machine learning models for estimating motor vehicle premiums in insurance industry.
4. Test and validate the models for effectiveness using sample data.

1.4 Research questions

1. What are the types of motor vehicle risks and how can they be classified?
2. Which factors can be used to effectively determine the level of motor vehicle risk and the premium to be paid?
3. Which are the appropriate machine learning models?
4. How well does the data of Kenyan motor vehicles fit to the established models of premium estimation?

1.5 Justification

The productivity of the insurance companies in Kenya is greatly dependent on savings that come from premium payment hence coming up with a model which will curb the underestimation of premium payment and cause losses for the insurance will be a plus to the insurance companies in Kenya.

Premium is increasingly recognized as the main source of sustainable competitive advantage in high-performance insurance companies hence a need for the insurance regulators to continuously undergo the process of premium management so that the resulting risk compensation can be sustained and wasteful expenses on exaggerated compensation payments is avoided. In Kenya, the insurance industry has made a lot of losses from certain insurance products, but the risk vs premium balancing was not given a lot of emphasis within the public sector and this has caused the level of profit to be lowered by 30% in over 60% of insurance companies (Manyara, 2016).

A good management of premium payment within insurance companies is an advantage to the insurance industry since accurate estimation of customer payment and customer compensation contributed to the sustainability of the different insurance companies. The proposed model for estimating premium will enhance the management of premium payment rates making It possible to attract the correct markets within paying customers and hence retaining them. This will lead to an increase in insurance company's productivity since the customers will be paying accurately determined rates and the insurance company can afford to pay them the compensation when the time comes.

The study will empower insurance regulators within the insurance industry with knowledge in the importance of accurate insurance premium calculation.

The study will add to the existing literature on premium estimation in the insurance business, factors that are likely to cause rates to go up or down and other measures that can be taken to ensure that the insurance premium that is being paid by the customer is not too much or too little.

Premium forecasting is a key role of any insurance regulation. It's vital for an insurance regulatory body to be able to predict future changes in premium rate since this affects both the quality of life and the ease of conducting business in a country. The results of this study are expected to:

- Provide the insurance regulatory bodies with information that can improve the estimation of motor vehicle premium. It is expected that further work can extend the estimation to include other types of premiums
- Improve policy making by improving the accuracy of such risk estimation for purposes of planning, execution and implementation.
- Influence insurance companies in accurate decision making. The developed methodology and model can be used freely by insurance companies to collect, record, store, and process data to produce information for premium determination decision makers.
- Improve the quality of products and services offered to customers. This is by gaining an edge in today's competitive market using the ability to quickly and efficiently explore, understand and act on the data to improve price competitiveness.

1.6 Chapter Summary

The rest of the thesis is presented as follows: Chapter 2 focuses on review literature related to insurance risk and premium determination. Different approaches and challenges are identified and discussed. Chapter 3 covers research methodology: research design, population size and sample, data collection instrument and data analysis approach. Chapter 4: presents the results; general characteristics of the data, descriptive statistics and results of a generalized linear model. Chapter 5 covers the discussion of results and conclusion of the study, research contributions and recommendations

CHAPTER TWO

2.0 Literature Review

The literature review analyses literature done by previous researchers that is relevant to insurance calculation modelling. The focus of this study is on insurance companies in Kenya. This chapter will explore the various techniques that have been used in premium determination. It will examine the challenges facing insurance companies on premium determination, calculation and compensation paying. It will review existing models, and the technology applied to existing models. Empirical review and analysis the use of generalized linear models in risk management and provide the conceptual framework of the thesis.

2.1 Introduction

Insurance plays a vital role in economic development. The following are some of the contribution of insurance sector to economic growth and employment:

- Reduces the capital firms need to operate
- Fosters investment and innovation by creating an environment of greater certainty
- Offering social protection alongside the state, facilitate firms' access to capital
- Promotes sensible risk management measures
- Fostering stable consumption throughout life and mobilizing savings (CEA, 2006).

In this regard insurance is defined as a social device providing financial compensation for the effect of misfortune (Hansell, 1974). The insurance industry in Kenya has untapped growth opportunities, and insurers are always seeking ways to increase their market shares. In a recent AKI annual report (2015) "The Insurance Industry recorded gross written premium of

KES 173.79 billion in 2015, corresponding to a 10.5% increase from the previous year. The gross earned premium increased by 9.8% to stand at KES 146.16 billion in 2015. The industry recorded a profit of KES 11.57 billion before tax in this same year. The industry asset base in 2015 increased by 11.5% to stand at KES 465.98 billion compared to KES 417.76 billion in 2014. The overall insurance penetration in 2015 was 2.79% compared to 2.93% in 2014. In 2014, penetration had been affected by GDP growth. The low penetration is an indication of untapped opportunities for insurance business in areas such as oil and gas, real estate, infrastructure, micro insurance and agriculture. The industry is working towards increasing this penetration to ensure that businesses across all sectors are covered and that more Kenyans are insured". (pg. 10)

"The insurance sector in East Africa is steadily developing, with its gross premiums totalling to $2.35b across the region. Kenya's premiums form 75% of the total, and it leads the region with a 3% penetration. However, insurance in the region is still underdeveloped as compared to the rest of the world. Only 3% of the East African population has health insurance. Life insurance remains particularly underdeveloped and despite the strong footing of international insurers in the region there is still room for growth". (pg. 22)

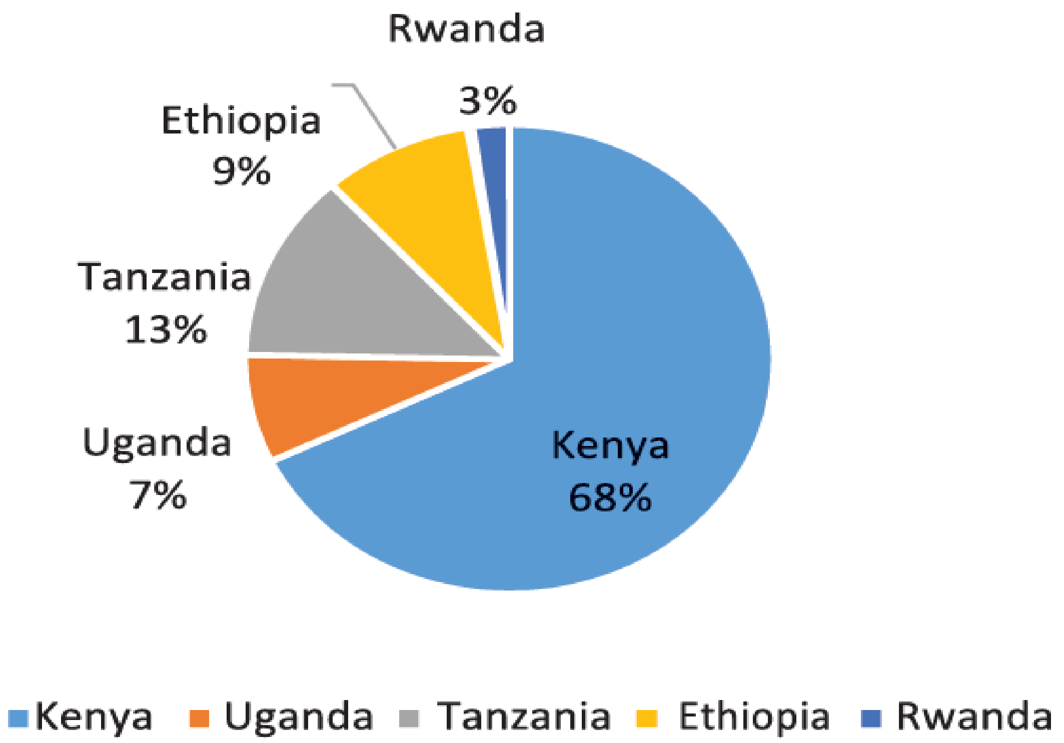

Figure 2.1: Gross written Premium in East Africa (From AKI Report 2015)

Although Kenyan insurance companies are the most well performing on almost the entire East Africa, not all insurance products are generating high profits for the insurers. It has been discovered that motor vehicles generate losses every year because the amount of premium paid is very low compared to the cost of compensating damages from accidents. Transport plays a crucial role in facilitating socio-economic development. Accordingly, road transport plays the dominant role in freight and passengers transportation as compared with the other transportation sub sector (Temesgen Zeleke, 2004). As of 2004, 99.3% of the passengers were used road transport services and 87.6% of the total freight was carried by motor vehicles (Ministry of Transport, 2005).

The rapid economic growth has also an effect for dramatically increased newly registered vehicles, during the past four years (2012-2014) personal cars have increased by 48% and the total number of all kinds of vehicles registered has now surpassed the 400,000 mark. All this has its own social, economic and environmental effects-both negative and positive (Birhanu & Mekonnen, 2014). Thus, the contribution of motor vehicle for transportation within a city and between cities is immense and now a day's transportation without the use of motor vehicle is unthinkable, especially in developing countries where other means of transport are not widely available. However, all its advantage that save energy, money and time, they adversely affect the lives and properties of citizens unless managed well.

Figure 2.2 shows a report of profits and losses for the different insurance products (AKI Report 2015, pg. 75). Private motor vehicles have in the recent years increased the amount of risk suffered of all insurance classes. Commercial motor vehicle damage forms a very high-risk class, more than five times the risk of insurance classes such as industrial fires. It is even more than 2 times the risk from medical treatments.

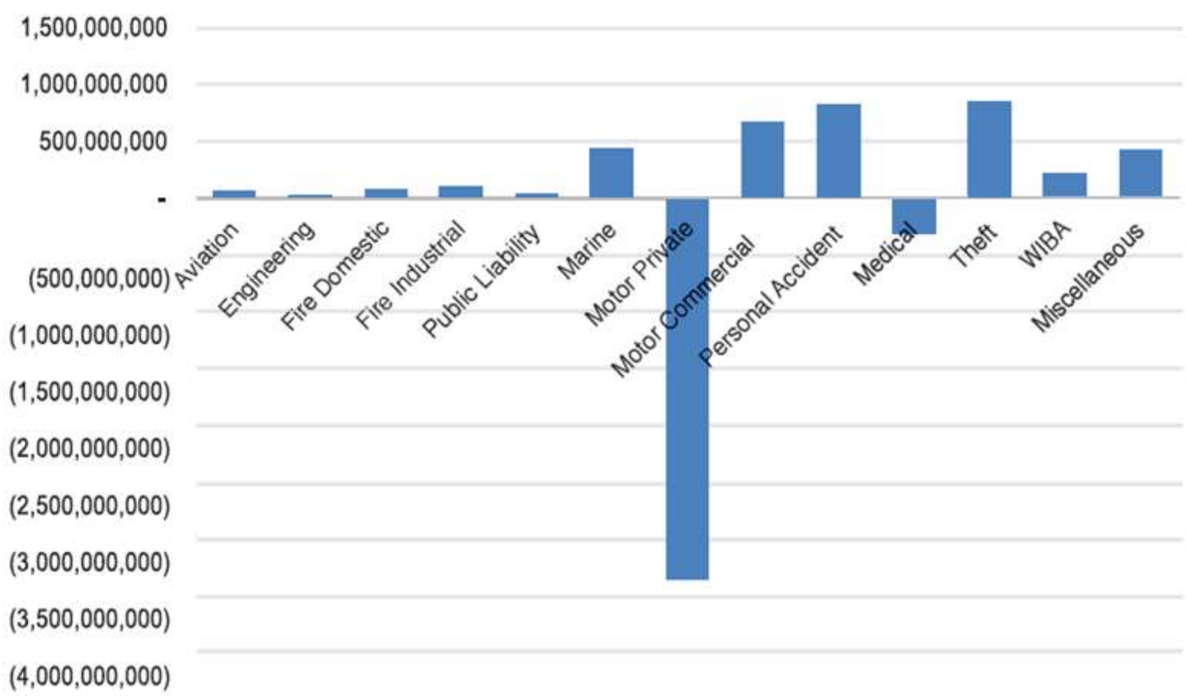

Figure 2.2: : Reported KSH loss/profit per class of insurance (Adopted from AKI Report 2015)

2.2 Motor Vehicle Premium Determination Techniques

Insurance companies operate under the main concept of pooling, i.e. the sharing of losses. Pooling spreads the losses incurred from a few customers over the entire group, such that an average loss value is substituted for the actual loss incurred. Furthermore, when losses are pooled, many exposure units are grouped together in a manner that it is possible to apply the law of large numbers to provide a substantially accurate prediction of future losses (Rejda, 2015). "The law of large numbers" implies that the greater the number of exposure units, the more certain insurers can be in estimating the premiums. This is because they are more capable of accessing the estimate and size of future loss payments. Using this estimate they can work out an appropriate charge that enables them to compensate for the losses (Bickelhaupt, 1983).

A "payment of fortuitous loss" is a payment that has not been foreseen, an unexpected payment that results from chance. In other words, the loss must be accidental. The law of large numbers is based on payments of fortuitous losses which occur randomly. The more such losses occur, the more disadvantaged an insurer becomes (Redja, 2015).

As the Insurance industry continues to grow in Kenya, research by the Association of Kenya Insurers (AKI) has shown that customers often pick an insurance company as purely based on the premium to be paid (AKI Annual Report, 2015). This has made the need for accurate ratemaking more important than ever. The SAS Institute, (SAS, 2011) stated that multiple factors can be used to determine premium rates, and that as competition increases, insurers are introducing new and innovative rate structures. The critical question in ratemaking is, "What risk factors or variables are important for predicting the likelihood, frequency and severity of a loss?" Although there are many obvious risk factors that affect rates, subtle and nonnutritive relationships can exist among variables that are difficult, if not impossible, to identify without applying more sophisticated analyses.

In addition, insurers must identify the risks they don't wish to underwrite as well as answer such questions as: "Are 30-year-old drivers less expensive to insure than 18-year-old drivers?" and "What is the relationship between claims severity and a driver's educational background?" Traditional univariate analysis methods are outdated, and insurers have turned to multivariate statistical techniques such as generalized linear modelling to understand the relationships between multiple risk variables. Finally, insurers need to consider marketing costs, conversion rates and customer buying behavior in accurately pricing insurance products. It is becoming more evident that the key to gaining an edge in today's competitive market is the ability to quickly and efficiently explore, understand and act on the data to improve price competitiveness.

2.3 Europe and East African Motor Insurance Determination.

The fundamental difference between motor vehicle insurance in Europe versus East Africa lies in the object of focus for the insurance. While European motor insurance regulations generally require the driver to have motor insurance prior to driving a motor vehicle, in East Africa the motor insurance law requires that the motor vehicle should have insurance before being used (Campbell et al., 2012). This difference in principle will result in different variables that can be used for premium determination, even if the East African law was revised to be equivalent to the European law. As such, the multivariate statistical techniques, such as generalized linear modelling, that are needed to understand the relationships between the multiple risk variables at play within the European context are based on the individual.

Evidence from linear modelling in Europe considers that 16-year-olds are high-risk drivers. Kenya is a country in East Africa and motor vehicle usage is an important factor for prediction of likelihood, frequency and severity of a loss. Private motor insurance applicants have taken advantage of the low rates in the private motor insurance class and have used they cars for commercial purposes. As a result of the negative skew in premium-claim, many insurance firms have resorted to crude strategies for dealing with the unsustainable loses, such as by completely locking out certain motor vehicles from their insurance offers. Another complaint relates to the biased allocation of premium that has often overlooked the key factors used in premium insurance determination before arriving at a standard premium rate cover. An example used; motorcycles have the highest risk factors but they have the lowest premium rates of 2.5%. For example, Uganda – a neighboring East African country from Kenya – has acknowledged that motorcycles have high risk associated with them. Thus, although the Ugandan law is very similar to Kenya with respect to motor insurance cover, Uganda's premium rates are at 10%, which is much higher than Kenya's 2.5%.

The Insurance Regulatory Authority (IRA) of Uganda has published the minimum premium rate as shown in Table 2.1

MINIMUM PREMIUM RATES

	RECOMMENDED RATE
A. MOTOR: Rating based on usage of vehicle.	
1. **MOTORCYCLES** **Excess:** Own Damage: 10% of claim, minimum Shs.100,000/= Theft & Total Loss: 15% of claim, minimum Shs.100,000/=	10%
2. **MOTOR PRIVATE** Saloons, Station Wagons, & Pick Ups (for private use only) **Excess:** Own Damage: 10% of claim, minimum Shs.100,000/= Theft & Total Loss: 15% of claim, minimum Shs.100,000/=	4%
3. **MOTOR COMMERCIAL** (For fare paying passengers & goods (own goods & general cartage), goods for sale including raw materials)	
i) Saloons, Station Wagons, & Pick Ups (for commercial use only) **Excess:** Own Damage: 10% of claim, minimum Shs.100,000/= Theft & Total Loss: 15% of claim, minimum Shs.100,000/=	5%
ii) Lorries, Prime Movers/ Tractor Heads, Trailers & Haulers **Excess:** Own Damage: 10% of claim, minimum Shs.100,000/= Theft & Total Loss: 15% of claim, minimum Shs.100,000/=	6%
iii) Tankers a) Petrol, Gas, Chemical tankers b) all other tankers **Excess:** Own Damage: 10% of claim, min. Shs.200,000/= Theft & Total Loss: 15% of claim, min.Shs.200,000/=	7.5% 6%
iv) Buses: (Minibuses/ Taxis, coasters & buses) a) PSV. b) PMO/ Corporate & School buses **Excess:** Own Damage: 10% of claim, min. Shs.200,000/= Theft & Total Loss: 15% of claim, min. Shs.200,000/=	7.5% 6%
v) Special Types such as Ambulances , Hearses, Bullion Vans, Fire Fighting Vehicles, Farm Vehicles, Motor recovery vehicles, etc **Excess:** Own Damage: 10% of claim, minimum Shs.100,000/= Theft & Total Loss: 15% of claim, minimum Shs.100,000/=	4%
vi) Road Registered Mobile plants and construction vehicles i.e. concrete mixers, bull dozers, excavators, road rollers, cranes and others of the like nature **Excess:** Own Damage: 10% of claim, min. Shs.500,000/= Theft & Total Loss: 15% of claim, min.Shs.500,000	3%
vii) Motor Trade Risks a) Road Risks (Limit of liability per number plate) b) Internal Risks (Limit of Liability) **Excess:** 10% of claim, min. Shs.250,000/=	5%

Figure 2.3: Minimum premium rates in Uganda (Adopted from IRA of Uganda Source; http://ira.go.ug/minimum-premium-rates.pdf)

2.4 Insurance Pricing Options

The term "Premium" is defined as the amount paid by the insured to the insurer for the protection the insurer undertakes to give for the insured event. The amount of premium to be paid for a particular insurance contract depends on the terms of the contract and its value which is usually computed from the amount of risk. The premium is computed at the individual risk level, and price variations dependent on certain parameters like type of vehicle-body type and purpose of vehicle, i.e. private, commercial, taxi, rental etc. There are two options commonly adopted in motor insurance premium setting for the particular risk, i.e. either computed at the individual level based on past loss ratio or claim experience of the pool or by insurer association in each respective country and/or by regulatory body or the government. Accordingly, the market will have uniform pricing system which is centrally set by government or different price that will be freely set by individual insurer in various countries. Each of these exclusively or simultaneously exists with their own advantages and disadvantages. The detail of each pricing options is presented as follows:

2.4.1 Centrally Determined Price Option

Most regulators have placed various restrictions on insurance policy coverage and rates, particularly for non-life motor vehicle insurance. Statutory tariffs are commonly imposed on compulsory insurance and strictly regulated by national legislation due to its sensitive nature. They prevent unsustainable price wars (Ernst & Young, 2011). Unlike mature markets, prices are free and only subject to tough competition in most developing countries, compulsory premiums are statutory, and subject to government oversight, either directly or through some more complex governance process. In this case, the government is in charge of setting statutory in the form of setting of minimum and maximum or only maximum prices (Gönülal, 2010). In tariff pricing option, motor insurance premium rate was prepared in table format which will be applicable to all insurance companies and open to the public. The premium is calculated according to the risk + office expenses + other charges (Ben-Shahar & Porat, 2016). The rationale behind statutory prices is, generally, a combination of arguments;

- The primary public policy concern underlying this regulation relates to the willingness and

- Ability of consumers to observe and monitor the financial health of their insurer, especially when insurance is made compulsory.

Public policy in the motor vehicle insurance such as third-party insurance primarily seeks to overcome consumers' difficulties in observing and monitoring the financial health of their insurer, both before and during the lifetime of the insurance contract (Gönülal, 2010). Different countries have pursued motor insurance tariff liberalization at different times and in different ways. Insurance in the European Union (EU) has been deregulated fairly recently (mainly over the period 1968–94). Only a few years ago, just the car and the cover had an influence on the rate, but already we are seeing additional factors such as "Bonus/Malus" and "Age of driver" becoming common place. This trend towards a more complicated structure looks set to continue (Gönülal, 2010). In most countries statutory tariffs do not reflect claims costs and necessarily involve an element of cross subsidy and sometimes inhibit innovative pricing. As a result, this is an area where knowledge of the distinctive features is important and the trained of future market evolution will be a move towards partial de-tariff of MTPL insurance premium rating and full risk-based pricing typically correlates to market profitability and stability (Ernst & Young, 2011).

2.4.2 Free-Market Priced Option

In many developed and mature markets, prices are free and only subject to tough competition. In general, insurance risks can be very volatile, and the cost of meeting claims is constantly under pressure from inflation and other upward trends as a result the premium rating are very sensitive issue in insurance market. Most insurance policies are normally issued on a one-year renewable basis as a result in a competitive environment, an insurer can lose good business or gain bad business very quickly if its rates become out of line with the rest of the market. In this case, it is of vital importance for an insurer to keep rates under constant review and to amend them as necessary (Gönülal, 2010).

Competitive pricing will bring innovative product, new business idea, cheaper rates and variable discounts. However, the lower premiums force the insurer to be more selective to underwriting, and to accept only the better risks. It will also keep costs down and had economy administration and stricter claims procedures. This will have an impact on the general objectives of motor insurance (Einav et al., 2010). In addition, many insurers have been suffering losses after charging inadequate premiums. Thus, competition is not the solution to every problem. In some areas, the introduction of competition does not lead naturally to market-driven benefits. The most obvious of these are situations in which the collective community stands to benefit from an investment in which the effects go beyond an individual insure (Gönülal, 2010).

This is the issue most country regulatory bodies are concerned and keep prices is under close scrutiny. Particularly in motor insurance the government needs to adopt various strategies based on the actual case in order to maintain proper balance between “Nontariff” and “tariff” pricing systems.

2.5 Challenges Facing Kenya Motor Liability Insurance

Motor insurance in Kenya is governed by the Insurance Motor Vehicle Third Party Risks Act. Chapter 405 of the Kenya insurance Act provide compulsory insurance to protect the public for road traffic injuries arising from the use of motor vehicles and to forestall the effects of adverse selection on the insurers (Kenya Insurance law, 1946).

Following the act, all Kenya insurance companies started to underwrite Public Service Vehicles (PSV’S) besides other conventional insurance cover to ensure compliance by motor vehicle owners. These companies have however, faced enormous challenges in the underwriting of PSV insurance business and a situation that has from time to time threatened to destabilize the entire insurance industry. This has made insurers reluctant to underwrite this business because of high loss ratios. In due time, the Government introduced compulsory motor pool in 1985 which also failed and was abolished in 1989 (Macharia, 2009).

The insurance industry in Kenya was in a crisis particularly third-party motor insurance since the year 1989. It has been under scrutiny with little or no action in all aspects. Among others; the main contributory factor has been the premium rate determination and monitoring, lack of integrated data and co-operation between the insurance industry, increased fraud and malpractice. This is perpetrated by a syndicate of fraudsters comprised of the traffic enforcement agencies, ambulance chasing lawyers, medical doctors, private investigators, insurance companies’ staff, shareholders, claimants, law enforcement agencies and the

Judiciary. This is due to lack of a structured benefits scheme for various injuries, the courts were awarding excessive damages on motor vehicle accident compensation claims and insurance company had unlimited liability on accident causing damages with a limited capital base (Macharia, 2009).

For these and other reason Kenya insurance industries were faced wave of collapse and a crisis in the country, as there was no insurance provider for motor insurance, particularly PSV “matatu”, which is unregulated in nature. The collapse of insurance companies has diverse implications on innocent third parties, other insurers, policyholders, suppliers of goods and services and the industry at large. This continues to stir erosion of consumer confidence towards the insurance industry (Macharia, 2009).

The effects of an insurance company’s collapse are now dealt with under the Policyholder’s Compensation Fund established by Insurance Regulations of 2004. The fund contributions are 0.25% of the premium payable by the policy holder per insurance policy and a similar amount by the insurer. The Fund was established for the primary purpose of providing compensation to the individual policyholders up to a maximum of 100,000 Kenya shillings, provided that insurance company that has been declared insolvent and was operational in January 2005. Its’ secondary purpose is that of increasing the confidence of the general public in the insurance sector (Makove, 2011). Indeed, liability insurance in Kenya was compulsory, the method of calculating compensation in respect of the person injured was pain and suffering which is more subjective and main reason for victims invariably seeking legal redress. In that regard a structured compensation liability schedule has been adopted that laying value (compensation percentage of maximum value) on various injuries and respective category of injuries (degree of disablement) to be compensated to an injured party (Kenya Insurance law, 2013).

Therefore, Kenya motor vehicle third party liability insurance has the major challenges for insurance companies, owner of motor vehicles, innocent third-party road victims and the policy makers’ in general. The situation is more serious due to the unregulated operation PSV’S and inadequate premium rate determination and subjective method for personal injury claim compensation. This problem will be expected to improve with the implementation of policyholder’s Compensation Fund and structured compensation liability schedule. In addition, the introduction of phased premium adjustments and establishing optimal regulatory system on PSV's operation is an ending solution. These require maintaining proper balance between the interest of the public and the insurance company in general is the focus of the lesson.

2.6 Insurance Prediction Methods using Information Technology

Different estimation techniques have been proposed in the past to be used to estimating the level of risk vs premium payment. The table below lists some of them.

Technique	Methods		
	1. Econometric Model	**2. Regression Model**	**3. Tree-based Classification**
Description	A system of interdependent regression equations to describe economic sales or profit activity. Parameters of the equations are determined simultaneously. These models predict quite accurately but are relatively expensive to develop	Relates risk to other economic, competitive or internal variables and estimates an equation using least squares method. Relationships are analysed statistically. They predict quite accurately, especially in short and medium term	Consist of tree-like nested if-then statements for predictors that partition data. This approach generates a structure of "nodes" and terminal "leaves", within which a model is used to predict the outcome. They predict very accurately, especially in short and medium term
Accuracy Short term (0-3 months) Medium term (3 months – 2 years) Long term (above 2 years)	Good to very good Very good to excellent Good	Good to very good Good to very good Good	Very good Very good Good
Identification of deviations	Excellent	Very good	Excellent
Typical applications	Estimation of premium by product classes, forecasts, margins	Estimation of premium by product classes, forecasts, margins	Estimation of premium by product classes, forecasts, margins
Data required	Monthly or quarterly history to obtain meaningful relationships. To have more observations than the no of independent variables	Same as in econometric models	Same as in econometric models

Cost of estimating	Moderate to low	Low	Very low
Time required to develop	2+ months	Depends on ability to identify relationships	< 2 months
References	Evan (1969)	Clelland et al. (1966)	Quinlan (2004)

From the table above, machine learning methods such as tree-based classification techniques are the best for predicting insurance premium. Several reasons explain their popularity: First, they are very easy to interpretable and communicate to a nontechnical audience. Second, they can handle both numeric and categorical predictors without any pre-processing. Third, they perform feature selection and can handle missing values explicitly, and any missing value is used as another level/numeric value (Kuhn and Johnson 2013). However, there are known drawbacks worth mentioning, such as model instability leading to possibly big changes in the tree structure given a small change in the data, and the suboptimal performance due to their naturally defined rectangular regions. Further, standard trees are prone to overfitting, since they may find splits in the data that are peculiar to the specific sample being analysed (Kuhn et al. 2015).

Random Forest models (Breiman, 2001) have been developed to overcome these problems. They make use of bagging approach, i.e. fitting different trees on bootstrapped data samples and can be used for both classification and regression problems. Furthermore, they are computationally attractive since a high number of independent decision trees on different bootstrapped data samples can be built at the same time during the training phase and the final predictions are obtained by averaging the individual scores of each tree. The trees are ideal for bagging, since they can capture the complex structures of interaction in the data and if developed sufficiently in depth, they have a relatively low distortion and, as the trees are notoriously noisy, benefit greatly from averaging.

2.7 Conceptual Framework

The figure below shows the most important criteria for analysis derived from the literature review, setting the theoretical framework and preparing for chapter 3. This figure structured the operationalization of my research with the study variables as will be explained in the section below and will become expounded in the following chapter.

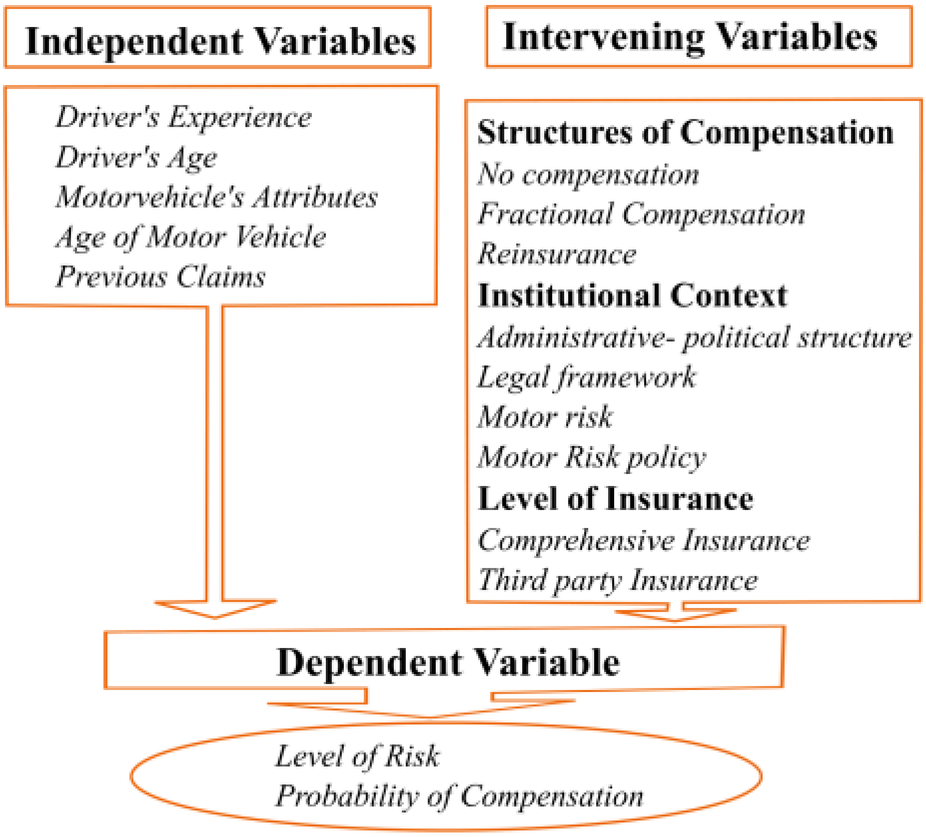

Figure 2.4: The project Conceptual framework

CHAPTER THREE

3.0 Research Design and Methodology

This Chapter presents the research approach and strategies for fulfilling the study objectives and conducting this study. The approach for collecting and analyzing qualitative data will be discussed within the specified scope of the study. The Chapter also deliberates on the data collection and sampling techniques that were applied for identifying the relevant participants. The chapter further reports of the research strategy for analyzing the qualitative data through measures of central tendency (mean, standard deviation) and generalized linear regression.

3.1 Research Approach

There exist two key approaches in research which clearly identify the investigation path should be followed (Saunders et al., 2009). The first approach is deductive, which is concerned with building a theory with hypotheses and seeking to test the validity. The second is the inductive approach which focuses on collecting empirical evidence and building a theory from the findings (Creswell et al., 2011).

Deductive research is the dominant approach in natural sciences where a theory is rigorously tested in a controlled context according to the prevailing laws of the environment which makes predictions about the outcome (Saunders et al., 2009). In inductive research, the researcher observes the patterns that are derived from empirical evidence, and infers the findings to the theory. This is called “theory building research” (Bhattacherjee, 2012). The methodology of this study makes heavy use of inductive research to collect new evidence from the primary data that is collected by the researcher. It must also be noted that deductive research has been applied, because the study uses advice from the literature review to establish a framework for collecting data and assessing the evidence.

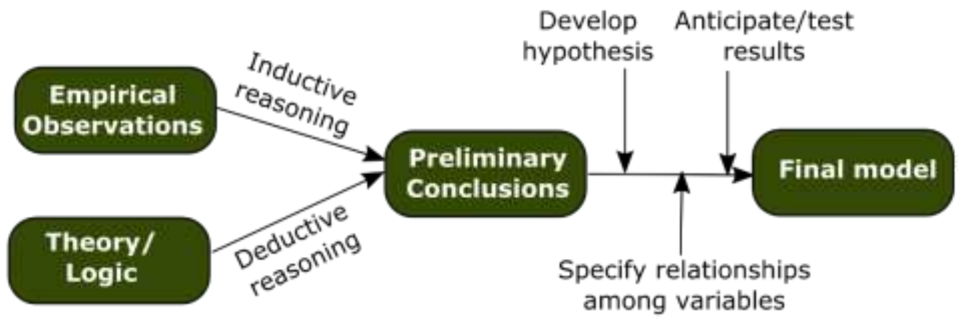

Figure 3.1: The applied research approach for the study

Since the research hypothesis of the study is subjective in nature, an inductive approach was followed using a qualitative method of research in order to build upon previously identified theories or create new ones by inferring from patterns formed from the observed findings using empirical data.

3.2 Research Strategy

The research has been conducted using multiple sources and types of primary data to inform the triangulation of evidence, increase the reliability of results and properly corroborate the data gathered from other sources (Yin, 2009). The case study research collected data using various methods such as questionnaires, document analysis, participant observation, and collecting insurance company statistics.

3.2.1 Addressing the Research Objectives

The main objective was ***to establish machine learning models for regulating motor vehicle premiums among insurance companies***. In order to meet this objective, specific objectives were addressed in the methodology section as follows:

Objective 1: *To identify and classify sources of risk in motor vehicle.* This was achieved through review of the literature and insurance company documents, as well as by collecting questionnaire responses from employees from the insurance sector. The sampling technique and size is explained in Section 3.4.

Objective 2: *To identify the variables for motor vehicle premium determination*. This was achieved by extensively reviewing the literature (Chapter 2) and critically assessing different expert recommendations, to establish a conceptual framework for data collection and analysis. This resulted in the identification of an effective conceptual framework suitable for analyzing the local insurance data. This framework is shown in Section 3.2. Descriptive analysis also helped to establish the variables.

Objective 3: *To establish machine models for estimating motor vehicle premiums in insurance industry.* Descriptive strategies were used to identify specific patterns and trends in the collected data from insurance companies. This informed the identification, design and implementation of suitable classification models using decision trees and a regression model for estimating motor vehicle insurance premium. The models are explained in Section 3.5 and findings from estimating these models are in Chapter 4.

Objective 4: *To test and evaluate the models for effectiveness using sample*. Using the results of the generalized linear regression model, the model was tested for success. This involved using statistics such as error rates from classification, R^2, and the Akaike Information Criterion (AIC), a measure of statistical quality control (Hu, 2007). Results of models' goodness-of-fit tests are shown in Chapter 4.

3.2.2 Case Study Designs

Four types of case study designs were considered within the domain of data collection and analysis, according to Sanders et al. (2011):

i) Single case (holistic design) – where the unit of analysis is one single subject, and where the subject is unique.

ii) Single case (embedded design) – this contains more than one unit of analysis present in an organisational context.

iii) Multiple case (holistic design) – where the phenomena being studied is compared with other cases across different organizations.

iv) Multiple case (embedded design) – where each case study is uniquely addressed according to the phenomena being studied.

A holistic design was applied to create models that offer analytical evaluation of

3.3 Data Collection

The collection of data consists of primary data from questionnaires and secondary data collected in Nairobi like official documents, reports, policy documents and scientific articles. Data used to test the model was collected from multiple different resources. The primary data was directly addressing the specific objectives 1 and 2. The secondary data helped in analyzing motor risk policies in Nairobi and also the role of insurance companies in the specific cases. Scientific articles e.g. journals gave insight to studies that have already been done before. The assumptions made were that the variables chosen for premium estimation were ultimate and true inputs to premium determination and that the data collection periods were optimum. Data collection was carried out by physically moving from organization to organization. Pre-established salient points and semi-structured questionnaires were used to uncover deeper meanings through a qualitative research design method using a multiple case study.

3.3.1 Sampling Techniques

Key-informant respondents from insurance company senior and middle level staffs were chosen through non-probability sampling approach; purposive or judgmental sampling method. This method allows the researcher to identify the right interviewee and collect relevant data for the study.

The target groups for the questionnaire-based survey were insurance companies' operation staffs and motor insurance policyholders. The data from the two respondent categories was drawn from different sampling designs for different situations. In this study, proportional random sampling technique was employed to identify respondents from insurance companies' operation staffs represented by their branch offices.

The criteria and methods utilized for respondents' selection and contact are as follows:

- Insurance company's branch offices, which has been in operation to underwrite motor insurance since September 2011 are selected proportionally.
- Insurance officers directly involved in motor policy underwriting and/or claims handling are randomly selected from proportionally chosen insurance companies' branch office.
- Anyone who buys motor insurance policy and met accidentally at proportionally selected insurance company's branch office counter during the month of November 2016.

- The reason why the researcher chose the month, November is due to the relatively less burden of the insurance officers and the convenience to collect data from the target groups.
- The prioritization of accidental technique over the other methods is because of the potential benefits that it requires minimum effort and manpower, saves time and promotes access to the most convenient group of respondents at a time.

3.3.2 Sample Size

Different methods and approaches of sample size determination exist, but there is no agreement on what the minimum sample size should be (Hailu et al., 2012). This study employed the formula suggested by (Singh& Masuku, 2014) to calculate the sample size as follows:

$$\mathbf{n} = \frac{\boldsymbol{N}}{\mathbf{1} + \boldsymbol{N}(\boldsymbol{e})^{\mathbf{2}}}$$

Where n is the sample size, **N** is the total population size, and **e** is the level of precision (sampling error).

During the survey study, there were 13 insurance companies operating in Nairobi. These insurance companies operate through 125 branch offices within Nairobi and are considered as a study population. The minimum sample size for this study for a population, N of 125 respondents at 95% and with the degree of variability, **e** at 0.05 was calculated as:-

$$n = 125/ [1 + 125(0.05)^2]$$

This implies that out of the intended population of 125, the study required responses at least 95 participants, and a total of 95 questionnaires were distributed. The data collection also incorporated sample motor insurance policyholders for the purpose of triangulation and validation of data collected from insurance companies' officers. The non-probability sampling technique - accidental sampling was employed to pick respondents of policyholders from proportionally selected sample insurance companies' branch office counter.

The data was collected during the month of November 2017 and it took four weeks to distribute the questionnaires and collect both the responses. Every week approximately three branch offices were managed to collect data from insurance officers and policyholders.

To address the research objectives and supplement the finding from primary source the researcher additionally used three years secondary panel data obtained from a sample of five insurance companies. The five insurance companies were selected through non-probability sampling - convenience method on the basis of higher number of motor policy certificate issued during the month of November 2015 and data availability, which account 65.3% of the total motor insurance certificate issued, i.e. representative of the insurance industry.

3.3.3 Qualitative Data Collection/ Variables

The main method of collecting data from individual participants (both the employees and policy holders met) was by administering questionnaires. Participants were asked descriptive questions about various aspects of motor vehicle insurance. The researcher organized for a meeting with various participants and had a short preliminary session with them. The questionnaires were randomly distributed to the participants who filled them and handed them to back to the researcher. The researcher also mailed the questionnaire to various participants, who filled them and mailed back to the researcher.

3.4 Structure of the Research Questionnaires

The questionnaires adopted in this study were divided into three sections:

3.4.1 General Demographic Details

This section describes the general statistics of the research population and presented for statistical reasoning. It consists of five questions to find out the demographic features of the respondents such as gender, age, marital status, level of education and work experience.

3.4.2 Attitude, Knowledge and Perception

This part of the questionnaire was designed to assess motor insurance compensation for each respondent category to evaluate the effectiveness of "Vehicle Premium Against Risks Proclamation" enforcement.

3.4.3 Non- Life Motor Insurance Premium vs Risk

This section of the questionnaire address issues to measure attitude, knowledge and perception of the insurance officers and policyholders towards motor insurance implementation and management.

3.5 Analysis Strategy

The data analysis method applied for this study comprises both quantitative and qualitative methods. Quantitative method was used with a view to properly address the research questions and qualitative analysis method was employed to analyzed response on open-ended questionnaires. The data collected through close-ended questionnaires were analyzed with descriptive statistics. The data are presented using tables and charts. The researcher also utilized various descriptive statistical techniques such as mean, percentage and frequency distribution tables. The Waikato Environment for Knowledge Analysis (Weka) was used to conduct decision trees classification. For modeling the regression analysis, SPSS version 16.0 software tool was used.

To determine the variation between premium payment and motor vehicle risk factors, a regression model was applied in the form:

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \beta_4 x_4 + \beta_5 x_5 + \beta_6 x_6 + \beta_7 x_7 + e$$

Where β_0 is the intercept of the regression, β_1 to β_5 are the slopes representing the expected y value, and e represents the errors that should be normally distributed with mean 0 and equal variance. The independent variables are as follows:

x_1 is a Policyholder (or driver) Experience

x_2 is the Policyholder Age

x_3 is the Policyholder Education

x_4 is the Previous History

x_5 is the Vehicle Type

x_6 is the Vehicle Age

x_7 is the Lease Condition

3.6 Ethical Considerations

The researcher asked and received permission from the administration/authorizing offices of the targeted organizations before conducting the data collection. The insurance firms were reassured of the privacy and confidentiality on the information which they provided.

In order to ensure that prospective respondents willingly participated in the study, a short preliminary session was held with each participant. They were informed of the significance of their contributions and the scope within which their responses would be applied for the research work. Their consent was sought before they were asked to participate in the study and the respondents will be requested to exercise honesty when filling the questionnaires. The field of name in the questionnaire was optional, and it was therefore not used in the analysis.

CHAPTER FOUR

4.0 Data Analysis and Interpretation of Results

This Chapter provides the results from the statistical analysis and interpretation of the provided results. Data was analyzed using Statistical Package for the Social Sciences (SPSS) version 16.0 and findings presented in form of tables, pie charts and bar graphs. Interpretation was done using descriptive tools such as the frequencies and percentages as well as measure of central tendencies which included the mean and standard deviation. The Chapter also provides results of machine learning through decision trees regression modeling using the sampled data.

4.1 Demographic Outlook (Objectives 1& 2)

Demographic findings were based on the age and gender of respondents who completed the survey, the number of years of experience in driving, the employment status and frequency of driving a motor vehicle.

The largest age groups of respondents were aged between 18 years and 30 years, making up more than half of the survey participants. 31 to 40 years had the second largest number of survey participants. These are also usually the age groups where most drivers belong. Only 3% of the survey participants were 51 years and above.

Table 4.1: Age Information

Age of the Respondents		
	Frequency	**Percent**
Below 18 years	4	5.3
18-30 years	39	51.3
31-40 years	21	27.6
41-50 years	9	11.8
51 years and above	3	3.9
Total	**76**	**100.0**

According to gender, more male participants were involved in the survey (59%) than female participants (41%).

Table 4.2: Gender Information

Gender of the Respondents		
	Frequency	**Percent**
Male	45	59.2
Female	31	40.8
Total	**76**	**100.0**

On years of driving experience, most of the participants had less than 1-year experience making up 43%. Most of these people filled the questionnaire on their way to take or renew their insurance policy. The next largest group had up to 3 years' experience (22%). This group is important because of the fact that they can share solid experience about insurance premium payment and claiming. Drivers with 3-5 years' experienced were the smallest group, making up 9%. This group is also important for experience.

Table 4.3: Driving Information

Driving Experience of the Respondents		
	Frequency	**Percent**
Less than 12 months	33	43.4
13 - 24 months	10	13.2
25 - 36 months	17	22.4
37 - 60 months	7	9.2
More than 60 months	9	11.8
Total	**76**	**100.0**

On employment status, about 5% were unemployed, with most being students. The largest number of participants (47%) was employed in white-collar jobs or blue-collar jobs. One person was a volunteer worker and another was suspended from his current job position.

Table 4.4: Employment Information

Employment Status of the Respondents		
	Frequency	**Percent**
Unemployed	4	5.3
Self-employed	23	30.3
Employed	36	47.4
Retired	11	14.5

Other	2	2.6
Total	**76**	**100.0**

Findings of frequency of driving a motor vehicle showed that the majority drive every day but only use their cars to go to work. This makes up 56.6%. 17% operate the motor vehicle throughout the day and use it as an income generating tool. 11% use the motor vehicle about once per week and 12% uses it once or twice per month. Only a few people (4%) use their motor vehicle very rarely but chose to insure it anyway.

Table 4.5: Driving Frequency Information

Employment Status of the Respondents		
	Frequency	**Percent**
1-5 times per year	3	3.9
1-2 times per month	9	11.8
Weekly	8	10.5
Daily	43	56.6
Throughout the day	13	17.1
Total	**76**	**100.0**

4.2 Results from Motor Vehicle Insurance Survey

From this section the results are objective findings that will be presented in aggregated form with mean and standard deviation. Some of these results will also be used in regression analysis.

Table 4.6: Driving Frequency Information

Employment Status of the Respondents			
	Sample (n)	**Mean**	**Standard Deviation**
Insurance payment per month	76	3,950	865
Approximate vehicle value	76	995,511	736.9
Age of vehicle (in months)	76	3.8	2.5
Duration of vehicle ownership (months)	76	2.7	2.6
No of accidents per year	76	0.9	0.4

On average, the participants surveyed pay KSh.3,950 per month, about KSh.50,000 per year. The highest standard deviation is on duration of vehicle ownership, which is almost the same as the mean indicating a lot of variation.

Figure 4.1 indicates that between 2015, 2016 and 2017, the newly registered vehicles of 0-12 months had the highest number of compensation from theft or accident. This indicates that the new cars have the highest risk. The risk reduces as the age of the vehicle increases, except for the last 2 categories of age where it is unclear and compensation values show fluctuating amount.

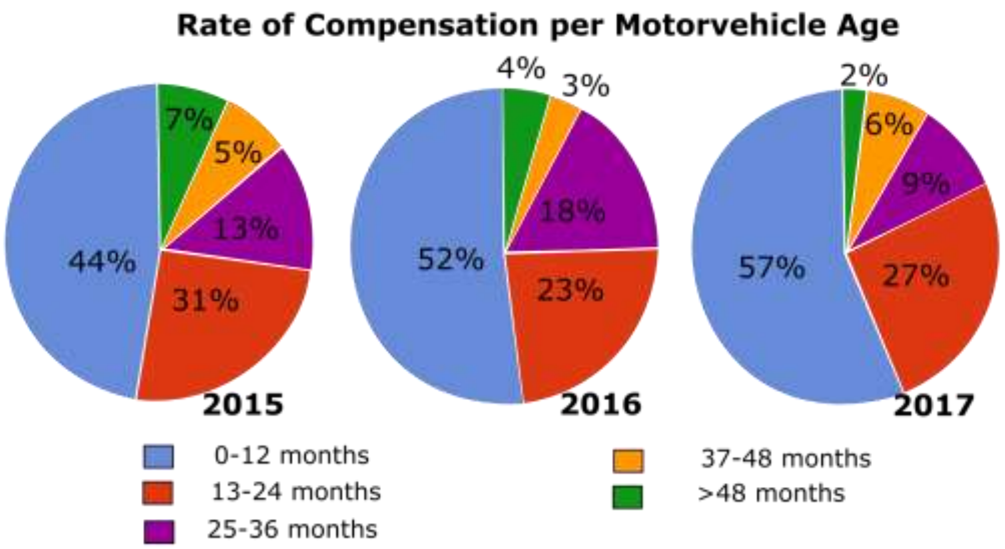

Figure 4.1: Motor vehicle compensation

Figure 4.2 shows the amount of money compensated by the insurance companies in millions of KSh. according to the type of motor vehicle. Luxury and sports cars received the highest compensation, either because of their high value or the likelihood to crash during a race. Private saloon cars and the truck and utility vehicles were almost similar in trends but also received high compensation. Minivans had the lowest compensation rate, especially in 2016 when less than 2M was paid out to compensate them.

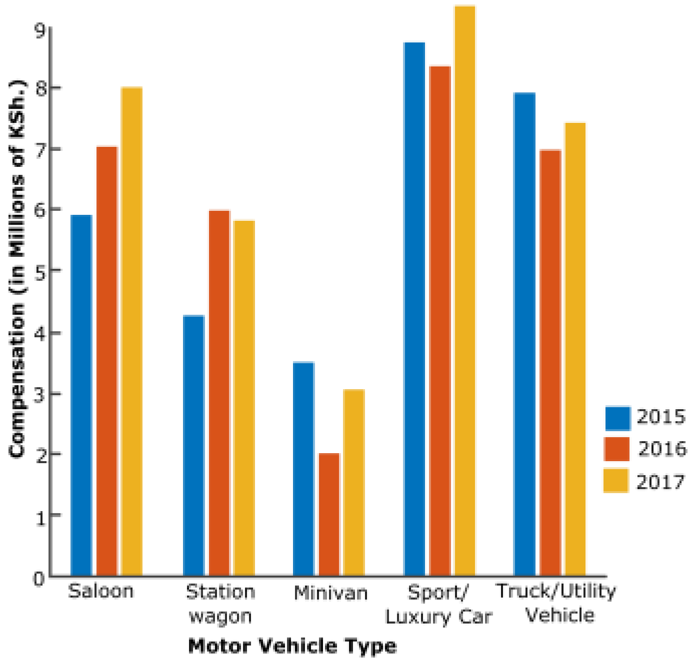

Figure 4.2: Compensation according to the type of motor vehicle

With data from the survey study, analysis of covariance was done for age against the amount compensated in KSh. This was done separately for male and female participants but both genders showed the same pattern. People were more likely to need compensation between 25 and 35 years old. Younger people also required high compensation, but after 60 years the compensation rate was very low.

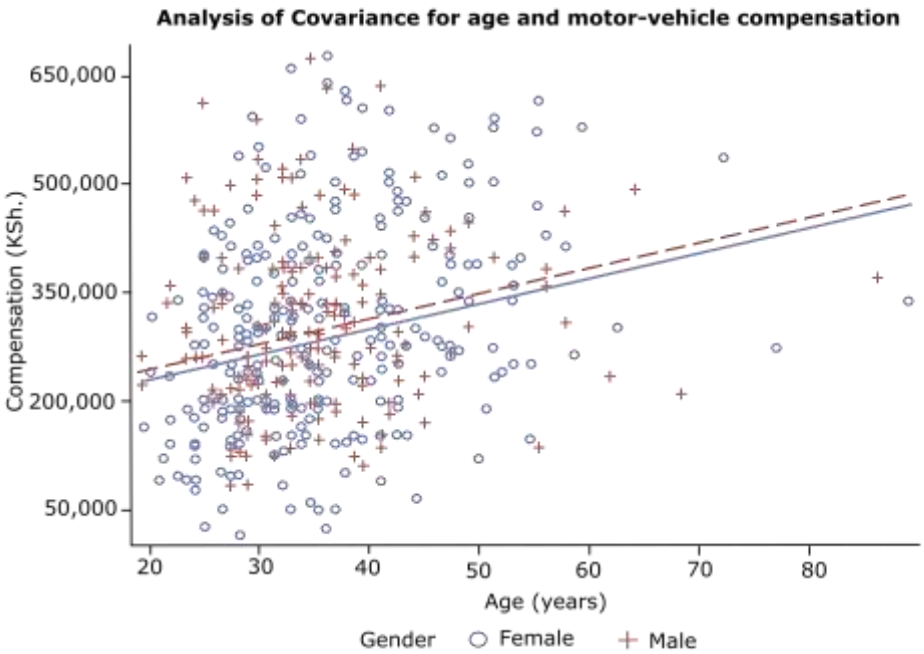

Figure 4.3: Compensation by age and gender

4.3 Classification and Regression Analysis

The next part of the analysis was to develop a predictive model to predict the probability of claim given some possible risk factors on the occurrence of claims from the insurance industry. One of the ways to model insurance risk is through classification with Decision Trees. According to Duncan (1979), an advantage of the Decision Tree is that it is easy to interpret. The Decision Tree also works even if there is nonlinear relationship between variables. Finally, it does not require linearity assumption and it is not sensitive to outliers.

The output from the J48 classifier on 4 fields of the insurance data using 10-fold cross-validation is as shown below:

```
Number of Leaves :  12
Size of the tree :  26
Time taken to build model: 0.03 seconds

=== Summary ===
Correctly Classified Instances        177               79.1333 %
Incorrectly Classified Instances      83                20.8667 %
Kappa statistic                          0.3807
Mean absolute error                      0.1773
Root mean squared error                  0.1885
Relative absolute error                 25.4768 %
Root relative squared error             27.7122 %
Total Number of Instances             260

=== Detailed Accuracy By Class ===
             TP Rate  FP Rate  Precision  Recall  F-Measure ROC Area  Class
             0.662     0.481      0.787     0.762     0.622      0.616    1
             0.519     0.338      0.797     0.619     0.555      0.616    0
Weighted Avg.0.591     0.411      0.792     0.691     0.589      0.616
```

Figure 4.4 shows the output from the classification for policyholders. The significant nodes ($p \leq 0.05$) were displayed with the root of the decision tree being divided by the most important variable being the compensation amount. The policyholder experience highly determines compensation amount. Policyholders who had over 3 years driving experience are less likely to receive compensation because of an accident where PolicyHolder.Experience > 3.1 Years (215, with 173 Yes and 43 No). Age is also a good determinant, but on the negative side of risk. Younger drivers with less than a diploma for education, are the most likely to require high compensation amount where PolicyHolder. Age ≤ 31 Years (198, with 133 Yes and 65 No).

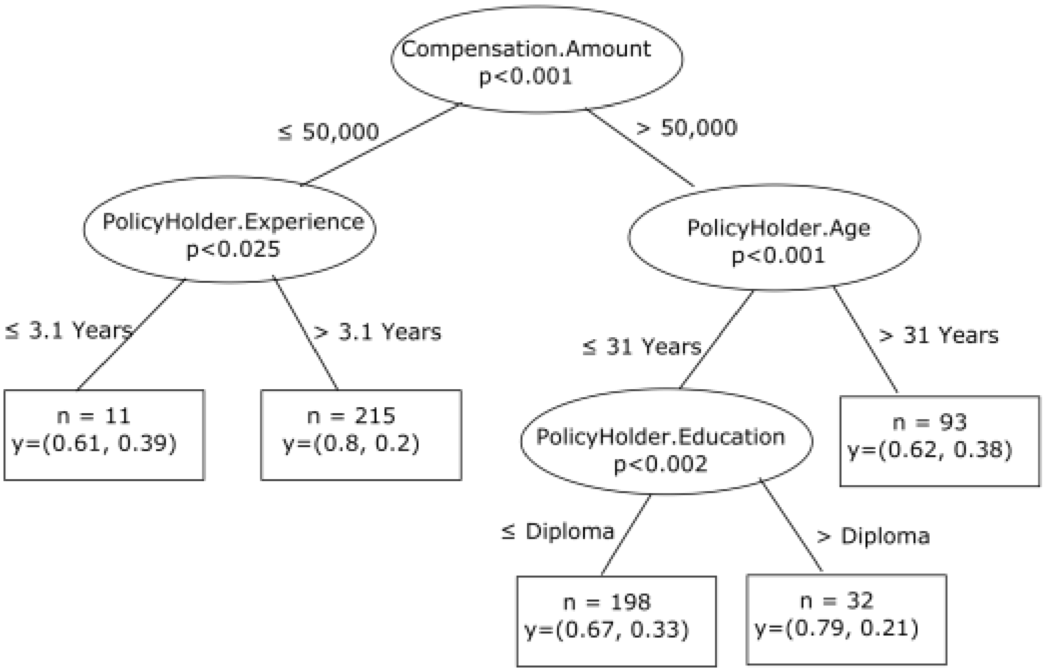

Figure 4.4: Classification output by Policyholder's characteristics

Figure 4.5 shows the output from the classification for motorvehicles. The significant nodes ($p \leq 0.05$) were displayed. The root of the decision tree is divided by the most important variable being the compensation amount. Motorvehicles with no previous accident history were less likely to need high compensation as a result of accidents where Previous.History ≤ 0 (203, with 142 Yes and 61 No). Motorvehicles such as sports cars had the highest risk of getting accidents and receiving high amount of compensation, especially for vehicles driven for not more than 1 year where Vehicle.Type ≤ 2 and Vehicle.Age ≤ 1 Year (76, with 51 Yes and 25 No).

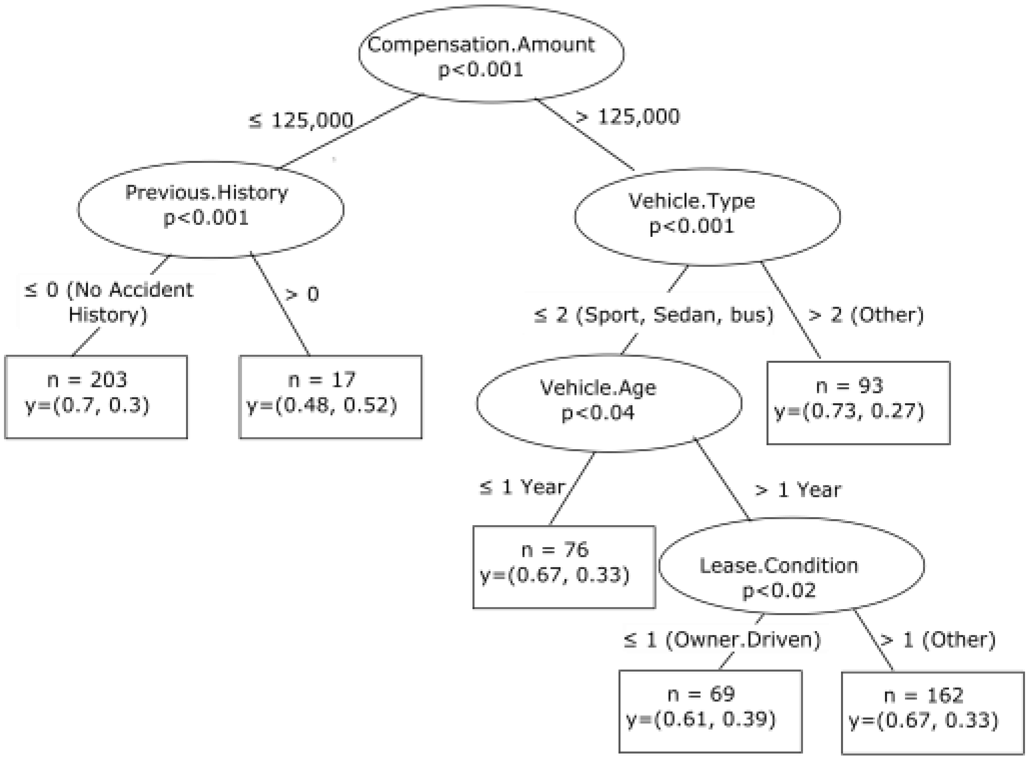

Figure 4.5: Classification output by Motorvehicle's characteristics

As shown on the confusion matrix on Table 4.7, policyholders who did not make a claim in the current year stand 49% chance of no claim in the current year and 7% chance otherwise. Policyholders who made a claim in the current year have a 33% chance of making a claim in the current year. The model is quite accurate as shown in the overall error level of 0.16%. The overall prediction of claim shows that the propensity for a policyholder to make a claim depends very much on the loss history of that claimant.

Table 4.7: Confusion matrix for Probability of Insurance Claim

Confusion Matrix			
Actual	**Predicted**		**Error**
	No	Yes	
No	0.49	0.07	0.12
Yes	0.10	0.33	0.23
Overall Error:			**0.16325**

However, decision trees have disadvantages. According to Smith and Tansley (2004) the decision tree model generally overfits. It means it does not perform well on validation sample, and it assumes all independent variables interact with each other, which is generally not the case every time.

Using results from the decision tree as guidance of which variables were good determinants for compensation, a regression model was created to test the dependent variable (Compensation) side by side with 7 independent variables: Policyholder Experience, Policyholder Age, Policyholder Education, Previous History, Vehicle Type, Vehicle Age and Lease Condition. Table 4.8 shows the estimation results. The previous history is the most significant variable with a t-statistic of 6.121 and *p*-value of 0.000. Policyholder's (or driver's) experience was negative to the probability for claim, meaning that the longer the policyholder had driven the motor vehicle, the less he/she was likely to be involved in an accident. The same with Motor vehicle's age. It seems that older vehicles were less involved in an accident than newer vehicles. According to the regression results, all variables were significant except the policyholder's education level which was found not to be significant to the likelihood of being involved in an accident and claiming from the insurance company.

Table 4.8: Regression Analysis

	Coefficient	Standard Error	t-statistic	p-value	95% confidence interval	
P.EXP	-0.022	0.002	4.123	0.000**	0.013	0.031
P.AGE	0.139	0.094	3.110	0.023*	0.060	0.218
P.EDU	0.307	0.105	1.977	0.061	0.139	0.581
HIST	0.358	0.198	6.121	0.000***	0.290	0.426
V.TYPE	0.128	0.057	2.844	0.021*	-0.096	0.354
V.AGE	-0.012	0.003	3.258	0.010**	-0.018	-0.006
L.COND	0.417	0.126	3.315	0.002**	0.398	0.421
n	260					
R^2	0.745					
Adjusted R^2	0.691					

To fulfil objective 4 which is to test and validate the models for effectiveness, the decision tree analysis applied the confusion matrix to discover a very low error rate. The relative error and the root mean square error were also below 30%. This means that the model was quite accurate. For the regression model, the R^2 value of 0.691 shows that the model predicts high above average. Even the adjusted R^2 which considered that multiple independent variables have been used in the model is still high. The 95% confidence intervals were not very far from the coefficient estimates; therefore the variables were all fitting the data. Therefore, the model is effective.

4.4 Conclusion

The general observation from the results of this Chapter is that several factors are important for determining risk in addition to the motor vehicle insurance premium value. If sufficient variables are used and a proper model is applied, the process of calculating premiums can be very accurate and beneficial to both the insured and the insurer. Such information is very important to starting companies, old companies that wish to implement or expand usage of insurance products, and to the insurance regulators for identifying sources and magnitudes of risk for different insurance products.

CHAPTER FIVE

5.0 Discussion and Conclusion

5.1. Introduction

This is the final Chapter of this thesis and covers a summary of the previous chapters based on the relevant theoretical and practical findings of this work. Furthermore, this chapter discusses the key study results and assesses the relevance of the results against other previous studies. It examines how the study objectives have been met, how existing gaps have been addressed in this study and the value addition of the study to the practices of estimating insurance risk and premium. In short, the author offers the theoretical and industrial implications of the study, with its limitations and propositions for future direction.

5.2. Summary of Findings

The primary objective of this study was to establish machine learning models for regulating motor vehicle premiums among insurance companies. Results from this primary objective and associated specific objectives can help to investigate consequences of action by policyholders in a buyer-seller relationship in the car insurance market.

The findings started with descriptive analysis of study participants, where more male participants were involved in the survey than females. Most of the participants had less than 1-year experience making up 43%. The next largest group had up to 3 years' experience. This group is important because of the fact that they can share solid experience about insurance premium payment and claiming. Drivers with 3-5 years' experienced were the smallest group, but this group was also important for the survey study because of sharing experience. The largest numbers of surveyed participants were employed in white-collar jobs or blue-collar jobs. Findings of frequency of driving a motor vehicle showed that the majority drive every day but only use their cars to go to work. Only a few people use their motor vehicle very rarely but chose to insure it anyway.

Newly registered vehicles of 0-12 months had the highest number of compensation from theft or accident. Luxury and sports cars received the highest compensation, either because of their high value or the likelihood to crash during a race. Private saloon cars and the truck and utility vehicles were almost similar in trends but also received high compensation. Minivans had the lowest compensation rate, especially in 2016.

From results of classification and decision trees, the most important variable was compensation amount. The policyholder experience highly determines compensation amount. Policyholders who had over 3 years driving experience are less likely to seek compensation than the drivers with less experience. Age is also a good determinant. Younger drivers with less than a diploma for education are the most likely to require high compensation amount. Cars with no previous accident history were less likely to need high compensation. Policyholders who made a claim in the current year have a 33% chance of making a claim in the current year.

The regression model was used to test how compensation is influenced by 7 independent variables: Policyholder Experience, Policyholder Age, Policyholder Education, Previous History, Vehicle Type, Vehicle Age and Lease Condition. From the key results, experience was negative to compensation, i.e., policyholders with more experience driving were less likely to be involved in an accident. The same with Motor vehicle's age, where older vehicles were less involved in an accident than newer vehicles. All the independent significantly influenced compensation except the policyholder's education level which was found not to be significant to the likelihood of being involved in an accident and claiming from the insurance company.

5.3 Discussion of Results

On average, policyholders pay KSh.3, 950 per month, about KSh.50, 000 per year. The highest standard deviation is on duration of vehicle ownership, which is almost the same as the mean indicating a lot of variation. These findings have been discovered in other studies such as Desalegn, A. (2014) and Poposki et al. (2015). This shows that in general insuring a motor vehicle is an expensive investment. It is important for policyholders to get value for their money, and understanding how key variables influence their ability to receive compensation can help a lot. Other key findings were that the newly registered vehicles of 0-12 months had the highest number of compensation from theft or accident and that Luxury and sports cars received the highest compensation. These two findings are also similar with findings from other researchers (e.g. Müller and Te, 2017; Stevenson et al., 2017; Lesmana et al, 2018) and shows that drivers of such vehicles should be charged a much higher premium than any other groups.

All the results from the first three specific objectives were agreeing with each other, and the results from both decision trees and regression model were quite similar in identifying the most influencing variables. As based on the specific objective 4 which was to test the models, it can be reported that all the models were quite accurate. For example, analysis of covariance was done for age against the amount compensated in KSh. This was done separately for male and female participants but both genders showed the same pattern. People were more likely to need compensation between 25 and 35 years old. Younger people also required high compensation, but after 60 years the compensation rate was very low.

The classification model using decision trees was quite accurate as shown in the overall error level of 0.16%. The decision tree analysis furthermore applied the confusion matrix to discover a very low error rate. The relative error and the root mean square error were also below 30%. This means that the model was quite accurate. The overall prediction of claim also identified that the propensity for a policyholder to make a claim depends very much on the loss history of that claimant. Results from the decision tree model guided the variables which were good determinants to be selected for the regression analysis, and results from the regression estimates matched the findings from decision tree classifications. For the regression model, the R^2 value of 0.691 proved that the model prediction was high above average. Even the adjusted R^2 which considered that multiple independent variables have been used in the model still recorded high values. The 95% confidence intervals were not very far from the coefficient estimates; therefore the variables were all fitting the data. Therefore, the model is effective.

Thus, the findings of this study are expected to help the stakeholders of the insurance industry, such as policymakers, industry professionals, and senior management practitioners, to coordinate the exchange relationship between policyholders and insurers better, considering the ever-evolving complexity and the dynamic nature of the business environment. This will improve integration, and cooperation, which will enhance business and provide sustainable competitive business advantage, offering flexibility and helping the businesses adapt to the ever-changing market.

Several research gaps have been addressed by results of this study. First, there are not many studies in Kenya that deal with uncovering the risk affecting insurance covers both from the policyholder's end and the insurance company's end. Some studies e.g. Tyson (2015) have studied life insurance markets in Sub-Saharan Africa. Thabet (2016) looked at hidden action problems affecting the car insurance market in Egypt. Freudenreich and Mußhoff (2018) looked at how technology adoption can be used to subsidize premium for maize farmers in Mexico. In general studies from the Kenyan context are very limited.

Secondly, many research studies have concentrated on life insurance and neglected non-life insurance which is equally risky and expensive. This study has therefore provided useful knowledge regarding variables affective non-life insurance compensation. Finally, many studies have only preformed descriptive analysis when measuring risk (Guest et al. 2017). This study has extended the analysis and applied machine learning methods in the form of decision trees and regression analysis to accurately estimate the insurance risk, thereby adding significant knowledge.

5.4. Study Limitations and Future Recommendations

Although the results from this study are encouraging, further work is needed before the procedures are used in practice. First, analysis using a larger sample of data is necessary as the study has used a limited sample. Secondly particular attention should be made to the following points.

(1) How well the assumptions behind the credibility model are fulfilled.

(2) Whether improved variables for the model can be found.

(3) How many years of claim statistics should be used, and whether other forms of credibility estimation (such as the evolutionary models suggested by Sundt, 1983) give better results.

(5) How well technical experts can assess the risk premium for car models.

(6) How the method can be applied to the different elements of insurance (fire, theft, third party etc.).

Finally, the results of using the procedures described here should also be compared with present practice, where estimates of risk premium from the identified variables are, in effect, subjectively combined. It is expected that a dramatic improvement will be discovered, and that the proposed procedures described in this thesis will enable the process of motor vehicle insurance classification to be put on a more objective footing.

5.5. Conclusion

This thesis has set out the methodology for combining different variables of risk premium for a particular motor vehicle from different sources of information. It is important to apply the findings to the real world effectively, as the existing theories fully support these. The outcome of this study will be useful for managers and practitioners in the insurance and related industries to improve their relationship with their trading partners, thus strengthening coordination, cooperation, and integration in their operations. This will grow their market share, increase their profitability, and make their business sustainable by offering them a competitive advantage in the market.

Application of the methods to a specific customer gives very positive results. Technical attributes such as car value appear to be helpful in estimating insurance risk, but the age and driving experience are much more significant determinants. It is also apparent that risk premium estimates for car models with few (or no) claim statistics can be dramatically improved. Another useful implication of this study for managers in the insurance sector is that they should develop a comprehensive and flexible pre-screening strategy that can be adapted to the changing market realities and policyholders' expectations, thus minimizing market dynamics. This will drastically reduce or even eliminate problems with inaccurate estimation in the Kenyan car insurance market, as the quality of pre-screening and the frequencies and magnitude of claims by policyholders are largely correlated.

REFERENCES

AKI Annual Report (2015). Association of Kenya Insurers Annual Report, 2015

Ben-Shahar, O., & Porat, A. (2016). Personalizing Negligence Law. NYUL Rev., 91, 627.

Bhattacherjee, A. (2012). Social science research: Principles, methods, and practices.

Bickelhaupt, D. L. (1983). General insurance . Homewood: Richard D. Irwin.

Campbell, J. C., Ikegami, N., & Gibson, M. J. (2012). Lessons from public long-term care insurance in Germany and Japan. Health Affairs, 29(1), 87-95.

Creswell, J. W., Klassen, A. C., Plano Clark, V. L., & Smith, K. C. (2011). Best practices for mixed methods research in the health sciences. Bethesda (Maryland): National Institutes of Health, 2013, 541-545.

Datamonitor (2015) Non-life insurance industry profile global. London: Datamonitor plc. Available at https://en.wikipedia.org/wiki/Datamonitor

Desalegn, A. (2014) Assessment of Motor Insurance Business on Financial Performance of Insurance Company, The Case Of Awash Insurance Company. PhD diss., St. Mary's University.

Diacon, S. (Ed.). (2016). A guide to insurance management. Springer.

Dominique-Ferreira, S. (2017). How important is the strategic order of product attribute presentation in the non-life insurance market?. Journal of Retailing and Consumer Services, 34, 138-144.

Duncan, R. (1979). What is the right organization structure? Decision tree analysis provides the answer. Organizational Dynamics, 7(3), 59-80.

Einav, L., Finkelstein, A., & Levin, J. (2010). Beyond testing: Empirical models of insurance markets. Annu. Rev. Econ., 2(1), 311-336.

Farrell, D., Wheat, C., & Mac, C. (2017). Paying a Premium: Dynamics of the Small Business Owner Health Insurance Market.

Freeck, J., Ford, A. E., Sloan, P., Schmidt, R. T., Angal, M. D., & Kothari, N. (2018). U.S. Patent No. 9,972,184. Washington, DC: U.S. Patent and Trademark Office.

Freudenreich, H., & Mußhoff, O. (2018). Insurance for Technology Adoption: An Experimental Evaluation of Schemes and Subsidies with Maize Farmers in Mexico. Journal of Agricultural Economics, 69(1), 96-120.

Frees, E.W., R.A. Derrig, and G. Meyers. 2014. Predictive Modeling Applications in Actuarial Science. Vol. 1. Cambridge University Press.

Frees, E.W., G. Meyers, and R.A. Derrig. 2016. Predictive Modeling Applications in Actuarial Science: Volume 2, Case Studies in Insurance. International Series on Actuarial Science. Cambridge University Press.

Gönülal, S. (2010). Motor third-party liability insurance.

Guest, R., Tran, Y., Gopinath, B., Cameron, I. D., & Craig, A. (2017). Psychological distress following a motor vehicle crash: evidence from a statewide retrospective study examining settlement times and costs of compensation claims. BMJ open, 7(9), e017515.

Hansell, D. S. (1974). Elements of Insurance... Macdonald & Evans.

Hu, S. (2007). Akaike information criterion. Center for Research in Scientific Computation, 93.

Kuhn, M., and K. Johnson. 2013. Applied Predictive Modeling. Springer

Lesmana, E., Wulandari, R., Napitupulu, H., & Supian, S. (2018, January). Model estimation of claim risk and premium for motor vehicle insurance by using Bayesian method. In IOP Conference Series: Materials Science and Engineering (Vol. 300, No. 1, p. 012027). IOP Publishing.

Müller, D., & Te, Y. F. (2017, December). Insurance premium optimization using motor insurance policies—A business growth classification approach. In Big Data (Big Data), 2017 IEEE International Conference on (pp. 4154-4158). IEEE.

Macharia, R. (2009). The Motor Insurance Industry in Kenya: Adopting the No-Fault Insurance System.

Makove, S. M. (2011, May). African policy approaches: Microinsurance in Kenya. In AIO–A2ii Regulators' Workshop.

Manyara, C. G. (2016). Combating road traffic accidents in Kenya: A challenge for an emerging economy. In Kenya After 50 (pp. 101-122). Palgrave Macmillan, New York.

Munge, K., & Briggs, A. H. (2013). The progressivity of health-care financing in Kenya. Health policy and planning, 29(7), 912-920.

Mwangi, M., & Murigu, J. W. (2015). The determinants of financial performance in general insurance companies in Kenya. European Scientific Journal, ESJ, 11(1).

Peters, J. A. J., Ferguson, D., Madigan, R., & McKenna, T. (2016). U.S. Patent Application No. 14/607,433.

Poposki, K., Kjosevski, J., & Stojanovski, Z. (2015). The determinants of non-life insurance penetration in selected countries from South Eastern Europe 1. Economics and Business Review, 1(3), 20.

Pozzolo, A. dal. 2011. "Comparison of Data Mining Techniques for Insurance Claim Prediction." Master's thesis, University of Bologna.

Rejda, G. E. (2015). Social insurance and economic security. Routledge.

Sanders, N. R., & Wagner, S. M. (2011). Multidisciplinary and multimethod research for addressing contemporary supply chain challenges. Journal of Business Logistics, 32(4), 317-323.

SAS (2011). SAS Institute Inc., Cary, NC, USA.

Smith, L., & Tansley, J. (2004). "Decision tree analysis." U.S. Patent Application No. 10/406,836.

Stevenson, M., Harris, A., Mortimer, D., Wijnands, J. S., Tapp, A., Peppard, F., & Buckis, S. (2017). The effects of feedback and incentive-based insurance on driving behaviours: study approach and protocols. Injury prevention, injuryprev-2016.

Sundt, B. (1983). Finite credibility formulae in evolutionary models. Scandinavian Actuarial Journal, 1983(2), 106-116.

Thabet, E. (2016). Hidden action problems: The case of insurers and business policyholders in the Egyptian car insurance market (Master's thesis, Høgskolen i Molde-Vitenskapelig høgskole i logistikk).

Tyson, J. E. (2015). Life Insurance Markets in Sub-Saharan Africa. London: Overseas Development Institute.

Yang, Y., Qian, W., & Zou, H. (2018). Insurance premium prediction via gradient tree-boosted Tweedie compound Poisson models. Journal of Business & Economic Statistics, 36(3), 456-470.

Yin, R. K. (2009). Case study research: Design and methods 4th ed. In United States: Library of Congress Cataloguing-in-Publication Data (Vol. 2).

APPENDIX A: QUESTIONNAIRE

This questionnaire is divided into four parts namely:

SECTION A: Demographic, Ownership and Driving Information

SECTION B: Motor Vehicle Insurance Information

SECTION C: Compensation Information

SECTION D: Challenges of Motor Vehicle Insurance

I am a student of MSc. in Information Systems Management doing my project on Motor vehicle risk and premium determination. Your participation in this survey will help me a lot to complete my research study. The survey responses will be treated with strict confidence and for academic purposes only. Result will be reported in aggregation and any private information will be coded.

Please answer the questions below by either ticking in the appropriate box or writing the appropriate answer.

SECTION A: Demographic, Ownership and Driving Information

(Please tick as appropriate)

1. What is your age bracket?

Below 18 years [ ] 18-30 years [ ] 31-40 years [ ] 41-50 years [ ] 51 years and above [ ]

2. Gender: *Male [ ]   Female [ ]*
3. Highest academic qualification:

None [ ] Certificate [ ] Diploma [ ] Undergraduate [ ] Postgraduate [ ]

4. Employment status:

Unemployed [ ] Self-employed [ ] Employed [ ] Retired [ ]

Other (specify) ……………………………………………………………

5. How long have you been driving?

Less than 1 year [ ] 1-2 years [ ] 2.1-3 years [ ] 3.1-5 years [ ] more than 5 years [ ]

6. Which vehicle do you drive?

Saloon [ ] station wagon [ ] minivan [ ] Sports/luxury car [ ] Truck or utility vehicle [ ]

7. How often do you drive your vehicle?

1-5 times per year [ ] 1-2 times per month [ ] weekly [ ] daily [ ] throughout the day [ ]

8. How many vehicles do you own?

None [ ] 1 [ ] 2-3 [ ] 4-5 [ ] more than 5 [ ]

9. How many months has it been since you purchased your vehicle?

Less than 12 [ ] 12-24 months [ ] 25-36 [ ] 37-60 [ ] more than 60 [ ]

10. How long (in months) has it been since the car was registered in Kenya?

Less than 12 [ ] 12-24 months [ ] 25-36 [ ] 37-60 [ ] more than 60 [ ]

SECTION B: Motor Vehicle Insurance Information

(Please tick as appropriate)

1. What company is the current policy with?

(specify) ………………………………………………………………

2. How many years have you been with the current company?

Less than 1 [ ] 1 [ ] 2 [ ] 3[ ] 4 and above [ ]

3. How much do you pay as insurance cover per month?

None [ ] 100-2,000 [ ] 2,001-5,000 [ ] 5,001-7,000[ ] 7,001 and above [ ]

4. Has there been any lapse in coverage?

5. If yes, indicate why this happened:

(specify) ………………………………………………………………

6. If yes, indicate the dates this happened:

(specify) ………………………………………………………………

7. What is the approximate value of the vehicle?

(specify in KSh.) ……………………………………………………

8. Which kinds of anti-theft devices do you have in the vehicle?

None [ ]   alarm only [ ]   tint [ ]   immobilizer [ ]   multiple devices[ ]

9. Have you lost your vehicle to theft in the past three years?

Yes [ ]   No [ ]

10. If any, indicate the year that this happened:

(specify) ………………………………………………………………

11. How many accidents have you had this year?

None [ ]   1 [ ]   2 [ ]   3[ ]   4 and above [ ]

12. If any, indicate the rate of vehicle damage:

None [ ]   very slight [ ]   Moderate [ ]   quite wrecked[ ]   write-off [ ]

13. Were you at fault?

No [ ]   I don't know [ ]   Yes[ ]

14. How many accidents did you have last year?

None [ ]   1 [ ]   2 [ ]   3[ ]   4 and above [ ]

15. If any, indicate the rate of vehicle damage:

None [ ]   very slight [ ]   Moderate [ ]   quite wrecked[ ]   write-off [ ]

16. Were you at fault?

No [ ]   I don't know [ ]   Yes[ ]

17. How many accidents did you have in 2015?

None [ ]   1 [ ]   2 [ ]   3[ ]   4 and above [ ]

18. If any, indicate the rate of vehicle damage:

None [ ]   very slight [ ]   Moderate [ ]   quite wrecked[ ]   write-off [ ]

19. Were you at fault?

No [ ]   I don't know [ ]   Yes[ ]

Questions 20-25 to be filled by those insuring their vehicles to be driven by another operator

20. Has your operator been convicted of a crime resulting from the use of a motor vehicle or been convicted of theft of a motor vehicle?

No [ ]   I don't know [ ]   Yes[ ]

21. Has your operator been convicted of fraud or intent to defraud involving an insurance claim or application for insurance?

No [ ]   I don't know [ ]   Yes[ ]

22. Has your operator been denied payment of a claim in excess of KSh. 50,000 under a motor vehicle policy due to fraud/intent to defraud involving a claim/application?

No [ ]   I don't know [ ]   Yes[ ]

23. Has your operator knowingly provided materially false/misleading information in connection with an application/renewal/claim under an insurance policy?

No [ ]   I don't know [ ]   Yes[ ]

24. Is any household member's driver's license currently suspended or revoked?

No [ ]   I don't know [ ]   Yes[ ]

25. Has any company declined or cancelled auto insurance due to non-payment for you or any listed driver in the past 2 years?

No [ ]   I don't know [ ]   Yes[ ]

SECTION C: Motor Vehicle Compensation Information

This section is to be filled by only those who indicated in Section B that they had accidents or theft (Questions 9-19)

1. If you lost your vehicle to theft in the past three years, were you compensated?

Yes [ ]   No [ ]

2. If yes, indicate the amount in KSh:

(specify) …………………………………………………………………

3. How long (in days) did it take to compensate?

0-2 [ ]   3-7 [ ]   8-21 [ ]   22-45[ ]   46 and above [ ]

4. If your vehicle had an accident this year, were you compensated?

Yes [ ]   No [ ]

5. If yes, indicate the amount in KSh:

(specify) …………………………………………………………………

6. How long (in days) did it take to compensate?

0-2 [ ]   3-7 [ ]   8-21 [ ]   22-45[ ]   46 and above [ ]

5. If your vehicle had an accident last year, were you compensated?

Yes [ ]   No [ ]

6. If yes, indicate the amount in KSh:

(specify) …………………………………………………………………

7. How long (in days) did it take to compensate?

0-2 [ ]   3-7 [ ]   8-21 [ ]   22-45[ ]   46 and above [ ]

8. If your vehicle had an accident in 2015, were you compensated?

Yes [ ]   No [ ]

9. If yes, indicate the amount in KSh:

(specify) …………………………………………………………………

10. How long (in days) did it take to compensate?

0-2 [ ]   3-7 [ ]   8-21 [ ]   22-45[ ]   46 and above [ ]

SECTION D: Challenges of Motor Vehicle Insurance

Using the Likert Scale provided, rate the extent of satisfaction with premium payment and vehicle compensation by ticking against the most appropriate response (1=Completely disagree, 2=Disagree, 3=Not sure, 4=Agree, 5=Completely agree). You can base this using your own experience or the experience learnt from others:

No.	Statement	1	2	3	4	5
1.	The cost of insurance premium is reasonable					
2.	It is easy to file a claim with your insurance					
3.	You have received bonuses and incentives from your insurance company on the insurance premiums					
4.	You have received benefits from the insurance policy that you currently have.					
5.	The insurance company pays promptly when a loss occurs					
6.	Insurance company representatives are very helpful for filing a claim and getting compensation.					
7.	The motor vehicle premium amount is charged fairly					
8.	The benefit has been much larger than the amount paid for insurance of the motor vehicle.					
9.	The cost of insurance goes very much lower each year.					
10.	You always pay comprehensive premium insurance for your car when new					
11.	It is not worth to pay for comprehensive premium insurance after 6-7 years of using a motor vehicle					

Other (specify)

……………………………………………………………………………………………

……………………………………………………………………………………………

……………………………………………………………………………………………

……………………………………………………………………………………………

Thank you for completing this questionnaire.