**DRUG TREATMENT AND LENGTH OF STAY**

**AT HOSPITAL FOR DIABETIC PATIENTS**

***A project report submitted at the end of the 8th semester in the partial***

***fulfillment of the Award of the Degree of***

**BACHELOR OF TECNOLOGY IN**

**COMPUTER SCIENCE & SOFTWARE ENGINEERING**

***Submitted By***

**ANURADHA SAHITHI PADAVALA (Regd. No: 316506402002)**

**VELUGULA BINDUSREE (Regd. No: 316506402006)**

**BHAVISETTY DEVA HARSHINI (Regd. No: 316506402010)**

***Under the Guidance of***

**Dr. A. MARY SOWJANYA**

**Assistant Professor, Department of CS & SE**

**Andhra University**

****

**DEPARTMENT OF COMPUTER SCIENCE & SYSTEMS ENGINEERING**

**ANDHRA UNIVERSITY COLLEGE OF ENGINEERING (A)**

**ANDHRA UNIVERSITY**

**VISAKHAPATNAM - 530003**

**2016-2020**

**DEPARTMENT OF COMPUTER SCIENCE & SYSTEMS ENGINEERING**

**ANDHRA UNIVERSITY COLLEGE OF ENGINEERING (A)**

**ANDHRA UNIVERSITY**

**VISAKHAPATNAM - 530003**

****

**CERTIFICATE**

This is to certify that it is a bonafidework done by **Miss. ANURADHA SAHITHI PADAVALA** bearing Regd. No: 316506402002 , **Miss. VELUGULA BINDUSREE** bearing Regd. No: 316506402006 and **Miss. BHAVISETTY DEVA HARSHINI** bearing Regd. No: 316506402010 students of **BACHELOR OF TECHNOLOGY COMPUTER SCIENCE & SOFTWARE ENGINEERING** from the **Department of Computer Science & Systems Engineering (A), Andhra University College of Engineering (A)** , has carried out the bonafidework entitled **“DRUG TREATMENT AND LENGTH OF STAY AT HOSPITAL FOR DIABETIC PATIENTS”**at Andhra University College of Engineering(A), Visakhapatnam .

Project Guide Head of the Department

**Dr. A. Mary Sowjanya Prof.Kuda Nageswara Rao**

Dept. of CS&SE, Dept. of CS&SE,

AUCE(A), AUCE(A),

Andhra University, Andhra University,

Visakhapatnam. Visakhapatnam.

# 

**DECLARATION**

I hereby solemnly declare that this project report entitled “ **DRUG TREATMENT AND LENGTH OF STAY AT HOSPITAL FOR DIABETIC PATIENTS**” is bonafidework done by **ANURADHA SAHITHI PADAVALA , VELUGULA BINDUSREE and BHAVISETTY DEVA HARSHINI** of **BACHELOR OF TECHNOLOGY COMPUTER SCIENCE & SOFTWARE ENGINEERING** submitted to the **Department of Computer Science & Systems Engineering**, **at Andhra University College of Engineering (A)** , Visakhapatnam , submitted in partial fulfillment of the requirements for the award of degree of **BACHELOR OF TECHNOLOGY COMPUTER SCIENCE & SOFTWARE ENGINEERING** during 2020 in **ANDHRA UNIVERSITY COLLEGE OF ENGINEERING (A)** . I assure that this project was not submitted in any other institution for the award of any other degree .

Place : Visakhapatnam

Date :

Signature

**ANURADHA SAHITHI PADAVALA**

**(Regd. No. 316506402002)**

**VELUGULA BINDUSREE**

**(Regd. No. 316506402006)**

**BHAVISETTY DEVA HARSHINI**

**(Regd. No. 316506402010)**

**ACKNOWLEDGEMENT**

I would like to express my deep sense of gratitude to my esteemed college "**Andhra University College of Engineering"** for providing all the possible resources required for successfully implementation and execution of the project work and for providing me an opportunity to fulfill my cherished desire.

I would like to express my thankfulness and deep sense of gratitude to my project guide **Dr. A. Mary Sowjanya**, Department of Computer Science& System Engineering, Andhra University College of Engineering(A), Visakhapatnam for her guidance and support throughout the project and leading me in completing the project effectively by giving me valuable suggestion.

It gives me boundless pleasure to avail myself this opportunity to express my deep sense of gratitude and whole hearted thanks to **Prof. Kuda Nageswara** **Rao,** Head of the Department of **Computer Science and Systems Engineering** for his continuous support completion of the project.

Finally, I thank all the members of the teaching and non-teaching staff of the Department of Computer Science and System Engineering for their support in the completion of the project.

**ANURADHA SAHITHI PADAVALA**

**(Regd. No. 316506402002)**

**VELUGULA BINDUSREE**

**(Regd. No. 316506402006)**

**BHAVISETTY DEVA HARSHINI**

**(Regd. No. 316506402010)**

# ABSTRACT

Over the past few years, there has been increased interest in data mining and machine learning methods to improve hospital performance and survival analysis is one of the methods which could be of greater help.

Survival analysis is a branch of statistics for analyzing the expected duration of time until one or more events happen, such as death in biological organisms and failure in mechanical systems. So , it is useful to predict few real life events like status of business in future , the duration of cure of a patient , education or career development , crowd funding .

In this project, the objective is to predict the length of stay using various diagnosis for diabetes using insulin, drugs and combination of drugs. As diabetes is a very common disease in all age groups; it can lead to heart disease as well as increasing the risks of developing few other disorders . So, diabetes is one of the most serious health challenges, even in developed countries.

Here a particular patient undergoes a treatment and the result of the treatment decides how much time it takes for a patient to be cured and in the process , a combination of drug or insulin is given to the patient and check the chances for readmission.This prediction will be done using statistical methods like survival analysis using R as a platform. This project uses Kapler meire method for analysis.

**TABLE OF CONTENTS**

**Page No**

1. **Introduction** 1
   1. Data Mining 1
   2. Machine Learning 1
   3. Data Analysis 2
   4. Survival Analysis 3
   5. Overview of diabetes 3
   6. Types of diabetes 4
   7. Types of classifiers 4
   8. Overview of R 5
      1. Analysis with R 5
      2. R Software environment 5
      3. R studio 5
      4. R packages used 5
   9. Existing System 6
   10. Proposed System 6
2. **Related Work** 7
3. **Methodology** 9
   1. Aim 9
   2. Problem Statement 9
   3. Proposed System 9
4. **Requirement Analysis** 10
   1. Functional Requirements 10
   2. Non Functional Requirements 10
   3. Hardware Requirements 10
   4. Software Requirements 10
5. **Design** 11
   1. UML Diagrams 11
   2. Types of UML diagrams 11
      1. Use Case Diagram 11
      2. Activity Diagram 12
      3. Sequence Diagram 13
      4. Deployment Diagram 14
6. **Implementation** 15
   1. Data Set Description 15
   2. Kapler Algorithm 17
   3. Classifier Performance 18
   4. Flowchart 19
7. **Sample Code** 20
8. **Software Testing** 61
   1. Types of testing 61
   2. Test cases 62
9. **Results** 64
10. **Conclusion & Future Work** 68
11. **References** 69

**LIST OF FIGURES**

**Page No**

* Figure 5.2.1 : Use Case Diagram 11
* Figure 5.2.2 : Activity Diagram 12
* Figure 5.2.3 : Sequence Diagram 13
* Figure 5.2.4 : Deployment Diagram 14
* Figure 9.1 : Acarbose 64
* Figure 9.2 : Acarbose + Insulin 64
* Figure 9.3 : Acarbose + Pioglitazone 65
* Figure 9.4 : Acarbose + Glyrubide 65
* Figure 9.5 : Acarbose + Metaformin 66
* Figure 9.6 : Acarbose + Rosiglitazone 66
* Figure 9.7 : Cox proportional fit for Acarbose 67
* Figure 9.8 : Survival analysis with classification for Acarbose 67

**LIST OF TABLES**

**Page No**

* Table 6.1 : Data Set Description 15
* Table 8.2.1 : Test Case 1 62
* Table 8.2.2 : Test Case 2 62
* Table 8.2.3 : Test Case 3 63
* Table 8.2.4 : Test Case 4 63

1. **INTRODUCTION :**
   1. **DATA MINING :**

Data mining is an [interdisciplinary](https://en.wikipedia.org/wiki/Interdisciplinary" \o "Interdisciplinary) sub field of [computer science](https://en.wikipedia.org/wiki/Computer_science" \o "Computer science) and [statistics](https://en.wikipedia.org/wiki/Statistics" \o "Statistics) with an overall goal to extract information (with intelligent methods) **Data mining** is the process of discovering patterns in large [data sets](https://en.wikipedia.org/wiki/Data_set" \o "Data set) involving methods at the intersection of [machine learning](https://en.wikipedia.org/wiki/Machine_learning" \o "Machine learning), [statistics](https://en.wikipedia.org/wiki/Statistics" \o "Statistics), and [database systems](https://en.wikipedia.org/wiki/Database_system" \o "Database system). from a data set and transform the information into a comprehensible structure for further use. Data mining is the analysis step of the "knowledge discovery in databases" process or KDD. Aside from the raw analysis step, it also involves database and [datamanagement](https://en.wikipedia.org/wiki/Data_management" \o "Data management) aspects, [datapreprocessing](https://en.wikipedia.org/wiki/Data_pre-processing" \o "Data pre-processing), [model](https://en.wikipedia.org/wiki/Statistical_model" \o "Statistical model) and [inference](https://en.wikipedia.org/wiki/Statistical_inference" \o "Statistical inference) considerat-ions, metrics, [complexity](https://en.wikipedia.org/wiki/Computational_complexity_theory" \o "Computational complexity theory) considerations, post-processing of discovered structures, [visualization](https://en.wikipedia.org/wiki/Data_visualization" \o "Data visualization), and [online updating](https://en.wikipedia.org/wiki/Online_algorithm" \o "Online algorithm).

* 1. **MACHINE LEARNING :**

**Machine learning** (**ML**) is the study of computer algorithms that improve automatically through experience.It is seen as a subset of [artificial intelligence](https://en.wikipedia.org/wiki/Artificial_intelligence" \o "Artificial intelligence). Machine learning algorithms build a [mathematical model](https://en.wikipedia.org/wiki/Mathematical_model" \o "Mathematical model) based on sample data, known as "[training data](https://en.wikipedia.org/wiki/Training_data" \o "Training data)", in order to make predictions or decisions without being explicitly programmed to do so. Machine learning algorithms are used in a wide variety of applications, such as [email filtering](https://en.wikipedia.org/wiki/Email_filtering" \o "Email filtering) and [computer vision](https://en.wikipedia.org/wiki/Computer_vision" \o "Computer vision), where it is difficult or in feasible to develop conventional algorithms to perform the needed tasks.

Machine learning is closely related to [computational statistics](https://en.wikipedia.org/wiki/Computational_statistics" \o "Computational statistics), which focuses on making predictions using computers. The study of [mathematical optimization](https://en.wikipedia.org/wiki/Mathematical_optimization" \o "Mathematical optimization) delivers methods, theory and application domains to the field of machine learning. [Data mining](https://en.wikipedia.org/wiki/Data_mining" \o "Data mining) is a related field of study, focusing on [exploratory data analysis](https://en.wikipedia.org/wiki/Exploratory_data_analysis" \o "Exploratory data analysis) through [unsupervised learning](https://en.wikipedia.org/wiki/Unsupervised_learning" \o "Unsupervised learning). In its application across business problems, machine learning is also referred to as [predictive analytics](https://en.wikipedia.org/wiki/Predictive_analytics" \o "Predictive analytics).

* 1. **DATA ANALYSIS :**

**Data analysis** is inspecting, [cleansing](https://en.wikipedia.org/wiki/Data_cleansing" \o "Data cleansing), [transforming](https://en.wikipedia.org/wiki/Data_transformation" \o "Data transformation) and [modeling](https://en.wikipedia.org/wiki/Data_modeling" \o "Data modeling) [data](https://en.wikipedia.org/wiki/Data" \o "Data) with the goal of discovering useful information, informing conclusion and supporting decision-making. Data analysis has multiple facets and approaches, encompassing diverse techniques under a variety of names, and is used in different business, science, and social science domains. In today's business world, data analysis plays a role in making decisions more scientific and helping businesses operate more effectively.

**Classification:** It is a Data analysis task, i.e. the process of finding a model that describes and distinguishes data classes and concepts. Classification is the problem of identifying to which of a set of categories (sub populations), a new observation belongs to, on the basis of a training set of data containing observations and whose categories membership is known.

**Regression:** Regression is a [data mining](https://www.lifewire.com/what-is-data-mining-4784169) technique used to predict a range of numeric values (also called continuous values), given a particular dataset. For example, regression might be used to predict the cost of a product or service, given other variables.Regression is used across multiple industries for business and marketing planning, financial forecasting, environmental modeling and analysis of trends.

**Prediction:** Prediction in data mining is to identify data points purely on the description of another related data value. It is not necessarily related to future events but the used variables are unknown. Prediction derives the relationship between a thing you know and a thing you need to predict for future reference.

* 1. **SURVIVAL ANALYSIS :**

Survival analysis deals with predicting the time when a specific event is going to occur. It is also known as failure time analysis or analysis of time to death. For example predicting the number of days a person with cancer will survive or predicting the time when a mechanical system is going to fail.

The R package named **survival** is used to carry out survival analysis. This package contains the function **Surv()** which takes the input data as a R formula and creates a survival object among the chosen variables for analysis. Then we use the function **survfit()** to create a plot for the analysis.

* 1. **OVERVIEW OF DIABETES :**

Diabetes is a disease that occurs when your blood glucose, also called blood sugar, is too high. Blood glucose is your main source of energy and comes from the food you eat. [Insulin](https://www.niddk.nih.gov/Dictionary/I/insulin), a [hormone](https://www.niddk.nih.gov/Dictionary/H/hormone) made by the [pancreas](https://www.niddk.nih.gov/Dictionary/P/pancreas), helps glucose from food get into your cells to be used for energy. Sometimes your body doesn’t make enough—or any—insulin or doesn’t use insulin well. Glucose then stays in your blood and doesn’t reach your cells.

Over time, having too much glucose in your blood can cause [health problems](https://www.niddk.nih.gov/health-information/diabetes/overview/preventing-problems). Although diabetes has no cure, you can take steps to [manage your diabetes](https://www.niddk.nih.gov/health-information/diabetes/overview/managing-diabetes) and stay healthy. Sometimes people call diabetes “a touch of sugar” or “borderline diabetes.” These terms suggest that someone doesn’t really have diabetes or has a less serious case, but every case of diabetes is serious.

* 1. **TYPES OF DIABETES :**

### Type 1 diabetes

If you have [type 1 diabetes](https://www.niddk.nih.gov/health-information/diabetes/overview/what-is-diabetes/type-1-diabetes), your body does not make insulin. Your [immune system](https://www.niddk.nih.gov/Dictionary/I/immune-system) attacks and destroys the cells in your pancreas that make insulin. Type 1 diabetes is usually diagnosed in children and young adults, although it can appear at any age. People with type 1 diabetes need to take insulin every day to stay alive.

### Type 2 diabetes

If you have [type 2 diabetes](https://www.niddk.nih.gov/health-information/diabetes/overview/what-is-diabetes/type-2-diabetes), your body does not make or use insulin well. You can develop type 2 diabetes at any age, even during childhood. However, this type of diabetes occurs most often in middle-aged and older people. Type 2 is the most common type of diabetes.

**1.7 TYPES OF CLASSIFIERS :**

**Random forest :** It is a classification algorithm consisting of many decisions trees. It uses bagging and feature randomness when building each individual tree to try to create an uncorrelated forest of trees whose prediction by committee is more accurate than that of any individual tree.

**Naive Bayes**: A collection of classification algorithms based on Bayes' Theorem. It is not a single algorithm but a family of algorithms where all of them share a common principle, i.e. every pair of features being classified is independent of each other.

**Decision Tree** : It is a tree in which internal nodes are labeled by features. The classifier categorizes an object xi by recursively testing for the weights that the features labeling the internal nodes have in vector xi, until a leaf node is reached. The label of this node is then assigned to xi.

**Logistic Regression :** It is basically a supervised classification algorithm. In a classification problem, the target variable(or output), y, can take only discrete values for given set of features(or inputs), X. Contrary to popular belief, logistic regression is a regression model.

**K-Nearest Neighbor :**  It is a simple algorithm that stores all available cases and classifies new cases based on a similarity measure (e.g., distance functions). KNN has been used in statistical estimation and pattern recognition already in the beginning of 1970's as a non-parametric technique.

**1.8 OVERVIEW OF R :**

**1.8.1 ANALYSIS WITH R :**

Step 1 – First approach to data

Step 2 – Analyzing categorical variables

Step 3 – Analyzing numerical variables

Step 4 – Analyzing numerical and categorical at the same time

**1.8.2 R SOFTWARE ENVIRONMENT :**

Itis a free software environment for statistical computing and graphics. It compiles and runs on a wide variety of UNIX platforms, Windows and Mac OS. R is an integrated suite of software facilities for data manipulation, calculation and graphical display.

**1.8.3 R STUDIO :**

It is an integrated development environment (IDE) for R. It includes a console, syntax-highlighting editor that supports direct code execution, as well as tools for plotting, history, debugging and workspace management.

**1.8.4 R PACKAGES USED :**

**Survival** : The survival package is the cornerstone of the entire R survival analysis edifice. Not only is the package itself rich in features, but the object created by the Surv() function, which contains failure time and censoring information, is the basic survival analysis data structure in R. Dr. Terry Therneau.

**Survminer** :The R package survival fits and plots survival curves using R base graphs. There are also several R packages/functions for drawing survival curves using ggplot2.

**Ranger** : The ranger might be the surprise in my very short list of survival packages. The ranger() function is well-known for being a fast implementation of the Random Forests algorithm for building ensembles of classification and regression trees. But ranger() also works with survival data.

**Ggplot :**  It is a data visualization package for the statistical programming language R. Created by Hadley Wickham in 2005, ggplot is an implementation of Leland Wilkinson's Grammar of Graphics—a general scheme for data visualization which breaks up graphs into semantic components such as scales and layers.

# 1.9 EXISTING SYSTEM:

Big data is the traditional method used in predicting the survival analysis of a diabetic patient. Big data refers to massive volumes of both structured and unstructured datasets with complex structures that are difficult to capture, store, format, extract, integrate, analyze, and visualize using traditional methods and tools.. Hence, the R tool is used efficiently in our work with big data.

# 1.10 PROPOSED SYSTEM:

In this project prediction of diabetes will be done using R

as its platform. Method used in the project is KeplerMeier method.Basic treat-ment will be given using drugs and kept under study for 30 days , and then decision is made whether readmission is required or not.

1. **RELATED WORK :**

* Matteo Gagliolo and Catherine Legrand [1] For algorithm selection,here adopted an online approach, in which models of the run time distributions of the available algorithms are repeatedly updated and used to guide the allocation of computational resources, while solving a sequence of problem instances. The models are estimated using survival analysis techniques,which allow to reduce computation time, censoring the run times of the slower algorithms.
* Data flair team[2] The statistical tasks of predictions have always been around which allow you to know about the future based on the patterns of the past history. One of such techniques that allow you to measure the duration of time till the occurrence of a future event is Survival Analysis using R. It is one of the most interesting ways of evaluating future occurrences.
* R bloggers[3]To choose the correct method . It is not easy to apply the concepts of survival analysis right off the bat. One needs to understand the ways it can be used first. This includes Kaplan-Meier Curves, creating the survival function through tools such as survival trees or survival forests and log-rank test.
* Github [4] Despite major advances in science and technology, diabetes continues to be a chronic disease, with a thirty-day readmission rate of around 20%, as compared to an average of 12% for the rest of the diseases. Additionally, readmissions cost hospitals a fair amount of money, so the end goal is to identify and reduce the possibility of a readmission. Prevention of patient readmission has been given a greater importance due to large cost involvement.

#### Reinaldo Zezela, MSc student Big Data Analytics, University of Derby [5] (for data set)The data set used for the purpose of this study is Pima Indians Diabetes Database of National Institute of Diabetes and Digestive and Kidney Diseases. This diabetes database, donated by Vincent Sigillito, is a collection of medical diagnostic reports of 768 examples from a population living near Phoenix, Arizona, USA.

* IEEE [6] Performance and accuracy of the applied algorithms is discussed and compared. Comparison of the different machine learning techniques used in this study reveals which algorithm is best suited for prediction of diabetes. This paper aims to help doctors and practitioners in early prediction of diabetes using machine learning techniques.
* Westministers research [7] Understanding the pattern of length of stay in institutional long-term care has important practical implications in the management of long-term care. Furthermore, residents’ attributes are believed to have signiﬁcant effects on these patterns. In this paper, we present a model-based approach to extract, from a routinely gathered administrative social care dataset, high-level length-of-stay patterns of residents in long-term care. This approach extends previous work by the authors to incorporate residents’ features. Two applications using data provided by a local authority in England are presented to demonstrate the potential use of this approach.

# 3.METHODOLOGY :

**3.1 AIM :**

The objectives of survival analysis include the analysis of patterns of event times.Here, the comparison of distributions of survival times in different groups of individuals and examining whether and by how much some factors affect the risk of an treatment.

**3.2 PROBLEM STATEMENT :**

Is readmission required?

Hospital re-admissions are a great expense for health care. Hospital readmission is defined as a patient who was discharged from a hospital and readmitted within a set time frame (i.e., 30 days).

**3.3 PROPOSED SYSTEM :**

**TREATMENT-1 :** If that value is greater than 0.5 then it indicates that patient has been cured.if not readmission will be required. After readmission patient will be treated with insulin along with drug and the mean survivability capability will be calculated.

**TREATMENT-2 :**If the value is greater than 0.5 further treatment is not required, else readmission required.

**TREATMENT-3:**Now patient will be treated with the combination of drugs and the mean survivability will be calculated.if the value is greater than 0.5 , the patient is said to be cured. If not it indicates that the treatment does not have any effect.

1. **REQUIREMENTS ANALYSIS :**

**4.1 FUNCTIONAL REQUIREMENTS :**

A **functional requirement** defines a function of a [system](https://en.wikipedia.org/wiki/System" \o "System) or its component, where a function is described as a specification of behavior between outputs and inputs. Functional requirements may involve calculations, technical details, data manipulation and processing, and other specific functionality that define what a system is supposed to accomplish .Behavioral requirements describe all the cases where the system uses the functional requirements, these are captured in [use cases](https://en.wikipedia.org/wiki/Use_case" \o "Use case).

**4.2 NON FUNCTIONAL REQUIREMENTS :**

A **non-functional requirement** (NFR) is a [requirement](https://en.wikipedia.org/wiki/Requirement" \o "Requirement) that specifies criteria that can be used to judge the operation of a system, rather than specific behaviors. They are contrasted with [functional requirements](https://en.wikipedia.org/wiki/Functional_requirement" \o "Functional requirement) that define specific behavior or functions.The plan for implementing non functional requirements is detailed in the system architecture, because they are usually architecturally significant requirements.

**4.3 HARDWARE REQUIREMENTS :**

SYSTEM : Intel Core

HARD DISK : 1TB

MONITOR : 15”LED

INPUT DEVICES : Keyboard,Mouse

RAM : 8GB

**4.4 SOFTWARE REQUIREMENTS :**

OPERATING SYSTEM : Windows 10

CODING LANGAGUE : R

TOOL : R Studio (version 1.1.463)

1. **DESIGN :**
   1. **UML DIAGRAM :**

A UML diagram is a diagram based on the UML (**Unified Modeling Language**) with the purpose of visually representing a system along with its main actors, roles, actions, artifacts or classes, in order to better understand, alter, maintain, or document information about the system.

* 1. **TYPES OF UML DIAGRAMS:**

**5.2.1 USE CASE DIAGRAM :**

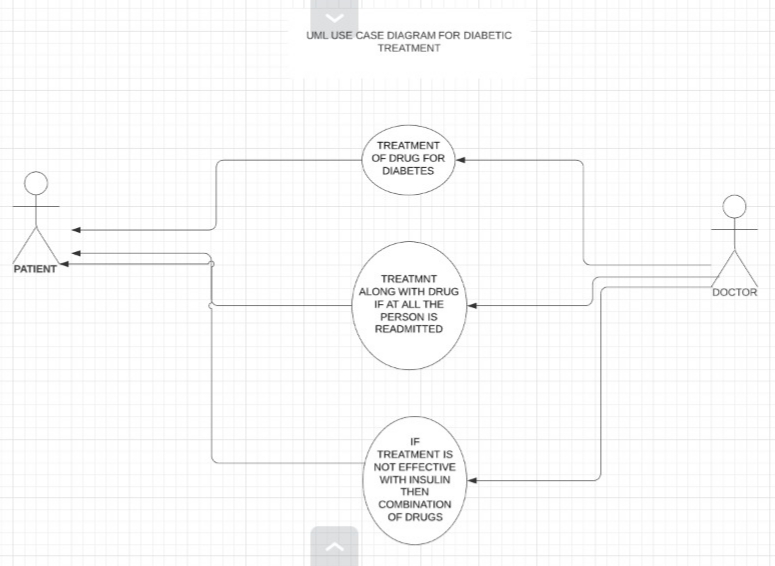
****

Figure 5.2.1 Use Case Diagram

**5.2.2 ACTIVITY DIAGRAM :**

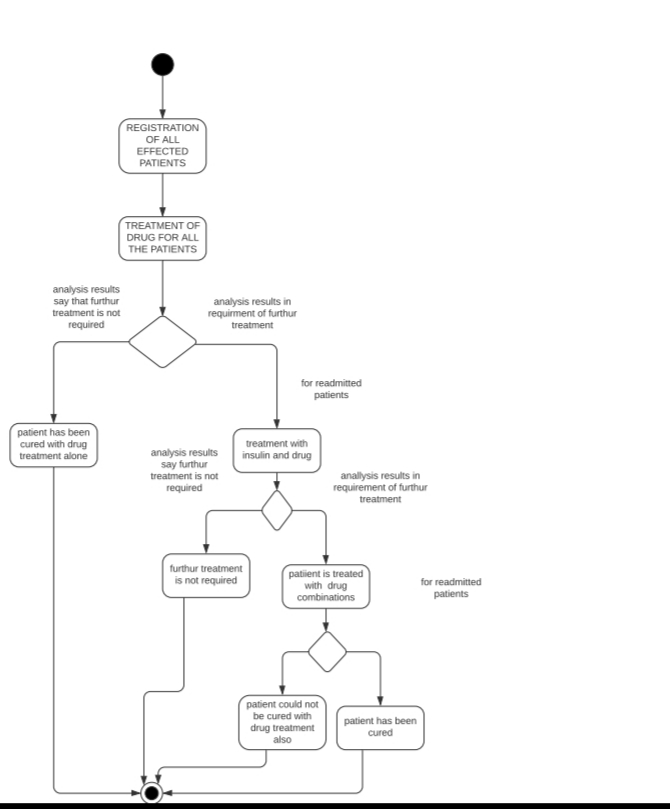
****

Figure 5.2.2 Activity Diagram

**5.2.3 SEQUENCE DIAGRAM :**

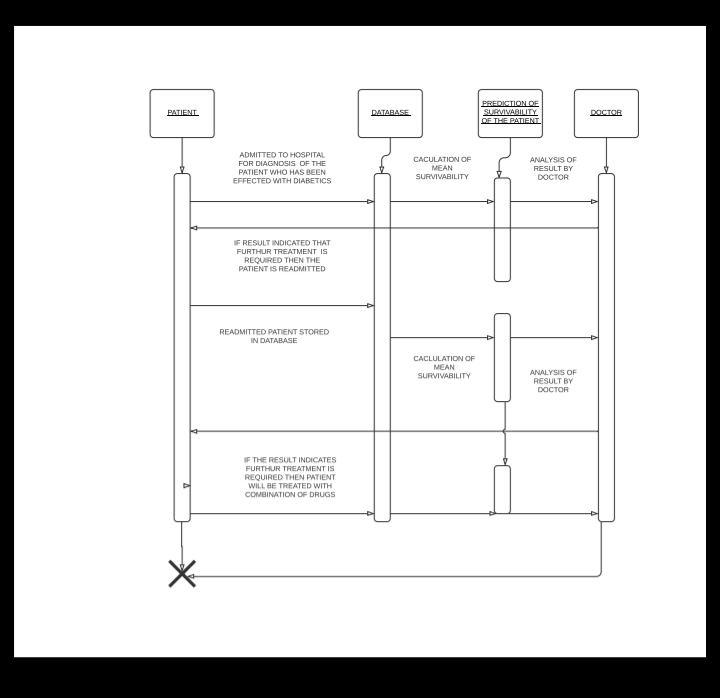
****

Figure 5.2.3 Sequence Diagram

**5.2.4 DEPLOYMENT DIAGRAM:**

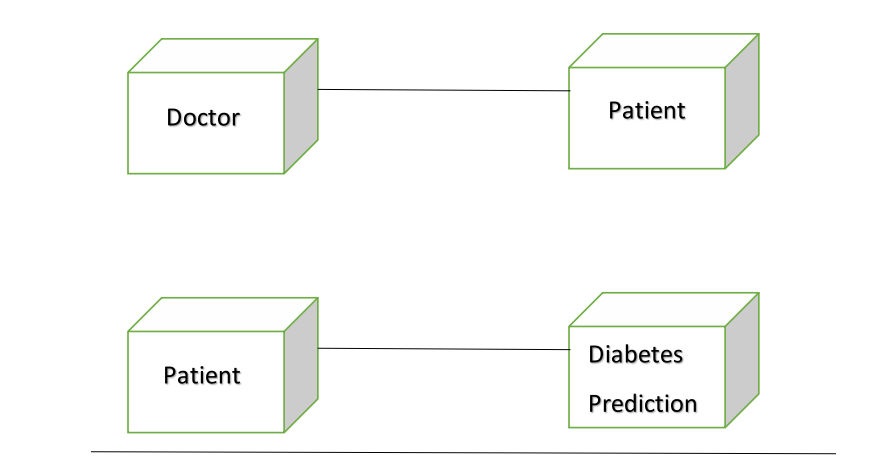


Figure 5.2.4 Deployment Diagram

1. **IMPLEMENTATION :**

**6.1 DATA DESCRIPTION :**

|  |  |  |
| --- | --- | --- |
| **FEATURE NAME** | **TYPE** | **DESCRIPTION AND VALUES** |
| Encounter id | Numeric | Unique identifier of an encounter |
| Patient number | Numeric | Unique identifier of a patient |
| Race | Nominal | Values:Caucasian,Asian,African,American,Hispanic and other |
| Gender | Nominal | Values:male,female and unknown or invalid |
| Age | Nominal | Grouped in 10-year intervals: [0,10), [10,20), ... , [90,100) |
| Weight | Numeric | Weight in pounds |
| Admission Type | Nominal | Integeridentifiercorrespondingto9distinctvalues,forexample ,emergency,urgent, elective, newborn, and not available |
| Dischargedisposition | Nominal | Integeridentifiercorrespondingto29distinctvalues,for example,dischargedto home,expired,andnotavailable |
| Admission source | Nominal | Integeridentifiercorrespondingto21distinctvalues,for example,physicianreferral, emergencyroom,andtransferfromahospital |
| Time in hospital | Numeric | Integernumberofdaysbetweenadmissionanddischarge |
| Payercode | Nominal | Integeridentifiercorrespondingto23distinctvalues,forexample,BlueCross\Blue Shield,Medicare,andself-pay |
| Medicalspecialty | Nominal | values,forexample,cardiology,internalmedicine,family\generalpractice,and surgeon |
| Number of lab procedures | Numeric | Number of lab tests performed during the encounter |
| Numberof procedures | Numeric | Number of procedures (otherthanlabtests) performed during the encounter |
| Numberof medications | Numeric | Number of distinct generic names administered during the encounter |
| Numberofoutpatient visits | Numeric | Number of outpatient visits of the patient in the year preceding the encounter |
| Numberof emergency visits | Numeric | Number of emergency visits of the patient in the year preceding the encounter |
| Number of in patient visits | Numeric | Number of in patient visits of the patient in the year preceding the encounter |
| Diagnosis1 | Nominal | The primary diagnosis (codedasfirstthreedigitsofICD9);848distinctvalues |
| Diagnosis2 | Nominal | Secondary diagnosis(codedasfirstthreedigitsofICD9);923distinctvalues |
| Diagnosis3 | Nominal | Additional secondary diagnosis(codedasfirstthreedigitsofICD9);954distinct values |
| Number of diagnoses | Numeric | Number of diagnoses entered to the system |
| Glucose serum test result | Nominal | Indicates the range of the result or if the test was not taken .Values:“>200,”“>300,” “normal,” and “none” if not measured |
| A1 c test result | Nominal | Indicates the range of the result or if the test was not take. Values: “>8” if the result was greater than 8%  ,“>7”iftheresultwasgreaterthan7%butlessthan8%,“normal” iftheresultwaslessthan7%,and“none”ifnotmeasured. |
| Change of medications | Nominal | Indicates if there was a change in diabetic medications(either dosage or generic name).Values: “change” and“ no change |
| Diabetes medications | Nominal | Indicates if there was any diabetic medication prescribed. Values: “yes” and “no” 0% Nominal  For the generic names: met for min, repaglinide, nateglinide, chlorpropamide, glimepiride, acetohexamide, glipizide, glyburide, tolbutamide, pioglitazone, rosiglitazone,acarbose,miglitol,troglitazone,tolazamide,examide,sitagliptin,insulin, |
| 24features for medications | Nominal | Met formin-rosiglitazone, and metformin-pioglitazone, the feature indicates whether thedrugwasprescribedortherewasachangeinthedosage.Values:“up”ifthedosage wasincreasedduringtheencounter,“down”if the dosage was decreased,“steady”if the dosage did not change, and “no” if the drug was not prescribed |
| Readmitted |  | Days to in patient readmission. Values:“<30”If the patient was readmitted in less than 30days, “>30”ifthepatientwasreadmittedinmorethan30days,and“No”for no record of readmission. |

Table 6.1 Dataset Description

**6.2 KAPLER ALGORITHM :**

**Step 1 :** Select a drug .

**Step 2 :** Consider “x” as the **mean survivability**  .

**Step 3 :** Now check for the value of x (if x>0.5 else x<0.5) .

**Step 4 :** If x>0.5 , then go to Step 10 .

**Step 5 :** Else if x<0.5 then treatment is required and go to Step 6 .

**Step 6 :** Now treatment is initiated by giving a combination of drug and insulin .

**Step 7 :** Check for the value of x (if x>0.5 else x<0.5) again .

**Step 8 :** If x>0.5 , then go to Step 10 .

**Step 9 :** Else if x<0.5 then combination is required .

**Step 10 :**  Treatment is not required .

**6.3 CLASSIFIER PERFORMANCE :**

**Classification metrics :**

We can use classification performance metrics such as Log-Loss, Accuracy, AUC(Area under Curve) etc. Another example of metric for evaluation of machine learning algorithms is precision, recall, which can be used for sorting algorithms primarily used by search engines.

**Survival metrics :**

The survival function gives the probability of survival past a given time point t. Hazard Function h(t): The hazard function gives the rate of failure in a given instant or time t, given that the observation in question has yet to experience the event.

**6.4 FLOWCHART :**

Select drug (declare x)

*Here , x = mean survivability*

If x > 0.5

If x < 0.5

If x < 0.5

If x > 0.5

Treatment with combination required

Treatment not required

Check for x

Give combination of

drug and insulin

Treatment required

Treatment not required

Check for x

Readmission is taken as

an event .

If x < 0.5

If x > 0.5

Check for x

Result not found with combination

Combination successful

1. **SAMPLE CODE :**

R version 3.6.3 (2020-02-29) -- "Holding the Windsock"

Copyright (C) 2020 The R Foundation for Statistical Computing

Platform: x86\_64-w64-mingw32/x64 (64-bit)

R is free software and comes with ABSOLUTELY NO WARRANTY.

You are welcome to redistribute it under certain conditions.

Type 'license()' or 'licence()' for distribution details.

Natural language support but running in an English locale

R is a collaborative project with many contributors.

Type 'contributors()' for more information and

'citation()' on how to cite R or R packages in publications.

Type 'demo()' for some demos, 'help()' for on-line help, or

'help.start()' for an HTML browser interface to help.

Type 'q()' to quit R.

[Previously saved workspace restored]

> library(survival)

> library(survminer)

Loading required package: ggplot2

Loading required package: ggpubr

Loading required package: magrittr

> library(dplyr)

Attaching package: ‘dplyr’

The following objects are masked from ‘package:stats’:

filter, lag

The following objects are masked from ‘package:base’:

intersect, setdiff, setequal, union

> library(ranger)

> library(ggfortify)

> library(ggplot2)

> dia <- read.csv("diabetic\_data.csv",TRUE,",")

> #DRUG : ACARBOSE

> dia <- dia %>%

+ mutate(diabetesMed = ifelse(diabetesMed == "NO",0,1))

> #KEPLER MEIER FIT FOR ACARBOSE DRUG

>

>

> acarbose\_km\_fit <- survfit(Surv(time\_in\_hospital,diabetesMed) ~ acarbose ,data = dia)

> acar\_test = mean(summary(acarbose\_km\_fit$surv))

> if(acar\_test > 0.5)

+ {

+ print("FURTHUR TREATMENT IS NOT REQUIRED")

+ print("PATIENT HAS BEEN CURED WITH THE DRUG ALONE")

+ }else{

+ #ACARBOSE WITH INSULIN

+ #KEPLER MEIER FIT

+

+ dia1 <- subset(dia , readmitted ==">30"|readmitted == "<30")

+ dia1 <- dia1 %>%

+ mutate(readmitted = ifelse(readmitted == ">30" ,0,1))

+ acra\_insulin\_fit <- survfit(Surv(time\_in\_hospital,readmitted) ~ acrabose + insulin ,data = dia1)

+ #VISUALISATION

+

+

+ acra\_with\_insulin\_test = mean(summary(acra\_insulin\_fit$surv))

+ if(acra\_with\_insulin\_test > 0.5)

+ {

+ print("PATIENT'S SURVIVABILITY HAS BEEN INCREASED AFTER TREATMENT WITH INSULIN AND FURTHER TREATMENT IS NOT REAQUIRED")

+ }

+ else{

+ print("PATIENT HAS TO BE TREATED WITH COMBINATION ALONG WITH INSULIN")

+ #COMBINATION OF DRUGS ALONG WITH INSULIN

+ #COMBINATION - 1

+ acra\_combi\_fit\_R <- survfit(Surv(time\_in\_hospital,readmitted) ~ acrabose + glipizide ,data = dia1)

+

+ acra\_combi\_test\_1 = mean(summary(acra\_combi\_fit\_R$surv))

+ #COMBINATION - 2

+ acra\_combi\_fit\_P <- survfit(Surv(time\_in\_hospital,readmitted) ~ glyrbuide + acrabose ,data = dia1)

+

+ acra\_combi\_test\_2 = mean(summary(met\_combi\_fit\_P$surv))

+ #COMBINATION - 3

+ acra\_combi\_test\_3 <- survfit(Surv(time\_in\_hospital,readmitted) ~ metformin + acrabose ,data = dia1)

+

+ acra\_combi\_test\_3 = mean(summary(acra\_combi\_test\_3$surv))

+ #COMBINATION - 4

+ acra\_combi\_fit\_4 <- survfit(Surv(time\_in\_hospital,readmitted) ~ acrabose + rosiglitazone ,data = dia1)

+

+ acra\_combi\_test\_4 = mean(summary(acra\_combi\_fit\_4$surv))

+

+ if(pio\_combi\_test\_1 > 0.5)

+ {

+ print("SURVIVABILITY CAPABILITY")

+ print(met\_combi\_test\_1)

+ print("COMBINATION OF ACRABOSE + PIOGLITAZONE HAS INCREASED THE SURVIVABILITY CAPABILITY")

+ }else if(met\_combi\_test\_2 > 0.5){

+ print("SURVIVABILITY CAPABILITY")

+ print(met\_combi\_test\_2)

+ print("COMBIMATION OF ACRABOSE + GLYBURIDE HAS INCREASED THE SURVIVABILITY CAPABILITY")

+ }

+ else if(met\_combi\_test\_3 > 0.5){

+ print("SURVIVABILITY CAPABILITY")

+ print(met\_combi\_test\_3)

+ print("COMBIMATION OF ACRABOSE + GLIPIZIDE HAS INCREASED THE SURVIVABILITY CAPABILITY")

+ }

+ else if(met\_combi\_test\_4 > 0.5){

+ print("SURVIVABILITY CAPABILITY")

+ print(met\_combi\_test\_4)

+ print("COMBIMATION OF ACRABOSE + GLIPIZIDE HAS INCREASED THE SURVIVABILITY CAPABILITY")

+ }

+ else {

+

+ print("COMBINATIONS ALSO DID NOT GIVE ANY RESULT")

+ }

+

+ }

+ }

Error in eval(predvars, data, env) : object 'acrabose' not found

> dia <- read.csv("diabetic\_data.csv",TRUE,",")

> #DRUG : ACARBOSE

> dia <- dia %>%

+ mutate(diabetesMed = ifelse(diabetesMed == "NO",0,1))

> #KEPLER MEIER FIT FOR ACARBOSE DRUG

>

>

> acarbose\_km\_fit <- survfit(Surv(time\_in\_hospital,diabetesMed) ~ acarbose ,data = dia)

> acar\_test = mean(summary(acarbose\_km\_fit$surv))

> if(acar\_test > 0.5)

+ {

+ print("FURTHUR TREATMENT IS NOT REQUIRED")

+ print("PATIENT HAS BEEN CURED WITH THE DRUG ALONE")

+ }else{

+ #ACARBOSE WITH INSULIN

+ #KEPLER MEIER FIT

+

+ dia1 <- subset(dia , readmitted ==">30"|readmitted == "<30")

+ dia1 <- dia1 %>%

+ mutate(readmitted = ifelse(readmitted == ">30" ,0,1))

+ acra\_insulin\_fit <- survfit(Surv(time\_in\_hospital,readmitted) ~ acarbose + insulin ,data = dia1)

+ #VISUALISATION

+

+

+ acra\_with\_insulin\_test = mean(summary(acra\_insulin\_fit$surv))

+ if(acra\_with\_insulin\_test > 0.5)

+ {

+ print("PATIENT'S SURVIVABILITY HAS BEEN INCREASED AFTER TREATMENT WITH INSULIN AND FURTHER TREATMENT IS NOT REAQUIRED")

+ }

+ else{

+ print("PATIENT HAS TO BE TREATED WITH COMBINATION ALONG WITH INSULIN")

+ #COMBINATION OF DRUGS ALONG WITH INSULIN

+ #COMBINATION - 1

+ acra\_combi\_fit\_R <- survfit(Surv(time\_in\_hospital,readmitted) ~ acarbose + glipizide ,data = dia1)

+

+ acra\_combi\_test\_1 = mean(summary(acra\_combi\_fit\_R$surv))

+ #COMBINATION - 2

+ acra\_combi\_fit\_P <- survfit(Surv(time\_in\_hospital,readmitted) ~ glyrbuide + acarbose ,data = dia1)

+

+ acra\_combi\_test\_2 = mean(summary(met\_combi\_fit\_P$surv))

+ #COMBINATION - 3

+ acra\_combi\_test\_3 <- survfit(Surv(time\_in\_hospital,readmitted) ~ metformin + acarbose ,data = dia1)

+

+ acra\_combi\_test\_3 = mean(summary(acra\_combi\_test\_3$surv))

+ #COMBINATION - 4

+ acra\_combi\_fit\_4 <- survfit(Surv(time\_in\_hospital,readmitted) ~ acarbose + rosiglitazone ,data = dia1)

+

+ acra\_combi\_test\_4 = mean(summary(acra\_combi\_fit\_4$surv))

+

+ if(pio\_combi\_test\_1 > 0.5)

+ {

+ print("SURVIVABILITY CAPABILITY")

+ print(met\_combi\_test\_1)

+ print("COMBINATION OF ACARBOSE + PIOGLITAZONE HAS INCREASED THE SURVIVABILITY CAPABILITY")

+ }else if(met\_combi\_test\_2 > 0.5){

+ print("SURVIVABILITY CAPABILITY")

+ print(met\_combi\_test\_2)

+ print("COMBIMATION OF ACARBOSE + GLYBURIDE HAS INCREASED THE SURVIVABILITY CAPABILITY")

+ }

+ else if(met\_combi\_test\_3 > 0.5){

+ print("SURVIVABILITY CAPABILITY")

+ print(met\_combi\_test\_3)

+ print("COMBIMATION OF ACARBOSE + GLIPIZIDE HAS INCREASED THE SURVIVABILITY CAPABILITY")

+ }

+ else if(met\_combi\_test\_4 > 0.5){

+ print("SURVIVABILITY CAPABILITY")

+ print(met\_combi\_test\_4)

+ print("COMBIMATION OF ACARBOSE + GLIPIZIDE HAS INCREASED THE SURVIVABILITY CAPABILITY")

+ }

+ else {

+

+ print("COMBINATIONS ALSO DID NOT GIVE ANY RESULT")

+ }

+

+ }

+ }

[1] "PATIENT'S SURVIVABILITY HAS BEEN INCREASED AFTER TREATMENT WITH INSULIN AND FURTHER TREATMENT IS NOT REAQUIRED"

> #VISUALISATION AND SUMMARY

> #DRUG ALONE

>

> autoplot( glipizide\_km\_fit,main = "ACARBOSE")

Error in autoplot(glipizide\_km\_fit, main = "ACARBOSE") :

object 'glipizide\_km\_fit' not found

> summary(acrabose\_km\_fit)

Error in summary(acrabose\_km\_fit) : object 'acrabose\_km\_fit' not found

> autoplot( acarbose\_km\_fit,main = "ACARBOSE")

> summary(acrabose\_km\_fit)

Error in summary(acrabose\_km\_fit) : object 'acrabose\_km\_fit' not found

> summary(acarbose\_km\_fit)

Call: survfit(formula = Surv(time\_in\_hospital, diabetesMed) ~ acarbose,

data = dia)

acarbose=Down

time n.risk n.event survival std.err lower 95% CI upper 95% CI

5 3 2 0.333 0.272 0.0673 1

6 1 1 0.000 NaN NA NA

acarbose=No

time n.risk n.event survival std.err lower 95% CI upper 95% CI

1 101458 14171 0.8603 0.001088 0.8582 0.8625

2 87287 17178 0.6910 0.001451 0.6882 0.6939

3 70109 17710 0.5165 0.001569 0.5134 0.5195

4 52399 13874 0.3797 0.001524 0.3767 0.3827

5 38525 9924 0.2819 0.001413 0.2791 0.2847

6 28601 7519 0.2078 0.001274 0.2053 0.2103

7 21082 5844 0.1502 0.001122 0.1480 0.1524

8 15238 4378 0.1070 0.000971 0.1052 0.1090

9 10860 2994 0.0775 0.000840 0.0759 0.0792

10 7866 2337 0.0545 0.000713 0.0531 0.0559

11 5529 1851 0.0363 0.000587 0.0351 0.0374

12 3678 1438 0.0221 0.000461 0.0212 0.0230

13 2240 1205 0.0102 0.000315 0.0096 0.0108

14 1035 1035 0.0000 NaN NA NA

acarbose=Steady

time n.risk n.event survival std.err lower 95% CI upper 95% CI

1 295 35 0.8814 0.01883 0.8452 0.9190

2 260 46 0.7254 0.02598 0.6762 0.7782

3 214 46 0.5695 0.02883 0.5157 0.6289

4 168 49 0.4034 0.02856 0.3511 0.4634

5 119 39 0.2712 0.02588 0.2249 0.3270

6 80 16 0.2169 0.02400 0.1747 0.2695

7 64 13 0.1729 0.02202 0.1347 0.2219

8 51 12 0.1322 0.01972 0.0987 0.1771

9 39 8 0.1051 0.01785 0.0753 0.1466

10 31 5 0.0881 0.01651 0.0611 0.1272

11 26 4 0.0746 0.01530 0.0499 0.1115

12 22 10 0.0407 0.01150 0.0234 0.0708

13 12 5 0.0237 0.00886 0.0114 0.0493

14 7 7 0.0000 NaN NA NA

acarbose=Up

time n.risk n.event survival std.err lower 95% CI upper 95% CI

1 10 2 0.8 0.1265 0.5868 1.000

4 8 1 0.7 0.1449 0.4665 1.000

5 7 1 0.6 0.1549 0.3617 0.995

6 6 3 0.3 0.1449 0.1164 0.773

7 3 2 0.1 0.0949 0.0156 0.642

8 1 1 0.0 NaN NA NA

> #DRUG + INSULIN

> glipi\_insulin\_fit <- survfit(Surv(time\_in\_hospital,readmitted) ~ acarbose + insulin ,data = dia1)

> plot(glipi\_insulin\_fit,main = "ACARBOSE + INSULIN")

> summary(glipi\_insulin\_fit)

Call: survfit(formula = Surv(time\_in\_hospital, readmitted) ~ acarbose +

insulin, data = dia1)

acarbose=Down, insulin=Up

time n.risk n.event survival std.err lower 95% CI upper 95% CI

6 1 1 0 NaN NA NA

acarbose=No, insulin=Down

time n.risk n.event survival std.err lower 95% CI upper 95% CI

1 6430 128 0.980 0.00174 0.977 0.984

2 5869 234 0.941 0.00301 0.935 0.947

3 4938 265 0.891 0.00415 0.882 0.899

4 3901 269 0.829 0.00529 0.819 0.840

5 2986 167 0.783 0.00609 0.771 0.795

6 2295 141 0.735 0.00693 0.721 0.748

7 1748 118 0.685 0.00783 0.670 0.701

8 1294 109 0.627 0.00891 0.610 0.645

9 930 76 0.576 0.00993 0.557 0.596

10 673 70 0.516 0.01119 0.495 0.539

11 469 31 0.482 0.01201 0.459 0.506

12 327 31 0.436 0.01338 0.411 0.463

13 205 32 0.368 0.01581 0.339 0.401

14 96 25 0.272 0.02022 0.235 0.315

acarbose=No, insulin=No

time n.risk n.event survival std.err lower 95% CI upper 95% CI

1 20599 581 0.972 0.00115 0.970 0.974

2 17607 777 0.929 0.00187 0.925 0.933

3 14021 825 0.874 0.00255 0.869 0.879

4 10418 645 0.820 0.00316 0.814 0.826

5 7548 493 0.767 0.00376 0.759 0.774

6 5572 364 0.716 0.00434 0.708 0.725

7 4041 313 0.661 0.00501 0.651 0.671

8 2885 225 0.609 0.00568 0.598 0.621

9 2018 149 0.564 0.00634 0.552 0.577

10 1441 119 0.518 0.00711 0.504 0.532

11 987 81 0.475 0.00794 0.460 0.491

12 647 71 0.423 0.00917 0.406 0.442

13 402 53 0.367 0.01070 0.347 0.389

14 185 48 0.272 0.01424 0.246 0.301

acarbose=No, insulin=Steady

time n.risk n.event survival std.err lower 95% CI upper 95% CI

1 13887 333 0.976 0.00130 0.973 0.979

2 12298 523 0.935 0.00217 0.930 0.939

3 9942 584 0.880 0.00300 0.874 0.886

4 7512 533 0.817 0.00382 0.810 0.825

5 5420 370 0.761 0.00453 0.753 0.770

6 3987 299 0.704 0.00526 0.694 0.715

7 2901 210 0.653 0.00594 0.642 0.665

8 2078 196 0.592 0.00682 0.579 0.605

9 1450 116 0.544 0.00756 0.530 0.559

10 1025 88 0.498 0.00839 0.481 0.514

11 728 41 0.470 0.00899 0.452 0.488

12 482 55 0.416 0.01047 0.396 0.437

13 281 37 0.361 0.01237 0.338 0.386

14 131 40 0.251 0.01689 0.220 0.286

acarbose=No, insulin=Up

time n.risk n.event survival std.err lower 95% CI upper 95% CI

1 5803 115 0.980 0.00183 0.977 0.984

2 5333 176 0.948 0.00298 0.942 0.954

3 4593 212 0.904 0.00409 0.896 0.912

4 3716 193 0.857 0.00508 0.847 0.867

5 2912 166 0.808 0.00604 0.797 0.820

6 2257 142 0.757 0.00701 0.744 0.771

7 1728 110 0.709 0.00793 0.694 0.725

8 1297 94 0.658 0.00895 0.640 0.676

9 952 70 0.609 0.00999 0.590 0.629

10 706 59 0.559 0.01114 0.537 0.581

11 490 42 0.511 0.01239 0.487 0.536

12 327 36 0.454 0.01413 0.428 0.483

13 190 27 0.390 0.01672 0.358 0.424

14 97 22 0.301 0.02102 0.263 0.346

acarbose=Steady, insulin=Down

time n.risk n.event survival std.err lower 95% CI upper 95% CI

1 20 1 0.950 0.0487 0.859 1

3 15 1 0.887 0.0762 0.749 1

acarbose=Steady, insulin=No

time n.risk n.event survival std.err lower 95% CI upper 95% CI

1 102 2 0.980 0.0137 0.954 1.000

2 93 1 0.970 0.0172 0.937 1.000

3 82 4 0.923 0.0283 0.869 0.980

4 63 2 0.893 0.0341 0.829 0.963

7 22 1 0.853 0.0513 0.758 0.959

9 13 1 0.787 0.0788 0.647 0.958

acarbose=Steady, insulin=Steady

time n.risk n.event survival std.err lower 95% CI upper 95% CI

1 27 1 0.963 0.0363 0.894 1.000

3 20 3 0.819 0.0829 0.671 0.998

5 13 2 0.693 0.1078 0.510 0.940

6 9 1 0.616 0.1202 0.420 0.903

8 7 1 0.528 0.1313 0.324 0.859

acarbose=Steady, insulin=Up

time n.risk n.event survival std.err lower 95% CI upper 95% CI

1 26 1 0.962 0.0377 0.890 1

2 23 1 0.920 0.0545 0.819 1

4 16 2 0.805 0.0898 0.647 1

acarbose=Up, insulin=No

time n.risk n.event survival std.err lower 95% CI upper 95% CI

5.000 3.000 1.000 0.667 0.272 0.300 1.000

acarbose=Up, insulin=Steady

time n.risk n.event survival std.err lower 95% CI upper 95% CI

acarbose=Up, insulin=Up

time n.risk n.event survival std.err lower 95% CI upper 95% CI

6 1 1 0 NaN NA NA

>

> #DRUG COMBI - 1

> met\_combi\_fit\_1 <- survfit(Surv(time\_in\_hospital,readmitted) ~ pioglitazone + acarbose ,data = dia1)

> plot(met\_combi\_fit\_1,main = "DRUG COMBINATION:PIOGLITAZONE + ACARBOSE")

> summary(met\_combi\_fit\_1)

Call: survfit(formula = Surv(time\_in\_hospital, readmitted) ~ pioglitazone +

acarbose, data = dia1)

pioglitazone=Down, acarbose=No

time n.risk n.event survival std.err lower 95% CI upper 95% CI

3 59 2 0.966 0.0236 0.921 1.000

4 55 5 0.878 0.0431 0.798 0.967

5 46 3 0.821 0.0515 0.726 0.928

6 31 2 0.768 0.0603 0.659 0.896

9 17 2 0.678 0.0802 0.537 0.855

10 12 1 0.621 0.0912 0.466 0.828

11 7 1 0.532 0.1134 0.351 0.808

14 3 2 0.177 0.1498 0.034 0.928

pioglitazone=No, acarbose=Down

time n.risk n.event survival std.err lower 95% CI upper 95% CI

6 1 1 0 NaN NA NA

pioglitazone=No, acarbose=No

time n.risk n.event survival std.err lower 95% CI upper 95% CI

1 43213 1093 0.975 0.000755 0.973 0.976

2 37984 1608 0.933 0.001240 0.931 0.936

3 30930 1746 0.881 0.001694 0.877 0.884

4 23589 1524 0.824 0.002121 0.820 0.828

5 17399 1098 0.772 0.002501 0.767 0.777

6 13017 885 0.719 0.002887 0.714 0.725

7 9612 707 0.666 0.003290 0.660 0.673

8 6969 580 0.611 0.003736 0.604 0.618

9 4917 389 0.563 0.004168 0.555 0.571

10 3519 304 0.514 0.004647 0.505 0.523

11 2445 180 0.476 0.005090 0.466 0.486

12 1623 181 0.423 0.005856 0.412 0.435

13 974 137 0.364 0.006895 0.350 0.377

14 465 125 0.266 0.009016 0.249 0.284

pioglitazone=No, acarbose=Steady

time n.risk n.event survival std.err lower 95% CI upper 95% CI

1 150 5 0.967 0.0147 0.938 0.996

2 132 2 0.952 0.0177 0.918 0.987

3 115 7 0.894 0.0270 0.843 0.949

4 84 4 0.851 0.0330 0.789 0.919

5 62 1 0.838 0.0352 0.771 0.910

6 38 1 0.816 0.0406 0.740 0.899

7 31 1 0.789 0.0471 0.702 0.887

8 23 1 0.755 0.0562 0.653 0.874

9 16 1 0.708 0.0697 0.584 0.859

pioglitazone=No, acarbose=Up

time n.risk n.event survival std.err lower 95% CI upper 95% CI

5 5 1 0.8 0.179 0.516 1

6 4 1 0.6 0.219 0.293 1

pioglitazone=Steady, acarbose=No

time n.risk n.event survival std.err lower 95% CI upper 95% CI

1 3324 64 0.981 0.00238 0.976 0.985

2 2945 99 0.948 0.00399 0.940 0.956

3 2400 136 0.894 0.00585 0.883 0.906

4 1809 109 0.840 0.00743 0.826 0.855

5 1341 92 0.783 0.00903 0.765 0.800

6 994 55 0.739 0.01024 0.719 0.760

7 724 40 0.698 0.01154 0.676 0.721

8 520 42 0.642 0.01349 0.616 0.669

9 381 17 0.613 0.01457 0.585 0.643

10 288 30 0.549 0.01710 0.517 0.584

11 205 12 0.517 0.01845 0.482 0.555

12 145 12 0.474 0.02065 0.436 0.517

13 94 11 0.419 0.02408 0.374 0.469

14 38 6 0.353 0.03202 0.295 0.422

pioglitazone=Steady, acarbose=Steady

time n.risk n.event survival std.err lower 95% CI upper 95% CI

3 20 1 0.950 0.0487 0.859 1

5 14 1 0.882 0.0795 0.739 1

pioglitazone=Up, acarbose=No

time n.risk n.event survival std.err lower 95% CI upper 95% CI

2 115 3 0.974 0.0149 0.9452 1.000

3 105 2 0.955 0.0195 0.9178 0.994

4 94 2 0.935 0.0238 0.8895 0.983

5 80 3 0.900 0.0303 0.8424 0.961

6 69 4 0.848 0.0382 0.7762 0.926

7 58 4 0.789 0.0454 0.7052 0.883

8 45 2 0.754 0.0497 0.6629 0.858

9 35 3 0.690 0.0578 0.5852 0.813

10 26 1 0.663 0.0613 0.5531 0.795

11 17 2 0.585 0.0749 0.4552 0.752

13 6 1 0.488 0.1087 0.3149 0.755

14 3 2 0.163 0.1376 0.0309 0.854

pioglitazone=Up, acarbose=Steady

time n.risk n.event survival std.err lower 95% CI upper 95% CI

> #DRUG COMBI - 2

> met\_combi\_fit\_2 <- survfit(Surv(time\_in\_hospital,readmitted) ~ glyrbuide + acarbose ,data = dia1)

Error in eval(predvars, data, env) : object 'glyrbuide' not found

> plot(met\_combi\_fit\_2,main = "DRUG COMBINATION:GLYRUBIDE + ACARBOSE")

> summary(met\_combi\_fit\_2)

Call: survfit(formula = Surv(time\_in\_hospital, readmitted) ~ glyburide +

pioglitazone, data = dia1)

glyburide=Down, pioglitazone=No

time n.risk n.event survival std.err lower 95% CI upper 95% CI

1 248 2 0.992 0.00568 0.981 1.000

2 234 4 0.975 0.01009 0.955 0.995

3 208 5 0.952 0.01429 0.924 0.980

4 177 7 0.914 0.01956 0.876 0.953

5 145 5 0.882 0.02342 0.838 0.930

6 115 9 0.813 0.03089 0.755 0.876

7 82 5 0.764 0.03610 0.696 0.838

8 64 4 0.716 0.04098 0.640 0.801

9 52 2 0.688 0.04379 0.608 0.780

10 38 1 0.670 0.04623 0.586 0.767

11 28 1 0.646 0.05040 0.555 0.753

12 19 1 0.612 0.05811 0.508 0.738

13 13 1 0.565 0.07018 0.443 0.721

14 2 2 0.000 NaN NA NA

glyburide=Down, pioglitazone=Steady

time n.risk n.event survival std.err lower 95% CI upper 95% CI

4 18 1 0.944 0.054 0.844 1

9 9 1 0.840 0.110 0.649 1

10 6 1 0.700 0.157 0.450 1

glyburide=Down, pioglitazone=Up

time n.risk n.event survival std.err lower 95% CI upper 95% CI

glyburide=No, pioglitazone=Down

time n.risk n.event survival std.err lower 95% CI upper 95% CI

3 50 2 0.960 0.0277 0.907 1.000

4 46 4 0.877 0.0472 0.789 0.974

5 39 3 0.809 0.0574 0.704 0.930

6 25 2 0.744 0.0687 0.621 0.892

9 12 2 0.620 0.0984 0.454 0.847

10 9 1 0.551 0.1090 0.374 0.812

11 5 1 0.441 0.1316 0.246 0.792

14 2 2 0.000 NaN NA NA

glyburide=No, pioglitazone=No

time n.risk n.event survival std.err lower 95% CI upper 95% CI

1 38972 993 0.975 0.000798 0.973 0.976

2 34195 1466 0.933 0.001313 0.930 0.935

3 27778 1604 0.879 0.001798 0.875 0.882

4 21112 1377 0.822 0.002249 0.817 0.826

5 15552 979 0.770 0.002646 0.765 0.775

6 11611 794 0.717 0.003054 0.711 0.723

7 8591 634 0.664 0.003478 0.657 0.671

8 6232 526 0.608 0.003951 0.601 0.616

9 4390 350 0.560 0.004405 0.551 0.568

10 3132 269 0.512 0.004906 0.502 0.521

11 2166 157 0.475 0.005369 0.464 0.485

12 1432 161 0.421 0.006197 0.409 0.434

13 853 123 0.360 0.007334 0.346 0.375

14 405 108 0.264 0.009574 0.246 0.284

glyburide=No, pioglitazone=Steady

time n.risk n.event survival std.err lower 95% CI upper 95% CI

1 2923 54 0.982 0.00249 0.977 0.986

2 2599 86 0.949 0.00420 0.941 0.957

3 2125 119 0.896 0.00618 0.884 0.908

4 1600 92 0.844 0.00781 0.829 0.860

5 1195 83 0.786 0.00956 0.767 0.805

6 887 53 0.739 0.01095 0.718 0.761

7 654 38 0.696 0.01233 0.672 0.720

8 471 40 0.637 0.01440 0.609 0.666

9 341 16 0.607 0.01554 0.577 0.638

10 255 24 0.550 0.01792 0.516 0.586

11 183 11 0.517 0.01942 0.480 0.556

12 128 8 0.484 0.02130 0.444 0.528

13 86 11 0.422 0.02548 0.375 0.475

14 36 6 0.352 0.03376 0.292 0.425

glyburide=No, pioglitazone=Up

time n.risk n.event survival std.err lower 95% CI upper 95% CI

2 102 3 0.971 0.0167 0.9383 1.000

3 92 2 0.949 0.0220 0.9073 0.994

4 81 2 0.926 0.0270 0.8746 0.981

5 70 3 0.886 0.0342 0.8218 0.956

6 60 4 0.827 0.0428 0.7474 0.916

7 49 2 0.794 0.0473 0.7060 0.892

8 39 2 0.753 0.0529 0.6560 0.864

9 30 3 0.678 0.0630 0.5647 0.813

10 23 1 0.648 0.0668 0.5296 0.793

11 15 1 0.605 0.0750 0.4744 0.771

13 7 1 0.518 0.1026 0.3517 0.764

14 4 2 0.259 0.1394 0.0904 0.744

glyburide=Steady, pioglitazone=Down

time n.risk n.event survival std.err lower 95% CI upper 95% CI

4.000 7.000 1.000 0.857 0.132 0.633 1.000

glyburide=Steady, pioglitazone=No

time n.risk n.event survival std.err lower 95% CI upper 95% CI

1 3818 99 0.974 0.00257 0.969 0.979

2 3381 136 0.935 0.00411 0.927 0.943

3 2787 136 0.889 0.00547 0.879 0.900

4 2161 130 0.836 0.00686 0.822 0.849

5 1588 105 0.781 0.00826 0.764 0.797

6 1184 75 0.731 0.00951 0.713 0.750

7 854 61 0.679 0.01093 0.658 0.701

8 605 43 0.631 0.01238 0.607 0.655

9 425 35 0.579 0.01414 0.552 0.607

10 315 28 0.527 0.01587 0.497 0.559

11 232 19 0.484 0.01739 0.451 0.519

12 160 17 0.433 0.01951 0.396 0.473

13 96 12 0.379 0.02247 0.337 0.425

14 52 15 0.269 0.02866 0.219 0.332

glyburide=Steady, pioglitazone=Steady

time n.risk n.event survival std.err lower 95% CI upper 95% CI

1 376 10 0.973 0.0083 0.957 0.990

2 324 12 0.937 0.0130 0.912 0.963

3 255 18 0.871 0.0193 0.834 0.910

4 193 16 0.799 0.0247 0.752 0.849

5 133 9 0.745 0.0289 0.690 0.804

6 94 2 0.729 0.0304 0.672 0.791

7 62 2 0.706 0.0336 0.643 0.775

8 41 2 0.671 0.0398 0.597 0.754

10 26 5 0.542 0.0610 0.435 0.676

11 17 1 0.510 0.0652 0.397 0.656

12 13 4 0.353 0.0794 0.227 0.549

glyburide=Steady, pioglitazone=Up

time n.risk n.event survival std.err lower 95% CI upper 95% CI

7.000 9.000 2.000 0.778 0.139 0.549 1.000

glyburide=Up, pioglitazone=Down

time n.risk n.event survival std.err lower 95% CI upper 95% CI

glyburide=Up, pioglitazone=No

time n.risk n.event survival std.err lower 95% CI upper 95% CI

1 333 4 0.988 0.00597 0.976 1.000

2 313 4 0.975 0.00861 0.959 0.992

3 279 8 0.947 0.01284 0.923 0.973

4 230 14 0.890 0.01920 0.853 0.928

5 183 11 0.836 0.02387 0.791 0.884

6 150 10 0.780 0.02805 0.727 0.837

7 118 8 0.728 0.03178 0.668 0.793

8 92 8 0.664 0.03604 0.597 0.739

9 66 3 0.634 0.03838 0.563 0.714

10 46 6 0.551 0.04589 0.468 0.649

11 29 3 0.494 0.05162 0.403 0.607

12 19 2 0.442 0.05783 0.342 0.572

13 14 1 0.411 0.06173 0.306 0.551

glyburide=Up, pioglitazone=Steady

time n.risk n.event survival std.err lower 95% CI upper 95% CI

2 22 1 0.955 0.0444 0.871 1

5 13 1 0.881 0.0816 0.735 1

glyburide=Up, pioglitazone=Up

time n.risk n.event survival std.err lower 95% CI upper 95% CI

11 1 1 0 NaN NA NA

> #DRUG COMBI - 3

> met\_combi\_fit\_3 <- survfit(Surv(time\_in\_hospital,readmitted) ~ metformin + acarbose ,data = dia1)

> plot(met\_combi\_fit\_3,main = "DRUG COMBINATION : METFORMIN + ACARBOSE")

> summary(met\_combi\_fit\_3)

Call: survfit(formula = Surv(time\_in\_hospital, readmitted) ~ metformin +

acarbose, data = dia1)

metformin=Down, acarbose=No

time n.risk n.event survival std.err lower 95% CI upper 95% CI

1 258 4 0.984 0.00769 0.9695 1.000

2 241 6 0.960 0.01241 0.9360 0.985

3 211 7 0.928 0.01685 0.8957 0.962

4 177 8 0.886 0.02165 0.8447 0.930

5 126 10 0.816 0.02920 0.7606 0.875

6 99 9 0.742 0.03550 0.6753 0.815

7 77 8 0.665 0.04095 0.5890 0.750

8 51 6 0.586 0.04696 0.5013 0.686

10 26 4 0.496 0.05745 0.3955 0.623

11 20 2 0.447 0.06149 0.3410 0.585

13 9 2 0.347 0.07822 0.2234 0.540

14 6 3 0.174 0.08097 0.0696 0.433

metformin=Down, acarbose=Steady

time n.risk n.event survival std.err lower 95% CI upper 95% CI

metformin=No, acarbose=Down

time n.risk n.event survival std.err lower 95% CI upper 95% CI

6 1 1 0 NaN NA NA

metformin=No, acarbose=No

time n.risk n.event survival std.err lower 95% CI upper 95% CI

1 38117 997 0.974 0.000817 0.972 0.975

2 33403 1417 0.933 0.001329 0.930 0.935

3 27233 1549 0.879 0.001812 0.876 0.883

4 20811 1356 0.822 0.002266 0.818 0.827

5 15402 979 0.770 0.002667 0.765 0.775

6 11557 784 0.718 0.003070 0.712 0.724

7 8543 625 0.665 0.003491 0.658 0.672

8 6222 507 0.611 0.003950 0.603 0.619

9 4417 340 0.564 0.004393 0.555 0.573

10 3192 285 0.514 0.004910 0.504 0.523

11 2225 172 0.474 0.005383 0.463 0.485

12 1472 168 0.420 0.006178 0.408 0.432

13 882 117 0.364 0.007191 0.350 0.378

14 418 103 0.274 0.009395 0.257 0.293

metformin=No, acarbose=Steady

time n.risk n.event survival std.err lower 95% CI upper 95% CI

1 133 3 0.977 0.0129 0.953 1.000

2 117 1 0.969 0.0152 0.940 0.999

3 103 6 0.913 0.0266 0.862 0.966

4 79 2 0.890 0.0305 0.832 0.951

5 62 1 0.875 0.0332 0.812 0.943

6 40 1 0.853 0.0389 0.780 0.933

7 32 1 0.827 0.0460 0.741 0.922

8 24 1 0.792 0.0555 0.691 0.909

metformin=No, acarbose=Up

time n.risk n.event survival std.err lower 95% CI upper 95% CI

5 4 1 0.75 0.217 0.426 1

6 3 1 0.50 0.250 0.188 1

metformin=Steady, acarbose=No

time n.risk n.event survival std.err lower 95% CI upper 95% CI

1 7913 152 0.981 0.00154 0.978 0.984

2 7055 279 0.942 0.00272 0.937 0.947

3 5684 316 0.890 0.00384 0.882 0.897

4 4258 263 0.835 0.00488 0.825 0.844

5 3096 198 0.781 0.00586 0.770 0.793

6 2264 141 0.733 0.00678 0.719 0.746

7 1653 111 0.683 0.00776 0.668 0.699

8 1181 106 0.622 0.00907 0.605 0.640

9 825 67 0.572 0.01022 0.552 0.592

10 569 43 0.528 0.01137 0.507 0.551

11 392 19 0.503 0.01225 0.479 0.527

12 271 24 0.458 0.01414 0.431 0.487

13 171 28 0.383 0.01755 0.350 0.419

14 79 27 0.252 0.02349 0.210 0.303

metformin=Steady, acarbose=Steady

time n.risk n.event survival std.err lower 95% CI upper 95% CI

1 37 2 0.946 0.0372 0.876 1.000

2 33 1 0.917 0.0458 0.832 1.000

3 30 2 0.856 0.0598 0.747 0.982

4 22 2 0.778 0.0755 0.643 0.941

5 14 1 0.723 0.0883 0.569 0.918

metformin=Steady, acarbose=Up

time n.risk n.event survival std.err lower 95% CI upper 95% CI

metformin=Up, acarbose=No

time n.risk n.event survival std.err lower 95% CI upper 95% CI

1 431 4 0.991 0.00462 0.982 1.000

2 408 8 0.971 0.00817 0.955 0.987

3 366 14 0.934 0.01251 0.910 0.959

4 301 13 0.894 0.01622 0.863 0.926

5 242 9 0.861 0.01903 0.824 0.899

6 191 12 0.806 0.02337 0.762 0.854

7 145 7 0.768 0.02648 0.717 0.821

8 100 5 0.729 0.03021 0.672 0.791

9 74 4 0.690 0.03441 0.626 0.761

10 58 4 0.642 0.03941 0.569 0.724

11 37 2 0.607 0.04427 0.527 0.701

12 26 1 0.584 0.04834 0.497 0.687

13 16 2 0.511 0.06420 0.400 0.654

14 6 2 0.341 0.10727 0.184 0.632

metformin=Up, acarbose=Steady

time n.risk n.event survival std.err lower 95% CI upper 95% CI

9 1 1 0 NaN NA NA

> #DRUG COMBI - 4

> met\_combi\_fit\_4 <- survfit(Surv(time\_in\_hospital,readmitted) ~ acarbose + rosiglitazone ,data = dia1)

> plot(met\_combi\_fit\_4,main = "DRUG COMBINATION : ACARBOSE + ROSIGLITAZONE")

> summary(met\_combi\_fit\_4)

Call: survfit(formula = Surv(time\_in\_hospital, readmitted) ~ acarbose +

rosiglitazone, data = dia1)

acarbose=Down, rosiglitazone=No

time n.risk n.event survival std.err lower 95% CI upper 95% CI

6 1 1 0 NaN NA NA

acarbose=Down, rosiglitazone=Steady

time n.risk n.event survival std.err lower 95% CI upper 95% CI

acarbose=No, rosiglitazone=Down

time n.risk n.event survival std.err lower 95% CI upper 95% CI

4 25 1 0.960 0.0392 0.886 1

5 18 2 0.853 0.0792 0.711 1

8 7 1 0.731 0.1317 0.514 1

11 4 1 0.549 0.1866 0.282 1

acarbose=No, rosiglitazone=No

time n.risk n.event survival std.err lower 95% CI upper 95% CI

1 43662 1093 0.975 0.000748 0.974 0.976

2 38368 1617 0.934 0.001230 0.931 0.936

3 31231 1784 0.881 0.001688 0.877 0.884

4 23806 1510 0.825 0.002106 0.821 0.829

5 17620 1127 0.772 0.002489 0.767 0.777

6 13190 895 0.720 0.002871 0.714 0.725

7 9739 713 0.667 0.003269 0.660 0.673

8 7052 585 0.612 0.003713 0.604 0.619

9 5005 390 0.564 0.004134 0.556 0.572

10 3591 318 0.514 0.004620 0.505 0.523

11 2500 177 0.478 0.005038 0.468 0.488

12 1683 185 0.425 0.005776 0.414 0.437

13 1009 144 0.364 0.006814 0.351 0.378

14 479 129 0.266 0.008908 0.249 0.284

acarbose=No, rosiglitazone=Steady

time n.risk n.event survival std.err lower 95% CI upper 95% CI

1 2960 64 0.978 0.00267 0.973 0.984

2 2642 92 0.944 0.00434 0.936 0.953

3 2171 100 0.901 0.00593 0.889 0.913

4 1656 126 0.832 0.00803 0.817 0.848

5 1180 66 0.786 0.00941 0.768 0.804

6 868 49 0.741 0.01080 0.720 0.763

7 635 34 0.702 0.01218 0.678 0.726

8 470 37 0.646 0.01421 0.619 0.675

9 320 21 0.604 0.01601 0.573 0.636

10 232 15 0.565 0.01787 0.531 0.601

11 160 17 0.505 0.02108 0.465 0.548

12 90 7 0.466 0.02411 0.421 0.515

13 62 4 0.436 0.02683 0.386 0.492

14 27 6 0.339 0.04062 0.268 0.429

acarbose=No, rosiglitazone=Up

time n.risk n.event survival std.err lower 95% CI upper 95% CI

2 70 1 0.986 0.0142 0.958 1.000

3 66 2 0.956 0.0249 0.908 1.000

4 60 3 0.908 0.0358 0.840 0.981

5 48 1 0.889 0.0398 0.815 0.971

6 40 2 0.845 0.0486 0.755 0.946

7 35 4 0.748 0.0626 0.635 0.881

8 25 1 0.718 0.0669 0.598 0.862

10 17 3 0.591 0.0863 0.444 0.787

12 8 1 0.518 0.1024 0.351 0.763

13 6 1 0.431 0.1161 0.254 0.731

acarbose=Steady, rosiglitazone=No

time n.risk n.event survival std.err lower 95% CI upper 95% CI

1 159 5 0.969 0.0138 0.942 0.996

2 140 2 0.955 0.0167 0.922 0.988

3 123 8 0.893 0.0264 0.842 0.946

4 97 2 0.874 0.0289 0.819 0.933

5 74 1 0.862 0.0308 0.804 0.925

6 49 1 0.845 0.0348 0.779 0.916

7 41 1 0.824 0.0396 0.750 0.906

8 32 1 0.798 0.0460 0.713 0.894

9 23 1 0.764 0.0556 0.662 0.881

acarbose=Steady, rosiglitazone=Steady

time n.risk n.event survival std.err lower 95% CI upper 95% CI

4 7 2 0.714 0.171 0.447 1

5 4 1 0.536 0.201 0.257 1

acarbose=Steady, rosiglitazone=Up

time n.risk n.event survival std.err lower 95% CI upper 95% CI

acarbose=Up, rosiglitazone=No

time n.risk n.event survival std.err lower 95% CI upper 95% CI

5 5 1 0.8 0.179 0.516 1

6 4 1 0.6 0.219 0.293 1

>

> #COX\_PROPORTIONAL

> dia <- read.csv("diabetic\_csv.data",TRUE,",")

Error in file(file, "rt") : cannot open the connection

In addition: Warning message:

In file(file, "rt") :

cannot open file 'diabetic\_csv.data': No such file or directory

> dia <- read.csv("diabetic\_data.csv",TRUE,",")

> dia <- dia %>%

+ mutate(diabetesMed = ifelse(diabetesMed == "NO" ,0,1))

> #km\_fit <- survfit(Surv(time\_in\_hospital,diabetesMed) ~ acarbose,data = dia)

> autoplot(km\_fit)

> #COX\_PROPORTIONAL

> cox <- coxph(Surv(time\_in\_hospital,diabetesMed) ~ acarbose , data = dia)

> cox\_fit <- survfit(cox)

> autoplot(cox\_fit)

> #RANDOM FOREST

> r\_fit <- ranger(Surv(time\_in\_hospital,diabetesMed) ~ acarbose,data = dia,importance = "permutation",splitrule = "extratrees")

Computing permutation importance.. Progress: 0%. Estimated remaining time: 4 hours, 51 minutes, 5 seconds.

Computing permutation importance.. Progress: 1%. Estimated remaining time: 1 hour, 57 minutes, 9 seconds.

Computing permutation importance.. Progress: 2%. Estimated remaining time: 1 hour, 37 minutes, 17 seconds.

Computing permutation importance.. Progress: 3%. Estimated remaining time: 1 hour, 29 minutes, 54 seconds.

Computing permutation importance.. Progress: 3%. Estimated remaining time: 1 hour, 24 minutes, 45 seconds.

Computing permutation importance.. Progress: 4%. Estimated remaining time: 1 hour, 21 minutes, 44 seconds.

Computing permutation importance.. Progress: 5%. Estimated remaining time: 1 hour, 19 minutes, 29 seconds.

Computing permutation importance.. Progress: 6%. Estimated remaining time: 1 hour, 17 minutes, 57 seconds.

Computing permutation importance.. Progress: 7%. Estimated remaining time: 1 hour, 16 minutes, 53 seconds.

Computing permutation importance.. Progress: 7%. Estimated remaining time: 1 hour, 16 minutes, 7 seconds.

Computing permutation importance.. Progress: 8%. Estimated remaining time: 1 hour, 14 minutes, 49 seconds.

Computing permutation importance.. Progress: 9%. Estimated remaining time: 1 hour, 14 minutes, 8 seconds.

Computing permutation importance.. Progress: 10%. Estimated remaining time: 1 hour, 13 minutes, 1 seconds.

Computing permutation importance.. Progress: 11%. Estimated remaining time: 1 hour, 12 minutes, 6 seconds.

Computing permutation importance.. Progress: 11%. Estimated remaining time: 1 hour, 11 minutes, 14 seconds.

Computing permutation importance.. Progress: 12%. Estimated remaining time: 1 hour, 10 minutes, 24 seconds.

Computing permutation importance.. Progress: 13%. Estimated remaining time: 1 hour, 9 minutes, 36 seconds.

Computing permutation importance.. Progress: 14%. Estimated remaining time: 1 hour, 8 minutes, 48 seconds.

Computing permutation importance.. Progress: 15%. Estimated remaining time: 1 hour, 8 minutes, 26 seconds.

Computing permutation importance.. Progress: 15%. Estimated remaining time: 1 hour, 8 minutes, 34 seconds.

Computing permutation importance.. Progress: 16%. Estimated remaining time: 1 hour, 7 minutes, 56 seconds.

Computing permutation importance.. Progress: 17%. Estimated remaining time: 1 hour, 7 minutes, 22 seconds.

Computing permutation importance.. Progress: 18%. Estimated remaining time: 1 hour, 7 minutes, 29 seconds.

Computing permutation importance.. Progress: 19%. Estimated remaining time: 1 hour, 6 minutes, 57 seconds.

Computing permutation importance.. Progress: 19%. Estimated remaining time: 1 hour, 6 minutes, 7 seconds.

Computing permutation importance.. Progress: 20%. Estimated remaining time: 1 hour, 5 minutes, 22 seconds.

Computing permutation importance.. Progress: 21%. Estimated remaining time: 1 hour, 4 minutes, 34 seconds.

Computing permutation importance.. Progress: 22%. Estimated remaining time: 1 hour, 3 minutes, 43 seconds.

Computing permutation importance.. Progress: 23%. Estimated remaining time: 1 hour, 2 minutes, 54 seconds.

Computing permutation importance.. Progress: 23%. Estimated remaining time: 1 hour, 2 minutes, 5 seconds.

Computing permutation importance.. Progress: 24%. Estimated remaining time: 1 hour, 1 minute, 17 seconds.

Computing permutation importance.. Progress: 25%. Estimated remaining time: 1 hour, 0 minutes, 30 seconds.

Computing permutation importance.. Progress: 26%. Estimated remaining time: 59 minutes, 46 seconds.

Computing permutation importance.. Progress: 27%. Estimated remaining time: 59 minutes, 3 seconds.

Computing permutation importance.. Progress: 27%. Estimated remaining time: 58 minutes, 20 seconds.

Computing permutation importance.. Progress: 28%. Estimated remaining time: 57 minutes, 35 seconds.

Computing permutation importance.. Progress: 29%. Estimated remaining time: 56 minutes, 50 seconds.

Computing permutation importance.. Progress: 30%. Estimated remaining time: 56 minutes, 20 seconds.

Computing permutation importance.. Progress: 31%. Estimated remaining time: 55 minutes, 38 seconds.

Computing permutation importance.. Progress: 31%. Estimated remaining time: 55 minutes, 18 seconds.

Computing permutation importance.. Progress: 32%. Estimated remaining time: 54 minutes, 38 seconds.

Computing permutation importance.. Progress: 33%. Estimated remaining time: 54 minutes, 2 seconds.

Computing permutation importance.. Progress: 34%. Estimated remaining time: 53 minutes, 18 seconds.

Computing permutation importance.. Progress: 35%. Estimated remaining time: 52 minutes, 34 seconds.

Computing permutation importance.. Progress: 35%. Estimated remaining time: 52 minutes, 4 seconds.

Computing permutation importance.. Progress: 36%. Estimated remaining time: 51 minutes, 48 seconds.

Computing permutation importance.. Progress: 37%. Estimated remaining time: 51 minutes, 35 seconds.

Computing permutation importance.. Progress: 38%. Estimated remaining time: 51 minutes, 23 seconds.

Computing permutation importance.. Progress: 39%. Estimated remaining time: 51 minutes, 17 seconds.

Computing permutation importance.. Progress: 39%. Estimated remaining time: 51 minutes, 11 seconds.

Computing permutation importance.. Progress: 40%. Estimated remaining time: 50 minutes, 53 seconds.

Computing permutation importance.. Progress: 41%. Estimated remaining time: 50 minutes, 37 seconds.

Computing permutation importance.. Progress: 42%. Estimated remaining time: 50 minutes, 24 seconds.

Computing permutation importance.. Progress: 43%. Estimated remaining time: 49 minutes, 49 seconds.

Computing permutation importance.. Progress: 43%. Estimated remaining time: 49 minutes, 26 seconds.

Computing permutation importance.. Progress: 44%. Estimated remaining time: 49 minutes, 11 seconds.

Computing permutation importance.. Progress: 45%. Estimated remaining time: 48 minutes, 54 seconds.

Computing permutation importance.. Progress: 46%. Estimated remaining time: 48 minutes, 34 seconds.

Computing permutation importance.. Progress: 47%. Estimated remaining time: 48 minutes, 12 seconds.

Computing permutation importance.. Progress: 47%. Estimated remaining time: 47 minutes, 40 seconds.

Computing permutation importance.. Progress: 48%. Estimated remaining time: 47 minutes, 11 seconds.

Computing permutation importance.. Progress: 49%. Estimated remaining time: 46 minutes, 39 seconds.

Computing permutation importance.. Progress: 50%. Estimated remaining time: 46 minutes, 11 seconds.

Computing permutation importance.. Progress: 51%. Estimated remaining time: 45 minutes, 37 seconds.

Computing permutation importance.. Progress: 51%. Estimated remaining time: 45 minutes, 6 seconds.

Computing permutation importance.. Progress: 52%. Estimated remaining time: 44 minutes, 36 seconds.

Computing permutation importance.. Progress: 53%. Estimated remaining time: 44 minutes, 2 seconds.

Computing permutation importance.. Progress: 54%. Estimated remaining time: 43 minutes, 26 seconds.

Computing permutation importance.. Progress: 55%. Estimated remaining time: 42 minutes, 49 seconds.

Computing permutation importance.. Progress: 55%. Estimated remaining time: 42 minutes, 6 seconds.

Computing permutation importance.. Progress: 56%. Estimated remaining time: 41 minutes, 28 seconds.

Computing permutation importance.. Progress: 57%. Estimated remaining time: 40 minutes, 51 seconds.

Computing permutation importance.. Progress: 58%. Estimated remaining time: 40 minutes, 16 seconds.

Computing permutation importance.. Progress: 59%. Estimated remaining time: 39 minutes, 39 seconds.

Computing permutation importance.. Progress: 59%. Estimated remaining time: 38 minutes, 58 seconds.

Computing permutation importance.. Progress: 60%. Estimated remaining time: 38 minutes, 18 seconds.

Computing permutation importance.. Progress: 61%. Estimated remaining time: 37 minutes, 38 seconds.

Computing permutation importance.. Progress: 62%. Estimated remaining time: 36 minutes, 57 seconds.

Computing permutation importance.. Progress: 63%. Estimated remaining time: 36 minutes, 18 seconds.

Computing permutation importance.. Progress: 63%. Estimated remaining time: 35 minutes, 38 seconds.

Computing permutation importance.. Progress: 64%. Estimated remaining time: 34 minutes, 57 seconds.

Computing permutation importance.. Progress: 65%. Estimated remaining time: 34 minutes, 10 seconds.

Computing permutation importance.. Progress: 66%. Estimated remaining time: 33 minutes, 17 seconds.

Computing permutation importance.. Progress: 67%. Estimated remaining time: 32 minutes, 25 seconds.

Computing permutation importance.. Progress: 67%. Estimated remaining time: 31 minutes, 40 seconds.

Computing permutation importance.. Progress: 68%. Estimated remaining time: 30 minutes, 59 seconds.

Computing permutation importance.. Progress: 69%. Estimated remaining time: 30 minutes, 19 seconds.

Computing permutation importance.. Progress: 70%. Estimated remaining time: 29 minutes, 38 seconds.

Computing permutation importance.. Progress: 71%. Estimated remaining time: 28 minutes, 57 seconds.

Computing permutation importance.. Progress: 71%. Estimated remaining time: 28 minutes, 15 seconds.

Computing permutation importance.. Progress: 72%. Estimated remaining time: 27 minutes, 29 seconds.

Computing permutation importance.. Progress: 73%. Estimated remaining time: 26 minutes, 44 seconds.

Computing permutation importance.. Progress: 74%. Estimated remaining time: 26 minutes, 0 seconds.

Computing permutation importance.. Progress: 75%. Estimated remaining time: 25 minutes, 16 seconds.

Computing permutation importance.. Progress: 75%. Estimated remaining time: 24 minutes, 28 seconds.

Computing permutation importance.. Progress: 76%. Estimated remaining time: 23 minutes, 42 seconds.

Computing permutation importance.. Progress: 77%. Estimated remaining time: 22 minutes, 53 seconds.

Computing permutation importance.. Progress: 78%. Estimated remaining time: 22 minutes, 7 seconds.

Computing permutation importance.. Progress: 79%. Estimated remaining time: 21 minutes, 20 seconds.

Computing permutation importance.. Progress: 79%. Estimated remaining time: 20 minutes, 33 seconds.

Computing permutation importance.. Progress: 80%. Estimated remaining time: 19 minutes, 46 seconds.

Computing permutation importance.. Progress: 81%. Estimated remaining time: 18 minutes, 59 seconds.

Computing permutation importance.. Progress: 82%. Estimated remaining time: 18 minutes, 12 seconds.

Computing permutation importance.. Progress: 83%. Estimated remaining time: 17 minutes, 26 seconds.

Computing permutation importance.. Progress: 83%. Estimated remaining time: 16 minutes, 39 seconds.

Computing permutation importance.. Progress: 84%. Estimated remaining time: 15 minutes, 52 seconds.

Computing permutation importance.. Progress: 85%. Estimated remaining time: 15 minutes, 5 seconds.

Computing permutation importance.. Progress: 86%. Estimated remaining time: 14 minutes, 17 seconds.

Computing permutation importance.. Progress: 87%. Estimated remaining time: 13 minutes, 30 seconds.

Computing permutation importance.. Progress: 87%. Estimated remaining time: 12 minutes, 42 seconds.

Computing permutation importance.. Progress: 88%. Estimated remaining time: 11 minutes, 53 seconds.

Computing permutation importance.. Progress: 89%. Estimated remaining time: 11 minutes, 5 seconds.

Computing permutation importance.. Progress: 90%. Estimated remaining time: 10 minutes, 17 seconds.

Computing permutation importance.. Progress: 91%. Estimated remaining time: 9 minutes, 28 seconds.

Computing permutation importance.. Progress: 91%. Estimated remaining time: 8 minutes, 40 seconds.

Computing permutation importance.. Progress: 92%. Estimated remaining time: 7 minutes, 52 seconds.

Computing permutation importance.. Progress: 93%. Estimated remaining time: 7 minutes, 4 seconds.

Computing permutation importance.. Progress: 94%. Estimated remaining time: 6 minutes, 15 seconds.

Computing permutation importance.. Progress: 95%. Estimated remaining time: 5 minutes, 26 seconds.

Computing permutation importance.. Progress: 95%. Estimated remaining time: 4 minutes, 37 seconds.

Computing permutation importance.. Progress: 96%. Estimated remaining time: 3 minutes, 49 seconds.

Computing permutation importance.. Progress: 97%. Estimated remaining time: 3 minutes, 1 seconds.

Computing permutation importance.. Progress: 98%. Estimated remaining time: 2 minutes, 13 seconds.

Computing permutation importance.. Progress: 99%. Estimated remaining time: 1 minute, 24 seconds.

Computing permutation importance.. Progress: 99%. Estimated remaining time: 36 seconds.

> death\_times <- r\_fit$unique.death.times

> surv\_prob <- data.frame(r\_fit$survival)

> avg\_prob <- sapply(surv\_prob,mean)

> #setup for ggplot

> kmi <- rep("KM",length(km\_fit$time))

> km\_df <- data.frame(km\_fit$time,km\_fit$surv,kmi)

> names(km\_df) <- c("Time","Surv","Model")

> #cox\_model\_generation

> coxi <- rep("Cox",length(cox\_fit$time))

> cox\_df <- data.frame(cox\_fit$time,cox\_fit$surv,coxi)

> names(cox\_df) <- c("Time","Surv","Model")

> #random forest

> rfi <- rep("RF",length(r\_fit$unique.death.times))

> rf\_df <- data.frame(r\_fit$unique.death.times,avg\_prob,rfi)

> names(rf\_df) <- c("Time","Surv","Model")

> plot\_df <- rbind(km\_df,cox\_df,rf\_df)

> p <- ggplot(plot\_df, aes(x = Time, y = Surv, color = Model))

> p + geom\_line()

1. **SOFTWARE TESTING :**

**VERIFICATION :** The process of checking that a software achieves its goal without any bugs. It is the process to ensure whether the product that is developed is right or not. It verifies whether the developed product fulfills the requirements that we have.

**VALIDATION :** The process of checking whether the software product is up to the mark or in other words product has high level requirements. It is the process of checking the validation of product i.e. it checks what we are developing is the right product . It is the Dynamic Testing.

**BLACK BOX TESTING :** Black box testing is a type of software testing in which the functionality of the software is not known. The testing is done without the internal knowledge of the products.

**WHITE BOX TESTING :** White box testing techniques analyze the internal structures the used data structures, internal design, code structure and the working of the software rather than just the functionality as in black box testing. It is also called glass box testing or clear box testing or structural testing.

**8.1 TYPES OF TESTING :**

**INTEGRATION TESTING :** Testing of all integrated modules to verify the combined functionality after integration is termed as[Integration Testing](https://www.softwaretestinghelp.com/what-is-integration-testing/).

Modules are typically code modules, individual applications, client and server applications on a network, etc. This type of testing is especially relevant to client/server and distributed systems.

**REGRESSION TESTING :** Testing an application as a whole for the modification in any module or functionality is termed as Regression Testing. It is difficult to cover all the system in [Regression Testing](https://www.softwaretestinghelp.com/regression-testing-tools-and-methods/), so typically [Automation Testing Tools](https://www.softwaretestinghelp.com/automation-testing-tutorial-1/) are used for these types of testing.

**UNIT TESTING :** Testing of an individual software component or module is termed as [Unit Testing](https://www.softwaretestinghelp.com/unit-testing/). It is typically done by the programmer and not by testers, as it requires detailed knowledge of the internal program design and code. It may also require developing test driver modules or test harnesses.

**8.2 TEST CASES :**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Test Case No** | **Patient ID** | **Treatment 1** | **Result**  **(Pass/Fail)** | **Treatment Status (Success/Fail)** |
| 1. | 101 | Give glipizide | Pass | Success |
| Fail | Fail (treatment to be continued) |

Table 8.2.1 Test Case 1

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Test Case No** | **Patient ID** | **Treatment 2** | **Result**  **(Pass/Fail)** | **Treatment Status (Success/Fail)** |
| 2. | 101 | Give a drug combination (glipizide+metaformin) | Pass | Success |
| Fail | Fail (treatment to be continued) |

Table 8.2.2 Test Case 2

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Test Case No** | **Patient ID** | **Treatment 1** | **Result**  **(Pass/Fail)** | **Treatment Status (Success/Fail)** |
| 1. | 101 | Give metaformin | Pass | Success |
| Fail | Fail (treatment to be continued) |

Table 8.2.3 Test Case 3

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Test Case No** | **Patient ID** | **Treatment 2** | **Result**  **(Pass/Fail)** | **Treatment Status (Success/Fail)** |
| 2. | 101 | Give a drug combination (metaformin+insulin) | Pass | Success |
| Fail | Fail (treatment to be continued) |

Table 8.2.4 Test Case 4

1. **RESULTS :**

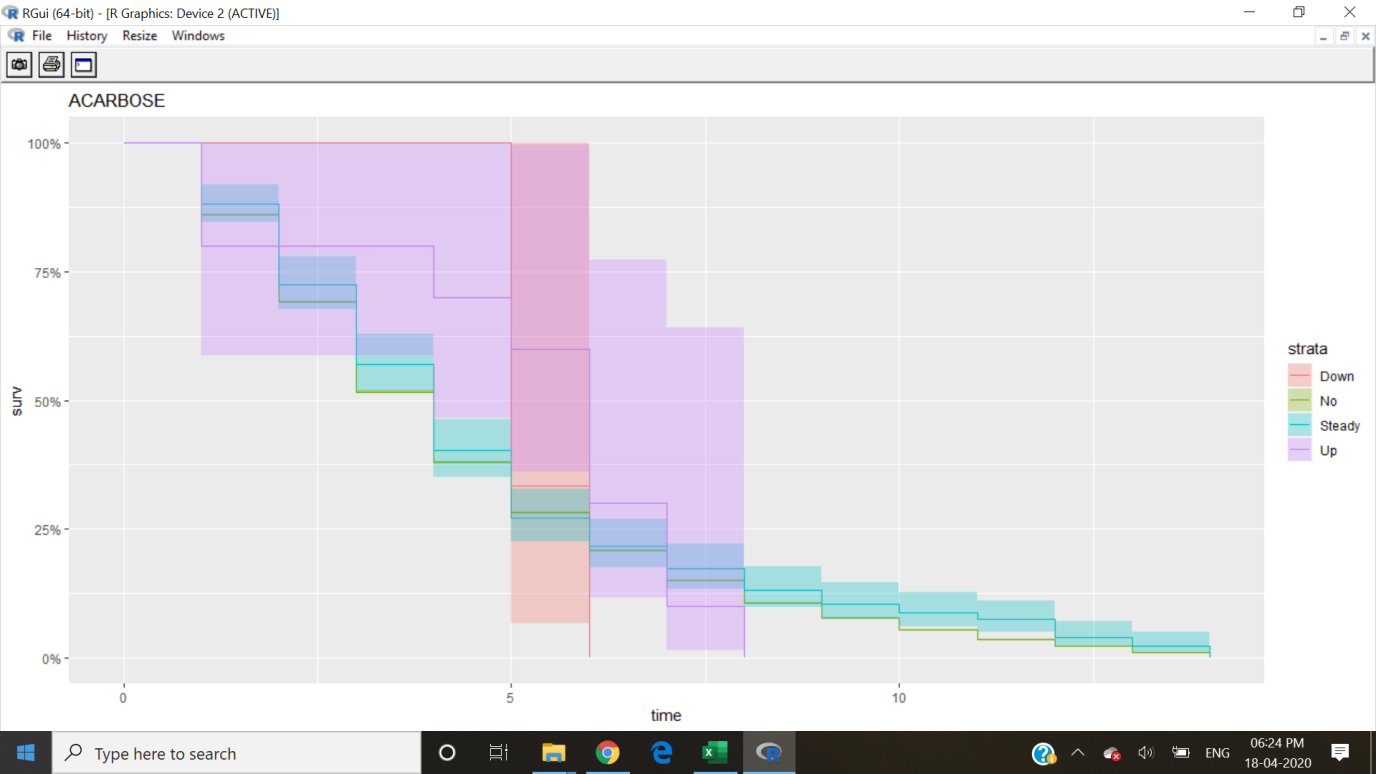
****

Figure 9.1 Acarbose

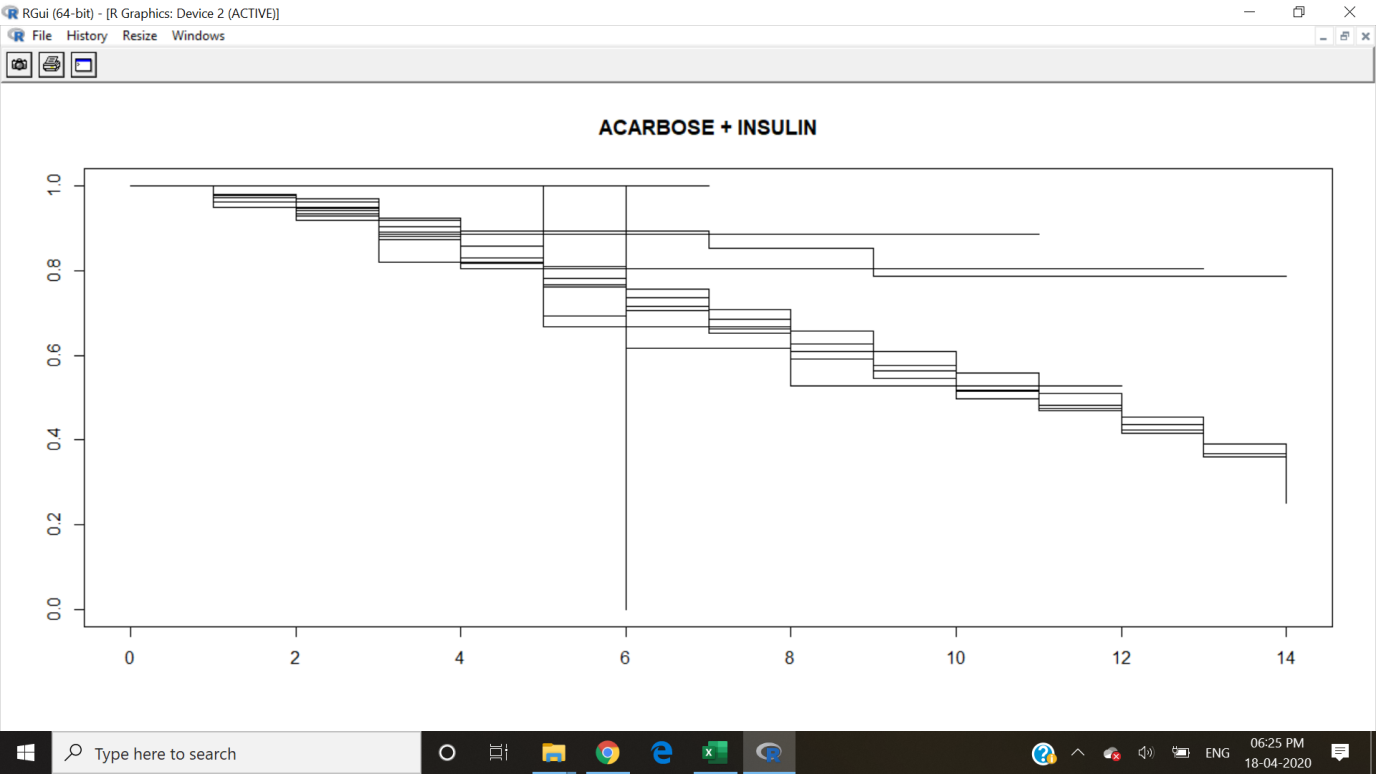
****

Figure 9.2 Acarbose + Insulin

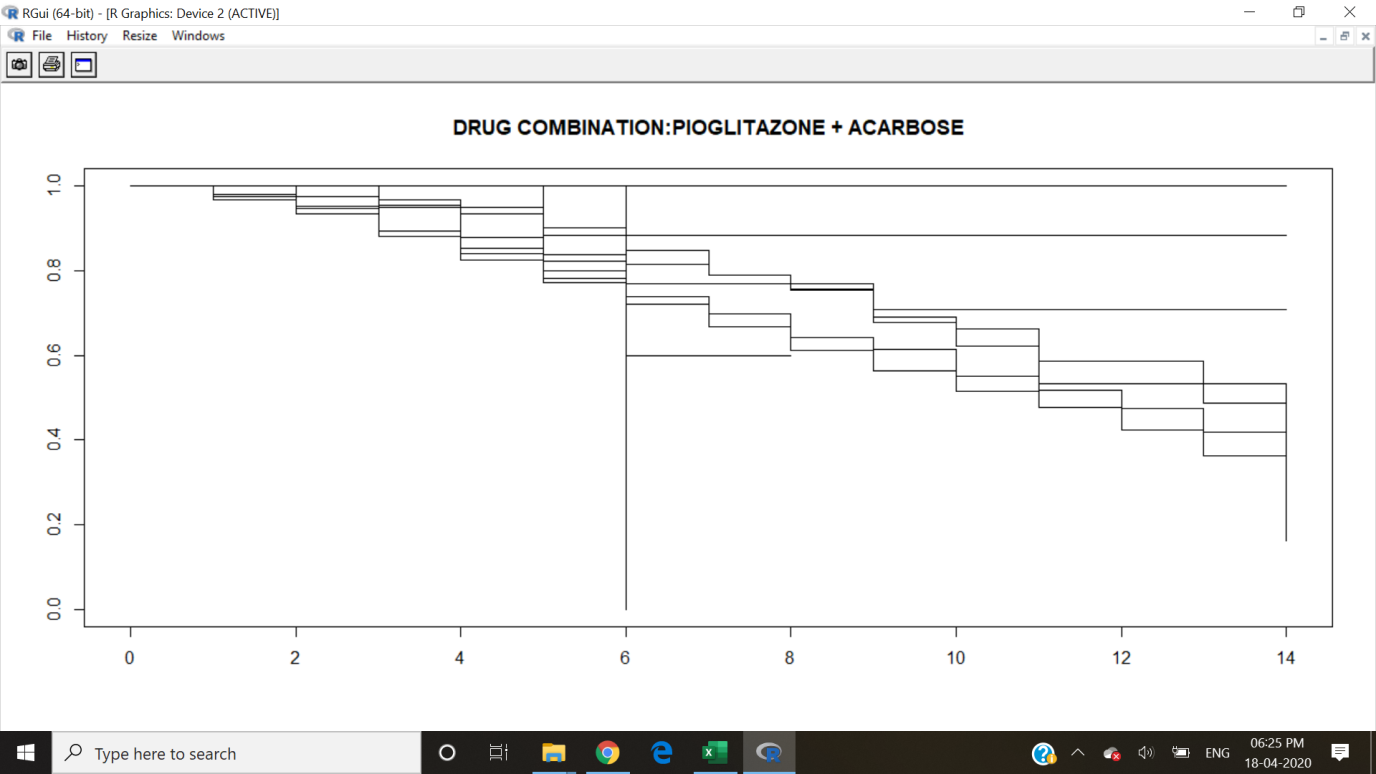
****

Figure 9.3 Acarbose + Pioglitazone

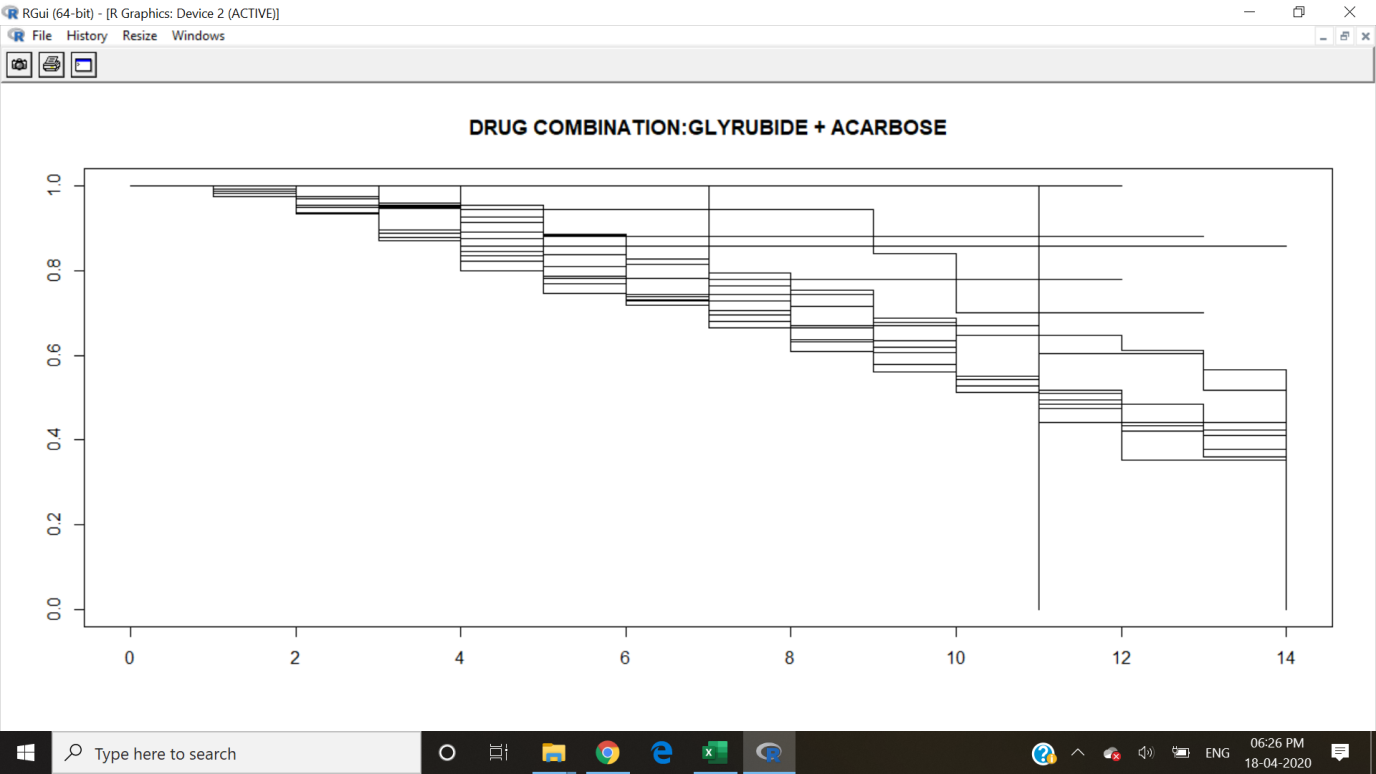
****

Figure 9.4 Acarbose + Glyrubide

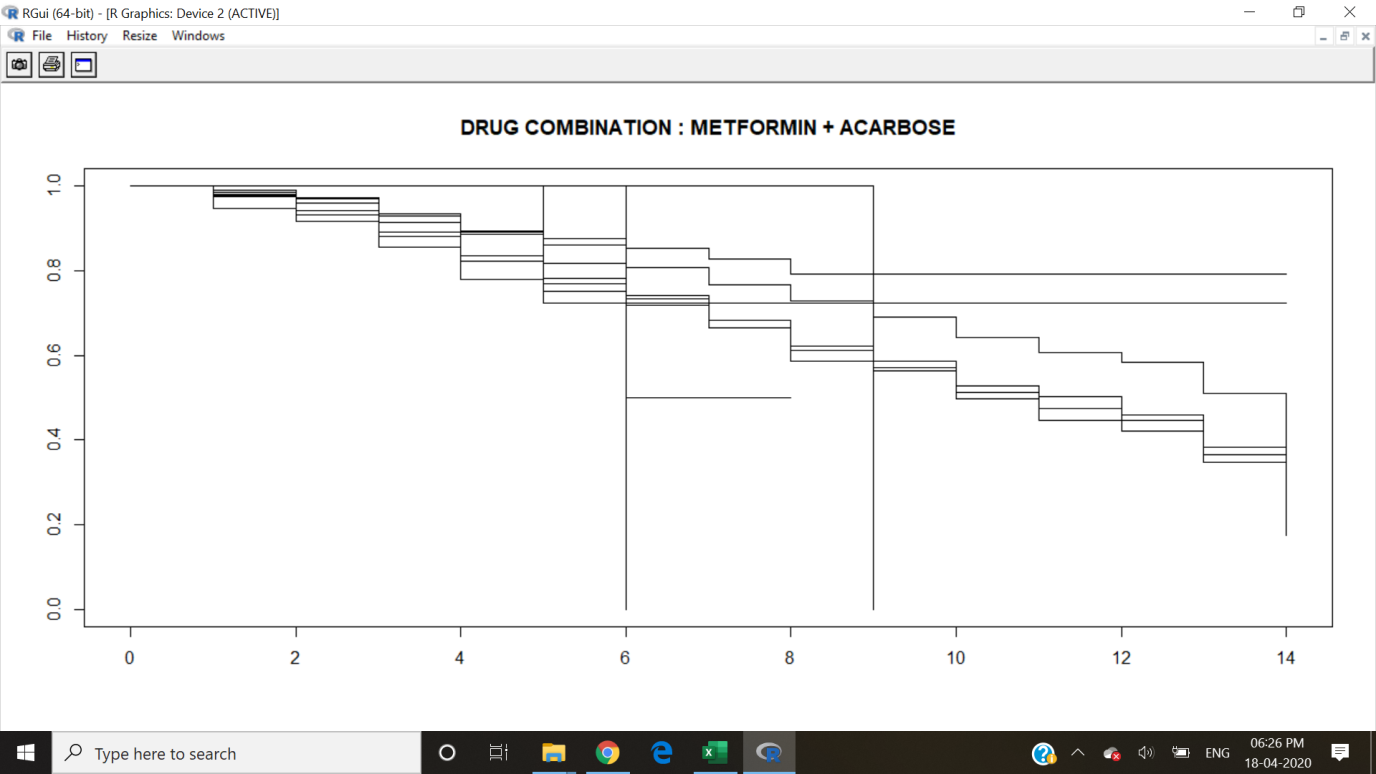
****

Figure 9.5 Acarbose + Metaformin

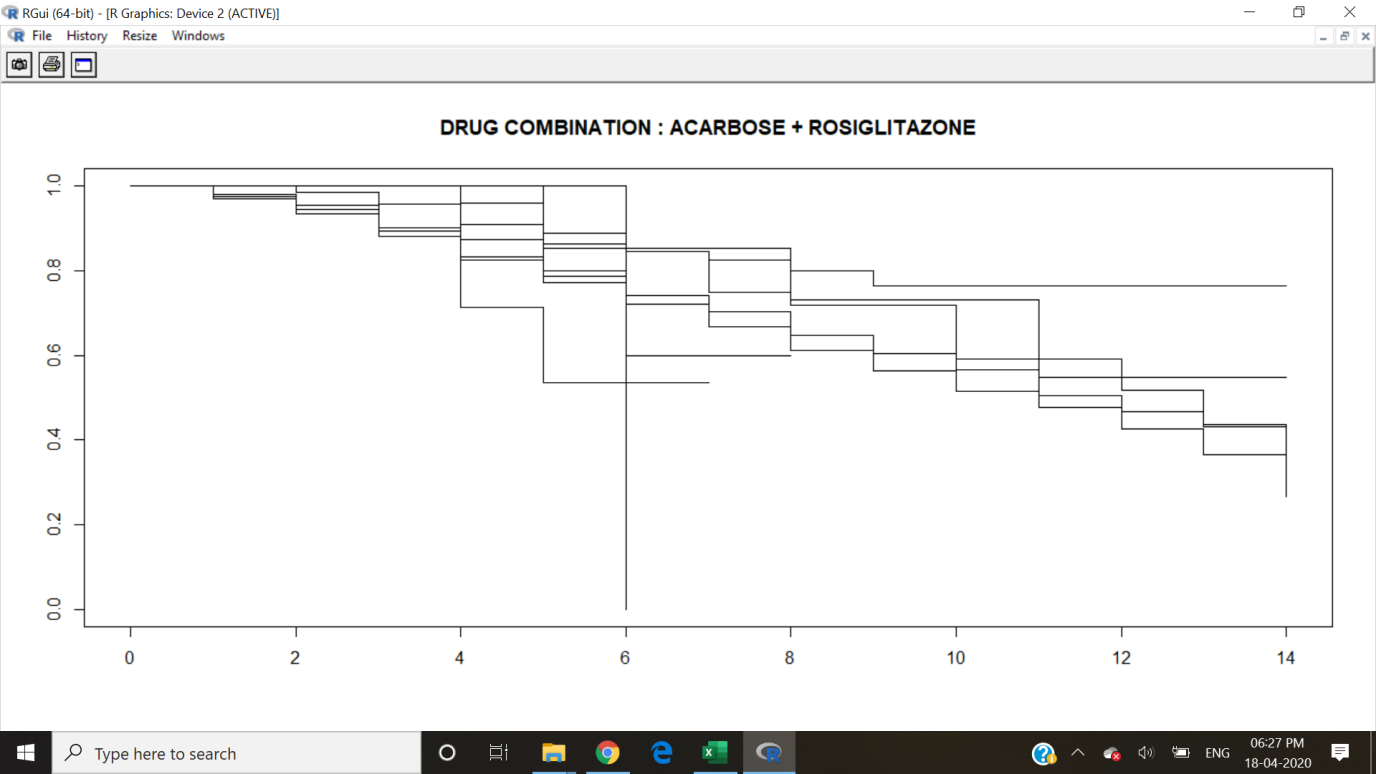
****

Figure 9.6 Acarbose + Rosiglitazone

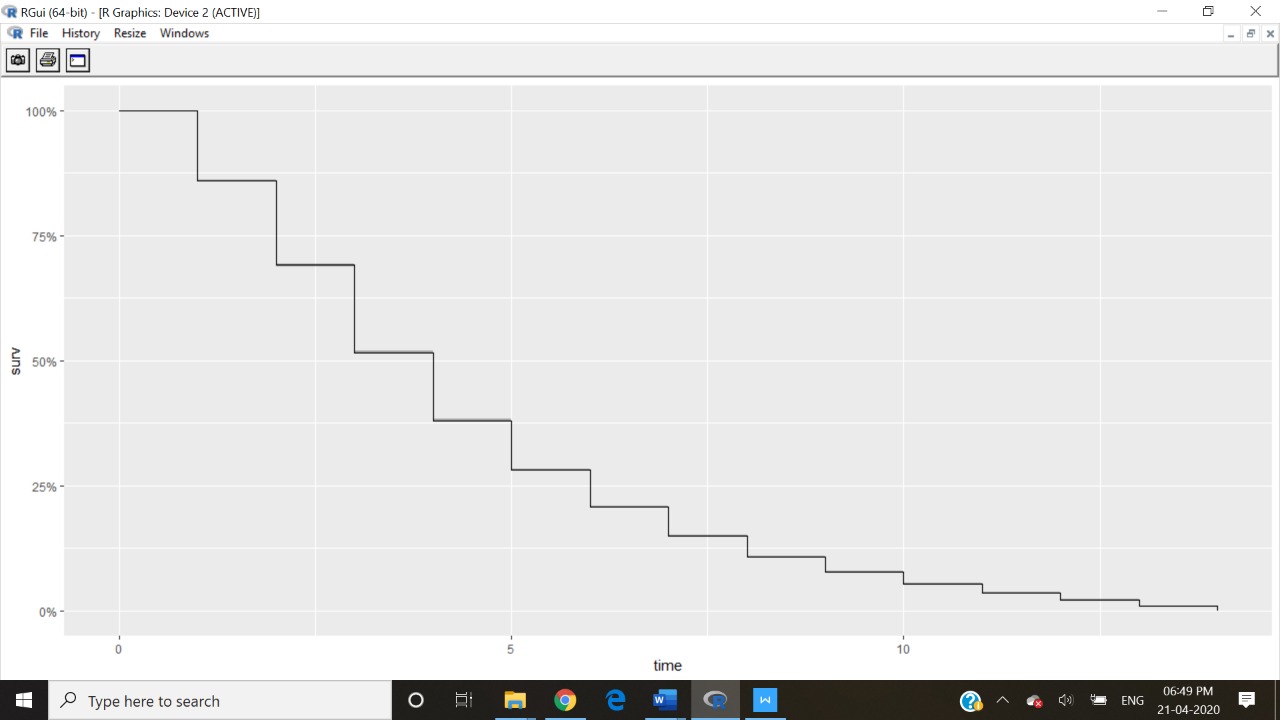
****

Figure 9.7 Cox proportional fit for Acarbose

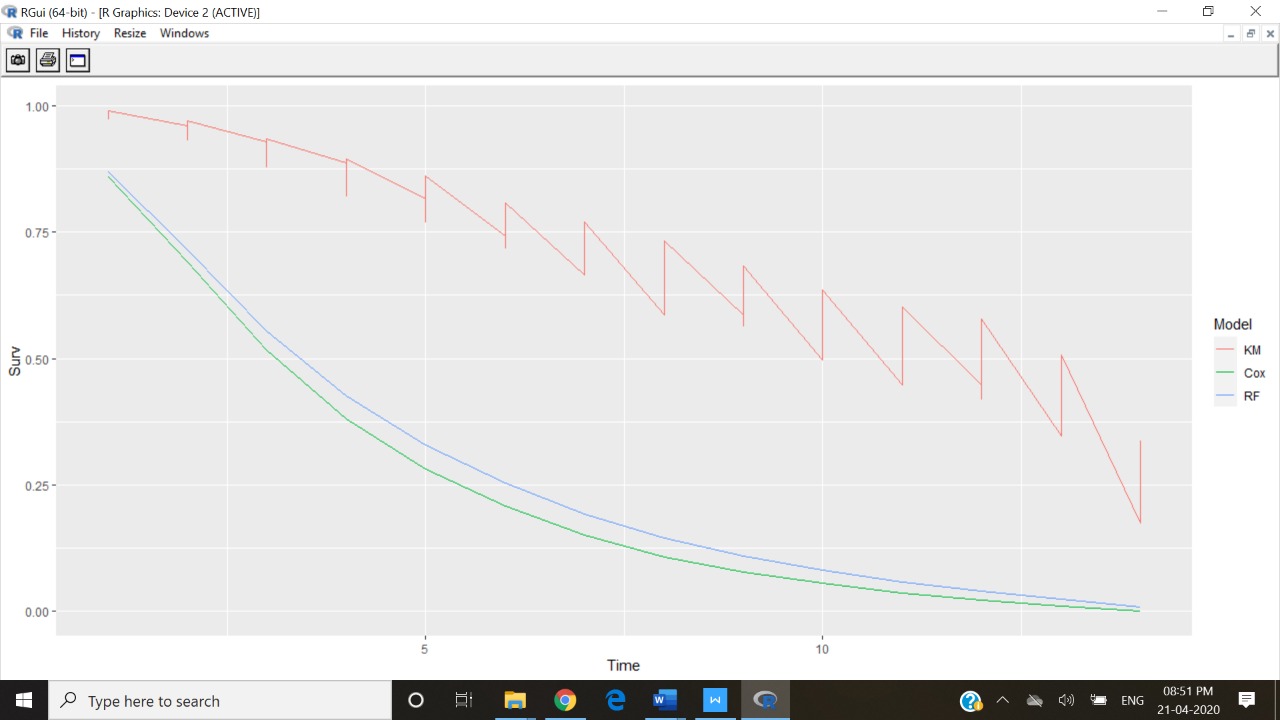
****

Figure 9.8 Survival analysis with classification for Acarbose

1. **CONCLUSION & FUTURE WORK :**

We check whether readmission is required or not and check the efficiency of survival analysis models and models combined with classification techniques like random forest method .The decision to obtain the process of survival analysis for patients with diabetes is a useful predictor of readmission rates which may prove valuable to reduce readmission rates and costs for the care of individuals with diabetes .The major diagnosis for patients is well known before readmission. The project can be much more efficient by using classification methods like Naive Bayes method or Kernel estimation .

1. **REFERENCES :**

[1].Matteo Gagliolo and Catherine Legrand <https://www.researchgate.net/publication/202051286_Algorithm_Survival_Analysis>

[2].Data flair team

<https://data-flair.training/blogs/survival-analysis-in-r/>

[3].This article was contributed by [Perceptive Analytics](https://www.perceptive-analytics.com/data-analytics-companies/" \t "https://www.r-bloggers.com/steps-to-perform-survival-analysis-in-r/_blank). Madhur Modi, Chaitanya Sagar

<https://www.r-bloggers.com/steps-to-perform-survival-analysis-in-r/>

[4].GITHUB<https://github.com/prabhuvashwin/Data-Analysis-on-Hospital-Readmission-Data>

[5].Reinaldo Zezela, MSc student Big Data Analytics, University of Derby

<https://rpubs.com/rzezela77/346228>

[6].IEEE<https://ieeexplore.ieee.org/abstract/document/8748992>

[7].<https://westminsterresearch.westminster.ac.uk/>

[8].Author [BichenZheng](https://www.sciencedirect.com/science/article/abs/pii/S0957417415003085?via=ihub" \l "!) [JingheZhang](https://www.sciencedirect.com/science/article/abs/pii/S0957417415003085?via=ihub" \l "!) [Mohammad Khasawneh](https://www.sciencedirect.com/science/article/abs/pii/S0957417415003085?via=ihub" \l "!)

[https://www.sciencedirect.com/science/article/abs/pii/S0957417415003085?via%3Dihub](https://www.sciencedirect.com/science/article/abs/pii/S0957417415003085?via=ihub)

[9].[Hyunyoung Baek](https://www.ncbi.nlm.nih.gov/pubmed/?term=Baek%20H%5bAuthor%5d&cauthor=true&cauthor_uid=29652932), Conceptualization, Writing – original draft, <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5898738/>

[10].OECD, *Average Length of Stay in Hospitals. Health at a Glance 2011*, OECD Indicators, OECD Publishing, <https://www.hindawi.com/journals/jhe/2016/7035463/>