

**JOMO KENYATTA UNIVERSITY OF AGRICULTURE**

**AND TECHNOLOGY (JKUAT)**

**DEPARTMENT OF COMPUTING**

**ICS 2406: COMPUTER SYSTEMS PROJECT**

**PROJECT PROPOSAL REPORT**

Title: A **WARRANTY FAILURE PREDICTIVE MODEL USING CLASSIFICATION  
ALGORITHIM (decision tree and logistic regression algorithms)**

Author:

NAME: AREBA BOSIBORI NORAH REG NO: SCT211-0095/2016

SUBMISSION DATE: SIGN:

COURSE: BSC COMPUTER SCIENCE

Supervisor 1:

Name: Prof. Waweru Mwangi Sign: -------------- ------ Date: ---------------

Supervisor 2:

Name: Prof. Silvester Kiptoo Sign: -------------------------- Date: -----------

ABSTRACT

Warranty is a critical element in the marketing of products. It provides assurance to customers that the manufacturer will provide compensation, through repair, replacement or refund, for purchased items that do not perform satisfactorily.

This has become more critical with the increase in consumer expectations that has occurred over recent years and the passage of legislation demanding better customer protection.

Offering warranty has serious implications for manufacturers. Although offering better warranty terms may give a manufacturer a marketing advantage over competitors, this entails an additional cost, namely that associated with the

servicing of warranty claims. Depending on the product and the manufacturer,

these costs typically vary from 1 to 10% of the sale price of the item, and may have

serious implications with regard to the manufacturer’s reputation and the

proﬁtability of the business. Warranty claims and supplementary data contain useful information about product quality and reliability. Analyzing such data can therefore be of benefit to manufacturers in identifying early warnings of abnormalities in their products, providing useful information about failure modes to aid design modification, estimating product reliability for deciding on warranty policy and forecasting future warranty claims needed for preparing fiscal plans. In the last two decades, considerable research has been conducted in warranty data analysis (WDA) from several different perspectives

INTRODUCTION

No manufacturer sets out to disappoint (let alone melt) a customer, but defects can exist so most provide a warranty to keep their customers satisfied. But anticipating warranty costs for a product or business can be challenging, even risky. Setting aside too little money or too many spare parts may lead to hardship if an unanticipated problem arises. Setting aside too much money ties up resources that might be used more effectively elsewhere. Many systems have been designed to support warranty systems in different companies for different products. Most warranty systems use warranty claims data to determine the failure characteristics of a product. Typically, the failure distribution and its parameters are determined using product manufacturing data for each month of production and the corresponding monthly failure counts derived from the warranty claims. If the data is collected systematically, the product ages at the times of failure can be derived. However, our experience shows that, in many cases, it may not be possible to know the failure ages of components. The information available each month might be limited to the volume of shipments and total claims or product returns. In such cases, the data hides the component age at the time of failure. This research shows that when the failure history information is incomplete, the failure distribution of the product can be determined using Bayesian analysis techniques applicable for handling incomplete data. Whear, Froning, and Noal, (2011) states that human beings always search for a methods tools or techniques that reduce the human effort for performing a specific task efficiently, in machine learning algorithms are designed in search a way that they learn by themselves on how to perform a task efficiently.

There are many applications in which failures are dependent upon usage, not time. For example, in the automotive industry, the failure behavior in the majority of the products is mileage-dependent rather than time-dependent. These kinds of products present a challenge for data analysis. For surviving units still working in the field, how do we know their usage (life) and how do we incorporate it into the estimation of the failure distribution?

The research I decided to undertake this project is on the reliability analysis that is the prediction of a warranty failure. Some factors that contribute to warranty failure are the analysis of warrant policies, cost analysis. Early detection of a significant warranty failure using the change point detection of failure rates can be relevant to the industry; this will help prevent customer churn from their products.

Analyzing their normal behavior and alerting sales professionals when a customer is likely to churn out can be of help to increase their revenue.

BACKGROUND INFORMATION

Research has been outgoing in most warranty systems and how predictive models can be used warranty analysis focusing on the product reliability with the warranty to either maximize their profits or to minimize the total costs (Henry, Collum, Al-Attar, & McLeod, 2011). Reliability is an important aspect of product perception and manufacturers are compelled to take corrective actions on the items failing within the warranty period. They only consider the quantitative aspects and interactions among decisions on reliability warranty, price, and production. However, the use of reliability of products distributes in different geographical regions exhibits a significant difference because the working conditions and service times differ from region to region (Barité, 2018). Analysis conducted based on extended services

Warranty-related decisions are made from the product life cycle and from the business perspective, which links technical and commercial issues.

Machine learning helps to identify the hidden patterns that hard not been previously recognized. Its computation is fast compared to the traditional rule-based. Depending on the particular application, the data that is analyzed can consist of either historical records or new information that has been processed for real time analytics uses. More advanced types of data analytics include predictive analytics which seeks to predict customers’ behavior, equipment failure and other future events and machine learning. Data analytics can be used in warranty analysis to gain insights in customer information and help in making decisions on warranties claims to customers .

Post-sale service is an integral part of the supply chain system. The main aim of this researches to reduce the warranty failure service cost and increase customer satisfaction. This research first aims to determine the root cause of the warranty failure and thereafter sends feedback to manufacturers to avoid future product failures. This task is accomplished by the proposed multi-agent framework which inherits number of specific task oriented autonomous agents such as service agent, fault

.

PROBLEM STATEMENT

Initially, warranty data analysis has been mainly focused on looking for unique methods to estimate warranty claims from warranty and to estimate product field reliability data with poor quality. Whereas, researches are focusing to develop early warning algorithms and to propose suggestions on design modification are these are extremely important for manufacturers as well as consumers

Unlike warranty estimation warranty prediction based on failure rates, that only relate to the finance aspects of a manufacturer and the consumer, an effective early failure prediction system for predicting warranty and effective design modification to reduce risk over consumers and significant property losses for manufacturers, organizations and individuals as well may be required

PROPOSED SOLUTION

The proposed solution is seeking to reduce the risk of failure analysis, hence appropriate differential classification will not only save the costs spent on unnecessary purchases of goods that have poor warranty terms but also save them from stocking products that will not have a higher return.

OBJECTIVE

**General**

The aim of this study is to develop a tool that can be used to discriminate warranty failure and success using classification models that uses the date of failure, Time of failure, Number of returns as the model parameters**.**

**Specific**

* To develop a warranty failure analysis tool using classification algorithms
* To analyze the factors associated with the failure using decision tree algorithm and logistic regression
* To reduce the risk of finding an appropriate classification algorithm that will save cost spent on poor warranty terms.

## Research questions

What are the main aspects of warranty failure?

Which are the previous used models that have met user requirements?

What have previous developers done to develop models in efficient and effective manner?

What are the concepts of ensuring machine learning is implemented on the model?

RESEARCH METHODS

Rapid Application Development

After considering different types of software development methodologies the one that emerged as the most appropriate method for this project was the rapid application development methodology. Rapid Application Methodology makes use of computer aided design tools which will be used in the development if the project. RAD will be the most appropriate for this project as RAD works best for projects whose scope is small and the work should be broken down into manageable chunks taking into consideration the short time frame of the project. It would be appropriate to select a software development methodology that allows for the project to be developed over a short space of time. In RAD model the functional modules are developed in parallel as prototypes and are integrated to make the complete product for faster product delivery.

RAD Model Design

RAD model distributes the analysis, design, build, and test phases into a very short, iterative development cycles. Below are phases of RAD.

Business Modeling

The business model for the product under development is designed in terms of flow of information and the distribution of information between various channels. A complete business analysis is performed to find the vital information for the business and what are the factors driving successful flow of information.

Data Modeling

The information gathered in business modeling phase is reviewed and analyzed to form sets of data objects vital for the business. The attributes of all data sets is identified and defined. The realization between these data objects are established and defined in details in relevance to the business model.

Process Modeling

The data objects sets are defined in the data modeling phase are converted to establish the business information flow needed to achieve specific business objectives as per the business model. The process model for any changes or enhancements to the data object set is defined in this phase. Process descriptions for adding, deleting, retrieving or modifying a data object are given.

Application Generation

The actual system is built and coding is done by using automation tools to convert process and data models into actual prototypes.

Testing and Turnover

The overall amount of testing time is reduced in RAD model as the prototype and independently tested during every iteration. However, on the other hand the data flow and the interfaces between the components need to be thoroughly tested with complete test coverage. It reduces the chances of major issues since most of the programming components have been tested.

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Activities** |  | **Jan** | | | | **Feb** | | | | **March** | | | | **April** | | | | | **May** | | | | | **June** | | | | | **July** | | | | |
| **Week** | **Hrs.** | **1** | **2** | **3** | **4** | **1** | **2** | **3** | **4** | **1** | **2** | **3** | **4** | | **1** | **2** | **3** | **4** | | **1** | **2** | **3** | **4** | | **1** | **2** | **3** | **4** | | **1** | **2** | **3** | **4** | |
| Project Identification | 3 |  |  |  |  | |  |  |  |  |  |  |  | |  |  |  |  | |  |  |  |  | |  |  |  |  | |  |  |  |  | |
|  |  |  |  |  |  |  |  |  |  | |  |  |  |  | |  |  |  |  | |  |  |  |  | |  |  |  |  | |
| Draft Proposal | 8 |  |  |  |  |  |  |  |  |  |  |  |  | |  |  |  |  | |  |  |  |  | |  |  |  |  | |  |  |  |  | |
| Writing |  |  |  |  |  |  |  |  |  |  |  |  | |  |  |  |  | |  |  |  |  | |  |  |  |  | |  |  |  |  | |
| Final Proposal | 8 |  |  |  |  |  |  |  |  |  |  |  |  | |  |  |  |  | |  |  |  |  | |  |  |  |  | |  |  |  |  | |
|  |  |  |  |  |  |  |  |  |  |  | |  |  |  |  | |  |  |  |  | |  |  |  |  | |  |  |  |  | |
| Literature Review | 8 |  |  |  |  |  |  |  |  |  | | |  | |  |  |  |  | |  |  |  |  | |  |  |  |  | |  |  |  |  | |
|  |  |  |  |  |  |  |  |  | |  |  |  |  | |  |  |  |  | |  |  |  |  | |  |  |  |  | |
| Data collection and analyses | 12 |  |  |  |  |  |  |  |  |  |  |  |  | |  |  |  |  | |  |  |  |  | |  |  |  |  | |  |  |  |  | |
|  |  |  |  |  |  |  |  |  |  |  |  | |  |  |  |  | |  |  |  |  | |  |  |  |  | |  |  |  |  | |
| System design | 48 |  |  |  |  |  |  |  |  |  |  |  |  | |  |  |  |  | |  | |  |  | |  |  |  |  | |  |  |  |  | |
|  |  |  |  |  |  |  |  |  |  |  |  | |  |  |  |  | |  |  | |  |  |  |  | |  |  |  |  | |
| System Development | 48 |  |  |  |  |  |  |  |  |  |  |  |  | |  |  |  |  | |  |  |  |  | |  |  |  |  | |  |  |  |  | |
|  |  |  |  |  |  |  |  |  |  |  |  | |  |  |  |  | |  |  |  |  | |  |  |  |  | |  |  |  |  | |
| Testing and Implementation | 12 |  |  |  |  |  |  |  |  |  |  |  |  | |  |  |  |  | |  |  |  |  | |  |  |  |  | |  |  | |  | |
|  |  |  |  |  |  |  |  |  |  |  |  | |  |  |  |  | |  |  |  |  | |  |  |  |  | |  |  |  |  | |
|  |  |  |  |  |  |  |  |  |  |  |  |  |  | |  |  |  |  | |  |  |  |  | |  |  |  |  | |  |  |  |  | |
| Project Report |  |  |  |  |  |  |  |  |  |  |  |  |  | |  |  |  |  | |  |  |  |  | |  |  |  |  | |  |  |  |
|  |  |  |  |  |  |  |  |  |  |  |  |  |  | |  |  |  |  | |  |  |  |  | |  |  |  |  | |  |  |  |  | |

BUDGET

|  |  |
| --- | --- |
| **ITEM** | **COST** |
| Internet bundles and modem | 3500 |
| Data collection | 2000 |
| Printing, copying, flash disk acquisition | 1000 |
| Airtime | 500 |
| **TOTAL** | **7000** |

**Justification**

Predicting warranty failure is critically important to businesses, the retailers, insurance firms, buyers and third‐party payer s alike. The ability to accurately assess outcomes in warranty with different firms by the retailer, however, has proven to be a difficult task and they are left to gamble between which company offers a good warranty terms which will prevent them from making losses is also still a challenge. Although a number of machine learning algorithms, including stimuli model in manufacturing process (M.Gallo *et, al* *2007*), fault detection **(**Q. P.He and J.Wang,*2007***)** and their combinationhave so far been developed for detecting failure, most have achieved little success, particularly when are employed in detecting the rate of item to be faulty other than those from which the score was derived. For example, predicting model for equipment fault detection in semiconductor manufacturing process **(**Munirathinam & Balakrishnan,*2016***)** recommended a better classification algorithm for predicting semiconductor fabrication in manufacturing process.

Most importantly, though, is that previously these prediction tools were largely derived using statistical analysis methods that fail to capture multi‐dimensional correlations that contain wear and tear information. In contrast, machine learning, which has long been used by other fields, including high‐energy physics (SECOM *2020*) to discriminate between signal and background, uses non‐parametric analysis methods to incorporate these interactions. As this approach offers theoretical advantages over ones used in the past, we hypothesized that it could be used to generate a model that more accurately predicts risk among products with failure than previously published scores.

It is also clear as day and night that there is no single algorithm that can be able to predict the warranty failure and prevent the retailers from unforeseen eventuality and thus left with only guess work. The proposed solution is seeking to reduce this risk by trying to prevent this problem. appropriate differential classification will not only save the costs spent on unnecessary purchases of goods that have poor warranty terms but also save them from stocking products that will not have a higher return

**Scope**

The study is limited to only identifying/classifying product warranty terms as ‘failure’, and ‘Success’. It does not aim to identify or classify other factor that may lead to poor outcome in purchase of product or other common dysfunction that normally dictate the need to purchase such as income status. Classification algorithms to be used are limited to classification decision tree (DT) and logistic regression. Datasets to be used for training will be limited to

**Chapter Two: Literature Review**

**2.1 Introduction**

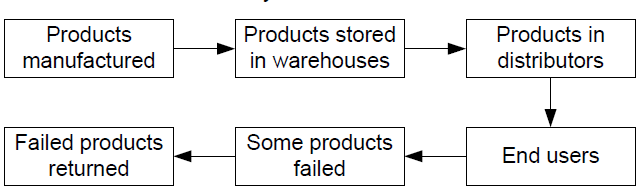
**M**any retailers don’t have ability to know that a particular product will sell well to some clients and others won’t do better because distinguishing failure, and a better buy or investment is still a challenge in product as we speak and is often not priority by retailer as well. However the aspect of correctly classifying warranty failure according to their respective product classes based on terms speak volume if targeted profit margin is to be realized. The factors that determine better warranty terms are determined by different factors, failure is mostly associated with duration between the purchase and the time when the default occurs. On the other hand better sell or success is largely determined by price and assurance that in the instance the default has occurred there will be repair at no cost.

According to (Inbrief, *2020*) warranty is defined as a form of guarantee that a manufacturer gives regarding the condition of it. It also outlines in what circumstances repairs will be made or refunds/exchanges allowed if the product does not perform as expected or described. On the same note warranty will only limit the situation in which the manufacturer will be obliged to fix a problem for example poor workmanship or majorly the defective parts. It is common practice especially in purchasing electric appliances which normally vary for a period of 12 months to two years but longer in relation to the price of the products. Under Consumer Rights Act 2015 (CRA 2015) states that even if someone has no warranty over goods that one has purchased but still they have statutory rights to return the product back within 30 days of purchase and get the full refund. After six months it’s upon the retailer to prove that you caused the problem with the product and if they can’t, they will have to repair or replace or refund you. This also present a challenge as there is no mechanism to ascertain that in practice. Broadly we can say the purpose warranty is to establish liability in the event of failure of an item (Blischke and Murthy, 1992)

Warranty is now a more than requirement in consumer and commercial transaction and can be used to serve many different purposes (wu and li, 2007). Many literature on warranty can be seen, for example in 1996,Djamulidin et al listed more than 1500 papers in this area alone and recently warranty has attracted even more attention (shaomin,2012) as can be seen in review papers (Murthy DNP and Djamaludin,2002) and (Thomas MU and Rao SS, 1999)

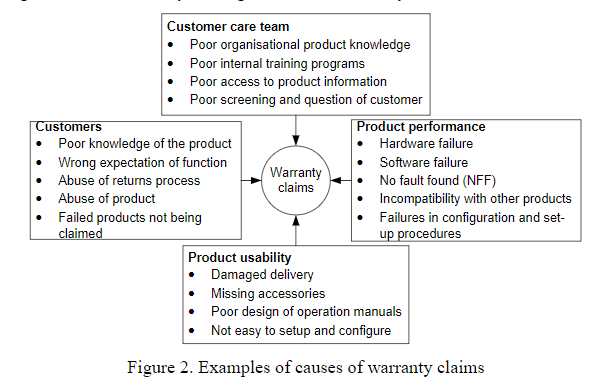
**2.2 Warranty claims**

A typical life cycle of warranty claim is shown below where the process normally starts from product manufacturing time and ends at the time when they are returned to the manufacturer



Source: shaomin 2012: warranty claim data analysis

Warranty claims can be caused by various different forms of failures. For example, the figure below shows some causes of warranty claims, which can be broadly categorized into four types of failures: hardware failures, software failures, human errors and organizational errors. It might be thought that most of publications in the reliability literature simply assume that warranty claims are due to hardware and/or software failures which is not actually the case. Such an assumption might not necessarily hold. For instance, an end-user might claim warranty although the item has not failed, or an end-user might not claim warranty although the item has already failed, see (Wu, 2011)



Source: shaomin 2012

**2.3 Significance of the study**

In manufacturing, failure to predict acceptance by buyers has a consequential effects, customer often go for those warranty that seems to favor them. Proper market acceptance depends on accurately identifying the factors and use of quality data and information at all levels of the organization system. This study serves as a starting point for classifying the warranty based on classification algorithm as the situation in production system to help the production companies make evidence informed decision and for proper planning in conjunction with the approved management processes. The study forms a basis for further research on evidence based management of production in general and specifically lead to generation of new ideasfor better and more efficient product management. The study will look at the use of purchase information data in marketing sector to predict the future market trend

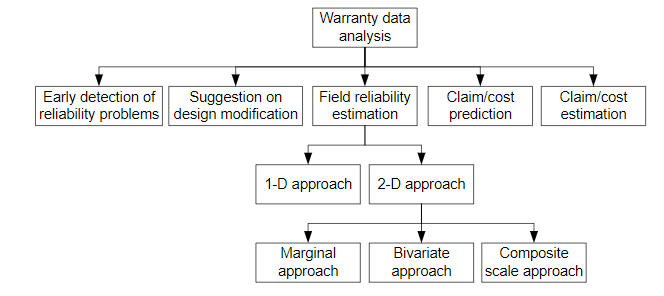
Findings and recommendations of the study would contribute towards the ongoing efforts of consumer satisfaction to develop better management operations system that would benefit industrial company workers identify their weakness and thus propose better ways that could help improve their efficiency through improved information use. The findings of the study will be used by production managers and will rely on haphazard personal experiences or subjective personal judgments or of friends/relative other than base their decisions and actions on concrete evidence and thus help re-invent themselves as problem solvers.

**2.4 Existing models that have been developed, limitations**

Different models have been proposed in classifying warranty the following highlight a few of the major algorithm that have been developed ranging from machine learning to more sophisticated deep learning algorithm

2.4.1 **Using data to predict Warranty Claim**

Warranty claim prediction can be broadly categorized into five areas according to (shaomin 2012), as shown in Figure 3.where, he reviewed existing research in these five areas as shown in the figure below



**2.4.1.1 Early Detection of reliable problem**

The intent of early reliability detection problems is to provide manufacturers with early opportunity to discover early indications of unexpected quality and reliability problems through warranty data analysis. In nutshell, this intention may be achieved through detecting abnormal alternate points in warranty data by using a variety of statistical techniques such as control charts or comparing probability distributions to the benchmarking distribution, or artificial intelligent techniques. This theory proposes use of text mining algorithm since the complaint comes from the customer side. However, products might be modified overtime and again; consequently, warranty claims can be due to a series of changing failure modes of the products. Supervised Text mining can only be used to find the marching pattern and possible to some extent to find the sentiment from the data which can only be classified as positive or negative which cannot really give prediction in production (zheng *et al*, 2019) otherwise this study holds that by using prediction algorithm rather than text analysis can really drive the information home due to low misclassification.

**2.4.1.2 Suggestion on design modification**

Whereas early detection analysis and design modification aims to detect abnormalities from warranty claim. On the contrary, in early detection analysis, period is an important factor: it focuses on techniques that can detect abnormalities at the earliest opportunity. As such, on-line monitoring techniques such as control may be useful. Design modification, on the other hand, emphases on using warranty databases to aid production engineers to change their system design and aims to improve the reliability and quality of their products. Hence, timeliness might be less important. This approach also advocates for the text mining suggested by (Buddhakulsomsiri and Zakarian,2010) to extract meaningful information from databases containing warranty claims using the elementary set concept and database manipulation techniques to search patterns and relationships among occurrences of warranty claims and create IF-THEN rules, where the IF portion includes a set of attributes representing product features (e.g. production date, repair date, mileage-at-repair, transmission, engine type, etc.) and the THEN portion includes a set of attributes representing decision outcome (e.g. problem related labor code).These rules are used to identify root causes of a particular warranty problem or to develop meaningful conclusions. Even though this algorithm has some advantages it however fails short hen some manufacturing companies have no existing databases.

**2.4.1.3. Field reliability estimation**

Field reliability estimation is appropriate for manufacturers when selecting a warranty policy, planting maintenance regimes and preparation of spare parts. When dealing with this it is important to keep in mind that 1) warranty claims may contain data that is not complete 2) warranty claims are only collected from early life of a product which might provide little information in terms of reliability and 3) we also need to know warranty also contain claim that are due to human factors.

Warranty policies can be categorized into one-and two-dimensional. A one-dimensional (1-D) policy is defined by an interval (age only or usage only) as warranty limit. A two-dimensional (2-D) policy is defined by a region in the two-dimensional plane: generally one dimension representing age and the other representing usage. For different types of products, usage can be different, for example, output-based (miles for cars, copies made for photocopier, etc.), time-based (fraction of the time used –air-conditioners, heaters, etc.), stress level (used continuously but different stress levels –air conditioners on hot or very hot days).

**2.4.1.3.1 One-dimensional (1-D) approach**

This majorly depends on usage based, for instance, the warranty limit for a copy machine can be a number of copies that the machine has made since its purchase.in the field of engineering usage based is more relevant and hence modelling of usage accumulation is essential part of reliability analysis

**2.4.1.3.1 .1 Age-based analysis**

In most literature, age-based field reliability estimation has not been the main focuses. The reason might be due to the fact that techniques on estimating field reliability are well established, given complete age information. Approaches to estimating the lifetime distribution include estimating mixed distributions, fitting the Weibull distribution based on a small number of warranty claims, estimating the lifetime distribution considering sales delay which all are based on statistical assumption and cannot be used for prediction purpose of warranty failure.

**2.4.1.3.1 .1.2.Usage-based analysis**

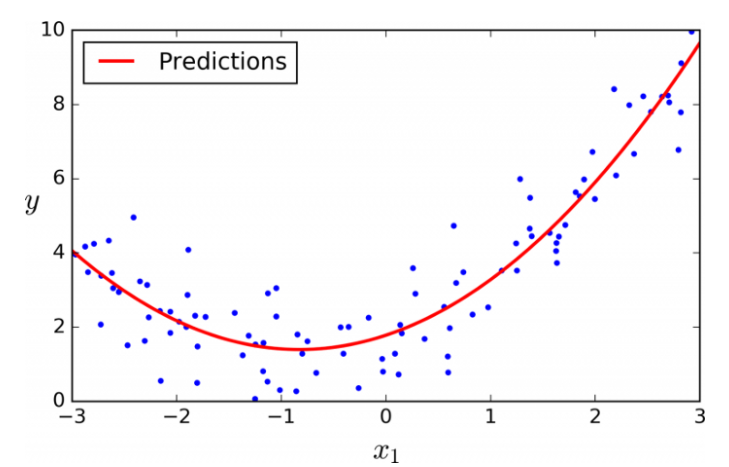
Good estimation of usage-based lifetime distributions for warranty claims requires complete information of the usage intensity of all items, including censored and claimed. However, distribution of the usage intensity items failing within the warranty limit might be different from those of products surviving the warranty. This causes a problem of obtaining censoring times (or usage intensity) for those products that have not reported to the product manufacturer. All the models suggested by the approach mentioned above follows the natural language processing which doesn’t have strong performance compared to classification supervised machine learning such as logistics regression and decision tree (DT).

**2.4.2 Improved machine learning model suggested to work in warranty prediction**

There are two models that have been suggested to improve the performance in warranty prediction which include the following

**2.4.2.1 Regression model to predict remaining useful lifetime (Rul)**

In using this model it is appropriate to use it on data that is static historic data and every event is labeled, again several different types of model must be part of the data set and its preferable to build such model when the degradation process is linear. The most appropriate machine learning algorithm suggested by this approach is ordinal least square (gareth et al, 2017) where linear regression is used as a simple to implement with output that is simple to interpret. Example where this is applicable is would be in a system that predict temperature, since it is a continuous value with estimate that would be simple to train. The figure below show how we can use to predict the temperature failure using the linear regression model

Source: Predicting machine learning maintenance

The linear regression model is very good model but only applicable for continuous data such as temperature and age, suppose that is not available and you are dealing whether the product will fail or succeed it turns out that it is not appropriate for this kind of problem.

**2.4.2.2 Deep learning machine learning models**

In using this approach the most common one is the long short term memory (LSTM) which is especially appealing for predictive maintenance and it employs the use of sequences. Time series data can be used to look back at longer period to detect failure patterns. It requires features such as timestamp and device identifiers. When using this approach the goal to predict at the time “t”, using data up to that time whether equipment will fail soon. The major short fall for this approach of machine learning is that it requires a lot of data and mostly it’s a computer intensive approach which requires a lot systems in place and a lot of skills.

**2.5 Comparative analysis of the models**

In explaining the concept of warranty failure in question, there are a lot of machine learning related studies both with wide-ranging and with improved scope. These studies offers in-depth of the leading theories with underlining aspects of their arguments. General, it can categorized firstly that supervised machine learning works best in these situation especially classification algorithm, secondly, A good model is that one that has low miss classification error and higher accuracy but as (Gareth et al, 2017) argue that it must not be highly biased and overfitting and a balance should be struck. Among all theories seen above, there is a general view that most of the solutions have not actually classified the warranty failure as “failure” or “success” and the only one that attempted that has not tested somewhere else, this study therefore aims to also bridge this gap as well. All these models cannot be implemented in masse due to many obvious constraints. In this regard, I intend to systematically study the most prominent classification theories by investigating the correct classification of warranty failure, competencies and skills which may provide a strong basis to reach an ultimate better production efficiency within the industries.

**2.6 Research gap**

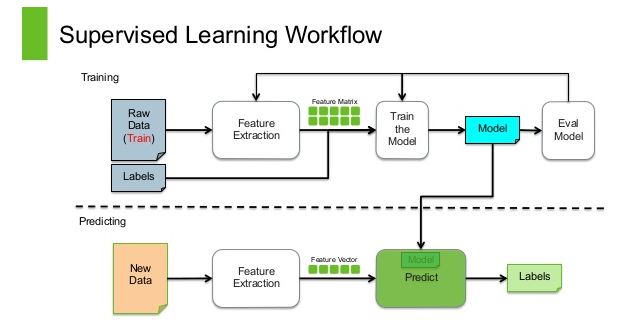
Proper warranty prediction requires high prediction accuracy to prevent the companies from making losses and in the long run make customer have informed decision. The aim of this study is to try to classify correctly the failure only and success only which can supplement other production protocols. The most observations to be drawn from previous sections encompass; terminologies and concept related to theories, history of several models that have been proposed in this field, and conclude what has not been done and remains as a glaring opportunity

The idea of machine learning in manufacturing can change the landscape of the production information systems pillar with the hope of bringing improvement, using a machine learning approach with sequential selection and cross-validation to identify the most informative items dramatically improved performance. Production system varies with different organization and there are some that are more advanced and those that are not advanced and are fully combining machine learning and utilizing production information system due to the available resources, expertise, and capital to implement solutions, but those that are not developed the developing this is not the case. There is no doubt, that adoption of machine learning in predicting warranty failure will improve the production information management by the production workers to provide targeted satisfaction to the customers.

**2.7 Conceptual framework**

The conceptual framework in this study uses the updates from (Alberto et al, 2017) comparative analysis model which are the key success factors for successful implementation of classification algorithm to compare support vector machine (SVM), linear discriminant analysis (LDA), non-logistic regression (NLR), and multilayer perceptron (MLP).Unlike their based approach this study will focus on logistic regression and decision tree (DT) to classify warranty failure as “failure” and “success” separately to see which model will perform well between the two in predicting the patient correctly in terms of sensitivity, specificity, ROC curve etc. the researcher has also modified the model and included other parameters for evaluation that are important for this study including management support, training, perceived usefulness. The best performing model will then be chosen in order to determine the adoption factor like region, state, area, city, consumer, product category and purpose

**Machine Learning Approach**



**CHAPTER THREE**

**RESEARCH METHODOLOGY**

**3.0. Introduction**

This chapter is going to discuss about the methodology to be used, the study population, sample size and sampling procedure.

**3.1 Data source**

In order to carry out this research I am intending to collect data for this research from the ASUS manufacturing repairs database of electronic products, which is a giant manufacturing company in electronics and is the rich source of data used for kaggle machine learning competition. This provides cross sectional information of products brought and the repairs that the customer received from this company.

**3.2 System Requirements**

**3.2.1 Functional Requirements**

The system should be able to:

1. Show information on the current different products failure prediction.
2. Provide a graph showing predictions for the time by which the product will fail.
3. Allow users to filter different products expected age at repair and consumer product category.
   * 1. **Non-functional Requirements**

The system should be:

1. Accurate – The best performing model should be that one that predict product warranty failure with an acceptable level of accuracy. (Target > 82%).
2. Robust – The appropriate model should still function even with missing (incomplete) data in the training data set.
3. User Friendly - It should provide an intuitive interface with relevant information at a glance, and have simple time filtering controls for the predictions.

**3.2.3 System Components**

The system shall have the following components:

**Data Model** Holds the historical data used by the system.

**Predictive Model.** This is a model that will be trained from historical data, then tested to check prediction accuracy. It gets its data from the Data Model, and uses Linear the best performing model for learning between which will be selected between Decision tree (DT) and logistic regression in terms of better performance.

**Prediction Engine** This gets its data from the predictive model, then uses it to make predictions on warranty claim.

**Web Portal:** This is a front-facing portal to be used by the managers and the retailers. It gets predictions from the prediction engine, then visualizes the predictions in graphical charts. It also provides information on the current product status.

**3.3. Research Design**

Research design used for this study will be descriptive survey. A study of the risk factors of warranty prediction is described through a multiple logistic regression and decision tree model.

The choice of the design survey is considered appropriate because it allows verifying whether the studied factors are statistically significant or not. The description through a multiple logistic regression model and decision tree is preferred because the dependent variable is dichotomous (having failure and not failure in warranty) and the independent variables are either continuous or categorical. The use of survey therefore is considered to be more appropriate in terms of resources, time and the overall objective of the study. Quantitative research design will be used in this study to explore and understand warranty prediction and to give systematic empirical investigation of social phenomena using statistical or numerical data or computation techniques (quantitative).

**3.4. Area Of study**

The study will be conducted in ASUS product manufacturing especially the electronics which is the giant in terms of manufacturing of these products .This provides cross sectional information of products brought and the repairs that the customer received from this company.

**3.5. Study population**

This study will be conducted to find out whether factors like region, state, area, city, consumer, product category and purpose cause jointly or partially is associated with warranty failure. Population is a large collection of all subjects from where a sample is drawn (Zikmund, Babin Carr and Griffin 2012). The target population or the unit of observation is a group of individuals, or objects that a sample is drawn for measurement (Kombo and Troomp 2009). A three year period secondary data from ASUS electronic manufacturing product manufacturing containing the customer’s record is used. This period extends from February 2005 to December 2009.

* Data considered in this study are relevant for the following reasons:
* First, the area is considered to be more appropriate in terms of financial resources, time and the overall objective of the study.
* Second, the products is from a manufacturing company and works with the customers living in different lifestyle and all economic categories are represented in the sample.
* Last, the data are recent and may reveal current situation of warranty failure.

**3.6. Sample size and sampling procedure**

Sampling is a process of selecting a number of individuals or objects from a population such that the selected group contains elements representative of the characteristics sought in the entire population. ASUS and the three years 2005, 2006, 2007, 2008 and 2009 were purposively selected according to the objectives of the study. The target population of the study includes in total 359 products from ASUS (2005-2009) dispatched in the following six different sectors as:

|  |  |
| --- | --- |
| **State** | **Total** |
| Andhra Pradesh | 59 |
| Assam | 2 |
| Bihar | 19 |
| Delhi | 22 |
| Goa | 5 |
| Gujarat | 29 |
| Haryana | 1 |
| Himachal Pradesh | 2 |
| Jammu and Kashmir | 2 |
| Jharkhand | 4 |
| Karnataka | 30 |
| Kerala | 28 |
| Madhya Pradesh | 4 |
| Maharashtra | 39 |
| Orissa | 17 |
| Rajasthan | 4 |
| Tamil Nadu | 46 |
| Tripura | 2 |
| Uttar Pradesh | 24 |
| West Bengal | 19 |
| **Grand Total** | **358** |

The sample size for patients is determined using the Yamane (1967) formula. That is:



Where N is the population size and n is the is the sample size while e is the precision level

Concerning our case study; the total number of patients’ folders (N) is 358.Then, by the equation

However, systematic random sampling has been used to select the products sample to be included in the sample size of each State. With this sampling technique, each element has an equal probability of being selected, but combinations of elements have different probabilities. In a population of size N, if the sample desired is of size n, the sampling interval k=N/n; randomly, a number j between 1 and k is selected, and then every kth element thereafter is taken. It means elements j + k, j +2k, j+3k, j +4k, etc. until the sample of size n is completed.

**3.7 Documentary review**

The study will utilize the internet tools such as yahoo, google scholar and scholarly reviewed journals to carry out document analysis of literature and content written in warranty and production management to compare the current trend on other comparative studies carried out on customer data base.

**3.8. Data Collection and Analysis**

Data collection is the process of gathering and measuring information related to the study that helps in answering the research questions. The methods are varied in terms of time, cost of money or other resources at disposal of researcher .The study would have considered all the retail company in the country but ASUS Company was sampled out purposively for the research since it has good data which has also been used by kaggle completion for machine learning and for this purpose the secondary data will be used. Those data were taken from customer database record related to purchase records and repair, covering a four-year period (2005 to 2009).

The following information from the data base of customers was taken: age, gender, occupation status, smoking, alcohol consumption, Cholesterol level, hypertension, family history and having diabetes or not. Python version 3.8 will be used to process data during analysis.

CHAPTER FOUR

SYSTEM DESIGN

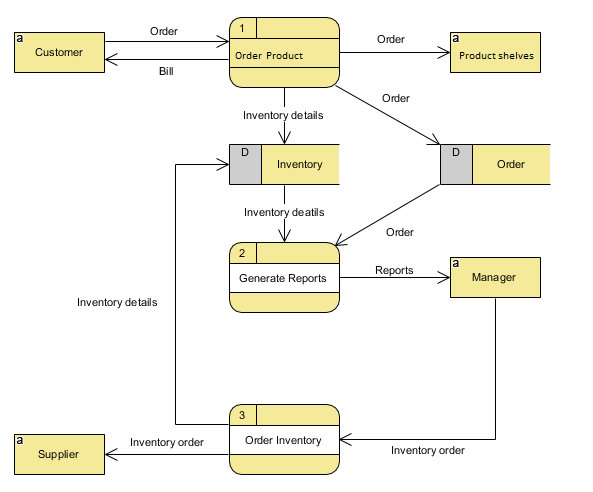
4.1 Data flow diagrams

Figure 4.1 system data flow diagram

4.2 Sequence Diagrams

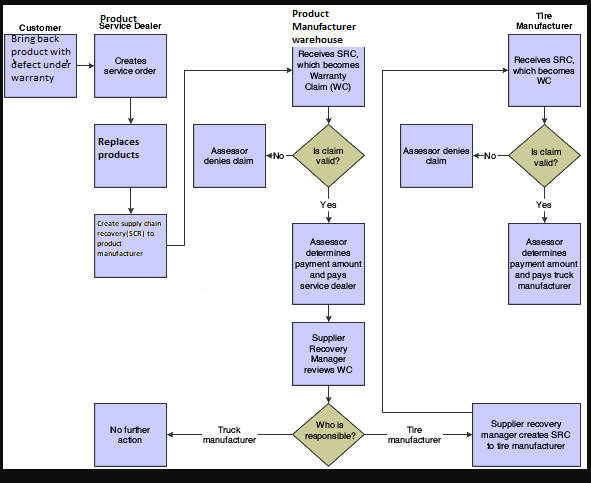
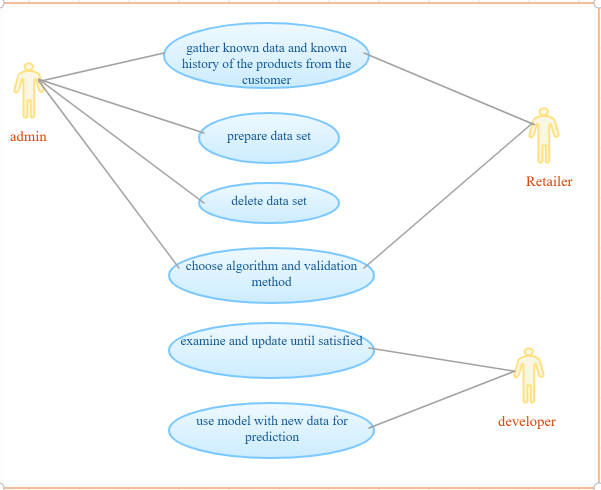


Figure 4.2 Client requesting warranty claim from the Product service dealer and how it is affected

( photo editor)

4.3 Use Case Diagram



(photo edior)

Warranty Failure MIS overall functionality system flowchart

View Warranty Failure Rate

Warranty Failure Calculatons

View Recommendations

Perform Assessment

User Login

Fig 4.4: overall functionality diagram

(Ms.Word)

**Classes and objects**

Warranty

type: string

warrantyAge:Number

DateReleased: Date

Paid()

Rejected()

Product

name:string

type:string

DatePurchased: Date

Paid()

Returned()

Manufacturer

ID:string

DateManufactured:Date

Returned()

Released

Retailer

ID:string

Name:string

Cash()

Cheque()



Fig 4.5: System’s classes, objects and attributes

(Ms Word)

Class diagram for warranty Failure

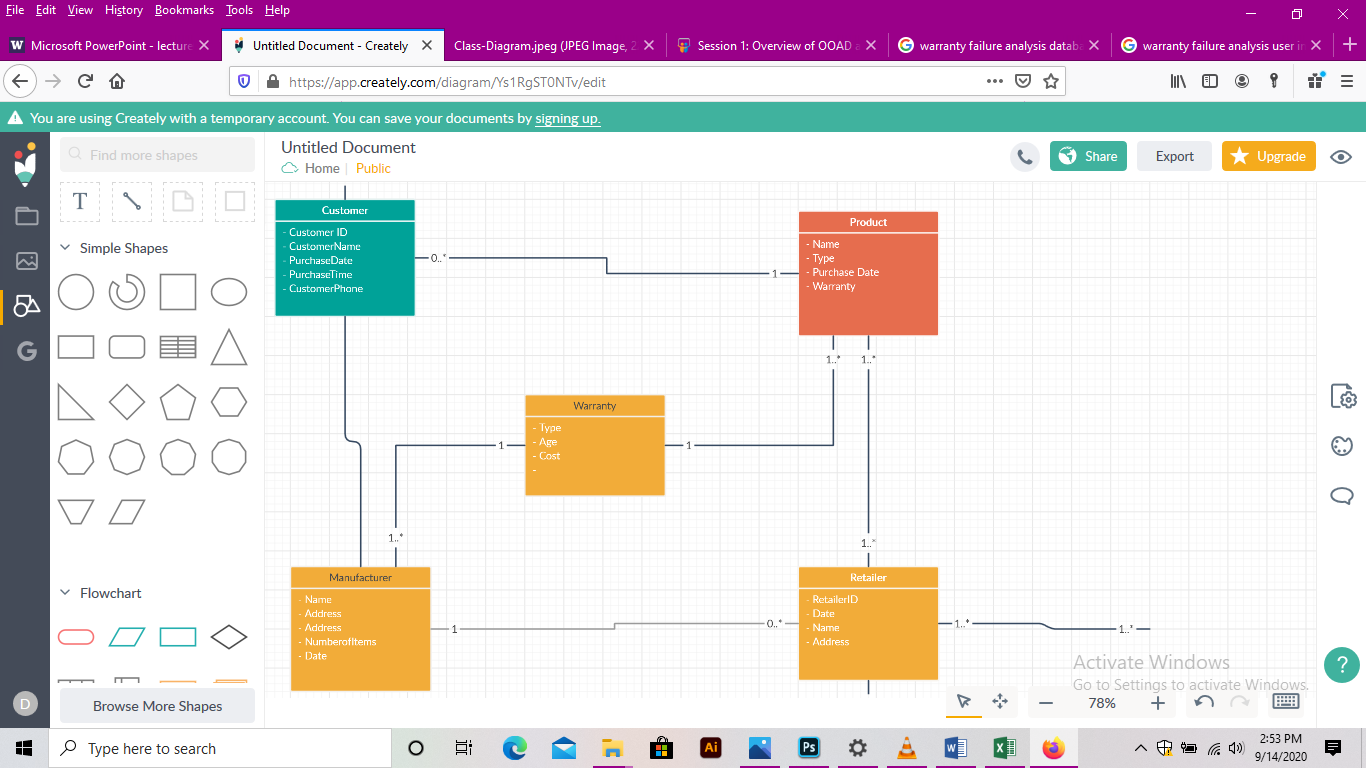
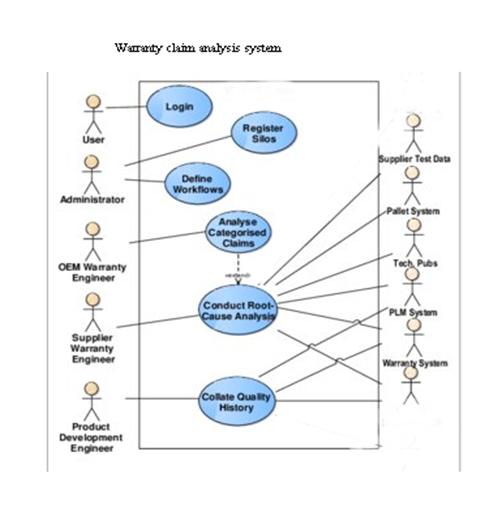


Fig 4.6: Class conceptual diagram

(Smartdraw)

Root Cause of product failure to determine a warranty claim

  
Fig 4.6:

(Smartdraw)

PLM- Product Lifecycle Management

OEM- Original Equipment Manufacturer

Database design prototype

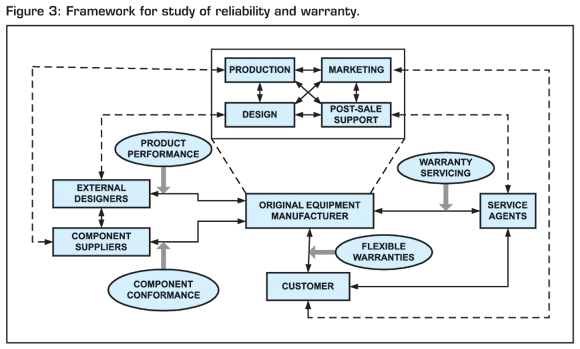


Fig 4.6: System DB design prototype

(Smartdraw)

**Sample user interface design**

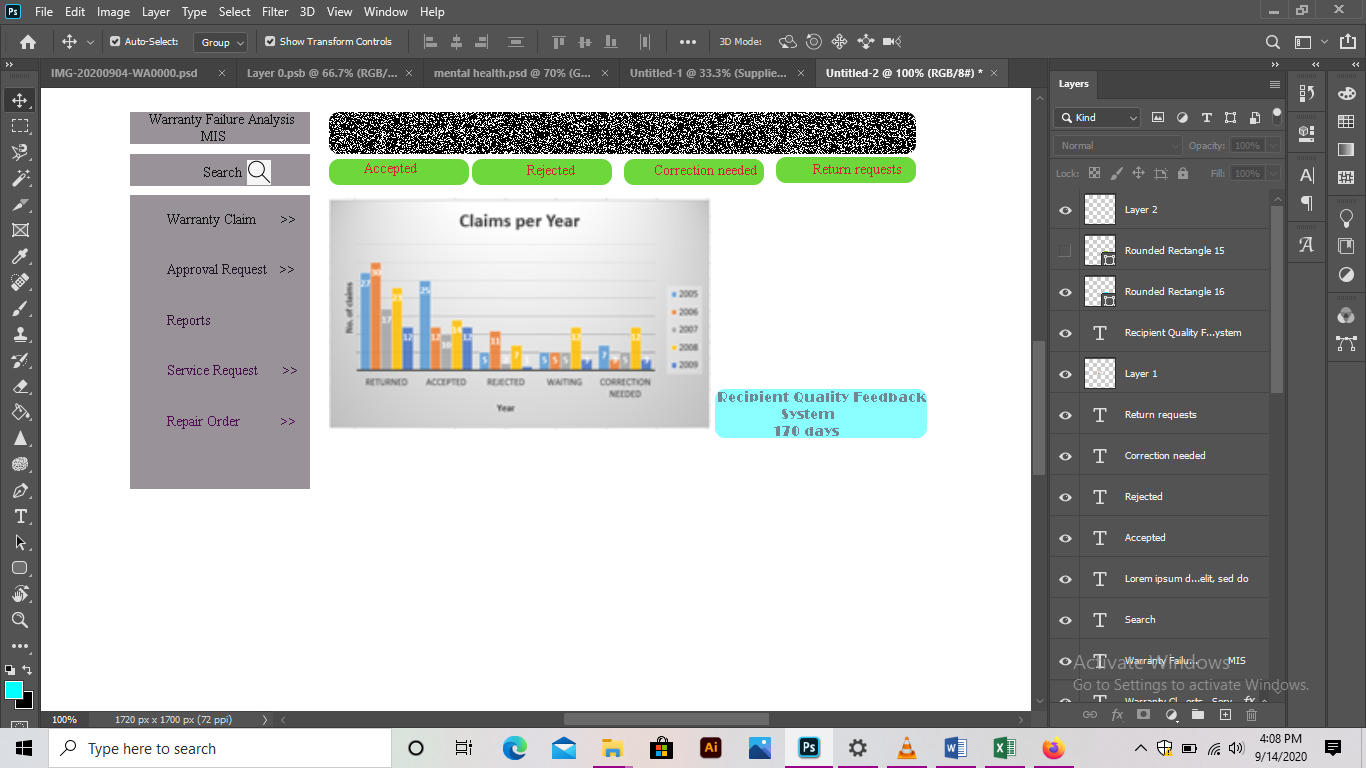


Fig 4.7: User interface sample

(Adobe Illustrator)