UNIVERSITY OF NAIROBI

SCHOOL OF COMPUTING & INFORMATICS

A DECISION SUPPORT MODEL FOR PREDICTING AVOIDABLE RE-HOSPITALIZATION OF BREAST CANCER PATIENTS IN KENYATTA NATIONAL HOSPITAL

BY

CHRISTOPHER OYUECH OTIENO
P58/75764/2012

SUPERVISOR: DR. ROBERT OBOKO

NOVEMBER 2014.

A RESEARCH REPORT SUBMITTED IN FULFILLMENT OF MASTERS DEGREE IN COMPUTER SCIENCE

TABLE OF CONTENTS

DECLARATION

Student
I hereby declare that the work in this thesis is my own except for quotations and summaries which have been duly acknowledged. The thesis has not been accepted for any degree and is not concurrently submitted for award of other degree.

Christopher Oyuech Otieno,
Admission No: P58/75764/2012,

Signature…………………………

Date……………………………….

Project supervisor
This project report has been submitted in fulfillment of the requirements of Master of Science in Computer Science in the University of Nairobi with my approval as the University Supervisor.

Dr. Robert Oboko,
University of Nairobi,
School of Computing and Informatics.

Signature…………………………………………………..

Date………………………………………………………...

ACKNOWLEDGEMENT

I would like to express my deepest gratitude to my University supervisor, Dr. Robert Oboko, for his excellent guidance, caring, patience, and providing me with an excellent atmosphere for doing this research.I would like to thank Dr Longino Mcheusi for been my clinician supervisor in the entire study.

My appreciation goes to Professor M.L Chidia, the General Secretary of KNH/UON-ERC for his guidance on prototype development and general approach on the study. I am grateful to the entire research committee of KNH/UON-ERC who found time to go through the research proposal and their recommendations on a numbers of drafts on how to execute the study.

My sincere gratitude goes to the following oncologist who gave their time and input on varies questions concerning the study. I am grateful to Dr Nyangasi,Dr Masolo,Dr Matungi,Dr Wilson Thinwa Kareu and Dr Calvin Abuya all from KNH Cancer Treatment Centre.

Last but not the least; I am grateful to following nurses; Elizabeth of ward 5B, Mrs Kanyi of Casualties and Ruth secretary of head of Cancer Department you were the best.

As always, the greatest debt one owes is to one's colleagues, friends, and family. In my case, this debt is large. Among the many here, I would especially like to thank my late father Mr. Elijah Otieno Sako,as I remember that unimaginable " painful night" you are a HERO!.You were a constant source of encouragement; this project is dedicated to you.

LIST OF FIGURES

Figure 13: Configuration ready for execution.

LIST OF TABLES:

LIST OF ABBREVIATIONS

ACGs……………………………………	Ambulatory Cost groups
DCGs…………………………………….	Diagnostic Cost Groups
BRP………………………………………	Business re-engineering process
KHN………………………………………	Kenyatta National Hospital
IOM's………………………………………	The Institute of Medicine's (IOM's)
WHO………………………………………	World Health Organization
ROC………………………………………	Receiver Operating Characteristic
PCTs………………………………………	Primary Care Trusts
HCC………………………………………	Hierarchical coexisting conditions
ICD9………………………………………	International Classification of Deceases Version 9
PARR………………………………………	Patients at Risk of Re-Hospitalization
CIHM………………………………………	Centre for Innovation in Health Management
DSM………………………………………	Decision Support Model

LIST OF APPENDICES

DEDICATION

I dedicate this work to my late parents; Elijah Otieno Sako;Collector Akinyi Otieno and my step mother Peninah Awuor Otieno. This study has also been dedicated to my lovely brothers and sisters for their support through the research period. I love you all.

Abstract

The study is about a clinical predictive model meant to support clinical decision-making of oncologists, which would have an immense potential benefit to their performance, provision of quality care and, better patient outcomes. It is accepted that one can take better care of a patient if one has superb knowledge (theory) about the clinical matters in question. For example, it could be said that with more information and knowledge, a clinician has a better chance of solving a clinical problem in favor of the patient, the hospital and himself. However, the problem nowadays is that global knowledge about a topic is often overwhelming for a clinician to process at the point of care or in urgent situations.

The study's predictive model incorporates patient-specific data which are well-structured and current knowledge base or evidence-based guidelines, thus serving the clinician by enhancing throughout use of her clinical decision-making process. Such support by this clinical model on basic cognitive processes involved in medical thinking to some extent relieves the clinician and provides him with new, better-formed and possibly superior methods to take best care of the ill. The clinical decision model characteristics' are related to clinical effectiveness, functionality, error prevention, potential for acceptance in the clinical world, system portability, cost effectiveness among others. It is important to fully understand its development and modalities adopted.

In this research study,6 samples records of breast cancer patients who were re-hospitalized in year (2009 to 2013), 30 were used. A Likert scaled questionnaire that was answered by clinicians, and 14 interview questionnaires on re-hospitalized patients are analyzed and the resulting information is used to develop a predictive decision support model.

The finished model is used to support clinicians decide upfront on vulnerable patients who are likely to be re-hospitalized. The developed decision tree model was validated by the test dataset which had been split from the same data that had been collected. The validated decision prediction model demonstrates sensitivity of 68.714%, a specificity of 71.42% and an area under the receiver operating characteristic curve of 0.908 for any prediction done.

This research study involved a real life application problem of predicting breast cancer patients to be re-hospitalized, not to be re-hospitalized or both cases.

Since the clinical decision support prediction model uses three sources of data from the hospital, patients to be discharged can be readily stratified into risk groups. This simple prediction model for evaluating patients before discharging may provide clinicians with a practical tool for counseling families and making management decisions just before patient discharge.

CHAPTER ONE: INTRODUCTION

1.1 Background

Avoidable re-hospitalizations result from care failures in the period immediately, before or after transition from hospital to the next source of care. These care failures result into clinical deterioration that leads to subsequent hospital visits, known as re-hospitalization. Preventable risks for re-hospitalization of a discharged breast cancer patient are; the level of medication to which the patient is discharged; incomplete medical dose; Fat consumptions; abuse of antibiotics; use of oral contraceptives; lack of physical exercise; overweight and expose to stress of varies types.

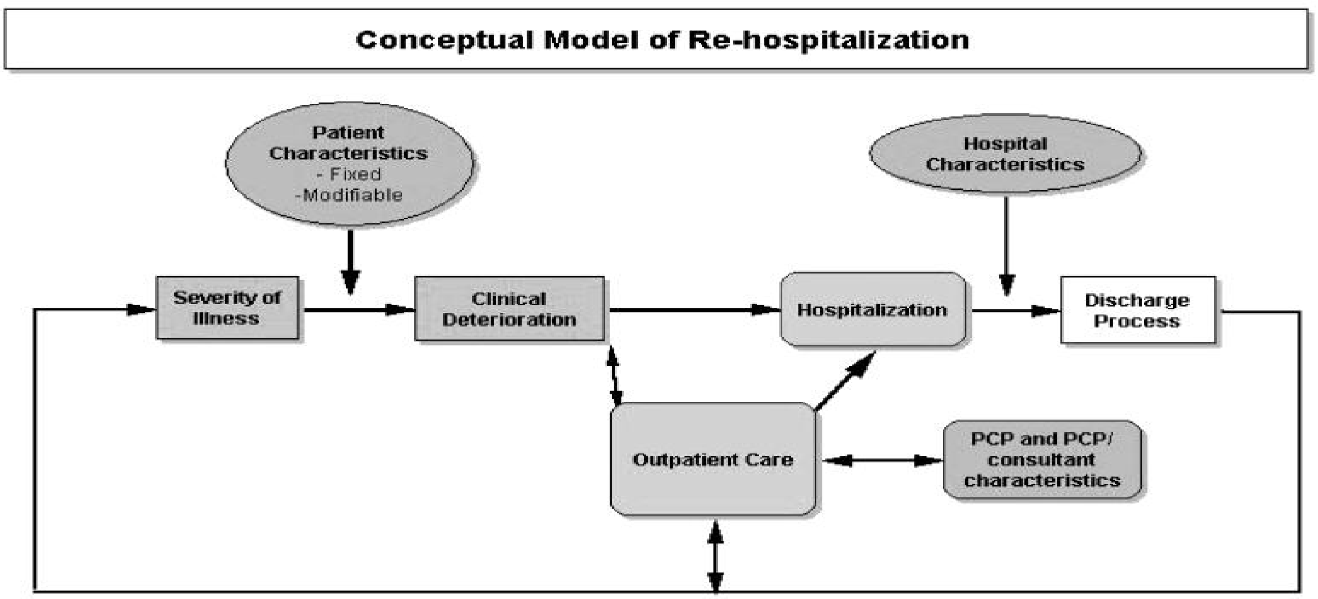


Figure 1:Re-engineering the Hospital Discharge Anthony, Chetty, et al, 2013.

Dr Alice Musibi, Medical Oncologist (2008) argues that Breast cancer is the deadliest and the most common cancer ailing women all over the world, for example in Australia 1 in 13 women will develop breast cancer at sometime in her life, in USA 215,990 women was found to have breast in 2004 and more common in older than younger women with an average age of 64 years.

She noted in the Cancer survey report which was conducted in Nairobi Kenya between years 2000-2003 that breast cancer was leading with 22.9% followed by cervical cancer with 19.3 %, as shown in figure 2 below. Mean age of diagnosis was 45 years.

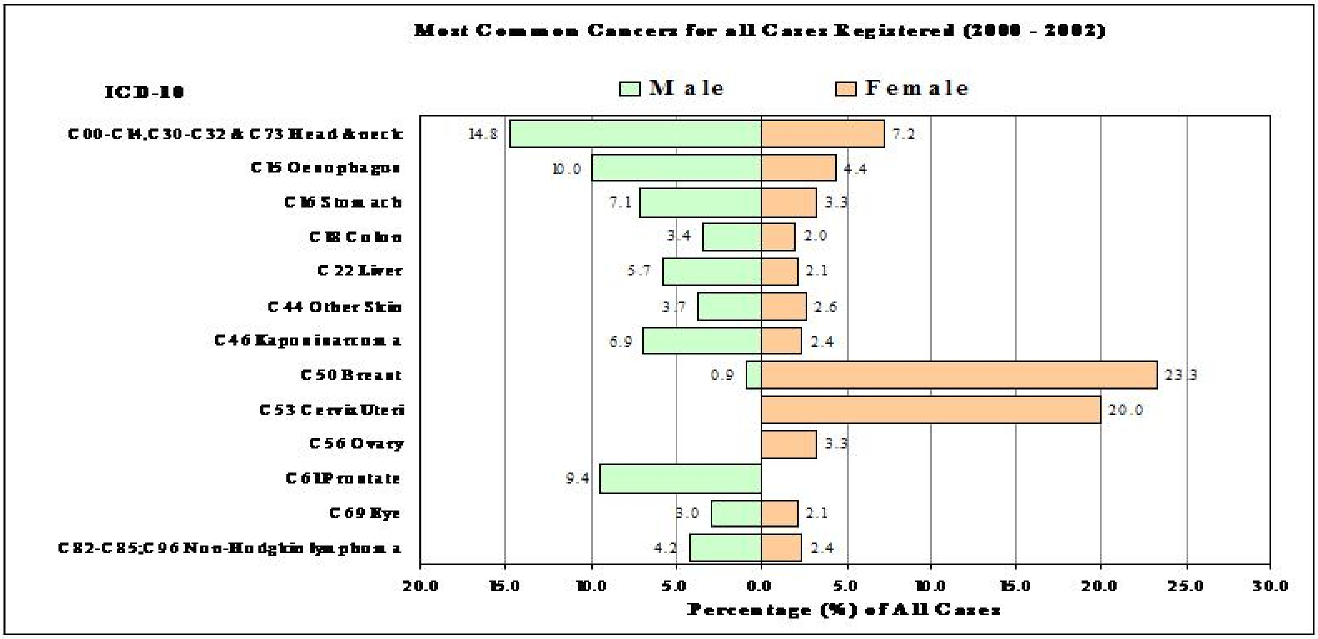


Figure 2: Breast cancer is the main cause of cancer incidence in Nairobi Kenya, Cancer Survey Report, Musibi 2008.

Breast cancer is the most common cancer affecting women in Kenya, whose healthcare costs imposes an increasingly burden on the government while the quality of care provided is arguably not adequate.

Nyogesa-Watt (2007) reported that there are few public and private hospitals in Kenya providing radio therapy services and patients have to travel across the country, some as far as 600 kilometer away to access such medical services. He noticed that Kenya has only 10 oncologists, a number far less for a country of over 40 million people.

Another scholar, Dr.Ian Hampson (1999), from The University of Manchester's Institute of Cancer Sciences who oversaw a cancer research in Kenyatta National Hospital (KNH) noted that the available radio therapy center handles over 3800 patients in a year which is below the needs of the country. He exclaimed that patients refereed from other periphery hospitals have to wait for months before accessing medical services sometimes leading to preventable death (1999).

There is an urgent need for an upfront predictive model, in the Cancer and Oncology Department of Kenyatta National Hospital (KNH) and other hospitals to classify patients been discharged into different preventive risks levels vulnerable for re-hospitalization. The risk level are; high risk level, moderate risk level and low risk level.

Risk classification level can be illustrated using Kaiser Permanente's risk triangle. In the Kaiser Permanente's risk triangle, it is the individuals at the top of triangle who are most at risk of emergency admission. Case management programmes attempt to target these individuals to prevent them being re-admitted. However, there is some debate as to whether this is the most appropriate area of the triangle on which to concentrate resources. It has been suggested that once an individual has reached this level of risk, an intervention is likely to be too late to prevent admissions. It may be of more value (both in financial and health outcomes terms) to identify those individuals in the lower two strata who are likely to move into the high risk/high cost level, Michael .et al (2005).

The urgently needed predictive model would inform clinicians of avoidable risks categories that a particular patient would adopt upon discharge so as to strategies clinical management targeting vulnerable ones, as Jaimie Oh, (2012) argues,” Medicare advantage patients experience fewer re-hospitalizations”.

1.2 The Research Problem

The Institute of Medicine’s (IOM’s) 1999 publication reports that the number of deaths due to preventive risk in hospitals is estimated to be between 44,000 and 98,000 per year and that more people die each year from such preventive risks than from car accidents (43,458), breast cancer (42,297) or AIDS (16,515).

As hospitalists provide more in-patient care, the transition from hospital care to primary care is a hand-off activity that provides opportunity for a high rate of preventive risk (Forster, et al 2012). Such preventive risk may lead to avoidable re-hospitalization, death or both. Re-hospitalization is not a hospital or a patient problem, it is a community problem, and ensuring that all sectors of the community are involved and work together to make care transitions effective, is important.

The WHO report (2014) highlighted that Cancer is the leading cause of death worldwide, accounting for 8.2 million deaths in 2012.In Kenya, different categories of cancer exist, among them breast cancer; the main leading cancer incidence in the country.

The study shows that preventive re-hospitalization of breast cancer incidence is rising in Kenyatta National Hospital (KNH) and exerts pressure in an already strained system (Musibi, 2013).

Andy Miller (2013) argues that avoidable re-hospitalization cost medicare billions of dollars a year and since most of the re-hospitalization is deemed preventable, it is likely that carefully designed interventions that target vulnerable patients would successfully reduce rates of subsequent hospital utilization.

So there is need for means of identification of discharged breast cancer patient who are vulnerable for re-hospitalization so that better and targeted management measures can be put in place in the wrack of time.

1.3 Purpose of the Study

The aim of this study is to develop a clinical decision support model to classify breast cancer patient vulnerable for re-hospitalization into different risk level from their clinical deterioration extract variables or risks attributes. The Hospital's clinicians would then takes appropriate clinical management of classified patients whose clinical deterioration would be classified as highest risk for re-hospitalization.

1.4 Objectives

a. Broad Objective

To implement a predictive support decision model from re-hospitalized breast cancer dataset to help clinicians decide upfront vulnerable patients who are likely to be re-hospitalized so as strategize interventions targeted at them to decongest wards and related health resources.

b. Specific Objectives

1. To identify preventive risk attributes that cause re-hospitalization of a discharged breast cancer patient.
2. To collect identified preventable risk attributes that cause re-hospitalization from patients and or past re-hospitalized breast cancer patient's administrative data.
3. To preprocess collected risks attributes data that causes preventable re-hospitalization of a discharged breast cancer patient into a database set.
4. To upload processed database set into the rapid miner 6.0 software.
5. Configure and apply related algorithms (such as ID3, C.45), operations (such as cross validation, pruning etc) to give a visual decision tree model.
6. To do analysis and evaluation of the resultant decision model through confusion matrix, level of confidence and the ROC curve and declare validly of the classifier model as per the standards.
7. Map the resultant clinical decision support model into web base application.

1.5 Research Outcome and Significance of the Study

1. Predicted breast cancer patient vulnerable for re-hospitalization are significant to hospitalist who devices upfront strategies of management targeting them hence improved quality of health care, reduced cost of health care, and more health resources available for new patient's hence improved patient volume served per unit time.
2. Predicted breast cancer patient vulnerable for re-hospitalization,may assist insurance company to calculate precise premium based on such risk vulnerability.
3. Low re-hospitalization information by some hospitals is used by some western governments to gives incentive to such hospitals for registering low rate of re-hospitalization.
4. Identification of preventable risk factors helps in sensitizing public on such hence personal initiative on reduction of breast cancer cases.
5. In strategies management targeting vulnerable patient for re-hospitalization, there is reduction of overall cancer incidence by 40% margin as reported by WHO (2014).

1.6 Scope of the Study

The study predicts upfront preventive re-hospitalization of a discharged breast cancer patient via clinical decision support model. Other utilities of prediction such as stratified sampling, ID 3/C4.5 algorithm and clinical knowledge concepts are integrated at different stages of research development. The clinical decision support model is then mapped into a web base application.

1.7 Assumptions of the research study

1. Preventive risks attributes are based on the assumption that a patient's illness burden better characterizes the patient's need for health services than only the presence of a specific primary disease of interest.
2. A clear delineation of discharging responsibilities as reengineered (figure 3) often does not exist and lack of communication results in repetition and gaps.

1.8 Study limitations

a) Before embarking on the study it is crucial to have a clear definition of the concept to be predicted, and historical examples of the concept.

b) Error or any clinical mistake made by the clinician, patients or the investigator in answering questionnaires compromises the overall quality of the research study significantly.

c) In general, one cannot make progress without a dataset for training of adequate size and quality.

d) Over-fitting occurs when a classification model describes random error or noise instead of the underlying relationship. A model which has been over fitted will generally have poor predictive performance, as it can exaggerate minor fluctuations in the data.

e) For the prediction to be successful, the training data must be representative of the test data.

   Typically, the training data come from the past, while the test data arise in the future. If the re-hospitalization to be predicted is not stable over time, then predictions are likely not to be useful. Here, changes in the general economy, lifestyle, and in social attitudes towards breast cancer, are all likely to change the behavior of patients in the future. The model needs constant update with time.

f) The predictive model can lead clinicians to an ever-increased focus on optimizing predictive power at the expense of understanding the broader situation of theory building and richer content of attributes on avoidable re-hospitalization.

g) Clinicians should be aware of the model temptation to shift away their attention from the real problem of concept building.

h) Clinicians should also be aware that the model doesn't read their minds but work on the “sword of data” and that the model is supportive but they make the actual decision.

1.9 Definitions of the important terms

Avoidable re-hospitalization: The process of being hospitalized again from reasons that can be prevented.

Re-hospitalization: To be readmitted to hospital within 30-45 days for the same illness a patient had initially.

Predictive Model: A form of data-mining technology that works by analyzing historical and current data and generate a model to help predict future outcomes.

Data Mining: Sometimes called data or knowledge discovery is the process of analyzing data from different perspectives and summarizing it into useful information.

Risk stratification: The identification of a patient's health risk category for planning, developing and implementing a personalized patient care plan by the care team, in collaboration with the patient.

Patient-oriented interventions: Personalized patient care plan on clinical interventions to improve patient's quality of health.

The clinical knowledge: Clinical experience on a particular disease.

Statistical model: A formalization of relationships between variables in the form of mathematical equations.

Activation function: A function used to transform the activation level of a unit (neuron) into an output signal.

Primary care physicians (PCPs): He/She is a physician who provides both the first contact for a person with an undiagnosed health concern as well as continuing care of varied medical conditions, not limited by cause, organ system, or diagnosis.

Co morbidity: Two or more disorders or illnesses occurring in the same person. They can occur at the same time or one after the other.

Business re-engineering process: It's the analysis and redesign of workflows within and between discharging processes in order to optimize end-to-end processes and automate non-value-added tasks

Ambulatory cost groups (ACGs): The overal patient's illness burden that characterizes the patient's need for health services than the presence of specific diseases of interest.

Diagnostic Cost Groups (DCG): A population-based classification and risk adjustment methodology.

CHAPTER TWO: LITERATURE REVIEW

2.1 Breast Cancer Problem and Treatment in Kenya

According to the National Cancer Institute of Kenya reports in 2012, 22,000 people out of the 28,000 or at least three out of four cancer cases diagnosed in Kenya succumbed to the disease. The high rate of morbidity is attributed to late diagnosis and lack of knowledge about cancer. Millions of poor cancer patients across up-country Kenya continue to suffer due to a dysfunctional public healthcare system even as the country ramps up its battle against the killer disease with a new law to boot. From Kisumu on the shores of Lake Victoria in Western Kenya and Eldoret in the Rift Valley, to Mombasa at the Coast, the plight of the poor cancer patient who cannot afford treatment in private hospitals, is the same. Faced with growing incidences of the disease in recent days with no commensurate expansion of facilities, medical facilities in these areas are groaning under the weight of the many patients that it can barely provide satisfactory services to.

The state of Kenya's arsenal against the disease dampens the optimism that greeted the enactment of the Cancer Prevention Act 2012, which was assented into law by President Mwai Kibaki having been passed by Parliament earlier in the year. The law was expected to revolutionize and re-energize the country's efforts against cancer that has lately ailed several prominent individuals, some of whom it lost their lives. In Kenya cancer has gained notoriety for its high profile victims, including the country's resently two health ministers – Public Health minister Beth Mugo and Medical Services minister Prof Anyang' Nyong'o who had to seek treatment abroad (Butunyi, 2013).

For many poor Kenyans, seeking cancer treatment abroad is an unattainable dream. Thus, they have to rely on public health facilities, which are however plagued by shortcomings such as shortage of experts and physical infrastructure, poor records on the disease at the two hospitals with registries - Kenyatta National Hospital and the Moi Teaching and referral Hospital. Among the innovative strategies that weree set to be introduced by the new law which is yet to be operationalized was teaching school children on how to prevent cancer and introduction of penalties against care providers who do not submit data on the disease (Butunyi, 2013). At the moment, there are insufficient facilities, poor records on prevalence, frequency and geographical distribution of the disease, as well as specialized medical professionals to lead the fight against the disease in Kenya. However, the sorry state of cancer facilities and equipment is not just a Kenyan problem; the rest of the East African region ails from the same shortcomings.

2.2 A clinical decision-support model

A clinical decision-support model is a computer program designed to help health professionals make clinical decisions. In a sense, any computer system that deals with clinical data or medical knowledge and is intended to provide decision support. Three types of decision-support are ranging from generalized to patient specific such as generating alerts and reminders; diagnostic assistance; therapy critiquing and planning; Image recognition and interpretation.

2.3 Characteristics of a Clinical Decision-Support Models

Decision Support model is a tool for information management for example in hospital information systems, bibliographic retrieval systems and specialized knowledge-management workstations.

Decision Support Model is a tool that provides data and knowledge needed, but also they do not help to apply that information to a particular decision task (particular patient).

Decision Support model is a tool for focusing attention for example: Clinical laboratory systems that flag abnormal values or that provide lists of possible explanations for those abnormalities, Pharmacy systems that alert providers to possible drug interactions or incorrect drug dosages

Decision Support model are designed to remind the physician of diagnoses or problems that might be overlooked.

Decision Support model is a tool for Patient-Specific Consultation for example its provide customized assessments or advice based on sets of patient-specific data and Suggest differential diagnoses, advice about additional tests and examinations, treatment advice.

Clinical decision support systems are active knowledge systems which use two or more items of patient data to generate case-specific advice.Main components of a Clinical decision support system are medical knowledge, patient data and case-specific advice.

The core function of a clinical decision support model is to determining what is true about a patient (e.g. correct diagnosis) so as to determining what to do or not.

Clinical decision support systems are model for giving advice such as a passive role for example when a physician uses the model when advice needed or active role when the system gives advice automatically under certain conditions (Saleem, 2008).

2.4 Clinical Decision Support Models for Prediction

Literature reviewed on the prediction models are the De Dombal's system for acute abdominal pain (1972) which was developed at Leeds University .The decision making was based on the naive Bayesian algorithm approach. The model was automated to reason under uncertainty and was designed to support the diagnosis of acute abdominal pain.

INTERNIST-I (1974) another prediction model is a rule-based expert system designed at the University of Pittsburgh. It was meant to diagnosis of complex problems in general internal medicine. It uses patient observations to deduce a list of compatible disease states. It was used as a basis for successor systems including CADUCEUS and Quick Medical Reference (QMR).

MYCIN (1976) is a another prediction model which is rule-based expert system designed to diagnose and recommend treatment for certain blood infections (extended to handle other infectious diseases).Clinical knowledge in MYCIN is represented as a set of IF-THEN rules with certainty factors attached to diagnoses.

2.5 Decision Support Models for Predicting Re-hospitalization

Among the literature reviewed is a joint project on behalf of the 28 strategic health authorities, to produce a risk prediction system for use by PCTs (Primary Care Trusts) to identify patients who are at high risk of hospitalization. This project was implemented by Essex Strategic Health Authority and used tools such as ambulatory cost groups (ACGs), Diagnostic Cost Groups (DCG) and Hierarchical Coexisting Conditions (HCC) Model. ACGs adopts ICD9-CM coding system and based on the assumption that a patient's illness burden better characterizes the patient's need for health services than only the presence of a specific disease.

Diagnostic Cost Groups (DCG) tool was integrated in the model to predict future costs of Medicare for population based on the 'worst' inpatient diagnosis recorded in a time period. Ambulatory diagnosis and the cumulative effect of multiple conditions in predicting total medical expenditure; also known as the DCG/HCC (Hierarchical Coexisting Conditions) formed the core feature of the model,(Rosen AK ,2001).Predictive Regression model was the main analytic tool.

Another literature reviewed is Adjusted Clinical Group (ACG) developed by Johns Hopkins University (Bloomberg School of Public Health, 2009). Aggregation of co morbidities diagnosis forms the major methodology. ACG identifies patient, groups and population that have high probability of hospitalization in future from aggregation of co morbidities. It present morbidity burden of a population, subgroups or patients hence capacity to predict resource use /cost for quality of health. It also supports identification of people with specified disease, such as HIV. ACG is a good resource management tool.

Another relevant literature reviewed is Patients at risk of Re-hospitalization (PARR1 and PARR2) algorithms by Health Dialog Analytic Solutions. The algorithms are patient specific that produces "risk score" for the probability of future readmission from patient past readmission records. The algorithm is used "real time" (while the patient is hospitalized) with resent readmission records and diagnostic information. PARR1 (Patients at Risk of Re-Hospitalization version 1) and PARR2 (Patients at Risk of Re-Hospitalization version 2) algorithms indicates high readmission rates for patients who have experience readmission before and less for those who have never. No general database to draw inferences except specific patient past records. It has shortcomings as it cannot comprehensively define the risk of readmission to hospital, Schoenmaker & Russo,(1993).Its Underestimate the total number of high risk patients, as it screens patients by using a single criterion which may neglect other potentially important risk factors. It lacks accuracy. For example, individuals who are at risk one year may not be at risk the next year and vice versa, Dove, Duncan & Robb (2003).

Another study is from Centre for Innovation in Health Management (CIHM), a consulting company gathering expert's information from health sector, public sector, organizational change consultancy and academics. CIHM is based at the University of Leeds and is currently engage in a project to develop a model to predict readmission of a discharged patient from regression model. The project shall be developed through decision trees and used by the general practitioners to decide a patient-oriented intervention, taking into account the clinical knowledge and outputs generated from Risk classification Tool.

Risk classification tool is using regression model whose aim is to stratify patients risk in terms of their future re-hospitalization; thus intervention design that can be responsive to the patient's risk. Predictive modeling is one of the tools of risk stratification. A predictive model is a statistical model whose output is a risk score for each patient, which is the probability of re-hospitalization in the future. General Practitioners, nurses and pharmacists participate in the project since they are confronted with thousands of treatment and referral decisions every day, and their clinical knowledge is a factor.

Another work reviewed is predicting readmissions of Heart failure via decision tree by James Natale and Shengyong Wang of University of Akron USA, (2013) .Rapid miner is the primary software for model creation. Confusion matrix is adopted for analysis of specificity and sensitivity.

Short comings of the study includes predictive model which is not focused on the discharging structure. It is therefore possible that the model predictability may not have been comprehensive since it doesn't focus on the discharging structure which has adverse events, slips and risk of re-admissibility embedded therein.

CHAPTER THREE: METHODOLOGY

3.1 The Research Design

The study design is largely exploratory hence the adoptions of the following methods:

(a) The survey of concerning literature; and

(b) The experience or clinician survey.

The survey of relevant literature such as hypothesis stated by earlier researchers is vigorously reviewed and the usefulness of such literatures evaluated to establish the ground for the study design. Experience survey is the surveying of clinicians on practical experience on the same problem to be studied.

The objective of the survey is to obtain insight into the relationships between variables and new ideas relating to the research problem. In this case, surveys of clinician(s) who are competent and can contribute new ideas are selected randomly to ensure representation of different types of experience and ideas. The selected clinicians are then given questionnaires to answer. This method provides rich and practical information on how the design study should be approached. Questionnaire to be answered by the clinicians can be seen at the appendices of this document.

Clinician's questionnaires are designs based on **Likert scale** techniques.

Questionnaires are designed with flexibility in mind in the sense that a clinician is allowed to raise issues and questions which the investigator has not previously considered. Copies of the questionnaires to be answered are sent to the clinician(s) well in advance. This gives them an opportunity for doing some advance thinking over the various issues involved so that, at the time of filling questionnaires, they are able to answer effectively. Thus, an experience clinician may enable the researcher to define the problem precisely and help in the formulation of the research hypothesis. This survey may as well provide information about the practical possibilities for conducting the research.

3.1.1 Study area description

The study is conducted in the Cancer Treatment Centre, Kenyatta National Hospital (KNH) under the supervision of clinician Longino Mucheusi, who is also the Training Coordinator and Assistant Chief Therapy Radiographer.

This study area handles treatment of cancer incidence including breast cancer on the daily basis. The department refers, admits and discharged breast cancer patients. The study area is extended to Ward 5B where the investigator interacted with the patients.

3.1.2 The Study Population

Targeted population includes all breast cancer patients re-hospitalized within a period of three-four months from the time of discharged of years 2007-2013.

The study state that 10,000 breast cancer patients were re-hospitalized between years 2007-2013.That is 2000 re-hospitalized patients are re-hospitalized yearly on average.

Inclusion Criteria for a patient

- Must be a breast cancer patient re-hospitalized within 30-45 days.
- Must have been a breast cancer patient and had experience re-hospitalization within 30-45 days.

Inclusion Criteria for a Clinician

- Must be a clinician experienced in oncology for not less than 1 year.

Exclusion Criteria for Patients

- Patient below 8 years is not considered.

Exclusion Criteria for a Clinician

- Any clinician who doesn't met clinician inclusion criteria above.

3.1.3 Sample Size: Disproportionate Sampling Design

The designed is used in cases where strata differ not only in size but also in variability (Variation or diversity).It is considered reasonable to take larger samples from the more variable strata and smaller samples from the less variable strata, which can then account for both (differences in stratum size and differences in stratum variability).This design is adopted since we have 3 strata (patient, oncologist and past record which differ in size and variability).Formula is as follows:

$$n_1/N_1\sigma_1 = n_2/N_2\sigma_2 = \ldots\ldots\ldots = n_k/N_k\sigma_k$$

where $\sigma_1, \sigma_2, \ldots$ and σ_k denote the standard deviations of the k strata, $N_1, N_2, \ldots, N_k$ denote the sizes of the k strata and $n_1, n_2, \ldots, n_k$ denote the sample sizes of k strata. This is called '*optimum allocation*' in the context of disproportionate sampling. The allocation in such a situation results in the following formula for determining the sample sizes different strata:

$$n_i = \frac{n \cdot N_i\,\sigma_i}{N_1\sigma_1 + N_2\sigma_2 + \ldots + N_k\sigma_k} \qquad \text{for } i = 1, 2, \ldots \text{ and } k.$$

For example we have Strata Oncologist O, Strata patient P and Strata Record R with Standards deviation O=15,P=18 and R=5 ,from Population of

$$O_p - 5000, \quad P_p - 2000, \quad R_p - 3000$$

Total sample size of n=49 which are allocated to different sample size strata as follows:

$$O_P = 5000, n_O = \frac{49(5000)(15)}{(5000)(15)+(2000)(18)+(3000)(5)} = 29.16 \text{ Number of Samples from oncologists}$$

$$P_R = 2000, n_P = \frac{49(2000)(18)}{(5000)(15)+(2000)(18)+(3000)(5)} = 14 \text{ Number of Samples from Patients}$$

$$R_P = 3000, n_R = \frac{49(3000)(5)}{(5000)(15)+(2000)(18)+(3000)(5)} = 5.8333 \text{ Number of Samples from the Past Records}$$

3.1.4 Sampling method

The study adopts probability sampling method, also known as random sampling/chance sampling. Under this sampling method, every patient, past record and oncologist in the population has an equal chance of inclusion in the sample size.

3.1.5 Recruitment and Consenting procedures

Recruitment advertisements

Recruitment advertisements are sent to prospective subjects (re-hospitalized patients and interested clinicians) to solicit participation. This also forms part of the consent process and must have KNH approval prior to use. Prospective participants who respond as willing to take part are contacted by the study investigator or the supervising clinician.

Direct recruitment of potential study participants.

The strategy is that the supervising clinician talks to fellow clinicians and the patients about the study and care is taken by the supervising clinician so that clinician or patient contacted don't feel pressured to participate.

Referrals.

Supervising clinician suggest referrals to the investigator about other potential clinician who can to take part in the study. Investigator then sends a "Dear Patient" letter or a "Dear Potential Study Participant" letter describing the study. The investigator and the supervising clinicians contacts details are included therein for the clinician and patients to give feedback of there willingness to participate.

3.1.6 Data collection

1. The data for the research is collected from survey questionnaires (see appendices) .The survey questionnaires are build from suitable questions modified from the related studies. In the questionnaires, Likert scale technique is used to determine if the respondent agreed or disagreed in a statement.
2. The clinician's survey questionnaires comprises of section A, of 12 questions; section B, of 9 questions; section C, of 11 questions; and section D, of 21 questions; on the clinician perception regarding re-hospitalization of a breast cancer patient.
3. Patient's survey questionnaires are made of section A, of 9 questions and section B, of 20 questions.

4. The clinician's survey questionnaire is distributed to the participating clinicians in the department of Cancer Treatment and Oncology KNH. The researcher interviews the patients and if the patient is in a position to write, then he or she is given the questionnaire to answer.

3.1.7 Data Preprocessing

Irrelevant attributes such as patient residential address, name, application ID, etc is removed. For example, the patient fathers name is irrelevant in predicting the future re-hospitalizations. Finally, the "Re-hospitalization" attribute is added to holds the predicted result, which can either be "readmitted", "no readmission" or "both". The following table (database) is constructed from clinician's questionnaire, patient's interview questions and past patient's records which have undergone data preprocessing.

Table 1: Database from clinician's, patient's questionnaire and past records.

Stress Burden	Complete Medical Dose	Smoking	Alcoholism	Pysical Exercise	Fat intake	Abuse Antibiotic	Abortion Or Miscariage	Keep Follow Up Appointments	Use Oral Contraceptive	Had Hormone Replacement Therapy	Appropriate Medication	Many love partners	Overweight(BMI)	Appropriate Discharge	Predicted Status
Pyschological Is	Not timely	Not a Smoke	yes	Not daily	No	Yes	Yes	Yes	No	No	Yes	No	No	Yes	No Re-Hospitalizatio
No Stress Burde	Timely	Not a Smoke	No	Daily	No	No	No	No	Yes	Yes	Yes	No	Yes	Yes	No Re-Hospitalizatio
Pysically & Pysc	Timely	Not a Smoke	No	Daily	No	No	Yes	Yes	Yes	Yes	NO	Yes	Yes	No	Re-Hospitalization o
No Stress Burde	Timely	Sneaky Smok	Yes	Not daily	Yes	No	No	No	No	No	NO	Yes	No	No	Re-Hospitalization o
Pysically & Pysc	Not timely	Committed S	No	Not daily	Yes	Yes	No	Yes	Yes	Yes	Yes	No	Yes	Yes	Re-Hospitalization o
No Stress Burde	Timely	Sneaky Smok	Yes	Daily	No	Yes	Yes	No	No	No	Yes	Yes	Yes	No	Re-Hospitalization o
Physical Isolate	Timely	Not a Smoke	No	Daily	No	No	No	Yes	Yes	Yes	No	Yes	Yes	Yes	No Re-Hospitalizatio
No Stress Burde	Timely	Committed S	No	Daily	Yes	No	No	Yes	Yes	No	No	No	No	Yes	No Re-Hospitalizatio
Pyschological Is	Not timely	Committed S	No	Not daily	No	Yes	Yes	No	Yes	No	Yes	No	No	No	Re-Hospitalization o
No Stress Burde	Timely	Not a Smoke	Yes	Daily	No	No	No	Yes	No	No	Yes	No	No	Yes	No Re-Hospitalizatio
Physical Isolate	Not timely	Sneaky Smok	No	Daily	No	Yes	No	Yes	No	Yes	No	Yes	Yes	Yes	Re-Hospitalization o
No Stress Burde	Not timely	Not a Smoke	Yes	Daily	Yes	No	No	No	Yes	No	No	Yes	Yes	No	Re-Hospitalization o
Pyschological Is	Not timely	Sneaky Smok	No	Not daily	No	No	No	No	Yes	No	No	No	No	Yes	No Re-Hospitalizatio
No Stress Burde	Not timely	Committed S	No	Not daily	No	Yes	No	Yes	No	Yes	Yes	Yes	Yes	No	No Re-Hospitalizatio
Pyschological Is	Not timely	Not a Smoke	Yes	Daily	Yes	Yes	No	No	No	Yes	Yes	Yes	Yes	No	Re-Hospitalization o
Physical Isolate	Not timely	Sneaky Smok	Yes	Daily	Yes	Yes	No	No	No	Yes	No	Yes	Yes	No	Re-Hospitalization o
Physical Isolate	Timely	Not a Smoke	No	Daily	Yes	No	No	Yes	No	Yes	Yes	Yes	No	Yes	No Re-Hospitalizatio
Physical Isolate	Timely	Committed S	No	Daily	No	No	Yes	No	Yes	Yes	Yes	No	Yes	Yes	No Re-Hospitalizatio
Pysically & Pysc	Not timely	Sneaky Smok	No	Daily	Yes	Yes	Yes	No	Yes	No	Yes	No	Yes	No	Re-Hospitalization o
Pyschological Is	Not timely	Committed S	Yes	Daily	Yes	No	No	Yes	No	Yes	Yes	Yes	No	Yes	No Re-Hospitalizatio
Pyschological Is	Not timely	Not a Smoke	Yes	Daily	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	No	Re-Hospitalization o
No Stress Burde	Timely	Not a Smoke	No	Daily	No	No	No	No	No	No	Yes	No	No	Yes	No Re-Hospitalizatio
Pysically & Pysc	Timely	Not a Smoke	No	Daily	No	No	Yes	Yes	Yes	No	Yes	No	Yes	Yes	No Re-Hospitalizatio
No Stress Burde	Timely	Sneaky Smok	Yes	Not daily	Yes	No	No	Yes	No	Yes	Yes	Yes	Yes	No	Re-Hospitalization o
Pysically & Pysc	Not timely	Committed S	No	Not daily	Yes	Yes	No	No	Yes	No	Yes	No	Yes	No	Re-Hospitalization o
No Stress Burde	Timely	Sneaky Smok	Yes	Daily	No	Yes	Yes	Yes	Yes	No	No	No	No	No	No Re-Hospitalizatio
Physical Isolate	Not timely	Sneaky Smok	No	Daily	No	Yes	No	No	No	No	No	No	No	Yes	No Re-Hospitalizatio
No Stress Burde	Not timely	Committed S	No	Not daily	No	Yes	Yes	Yes	No	No	Yes	Yes	Yes	Yes	No Re-Hospitalizatio
Pyschological Is	Not timely	Not a Smoke	Yes	Daily	Yes	Yes	No	No	Yes	Yes	Yes	Yes	Yes	No	Re-Hospitalization o
Physical Isolate	Not timely	Sneaky Smok	Yes	Daily	Yes	Yes	No	Yes	Yes	Yes	No	No	No	Yes	Re-Hospitalization o
Physical Isolate	Timely	Not a Smoke	No	Daily	Yes	No	No	No	No	Yes	No	Yes	Yes	Yes	No Re-Hospitalizatio
Physical Isolate	Timely	Committed S	No	Daily	No	No	Yes	Yes	Yes	No	Yes	Yes	Yes	No	No Re-Hospitalizatio
Pysically & Pysc	Not timely	Sneaky Smok	No	Not daily	Yes	Yes	Yes	Yes	Yes	Yes	No	No	No	Yes	Re-Hospitalization o
Pyschological Is	Not timely	Committed S	Yes	Not daily	Yes	No	No	No	Yes	Yes	No	No	No	Yes	Re-Hospitalization o
Pyschological Is	Not timely	Not a Smoke	Yes	Not daily	No	Yes	Yes	No	No	No	No	No	Yes	No	Re-Hospitalization o
No Stress Burde	Timely	Not a Smoke	No	Daily	No	No	No	Yes	Yes	No	No	Yes	No	Yes	No Re-Hospitalizatio
Pysically & Pysc	Timely	Not a Smoke	No	Daily	No	No	Yes	Yes	No	No	Yes	Yes	No	No	No Re-Hospitalizatio
No Stress Burde	Timely	Sneaky Smok	Yes	Not daily	Yes	No	No	No	Yes	Yes	Yes	Yes	Yes	Yes	Re-Hospitalization o
Pysically & Pysc	Not timely	Committed S	No	Not daily	Yes	Yes	No	No	Yes	Yes	No	Yes	No	No	Re-Hospitalization o
No Stress Burde	Timely	Sneaky Smok	Yes	Daily	No	Yes	Yes	Yes	No	No	Yes	No	No	Yes	No Re-Hospitalizatio
Physical Isolate	Timely	Not a Smoke	No	Daily	No	No	No	No	No	No	No	No	Yes	Yes	No Re-Hospitalizatio
No Stress Burde	Timely	Committed S	No	Daily	Yes	No	No	Yes	Yes	No	Yes	Yes	No	No	No Re-Hospitalizatio
Pyschological Is	Not timely	Committed S	No	Not daily	No	Yes	Yes	Yes	No	No	Yes	Yes	No	No	Re-Hospitalization o
No Stress Burde	Timely	Not a Smoke	Yes	Daily	No	No	No	No	Yes	No	No	No	Yes	Yes	No Re-Hospitalizatio
Physical Isolate	Not timely	Sneaky Smok	No	Daily	No	Yes	No	No	No	No	No	Yes	Yes	Yes	Re-Hospitalization o
No Stress Burde	Not timely	Not a Smoke	Yes	Daily	Yes	No	No	Yes	No	Yes	No	Yes	Yes	No	Re-Hospitalization o
Pyschological Is	Not timely	Sneaky Smok	No	Not daily	No	No	No	No	No	Yes	Yes	No	No	No	No Re-Hospitalizatio
No Stress Burde	Not timely	Committed S	No	Not daily	No	Yes	Yes	Yes	Yes	No	Yes	Yes	Yes	Yes	Re-Hospitalization o

3.2 System Development Methodology

3.2.1 Re-engineering Current Discharging System

A re-engineered discharging system gives a conceptual understanding of the causes of preventable risks and use of safety design concepts aimed at preventing and minimizing such preventable and re-admissible risks by detecting them upfront before harm occurs (figure 3).

Re-engineering considers both active and latent risks occurring at the time of hospital discharge.

Active re-hospitalization risks include those occurring at the time of hospital discharge during knowledge based decision-making performed at the point of care by Clinicians.

Active re-hospitalization risks are hospital characteristic related as shown in the conceptual model (figure1). Latent condition or risks are observed when there is system failure. Latent conditions are also clinicians and patients related as was shown in the conceptual model.

An example of a latent risk clinician related is when nurses and students are responsible for the discharge process and the harried nature of their work, as well as competing interests (e.g., new admissions requiring attention), results in the discharge of a patient not being considered a high priority, and can lead to an incomplete discharge process. Another example of latent risk patient related is on the lifestyle and non compliance to the discharging guidelines or regime. The Current discharging system is improved through Business Re-engineering Processes (BRP).

Types of re-hospitalization risks that occur at the time of hospital discharge is identified and re-engineered as shown below in figure 3 and figure 4 respectively.

Taxonomy of some avoidable Breast cancer Risk at time of hospital discharge

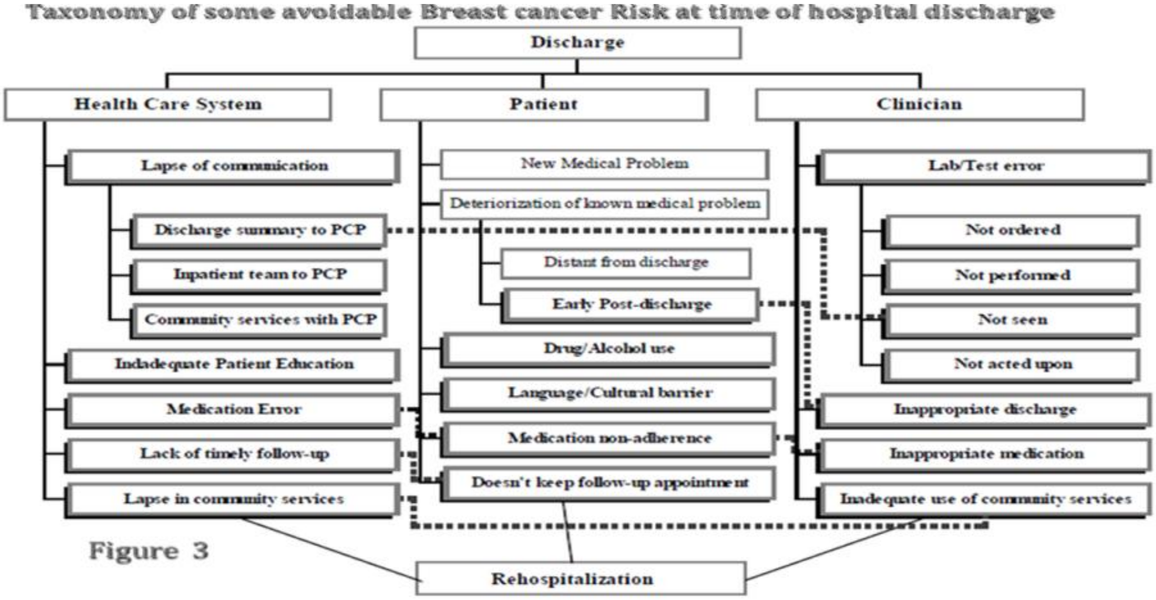


Figure 3

Highlighted boxes indicate risks potentially preventable with an intervention. While not detailed here, each type of risk can be further specified as exemplified by Lab/Test risks. JefreyL.Greenwald et al, 2007

The re-engineered taxonomy above, demonstrates how latent and active risks inter-relate, and highlights the importance of rule-based decision making what the supportive decision model will action. From the re-engineering, it is shown that hospital discharge is the moment when most re-hospitalization risks, lapses, and an adverse event happens.

Also to take note of at this point of discharging is that latent conditions (system failures) are combined with active failures consequently; patient may be discharged with a huge health burden which guarantees re-hospitalization within 30-45 days of discharge.

3.2.2 The Improved Discharging System

The elements displayed pertain to the decision about a patient's readiness for discharge.

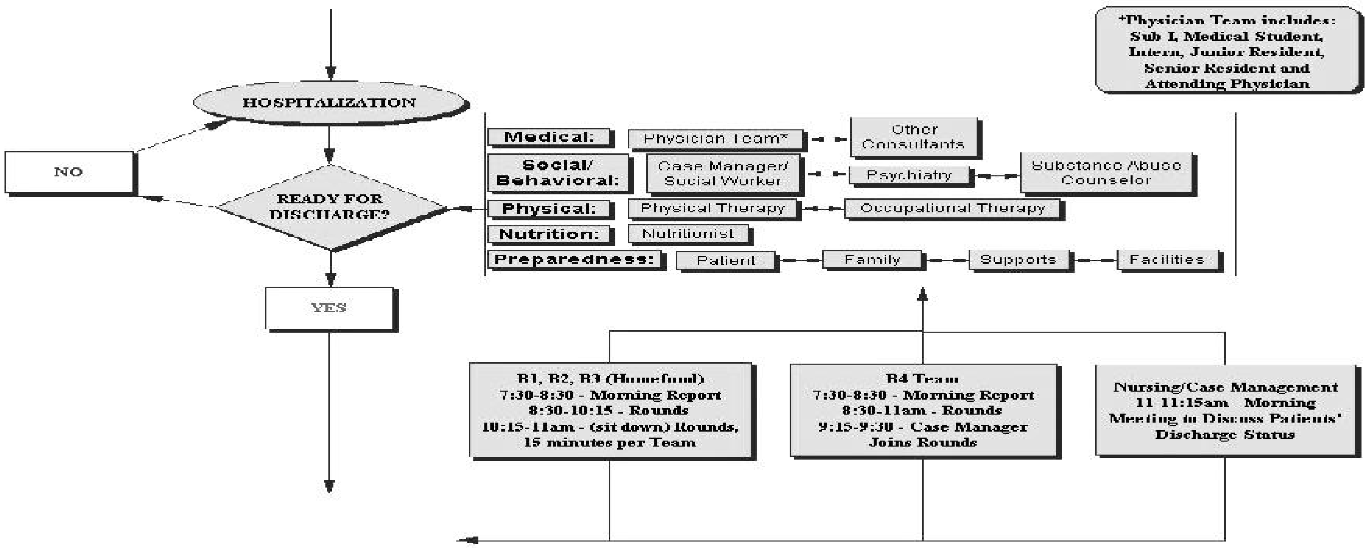


Figure 4: An Improved Discharging System, David Anthony, VK Chetty, et al, (2013).

3.2.3 Source of data

The source of data is from Breast Cancer Past Records, Ward 5B, the Oncologist and Patients in KNH. A research proposal of the entire study is submitted to the KHN/UON-ERC for review, suggestions and approval upon satisfaction. The approved proposal authorizes data access and this study went through the process.

3.2.4 Relevance of the data to the Clinical Decision problem

The identified data instances are represented by attribute-value pairs. For example, in smoking (committed, sneaking, or no smoking), in isolation (psychological, physical or no Isolation) and data attributes (alcoholism, abuse drugs, promiscuous) that have discrete output values (yes/no) befits this clinical decision problem perfectly well.

CHAPTER FIVE: IMPLEMENTING CLINICAL DECISION SUPPORT MODEL

5.1. Functional requirements:

For the time being, it is accepted that “automation, artificial intelligence and decision tree cannot replace human operations” (Terano,1991). A DSM does not automatically reach solutions. It helps (collaborates with) the decision-maker (clinician) for yielding the solution. The clinician is supported not replaced by model. The clinician is the main and final actor in the decision-making process.

For this reason, the clinician organically participate in the model implementation procedures for decision-making; by strong interacting with the DSM.The interactive decision models represent a promising solution for the poorly-structured problems. The implementer must allow the clinician himself to insert his own estimates for the key-parameters, to moderate the models and to obtain and judge the results in different conditions/constraints and under different criteria/objectives.

In principle, a decision solution transforms the initial state of the driven system into a target state.

In the poor-structured problems, all these three elements (initial state, decision solution and target state) have poor structure (uncertainty). Consequently, the basic objective of a DSM is to support the clinician (to offer aggregated information and knowledge) for structuring (clarifying) all the three elements.This clarifying activity needs iterative passing through all the steps of decision process and the DSM must support all of them, by providing the appropriate functions to help:

-Acquisition of information concerning the problem;

-Problem identification/definition (objectives, constraints, etc.);

-Establishing the model for solving the problem;

-Establishing the variants/options (preventable risk factors, risks attributes, etc.); (including socio-economic data, sustainability data and specific data to use the probabilistic and decision tree methods for coping with the uncertainty;

-Establishing the appropriate models for DSM analysis and evaluation;

-Friendly interfacing decision-maker with the DSS: access to the DSS functions; parameter exchange.

5.2. Operating Requirements:

-Flexibility, reliability and ease in use;

-System integrity and security;

-Adaptability to different user requests.

5.3. Implementation Requirements:

-Rapid miner version 6.0 for developing DSM ,Validations and Evaluations.

-Windows setup tools-version 5.7. for enabling window 7 operating system to accept installation of Django- rest framework.

-Free source software Python version -2.7.8 for coding

- Free source software Django-1.6.5 for enabling python codes into web base

-Django -rest-framework-master for enabling interaction of the web base to python codes

-Laptop computer core i3 and above, hard disk 250gb,ram 1gb.

5.4. The Implementation Stages

The implementation is divided into five stages. In the first stage, data attributes that causes re-hospitalization is collected then uploaded into the Rapidminer software. In the third stage, data preprocessing and visualization in the rapid miner is executed. In the fourth stage, Modeling and generation of a predictive decision trees based on ID3 and C4.5 algorithms is action. Lastly decision tree model is mapped into a web base database.

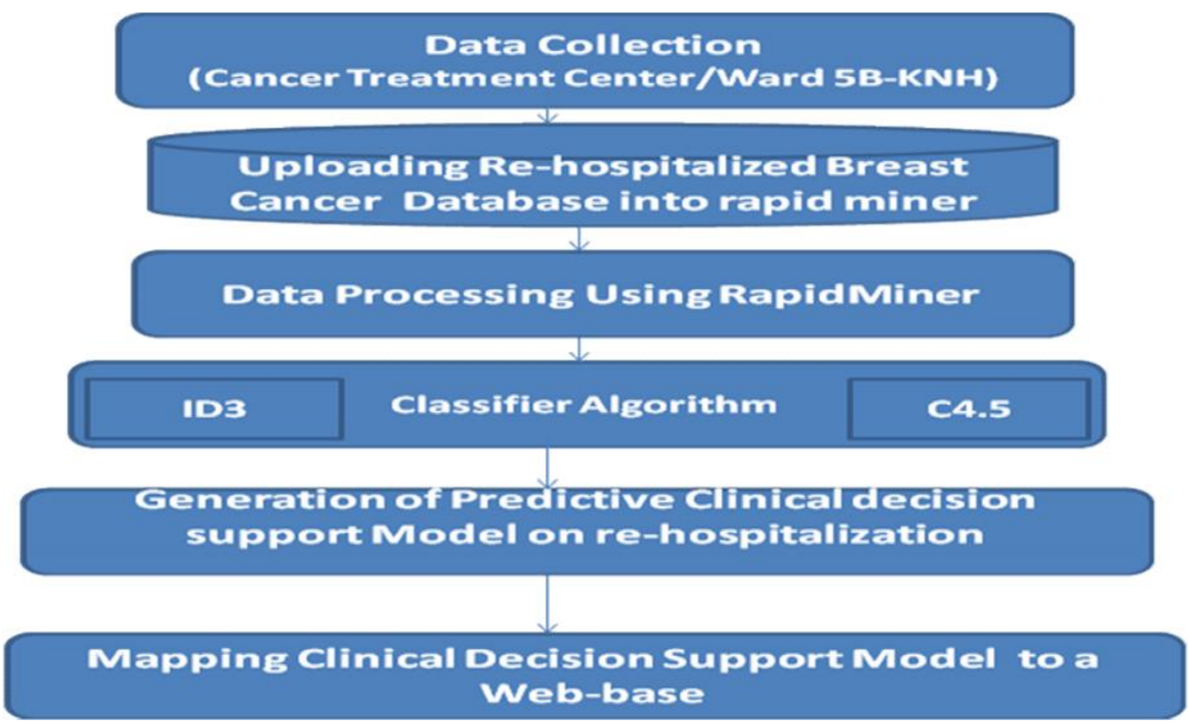


Figure 5: The Implementation Stages

5.4.1 Uploading Re-hospitalized Breast Cancer Database

The next step is to feed the preprocessed patient's database as input to RapidMiner as shown in the snapshot below.

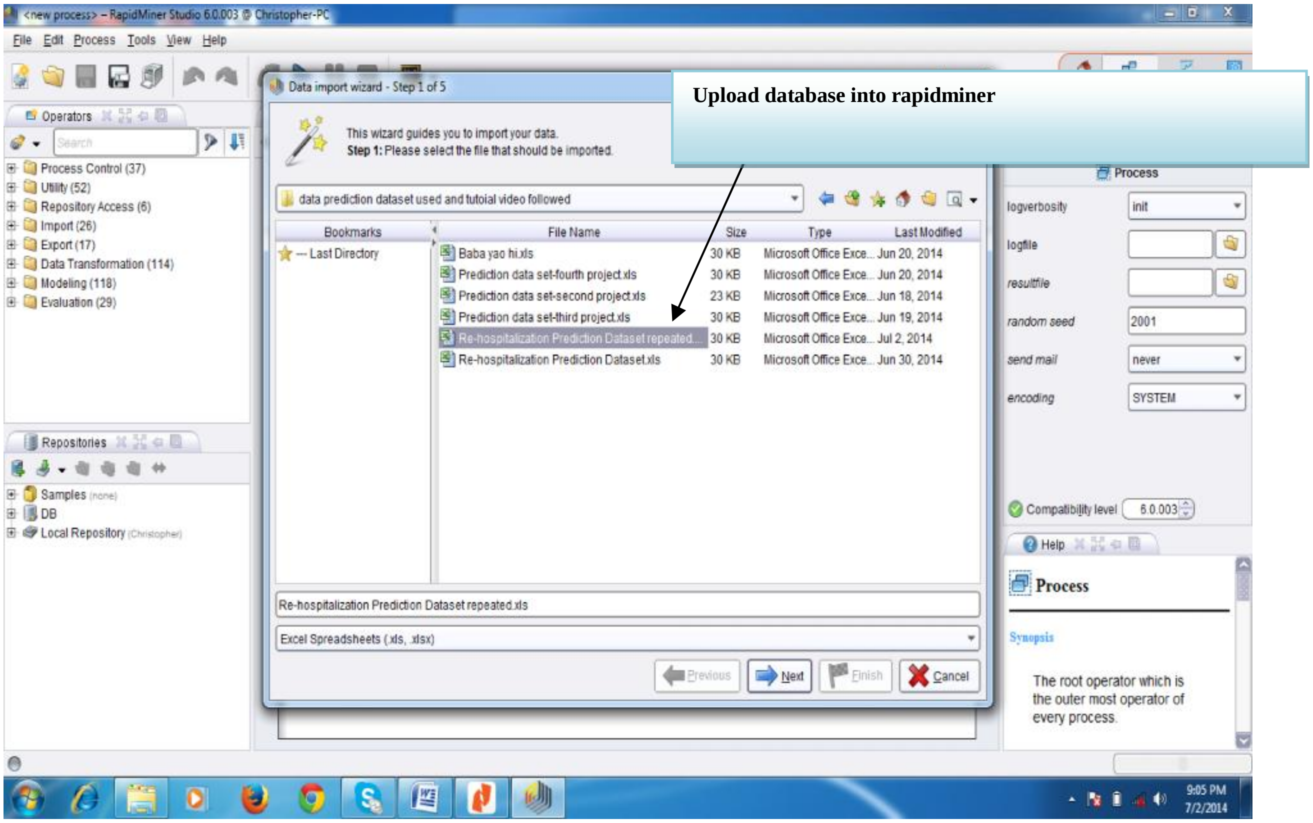


Figure 6: Uploading Re-hospitalized Breast Cancer Database

5.4.2 Data Processing and Visualization

The uploaded patient's database is then processed as shown in the following snapshot

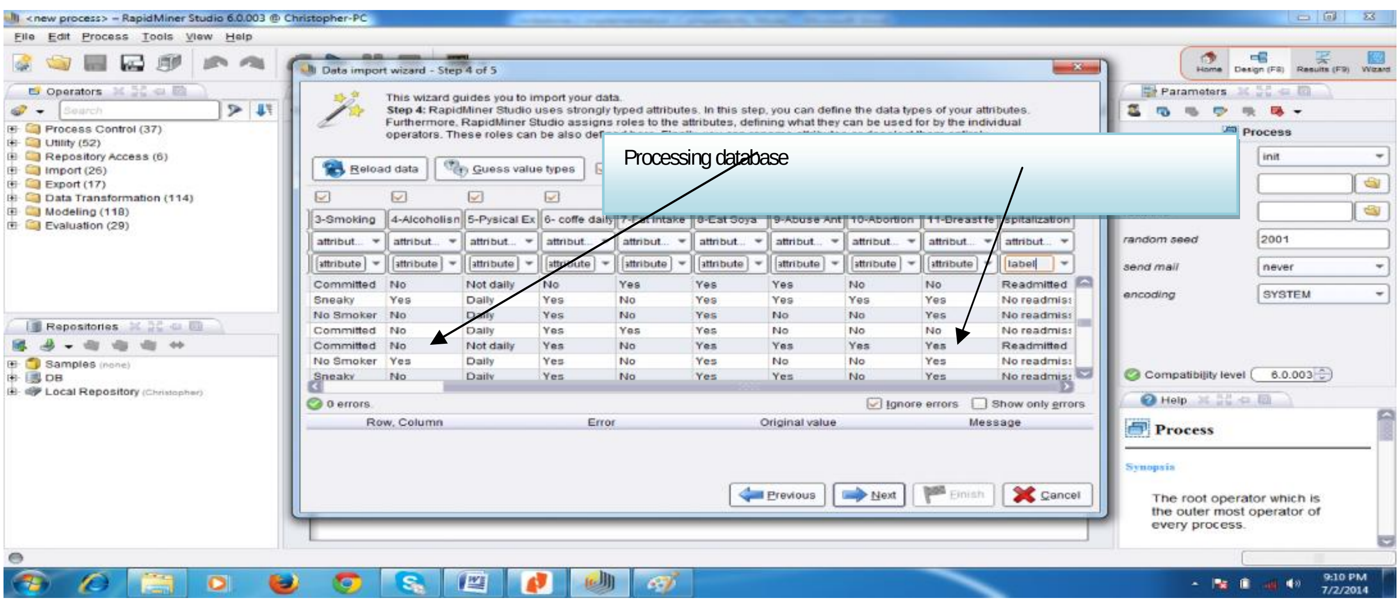


Figure 7: Data Processing and Visualization

5.4.3 The software design and implementation in summary

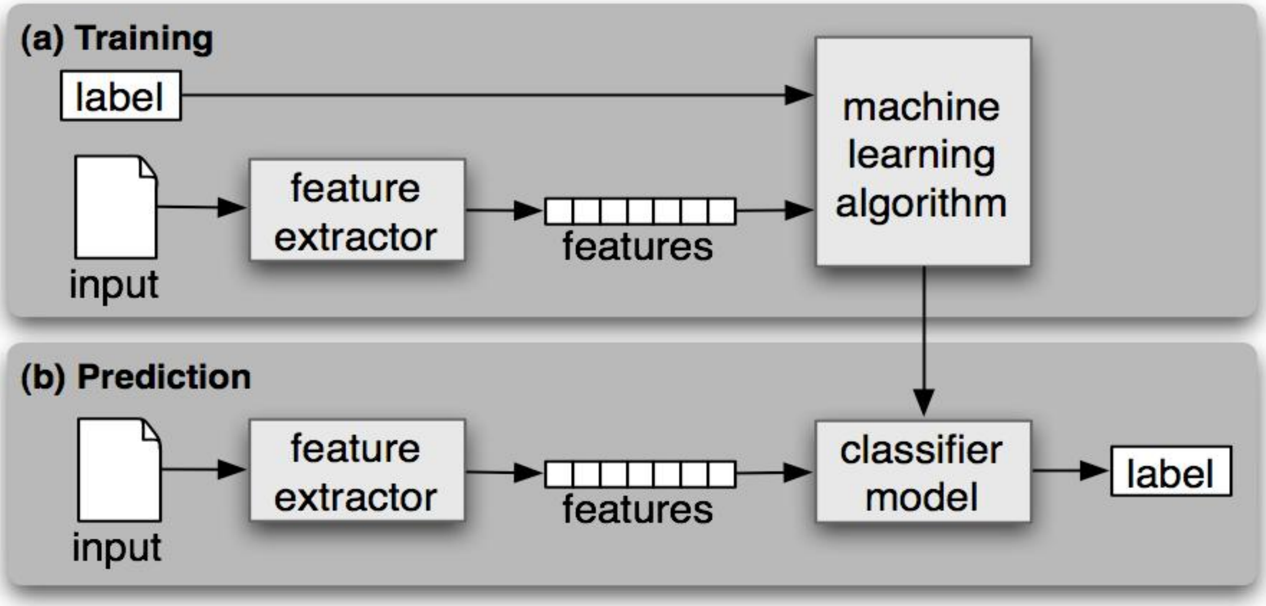


Figure 8: The software design and implementation in summary

5.4.4 Splitting the Patient Database

The software is designed to split the database into training and validation dataset. The hypothetical optimal decision tree learns the training dataset and gets validated using validation dataset. Patient's dataset is split in the ratio of 0.7and 0.3 for training and validation respectively. This is shown the screen shot below.

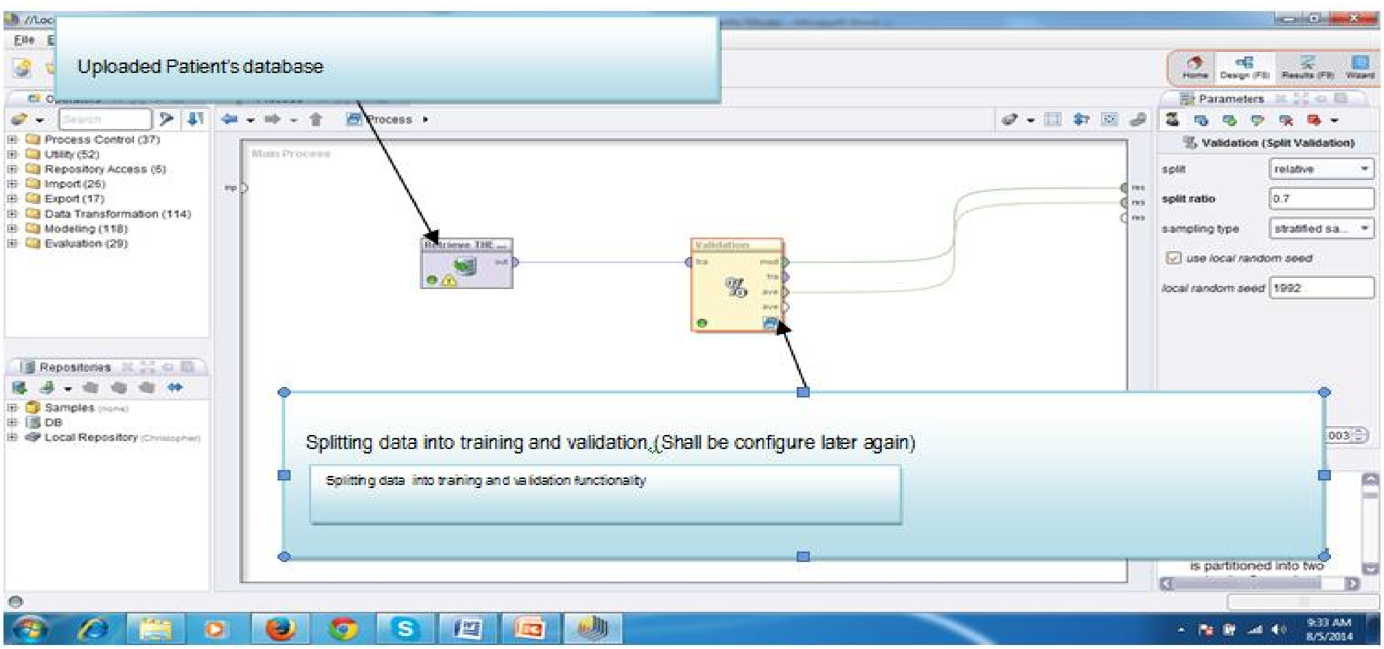


Figure 9: Splitting the Patient Database

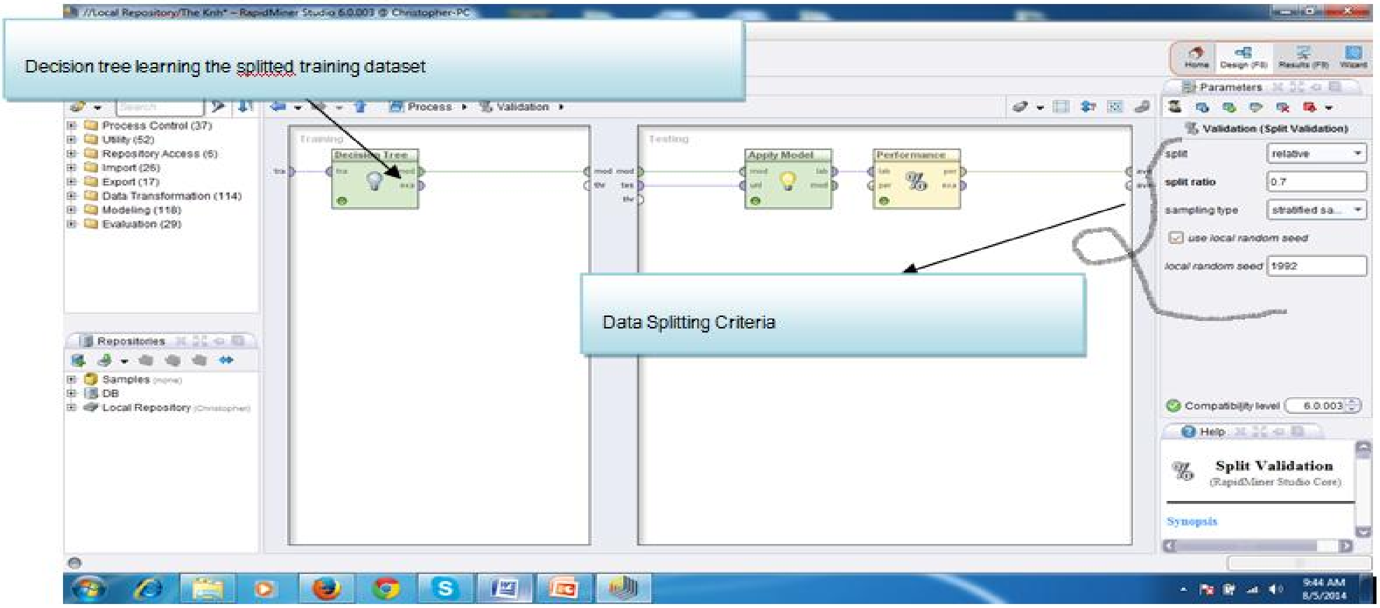


Figure 10: Decision learning and Training

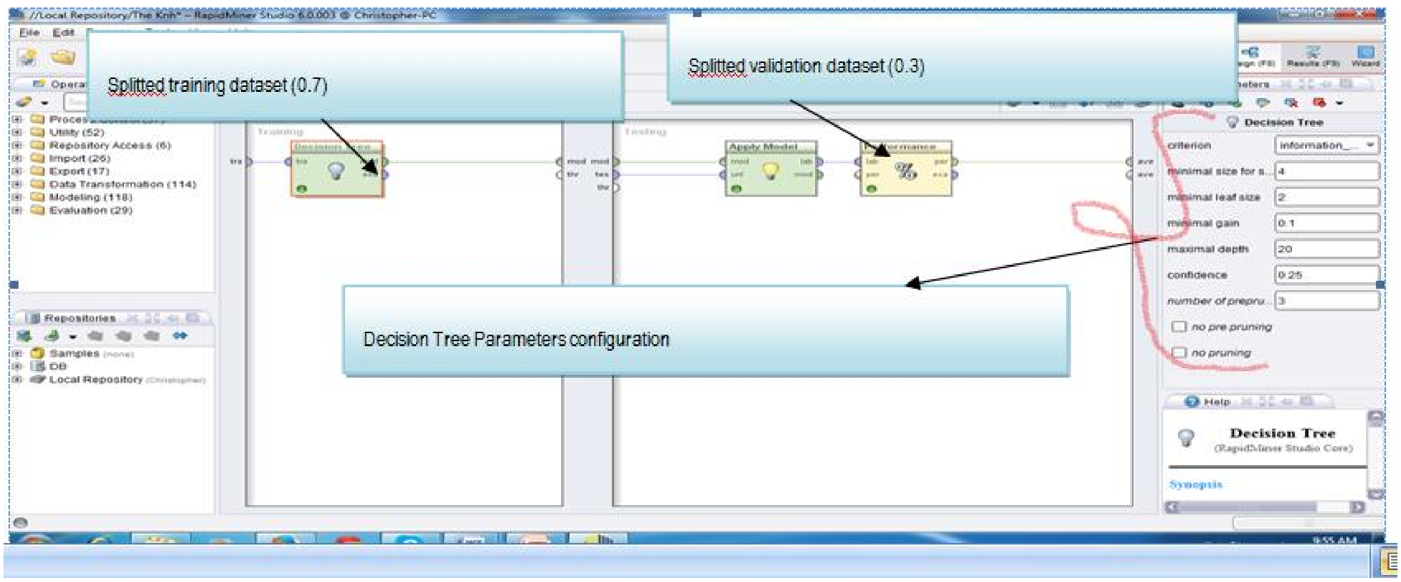


Figure 11: Splitting Training dataset to validate the model

5.4.5 Training the decision tree model (Algorithms)

We bring the decision tree model to learn training data as shown in figure 10 above. Note that this decision tree is configured as shown in the top left hand side on the parameters to have it reflect the hypothetical optimal tree.

ID3 Algorithm

This algorithm is said to be in use when check box on the “no pre pruning” and “not pruning” is enable as shown above.

C4.5 Algorithm

This is an improved ID3 algorithm and has fewer decision nodes as compared to ID3 algorithm whose "no pre pruning" and "not pruning" is not check to ensure pruning of the resulting decision tree. This algorithm is the one adopted in the above screenshot.

5.4.6 Validating learned Hypothetical Decision Support Model

Apply model and performance operators in the validating or testing dataset and configure the validating parameters.

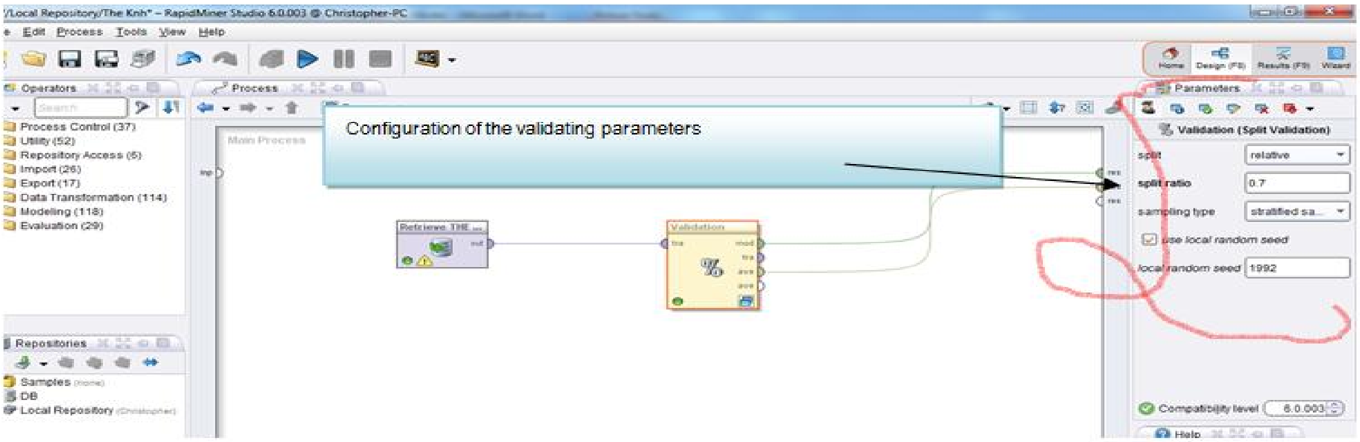


Figure 12: Validating learned Hypothetical Decision Support Model

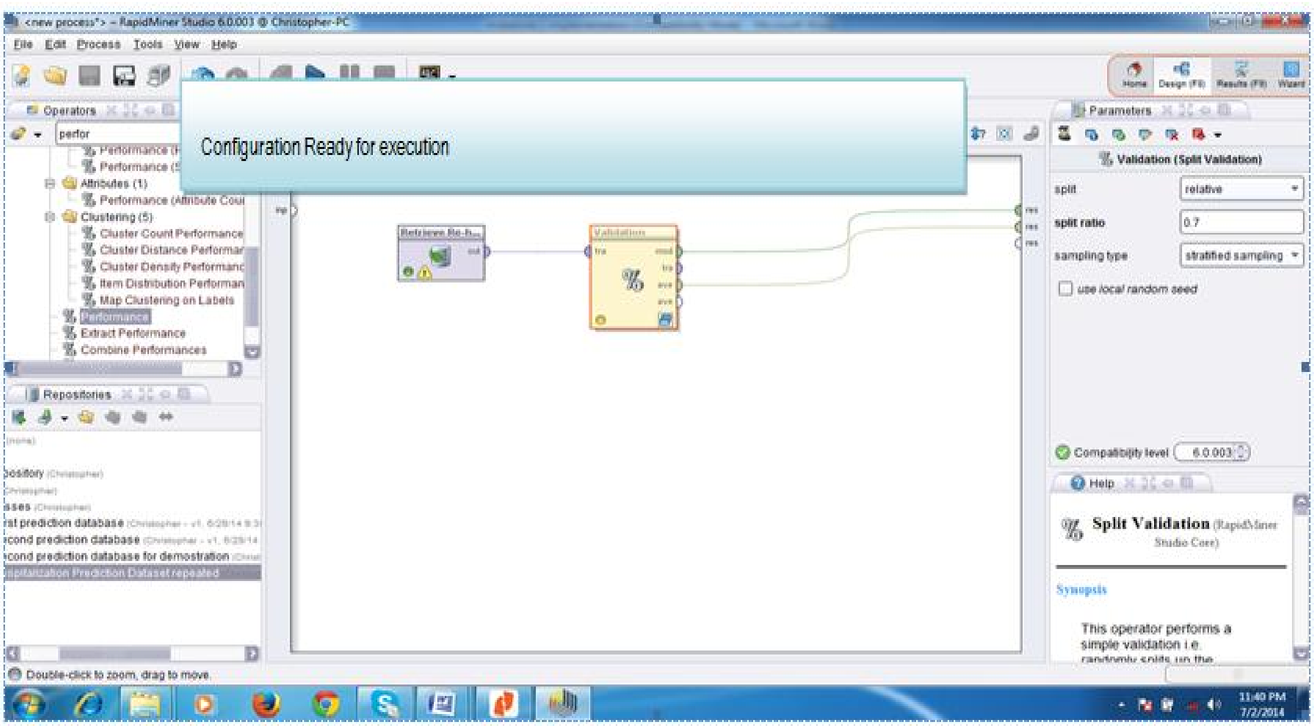


Figure 13: Configuration ready for execution

Click run.

The resulting clinical decision model is termed as an unpruned model and has high error rate as referred in the following observation by Lemma's theorem:

Definition (Overfitting)

Let D be a set of examples and let H be a hypothesis space. The hypothesis $h \in H$ is considered to overfit D if an $h' \in H$ with the following property exists:

$$Err(h, D) < Err(h', D) \quad \text{and} \quad Err^*(h) > Err^*(h'),$$

where $Err^*(h)$ denotes the true misclassification rate of h, while $Err(h, D)$ denotes the error of h on the example set D.

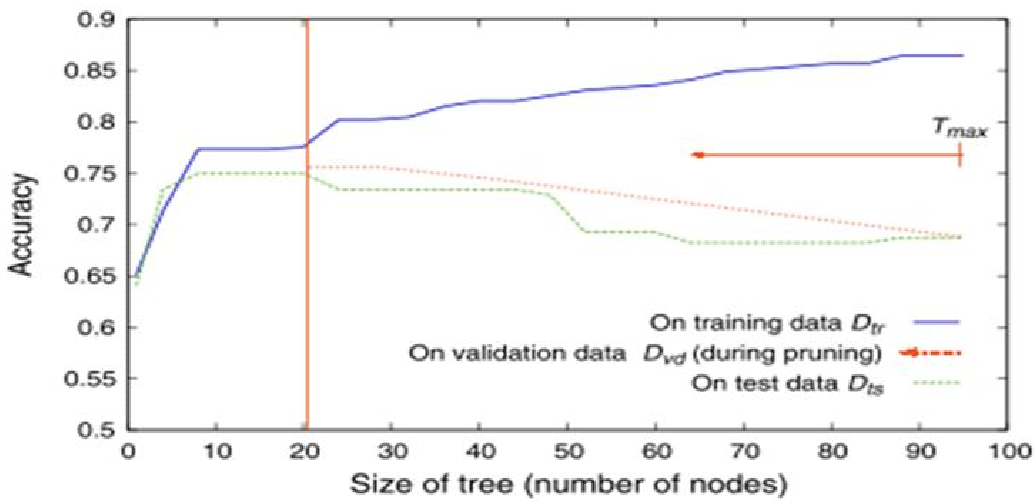
Reasons for overfitting are often rooted in the example set D:

- D is noisy
- D is biased and hence non-representative
- D is too small and hence pretends unrealistic data properties

Given a hypothesis Universal space H, and a hypothesis h(unproved decision tree) $h \in H$ which is a member of a Universal space. $h \in H$ is said to over fit the training data if there exist in the same H, an alternative hypothesis $h' \in H$ which is a member of H,Such that $h \in H$ the unpruned decision tree has a small error than an alternative hypothesis over the training examples, but $h' \in H$ an alternative hypothesis has a smaller error than $h \in H$ which is the unpruned decision tree over the entire distribution of instances (unseen examples). We are simply saying that $h \in H$ (unproved decision tree) has memorize the training example and can less classifier beyond the training example.

An alternative hypothesis $h' \in H$ has not memorize the training example hence has high error rate on memorizing the training examples than $h \in H$ (unproved decision tree) and is the reason why it has low error rate on unseen training example than $h \in H$ (unproved decision tree).

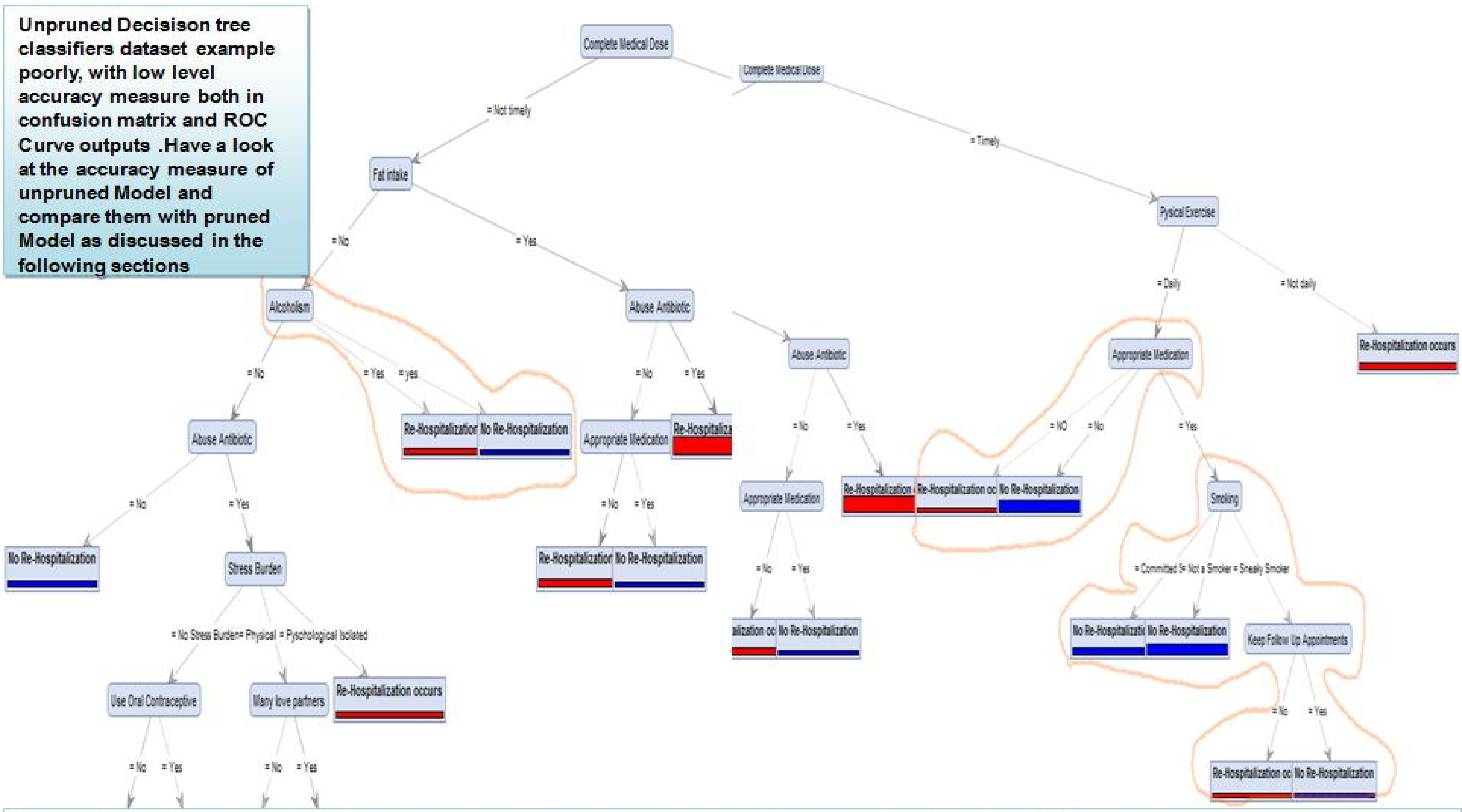
Behavior of the unpruned decision model (h) on both training and unseen(test) data set.h is doing very well on training data cause it has memorize the data



[Mitchell 1997]

Figure 14: behavior of pruned and unpruned model on new data

5.4.7 The Unpruned Clinical Decision support Model



Pruning is a technique that reduces the size of decision tree by removing sections of the tree that provide little power to classify instances. The dual goal of pruning is reduced complexity of the final classifier as well as better predictive accuracy by the reduction of overfitting and removal of sections of a classifier that may be based on noise or erroneous data(Tom Michel)

Figure 15: The Unpruned Clinical Decision support Model

5.4.8 The Pruned Clinical Decision support Model

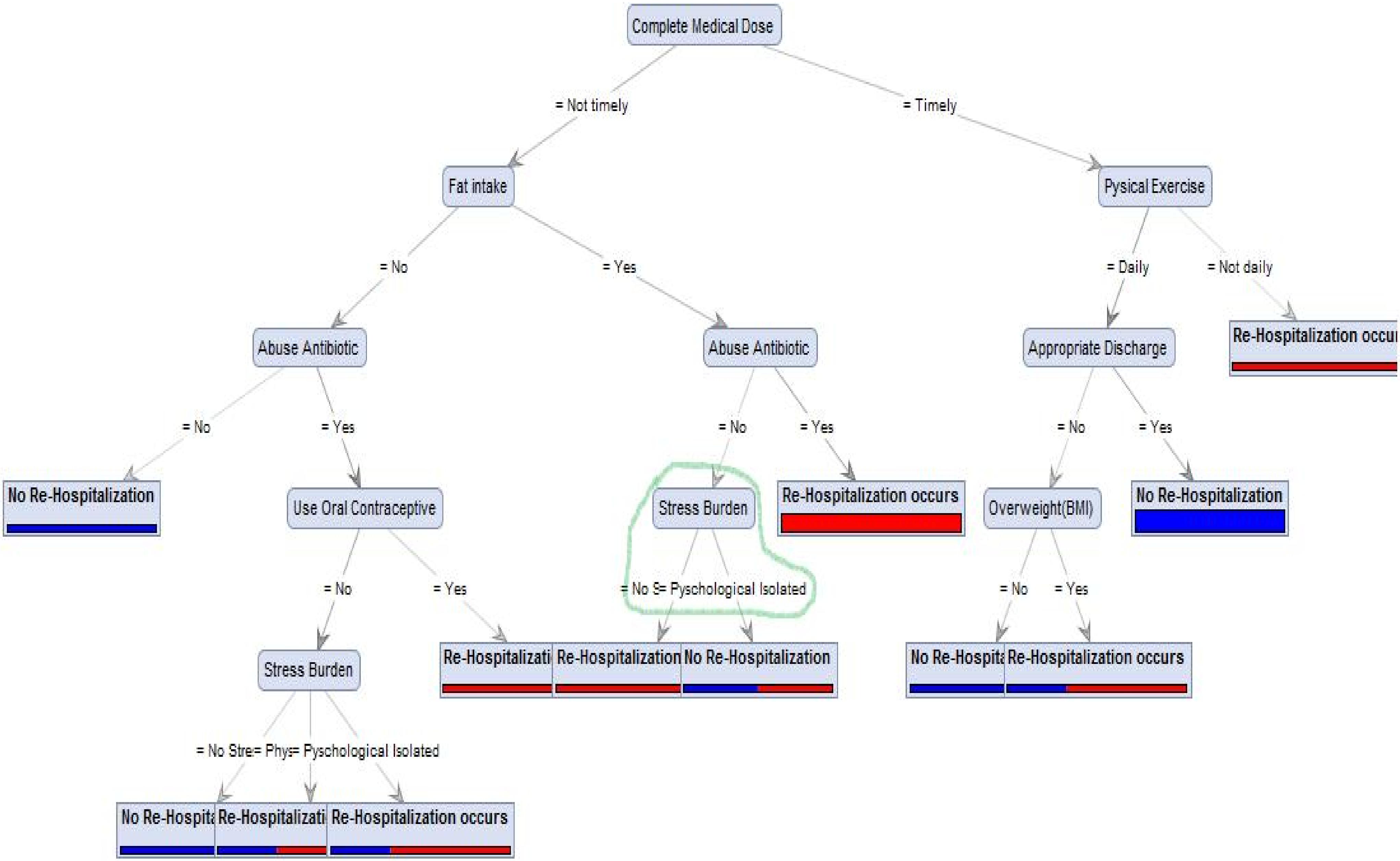


Figure 16: The Pruned Clinical Decision support Model

CHAPTER SIX: MODELING EVALUATION AND ANALYSIS OF RESULTS

6.1 The Model Accuracy Measures:

6.1.1 The Confusion Matrix

We begin by considering classification problems using only two classes. Formally, each instance ***I*** is mapped to one element of the set {p, n} of positive and negative class labels. A classification model (or classifier) is mapping from instances to predicted classes. Some classification models produce continuous output (e.g., an estimate of an instance ***I*** is a class membership probability) to which different thresholds may be applied to predict class membership.Other models produce discrete class label indicating only the predicted class of the instance. To distinguish between the actual class and the predicted class we use the labels {P1, N2} for the class predictions produced by a model.

6.1.2 The receiver operating characteristic (ROC)

ROC graphs have long been used in signal detection theory to depict the tradeoff between hit rates and false alarm rates of a classifier (Egan,1975; Swets et al., 2000). The medical decision making community has an extensive literature on the use of ROC graphs for diagnostic testing (Zou, 2002).Recent years have seen an increase in the use of ROC graphs in the machine learning community, due in partly to the realization that simple classification accuracy is often a poor metric for measuring performance (Provost and Fawcett, 1997; Provost et al., 1998) and that they have properties that make them especially useful for domains with skewed class distribution (test set) and unequal classification error costs.These characteristics have become increasingly important as research continues into the areas of cost-sensitive learning and learning in the presence of unbalanced classes(test set).

6.2 Unpruned Decision Model Accuracy Measures

The following is the matrix for unpruned decision model. These are two measurements accuracy, the confusion matrix and the ROC curve analysis. Unpruned model has poor performance generally as can be observed in the general accuracy measurements

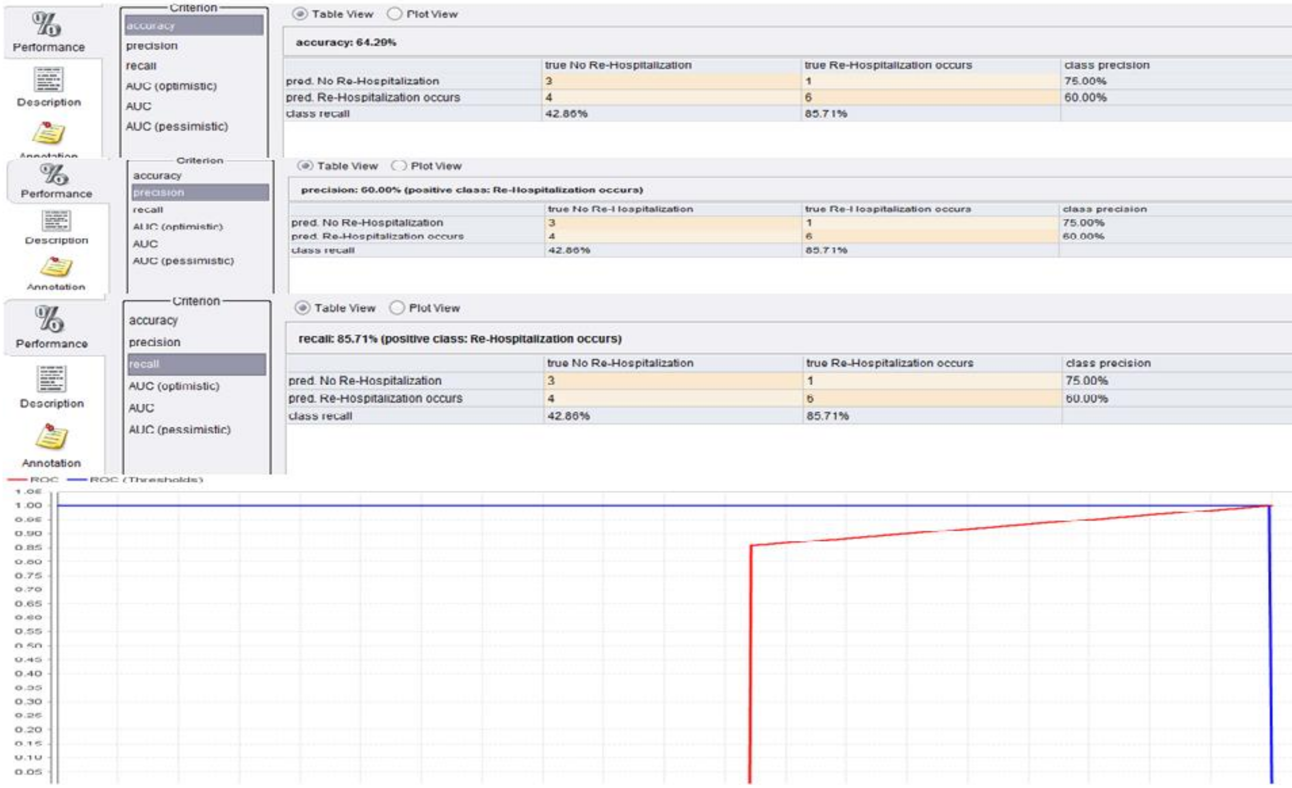
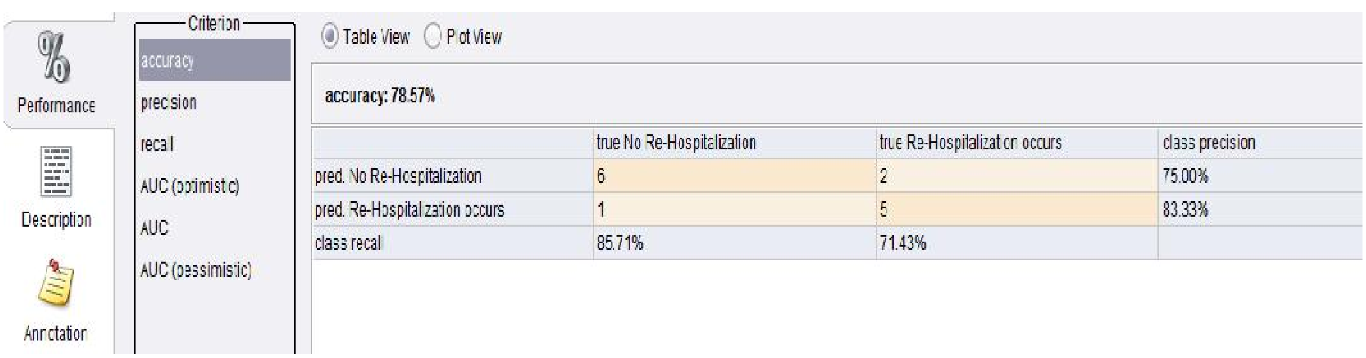


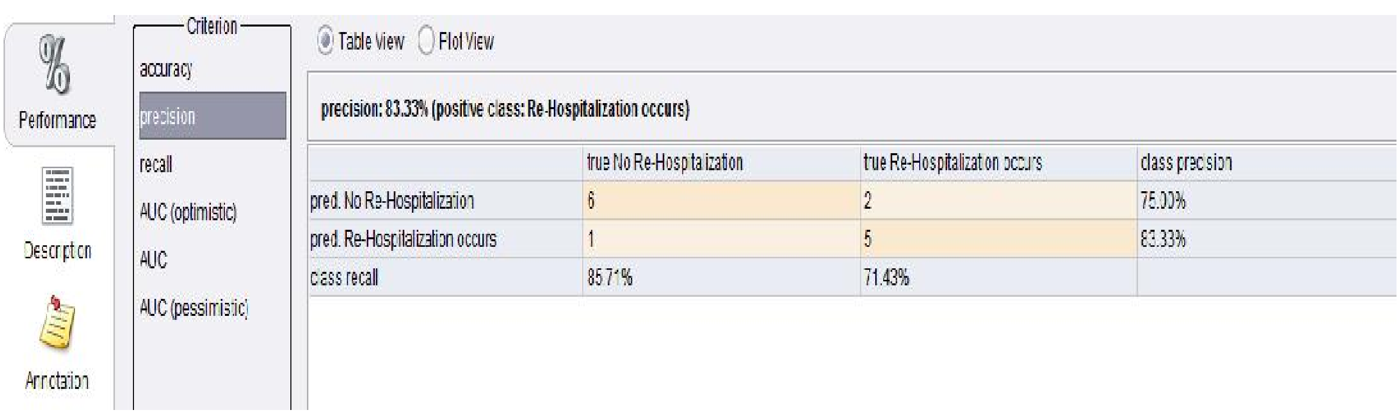
Figure 17: Measures Matrix and ROC Curve of Unpruned model

6.3 Pruned Decision Model Accuracy Measures

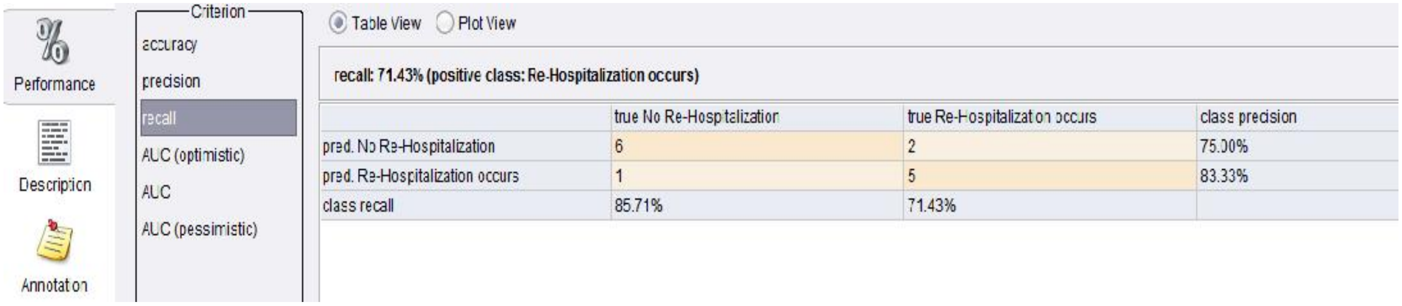
6.1.3 The Confusion Matrix of Pruned Decision Model



	true No Re-Hospitalization	true Re-Hospitalization occurs	class precision
pred. No Re-Hospitalization	6	2	75.00%
pred. Re-Hospitalization occurs	1	5	83.33%
class recall	85.71%	71.43%	



	true No Re-Hospitalization	true Re-Hospitalization occurs	class precision
pred. No Re-Hospitalization	6	2	75.00%
pred. Re-Hospitalization occurs	1	5	83.33%
class recall	85.71%	71.43%	



	true No Re-Hospitalization	true Re-Hospitalization occurs	class precision
pred. No Re-Hospitalization	6	2	75.00%
pred. Re-Hospitalization occurs	1	5	83.33%
class recall	85.71%	71.43%	

Patient class	Predicted by the model to be Re-hospitalized (P_1)	Predicted by the model not to be Re-hospitalized (N_2)	total	Recognition (%)
To be Re-hospitalized *(P_1)*	6 (TP)	2(FP)	8	75% *(precision)*
Not Re-hospitalized*(N_2)*	1 (FN)	5 (TN)	6	71.42%(*specificity)*
Total	7 (P)	7 (N)	14 (All)	78.57% (*accuracy*)

Figure 18: Measures Matrix in Rapid Miner and Manual Calculation of Pruned model

Confusion matrix and common performance metrics calculated from the model.

Precision = (TP)/(TP)+(FP)=6/8 = 75% ,also Precision = (TN)/(TN)+(FN)=5/6 = 83.33%

Recall = (TP)/(TP)+(FN)=6/7 = 85.71 % also similar to sensitivity.

Classifier Accuracy, or recognition rate: percentage of test set tuples that are correctly classified

Accuracy = (TP + TN)/All =6+5/14=78.57%.

Error rate: 1 – accuracy, or Error rate = (FP + FN)/All=2+1/14=0.2142857142857143

Sensitivity: True Positive recognition rate (Recall positive), Sensitivity = TP/P=6/7=85.714%

Specificity: True Negative recognition rate (Recall Negative), Specificity = 5/7=71.42%.

By convention, the performance of a classification model is usually summarized by the following two quantities related to the two types of errors: true-positive rate and false-positive rate.

In this context, the true-positive rate is the probability that a patient to be re-hospitalized is correctly classified as shall be re-hospitalized, and the false-positive rate is the probability that a patient who shall not be re-hospitalized is incorrectly classified as shall be re-hospitalized.

(The true-positive rate is also called sensitivity or recall and one minus the false-Positive rate is also called specificity.)

For an ideal classification rule, the true-positive rate is one and the false-positive rate is zero.

The magnitudes of acceptable false-positive rates and true-positive rates depend on the corresponding costs and perceived benefits of the institution concern.

6.1.4 The receiver operating characteristic (ROC) of Pruned Decision Model

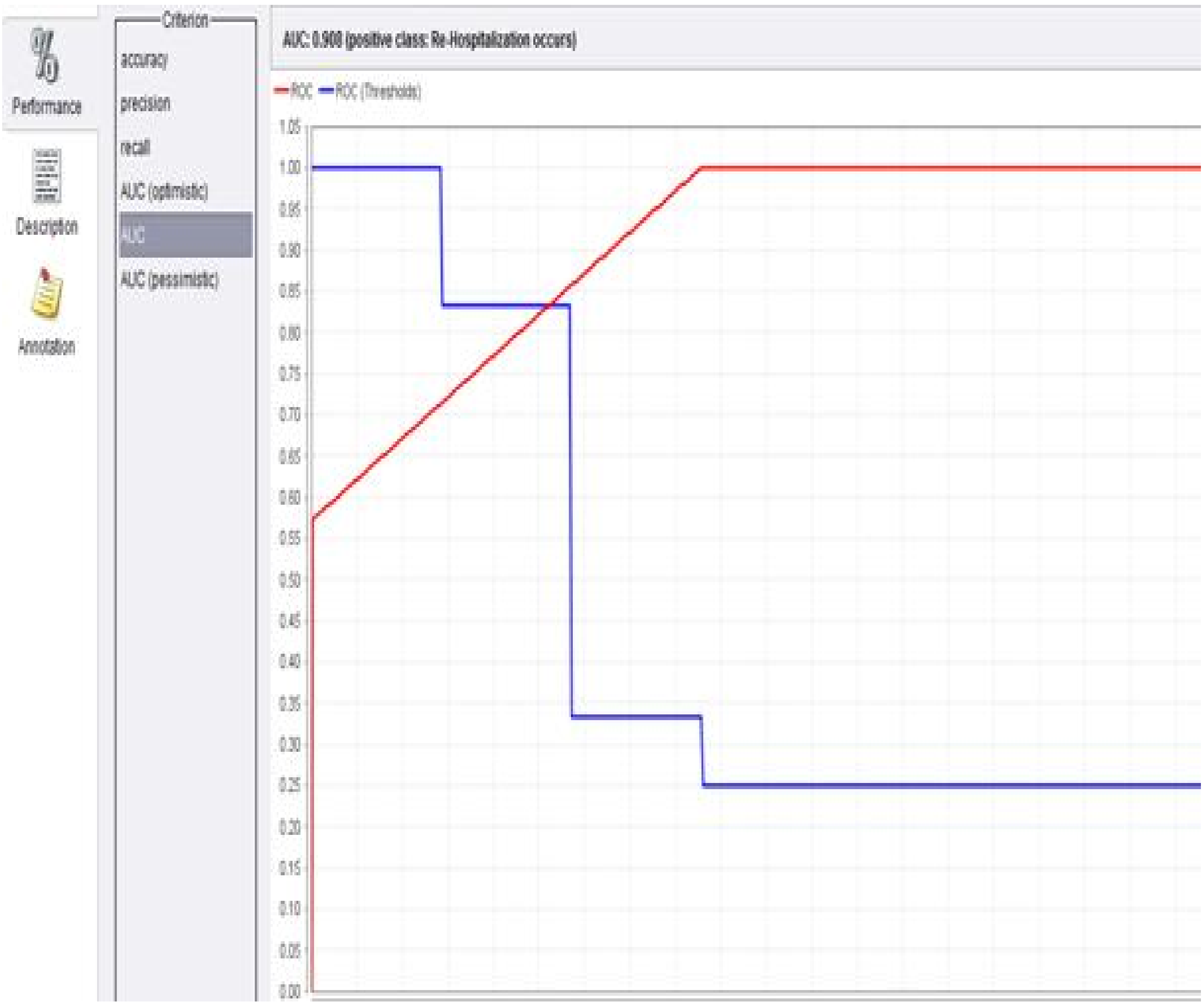


Figure 19: ROC Curve of the Pruned model

The ROC space

ROC graphs are two-dimensional graphs in which **TP** rate is plotted on the Y axis and **FP** rate is plotted on the X axis. An ROC graph depicts relative tradeoffs between benefit (true positives) and costs (false positives). Fig. 19 shows an ROC graph of a discrete classifier that outputs only a class label. Each discrete classifier produces an (**TP** rate, **FP** rate) pair corresponding to a single point in ROC space. The classifiers in Fig.19 are all discrete classifiers. Several points in ROC space are important to note. The lower left point (0, 0) represents the strategy of never issuing a positive classification; such a classifier commits no false positive errors but also gains no true positives. The opposite strategy, of unconditionally issuing positive classifications, is represented by the upper right point (1, 1).The point (0, 1) represents perfect classification.

Informally, one point in ROC space is better than another if it is to the northwest (**TP** rate is higher, **FP** rate is lower) appearing on the left-hand side of an ROC graph, near the X axis, may turn corresponds to one ROC point. Thus, a discrete classifier produces only a single point in ROC space. Fig 19 shows an ROC ‘‘curve’’ on a test set of 49 instances.

The instances are 25 positive and 24 negative.

Any ROC curve generated from a finite set of instances is actually a step function, which approaches a true curve as the number of instances approaches infinity.

The step function in Fig 19 is taken from a very small instance set so that each point’s derivation can be understood.

Although the test set is very small; we can make some tentative observations about the classifier.

It appears to perform better in the more conservative region of the graph; the ROC point at (0.0, 0.7) produces its highest accuracy (70%).

This is equivalent to saying that the classifier is better at identifying likely positives than at identifying likely negatives. Note also that the classifier best accuracy occurs at a threshold of P (0.45, 1.0) rather than at P (0.0, 1.0) as we might expect with a balanced distribution.

The classifier performs well in the entire classification except at P (0.0, 1.0).

More generally, the graph displays a cloud of (false-positive rate, true-positive rate) points, and the optimal ROC curve is the line connecting the points highest and farthest to the left.

The rationale for the optimal ROC curve is that:

a) One wants the highest true-positive rate for a given false-positive rate, and

b) One can specify a rule on the ROC line linking two (false-positive rate, true-positive rate) points by applying the rule for one point with some probability and the rule for the other point with one minus that probability.

However, in practice one would like one of the points on the optimal ROC curve to lie near the target false- and true-positive rates. For the reasons given above, interest is in the part of the ROC curve corresponding to a low false-positive rate when evaluating prediction.

The area under (ROC) curve is known as AUC.

This area, therefore, should be greater than 0.5 for a model to be acceptable; a model with AUC of 0.5 or less is worthless. Understandably, this area is a measure of predictive accuracy of model.

CHAPTER SEVEN: DESIGN AND IMPLEMENTION ON PYTHON WEB APPLICATION

The learnt predictive patterns for predicting patients' re-hospitalization is then implemented on web application. The predictive patterns are mapped to code in form of class methods in Python and translate code in the form of if-else ladders.

We then placed these ladders into Python class methods that accept only the splitting attributes - Isolation, Complete dose, Smoking, Alcoholism, Physical Exercise, coffee daily, fat intake, Eat Soya, Abuse Antibiotic, Abortion as method parameters. The class methods return the final result of that particular evaluation, indicating whether that patient would be re-hospitalized, will not or both when discharged.

Code Igniter- Django-1.6.5.

The web application of the Decision Support Model (DSM) is developed using a Python framework named Django-1.6.5.The application has provisions for multiple simultaneous clinician's registrations and logins. This ensures that the work of no two clinicians is interrupted during re-hospitalization evaluations. Figure (20 and 21) depict the DSM for registration and login interfaces respectively.

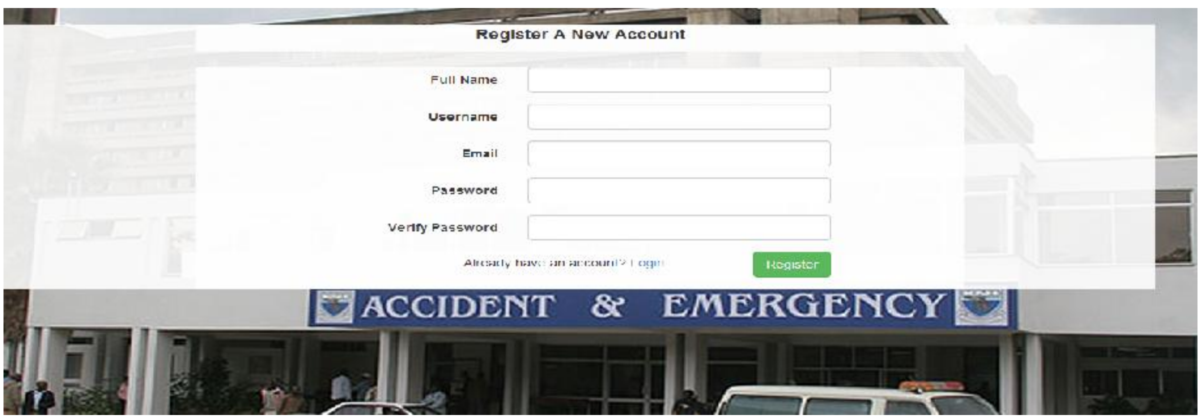


Figure 20: Registration Interface for Clinicians

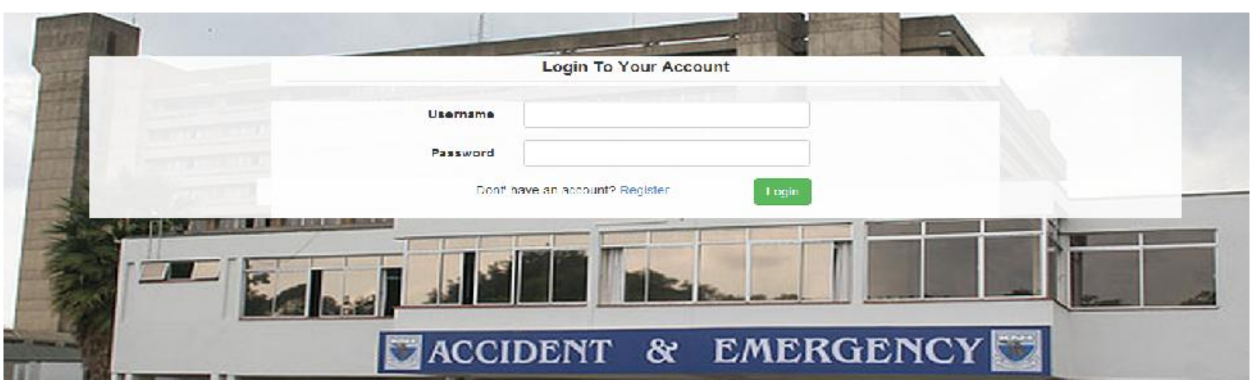


Figure 21: Login interface for registered Clinicians

Python web base application for Re-hospitalization Prediction

Once the DSM is mapped as class python methods, we built a web page for clinicians to do entry of the splitting values attribute of a patient as can be seen in figure 22 below. These values are used to predict avoidable re-hospitalization of a patient as either "Re-hospitalization occurs", or "no re-hospitalization"

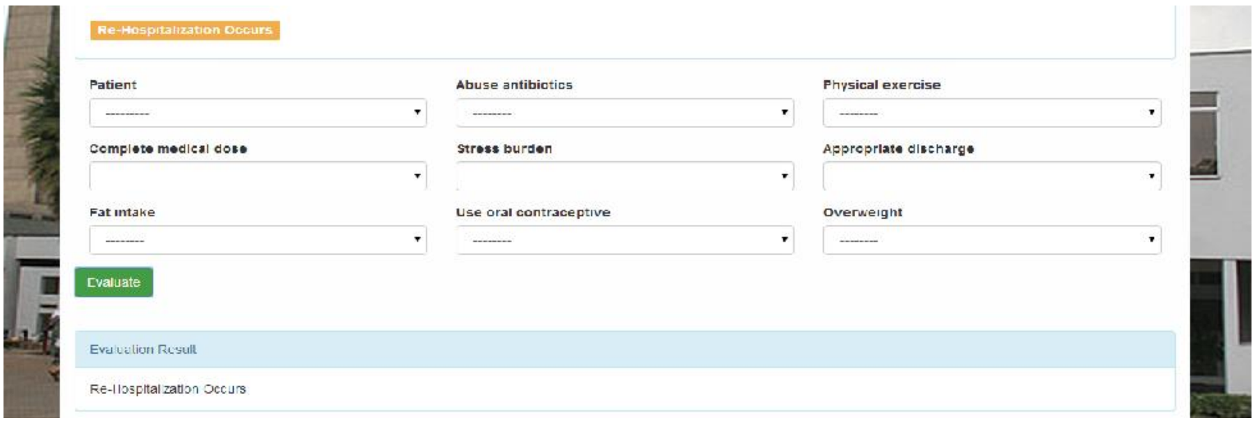


Figure 22: Web interface for attributes entry and results for re-hospitalization

Patient Name : Robert The Doctor

Patient Phone Number : +254726743396

Report

Date	Completed Medical Dose	Fat Intake	Abuse Antibiotics	Stress Burden	Use Oral Contraceptive	Physical Exercise	Appropriate Discharge	Overweight	Outcome
Aug. 17, 2014	Not Timely	Yes	No	No Stress Burden	No	Daily	No	Yes	No Re-Hospitalization
Aug. 17, 2014	Not Timely	Yes	Yes	Physical Isolation	Yes	Daily	No	Yes	Re-Hospitalization Occurs

print

Figure 23: Printable results for attributes entered on re-hospitalization.

Take note of the difference outcome on moderating values attributes on re-hospitalization on the same patient; Robert The Doctor.

CHAPTER EIGHT: DISCUSSION AND CONCLUSIONS

8.1. Discussions of the main Results

The prediction results of the decision tree modeling methods are demonstrated in figure 18 on the confusion matrixes. All the measurements parameters of the model recorded high level of accuracy which was confirmed by the Roc curve results. Therefore the overall accuracy of models is highly acceptable for example the model has a Precision of 75% for true positive and 83.33% for false positive with a Recall similar to sensitivity of 85.71 % .The general classier accuracy stood at 78.57% with an Error rate of 0.2142857142857143.Sensitivity of the model is recorded at 85.714% and its Specificity at 71.42%.The model is a poorly classifier at the initial stages for example while classifying 0.825 true positive and a good classifier in the entire classification stages. The DSM is largely a perfect classifier with high level of accuracy.

8.2. The Value of the study:

The study is solving the poorly structured problem of uncertainty in clinical decision making. This lowers avoidable mistakes, adverse events and even problem of thinking hard at the point of decision making hence reduction of subsequent resource usage or even death.

The poorly structured problem is transformed into a structured problem by stating the problem initial state, Solution state and the target state .When the problem is in these states; the curse of dimensionality in the decision making is lowered significantly because there are some rules and directives on how to reach the target solution.

The study also contributes at generating new theories. This model is valuable for theory building. The new types of data sets available are rich in detail; they include and combine information of multiple types (e.g.,temporal, cross-sectional, geographical, and textual), on a large number of observations, and with high level of granularity(e.g.,clicks or bids at the seconds level).Such data often contain complex relationships and patterns that are hard to hypothesize, especially given theories that exclude many newly measurable concepts.The model is designed to operate in such environments and detects new patterns and behaviors and help uncover potential new causal mechanisms, in turn leading to the development of new theoretical models.

The study helps in developing new measures since its support construct operationalization.

Operationalization implies internal and external theoretical grounding. This aspect is a more specific instance of new theory generation, since the development of new theory often goes hand in hand with the development of new measures (Compeau et al. 2007;28).

The study access relevance.It remains true that if we can predict successfully on the basis of a certain explanation (Statistical model), we have good reason and perhaps the best sort of reason, for accepting the explanation. The model is a useful tool for assessing the distance between theory (Statistical model), and practice. For example although explanatory (Statistical model) power measures can tell us about the strength of a relationship they do not quantify the empirical model's accuracy level in predicting new data.

In contrast, assessing predictive power of a theory can shed light on the actual performance of an empirical model. The model can therefore be used to assess practical relevance of a theory .Keil et al.

The model can be used to improve existing models.The model capture complex underlying patterns and relationships, and thereby improve existing explanatory statistical models.

It can be a Benchmark for accessing predictability power of other models to arouse appetite for further research. This clinical Predictive model plays an important role in quantifying the level of predictability of measurable phenomena (Ehrenberg and Bound 1993) by creating benchmarks of predictive accuracy.(There must be 3 models e.i predictive, theoretically grounded and newly researched model).Predictive accuracy benchmark is useful for evaluating the difference in predictive power of existing stable and grounded explanatory model. On one hand, an explanatory model that is close to the predictive benchmark may suggest that our theoretical understanding of re-hospitalization can only be increased marginally. On the other hand, an explanatory model that is very far from the predictive benchmark would imply that there are substantial practical and theoretical gains to be obtained from further research,(Collopy et al).Avery low level of predictability can spur the development of new measures, collection of data, and new empirical approaches. This model can also set benchmarks for potential levels of predictability of models. For example if newer models with more sophisticated data and/or analysis methods result in only small improvements in predictive power, then it indicates that the benchmark indeed represents the current predictability levels.

8.3. Limitations of the research

In general, one cannot make progress without a dataset for training of adequate size and quality.

It is crucial to have a clear definition of the concept to be predicted, and to have historical examples of the concept.

For the model to be successful, the training data must be representative of the test data. Typically, the training data come from the past, while the test data arise in the future. If the re-hospitalization to be predicted is not stable over time, then predictions are likely not to be useful. Here, changes in the general economy, lifestyle, and in social attitudes towards breast cancer, are all likely to change the behavior of patients in the future. The model therefore needs constant update with time.

The predictive model can lead clinicians to an ever-increased focus on optimizing predictive power at the expense of understanding the broader situation of theory building and richer content of attributes on avoidable re-hospitalization. Clinicians should be aware of the model temptation to shift away their attention from the real problem of concept building.

Clinicians should also be aware that the model doesn't read their minds but work on the “sword of data” and that the model is supportive but they make the actual decision.

8.4. The Conclusion

It is crucial to have a clear definition of the theory that prediction would be based on, and historical examples of the theory. Failure to have clear definition of theory, then you cannot do research design precisely and the study by extension.(that’s why expert are involved to do a clear theory definition)

Since we are using data from surveys, people don’t always provide accurate information. Not every patient will answer truthfully about (say) how many times they exercise — or how many alcoholic beverages they consume — per week hence skewed data resulting to over fitting of the model as witnessed in the unpruned model.

Data collected from different sources can vary in quality and format. Data collected from such diverse sources as surveys, Past records, data-entry forms, will have different attributes and structures. This may provide inconsistencies across merged data hence skewed distribution as also witnessed in the unpruned model.

One cannot make progress without adequate size and quality dataset for training or the model shall over fit the data as witnessed in the unpruned model.

Model Pruning is done iteratively, always choosing the node whose removal most increases the decision tree accuracy over the validation set. Pruning of nodes continues until further pruning is harmful.

This is a supportive model which don’t work independent of a clinician. The goal of the model is to make clinicians less wrong than they were, but not to assume that they will ever make clinicians 100 percent right.

The model has high accuracy in all the measures with slight lower prediction power at the initial stages due to insufficient data and skewed data distribution.

8.5. Future Works

Reasons for difficulties in implementing CDSS into everyday clinical practice come mainly from programmers' insufficient understanding of medical reasoning and decision analyses. Clinicians expect the model to read their minds and deliver exact verdict on the same problem.

However up to date, artificial intelligence and Machine learning don't read minds.They simply give causal relation and underlying patterns. There is need to provide mind reading model which is hereby suggested for further research.Further research is recommended for an autonomous "mind reading" CDSS.

Clinicians can over-increased focus on predictive power of the model at the expense of understanding the broader theory integrated in the model. Recommendation is that Clinicians should be made aware of the possible temptation of focusing more on predictive power that may shift away their attention from the real problem of concept/theory building. Again could it be that clinicians are expecting too much output(miracle) from their less input?.Further research is recommended.

For predictive model to be successful, the training data must be representative of the test data. Typically, the training data come from the past, while the test data arise in the future. If the re-hospitalization to be predicted is not stable over time, then predictions are likely not to be useful. Here, changes in the general economy, lifestyle, social attitudes towards breast cancer, are all likely to change the behavior of patients in the future. It is therefore recommended for a research on a model which will updates itself with the constant changes of lifestyle.

REFERENCE

1. Pamela J. Goodwin, TX Martin Stockler et al: Pain in Patients with Cancer, 2014. 1p.
2. Strategic Management of breast cancer: Alice Musimbi;2013. P. 10-35 .
3. Re-engineering the Hospital Discharge,David Anthony, VK Chetty, et al, 2013.
4. Wata, David Etale: Determinants of breast cancer treatment outcomes at Kenyatta National Hospital, 2012.1p.
5. Dr.C.Nyongessa: Kenya cancer statistic and national initiative report; 2010.p.1-2.
6. Dr.Ian Hampson, at al: Cervical Cancer Treatment Breakthrough using anti-HIV drugs discovered at Manchester University/Kenya National hospital 2013.p.1-3.
7. Natasha Curry et al. Predictive risks project; 2005.p.1-3.
8. Health Dialog Analytic Solutions. Patients at risk of re-hospitalization algorithm; 2006.
9. James Natale,Shengyong Wang.Prediction of Readmission of Heart Failure via decision tree model. University of Akron USA, 2006.
10. William C. R. et al:The Institute of Medicine's (IOM's) publication reports;1999.
11. Mary Ellen Dalton:President of the American Health Quality Association; 2013.
12. Antony, D. et al:Re-engineering the Hospital Discharge, Brown University; 2005.
13. Research Methodology. Kothari; 2004.p.78.
14. Shapiro DE. The interpretation of diagnostic tests. Stat Methods Med Res 1999;8:113–34
15. E. Coiera. The Guide to Health Informatics (2nd Edition). Arnold, London, October 2003.
16. Agency for Healthcare, Research and Quality/AHRQ (http://www.ahrq.gov/ and http://www.guideline.gov)

APPENDICES

APPENDIX 1:

Clinician's Questionnaires

SECTION A

Below are statements, circle the number that best suits your opinion on preventable risks factors that may cause Health deterioration hence avoidable re-hospitalization of a breast cancer patient. In question 17-18 you may give additional information. The numbers meanings have been explained below:

[1]. I completely disagree

[2]. I somewhat disagree

[3]. I cannot say

[4]. I somewhat agree

[5]. I completely agree

1	A committed smoker patient is more likely to be re-hospitalized within 30 days upon discharged	1	2	3	4	5
	Additional opinion:					
2	A Sneaking smoker patient is less likely to be re-hospitalized within 30 days upon discharged compared to a committed smoker	1	2	3	4	5
	Additional opinion:					
3	A patient Abusing Antibiotics is likely to be re-hospitalized	1	2	3	4	5
	Additional opinion:					
4	A patient with high intake of fat is likely to be re-hospitalized	1	2	3	4	5
	Additional opinion:					
5	A patient who have experiencing miscarriage or abortion is likely to be re-hospitalized	1	2	3	4	5
	Additional opinion:					
5	Alcoholic patient is likely to be re-hospitalized	1	2	3	4	5
	Additional opinion:					
6	A patient doing Physical exercises frequently is less likely to be re-hospitalized than a non doing physical exercise patient	1	2	3	4	5
	Additional opinion:					
7	A patient who consumes alcohol is likely to be re-hospitalized than a patient who doesn't.	1	2	3	4	5
	Additional opinion:					
3	A patient who have Low income family may experience re-hospitalization	1	2	3	4	5
	Additional opinion:					
4	Patient who eat soya regularity has less chances of re-hospitalization	1	2	3	4	5
	Additional opinion:					
5	patient who take balance diet has less chances of re-hospitalization	1	2	3	4	5
6	Patient who take coffee daily has less chances of re-hospitalization	1	2	3	4	5
	Additional opinion:					
7	A patient who don't Complete his/her dose is likely to be re-hospitalized	1	2	3	4	5
	Additional opinion:					
8	Isolated patient (psychologically) is likely to be re-hospitalized	1	2	3	4	5
	Additional opinion:					
9	Isolated patient (physically) is likely to be re-hospitalized					
	Additional opinion:					
10	A patient who did always breast feeding is less likely to be re-hospitalized	1	2	3	4	5
	Additional opinion:					
11	A patient who is fond of aborting is likely to be re-hospitalized	1	2	3	4	5
	Additional opinion:					
12	Promiscuous patient is likely to be re-hospitalized	1	2	3	4	5
	Additional opinion:					
13	Overweight especially in the waist defined as BMI (body mass index) over 25 is associated with increased risks of re-hospitalization.	1	2	3	4	5
14	A patient who had hormone replacement therapy is likely to be re-hospitalized than one who had not.	1	2	3	4	5
15	A patient who uses oral contraceptives (birth control pills) before the age of 20 is likely to be re-hospitalized than one who doesn't.	1	2	3	4	5
16	Birth and breast-feeding by the age of 20 may have less re-hospitalization than women who doesn't.	1	2	3	3	5
17						

SECTION B

Below are statements, circle the number that best suits your opinion on the slips, lapses, mistakes, and adverse events in the discharging structure that may cause Health deterioration hence avoidable re-hospitalization of a breast cancer patient. From question 10-16 you may give additional information not presented. The numbers meanings have been explained below:

[1]. I completely disagree
[2]. I somewhat disagree
[3]. I cannot say
[4]. I somewhat agree
[5]. I completely agree

1	Nurses and students are responsible for the entire discharge and new admissions.	1	2	3	4	5
	Additional opinion:					
2	A clear delineation of discharge responsibilities which does not exist.	1	2	3	4	5
	Additional opinion:					
3	lack of communication results in repetition and gaps	1	2	3	4	5
	Additional opinion:					
4	Inadequate patient education on his/her condition	1	2	3	4	5
	Additional opinion:					
5	Medication error may occur	1	2	3	4	5
	Additional opinion:					
5	Early Post-discharge	1	2	3	4	5
	Additional opinion:					
6	Language/Cultural barrier during discharging.	1	2	3	4	5
	Additional opinion:					
7	No follow-up appointment/arrangements	1	2	3	4	5
	Additional opinion:					
3	Absent/inadequate of nurses/case management meeting to discuss discharge status.	1	2	3	4	5
	Additional opinion:					
4	Insufficient preparedness for discharge on patient, family, support team and facilities.	1	2	3	4	5
	Additional opinion:					
5	No nutritionist prescription/Advice which is within patient reach to afford	1	2	3	4	5
	Additional opinion:					
6	No Substance abuse counselor or advice or watchdog	1	2	3	4	5
	Additional opinion:					
9	No Physical therapist advice or someone to ensure that physical exercises is done					
	Additional opinion:					
10		1	2	3	4	5
11		1	2	3	4	5
12		1	2	3	4	5
13		1	2	3	4	5
14		1	2	3	4	5
15		1	2	3	4	5
16		1	2	3	4	5

SECTION C

Below are statements, circle the number that best suits your opinion. The numbers meanings have been explained below: This is in regards to the reduction in overall Health deterioration hence avoidable re-hospitalization

[1]. I completely disagree
[2]. I somewhat disagree
[3]. I cannot say
[4]. I somewhat agree
[5]. I completely agree

1	Educate the patient about her or his diagnosis throughout the hospital stay	1	2	3	4	5
2	Make appointments for clinician follow-up and post discharge testing					
Make appointments with input from the patient regarding the best time and date for the appointment.		1	2	3	4	5
Coordinate appointments with physicians, testing, and other services.		1	2	3	4	5
Discuss reason for and importance of physician appointments.		1	2	3	4	5
Confirm that the patient knows where to go and has a plan about how to get to the appointment; Review transportation options and other barriers to keeping these appointments.		1	2	3	4	5
3	Discuss with the patient any tests or studies that have been completed in the hospital and who will be responsible for following up on the results	1	2	3	4	5
4	Organize post discharge services					
Be sure the patient understands the importance of such services.		1	2	3	4	5
Make appointments that the patient can keep.		1	2	3	4	5
Discuss the details of how to receive each service.		1	2	3	4	5
5	Confirm the medication plan.					
Reconcile the discharge medication regimen with that followed before the hospitalization.		1	2	3	4	5
Explain what medications to take, emphasizing any changes in the regimen.		1	2	3	4	5
Review each medication's purpose, how to take each medication correctly, and important adverse effects to watch out for.		1	2	3	4	5
Be sure the patient has a realistic plan for how to get the medications.		1	2	3	4	5
6	Reconcile the discharge plan with national guidelines and critical pathways.	1	2	3	4	5
7	Review the appropriate steps for what to do if a problem arises.					
Inform the patient about a specific plan for how to contact the primary care provider (or coverage) and provide contact numbers for evenings and weekends.		1	2	3	4	5
Inform the patient about what constitutes an emergency and what to do in cases of emergency.		1	2	3	4	5
8	Expedite transmission of the discharge résumé (summary) to the physicians (and other services, such as the visiting nurses, primary care physician), accepting responsibility for the patient's care after discharge.	1	2	3	4	5
9	Assess the patient's degree of understanding by asking for an explanation of the details of the plan in her or his own words.	1	2	3	4	5
10	In regards to question 9 above, contacting family members who will share in the care giving responsibilities.	1	2	3	4	5
11	Give the patient a written discharge plan at the time of discharge that contains					
The discharge medications, including what medications to take, how to take them, and how to obtain them.		1	2	3	4	5
Instructions on what to do if the condition changes.		1	2	3	4	5
Coordination and planning for follow-up appointments that the patient can keep.		1	2	3	4	5

SECTION D

Below are statements, circle the number that best suits your opinion. The numbers meanings have been explained below: Answer in regards to the reduction of Health deterioration hence avoidable re-hospitalization

[1]. I completely disagree

[2]. I somewhat disagree

[3]. I cannot say

[4]. I somewhat agree

[5]. I completely agree

1	At the time of discharge, you should schedule an appointment with a primary care provider at a time convenient to the patient.	1	2	3	4	5
2	During the patient discharge, nurses, trainee students can do the discharging.	1	2	3	4	5
3	You can discharge and do new admission sometimes at the same time.	1	2	3	4	5
4	You frequently discharge and do new admission at the same time.	1	2	3	4	5
5	A clear delineation of discharge responsibilities often does not exist.	1	2	3	4	5
7	Sometimes there exist lapses of communication in the discharging process.	1	2	3	4	5
8	Lapses of communication are more frequent in the explanation of the discharging summary to the primary care physician.	1	2	3	4	5
9	Lapses of communication is more frequent in the physician team to the primary care physician.	1	2	3	4	5
10	Inadequate patient education on his/her condition is frequent .	1	2	3	4	5
11	There is lack of timely follow up.	1	2	3	4	5
12	There is some medication error.	1	2	3	4	5
13	There is early some post discharge.	1	2	3	4	5
14	Patient who abuse drugs are likely to be re-hospitalized within 30 days upon discharge.	1	2	3	4	5
15	Patient who uses alcohol are likely to be re-hospitalized within 30 days upon discharge.	1	2	3	4	5
16	Non medication adherence patients are likely to be re-hospitalized within 30 days upon discharge.	1	2	3	4	5
17	Patient who don't keep timely follow up appointments are likely to be re-hospitalized within 30 days upon discharge.	1	2	3	4	5
18	Sometimes we experience lab/test error.	1	2	3	4	5
19	Sometimes we may forget to issue discharge summary.	1	2	3	4	5
20	Sometimes we may experience inappropriate discharge.	1	2	3	4	5
21	Sometimes we may prescribe inappropriate medication.	1	2	3	4	5

Appendix 2:

Patient's interview questionnaires

SECTION A

Below are statements, circle the number that best suits your opinion. The numbers meanings have been explained below: Questions are administered to a re-hospitalized patient

[1]. I completely disagree
[2]. I somewhat disagree
[3]. I cannot say
[4]. I somewhat agree
[5]. I completely agree

1	I was taught about my diagnosis during my hospital stay.	1	2	3	4	5
2	I have received a written discharge plan that is easy to read and understand.	1	2	3	4	5
3	I have follow-up appointments with my physicians.	1	2	3	4	5
4	I have received a written discharge plan that has the information I need to take care of myself at home.	1	2	3	4	5
5	I have been told about test results or studies that have not been completed before I go home.	1	2	3	4	5
6	I have a written list of my discharge medications and know which medications are new or changed.	1	2	3	4	5
7	If I need home health care, medical equipment, or other help or services after I go home, has been arranged.	1	2	3	4	5
8	When the nurses were teaching me, they asked me to explain what I had learned in my own words.	1	2	3	4	5
9	I understand what to do and who to call if a problem arises after I am home	1	2	3	4	5

SECTION B

Below are statements, circle the number that best suits your opinion. The numbers meanings have been explained below: Questions are administered to a re-hospitalized patient

[1]. I completely disagree

[2]. I somewhat disagree

[3]. I cannot say

[4]. I somewhat agree

[5]. I completely agree

1	How acceptable is it for a man (woman) to engage in sexual infidelity	1	2	3	4	5
2	How acceptable is it for a man (woman) to engage in emotional infidelity?	1	2	3	4	5
3	Have you ever had a partner commit infidelity and if so the type of infidelity the partner engaged in.	sexual	emotional	sexual and emotional		
4	Which were you more upset over?	sexual	emotional	sexual and emotional		
5	Have you even engaged in infidelity? if so which type of infidelity .	sexual	emotional	sexual and emotional		
6	How often do you engage in fidelity	1	2	3	4	5
7	What were the reasons why you engage in infidelity?	sexual dissatisfaction		emotional dissatisfaction		both
	Is there someone who smokes in your family?	1	2	3	4	5
8	How many people smoke in your family?	1-3	4-5	5-8	9	Above 10
9	Do you smoke?	1	2	3	4	5
10	Do you have someone who takes alcohol in your family?	1	2	3	4	5
11	Do you take alcohol	1	2	3	4	5
12	Do you take care of each other in your family (psychological and physically)	1	2	3	4	5
13	Do you feel taken care of in your family and given support whenever you need it(psychologically)	1	2	3	4	5
14	Do you feel taken care of in your family and given support whenever you need it(physically)	1	2	3	4	5
15	Do you do physical exercise	1	2	3	4	5
16	Do you have someone in your family who had miscarriage?	1	2	3	4	5
17	Have you ever miscarriage?	1	2	3	4	5
18	Do you have someone in your family who have done an abortion due to medical reasons or with no reasons	1	2	3	4	5
19	Have you ever done an abortion due to medical reasons or with no reasons	1	2	3	4	5
20	I enjoy eating soya	1	2	3	4	5

Appendix 3: Python Code class Mapping a Decision Model

A python class method with the if else ladder.

```
Decision Tree implementation Prototype
'''
import json
NODE_NAMES = ['Timely Medical Dose', 'Fat Intake', 'Abuse
Antibiotics',
            'Use Oral Contraceptive', 'Stress Burden', 'Physical Exercise',
            'Appropriate Discharge', 'Overweight',
            'No Re-Hospitalization', 'Re-Hospitalization Occurs']
class Node(object):
      def __init__(self, node_value):
      '''Initialize the node
      '''
      self.key = node_value
      self.positive_node = None # When the response is positive
      self.negative_node = None # When the response is negative
   def __repr__(self):
      return self.key
   def get_node_value(self):
      '''
      Sets the node' name
      '''
      return self.key
   def set_node_value(self, value):
      '''
      Returns the node's value
      '''
      self.key = value

   def set_positive(self, value):
      '''
      Sets the node returned when given a positive value
      '''
```

```
        if not self.positive_node:
            self.positive_node = Node(value)
        else:
            node = Node(value)
            current = self.positive_node
            node.positive_node = current
            self.positive_node = node

    def set_negative(self, value):
        '''Sets the node returned when given a positive value
        '''
        if not self.negative_node:
            self.negative_node = Node(value)
        else:
            node = Node(value)
            current = self.negative_node
            node.negative_node = current
            self.negative_node = node
    def get_positive_node(self):
        return self.positive_node
    def get_negative_node(self):
        return self.negative_node
    def is_tree(self):
        if self.negative_node and self.positive_node:
            return True
        return False
def addtree(data, tree=None):
    '''  Build the tree recursively
    '''if not tree:
        return
    positive_value = data[tree.key_dict][1]
    if positive_value:
        tree.set_positive(data[positive_value][0])
        tree.get_positive_node().key_dict = positive_value
```

```
        addtree(data, tree.get_positive_node())
        negative_value = data[tree.key_dict][2]
        if negative_value:
            tree.set_negative(data[negative_value][0])
            tree.get_negative_node().key_dict = negative_value
        addtree(data, tree.get_negative_node())
        return tree
    def buildtree(path, root_value="Timely Medical Dose"):
        f = open(path)
        data = json.load(f)
        root = Node(root_value)
        root.key_dict = "Timely Medical Dose"
        return addtree(data, tree=root)
    def evaluate(data, node):
        if not node.is_tree():
            return node
        value = data[node.get_node_value()]
        if value:
            leave = evaluate(data, node.get_positive_node())
        else:leave = evaluate(data, node.get_negative_node())
return leave
```