Software Requirements Specification

for

Loan Approval Prediction

Version 1.0 approved

Prepared by Group

Makerere University, College of Computing and Informatics Sciences, Software Engineering

1st July 2019

Table of Contents

Table of Contents ii

1. Introduction 1

1.1 Purpose 1

1.2 Document Conventions 1

1.3 Intended Audience and Reading Suggestions 1

1.4 Product Scope 1

1.5 References 1

2. Overall Description 2

2.1 Product Perspective 2

2.2 Product Functions 2

2.3 User Classes and Characteristics 2

2.4 Operating Environment 3

2.5 Design and Implementation Constraints 3

2.6 User Documentation 3

2.7 Assumptions and Dependencies 3

3. External Interface Requirements 3

3.1 User Interfaces 4

3.2 Hardware Interfaces 5

3.3 Software Interfaces 5

3.4 Communications Interfaces 5

4. System Features 7

4.1 Login 8

4.2 View Amalyzed Data 8

4.2 Application Details 8

5. Other Requirements 9

Appendix A: Glossary 9

Appendix B: Analysis Models 9

# Introduction

## Purpose

The purpose of this document is to present a detailed description of the Loan Approval prediction system. It will explain the purpose, features, interfaces, what the system will do and the constraints under which it must operate.

## Document Conventions

This document uses the Modern Language Association (MLA) format. Section headings indicated by whole numbers (1, 2, 3 ...) and sub-section headings numbered basing on the whole numbers their section headings such as 1.3 belongs to section 1 are written with a bold-faced text for emphasis. Italic text is used to label diagrams.

## Intended Audience and Reading Suggestions

This document is exclusively for the managing authority of Bank/finance company, whole process of prediction is done privately no stakeholders would be able to alter the processing.

## Product Scope

The system receives user loan application data from a form, processes it using the logistic regression model to determine user eligibility for a loan.

It predicts loan application eligibility based on the information from the train dataset through which the prediction model is generated.

## Reference

IEEE Std 1016 2009 (first version 1998)

# Overall Description

## Product Perspective

With the enhancement in the banking sector lots of people are applying for bank loans but the bank has its limited assets which it has to grant to limited people only, so finding out to whom the loan can be granted which will be a safer option for the bank is a typical process. So, this system tries to reduce this risk factor behind selecting the safe person so as to save lots of bank efforts and assets. This is done by mining the Big Data of the previous records of the people to whom the loan was granted before and on the basis of these records/experiences the machine was trained using the machine learning model which give the most accurate result. The main objective of this system is to predict whether assigning the loan to particular person will be safe or not.

The training data set is now supplied to machine learning model, on the basis of this data set the model is trained. Every new applicant’s details filled at the time of application form acts as a test data set. After the operation of testing, model predict whether the new applicant is a fit case for approval of the loan or not, based upon the inference it concludes on the basis of the training data sets.

## Product Functions

Major functions of the product

* Read the previously stored bank customers’ data into the system provided by the Data analyst as a train dataset.
* Graphically visualize the entered train dataset to clearly show data relationships.
* Upload new bank customer data to determine loan eligibility.
* Provide results output to the system user, whether loan has been approved or not.

## User Classes and Characteristics

* **Bank Customers**

These are the people who apply for bank loans. They are either new bank customers or old bank customers. An old bank customer must update his/her client information during the application whereas a new bank customer must create an account with the bank providing relevant information for requested loan.

* **Management**

Management accepts or rejects requested approved loans depending on present or future circumstances (bank and/or customer circumstances).

## Operating Environment

The system will operate effectively on a computer whose RAM is at least 2GB with an average processing speed of at least 1.6 GHz. The computer should have an operating system of either windows operating system of windows 7 and above or Linux on which R and RStudio has been installed. The system interface will require an up to date browser such as Chrome version: 59.0.3071 and above, Firefox version 52.0 and above.

## Design and Implementation Constraints

The system is constrained by a number of factors as highlighted below: -

1. Technology Constraints

The system will require use of web-based server technologies such as R Shiny Server, Statistical analysis, visualization tools such as the R language and it will not have a database.

1. Hardware Constraints:

The system will be supported by a computer system having a minimum of 1GB – 2GB RAM, 1-2 Cores (1.6 GHz), 1-2 GB Disk space (hard disk) and configuration for better performance 2GB – 4GB RAM.

1. Technology Constraints

The system will require use of web-based server technologies such as R Shiny Server, Statistical analysis, visualization tools such as the R language and it will not have a database.

## User Documentation

The documents below will be delivered with the system software.

1. System Design Document
2. System Report

## Assumptions and Dependencies

The system is developed in R using Rstudio and therefore requires R to be installed on the user’s computer. The user must have all the required R packages for the system to run. It requires the latest R version, and this applies to both windows and linux users.

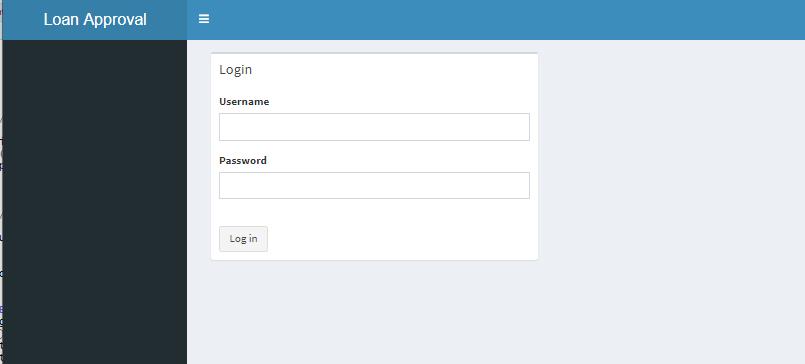
The chances of getting a loan is higher for:

* Applicants having a credit history.
* Applicants with higher applicant and co-applicant incomes
* Applicants with higher education level
* Properties in urban areas with high growth perspectives

# External Interface Requirements

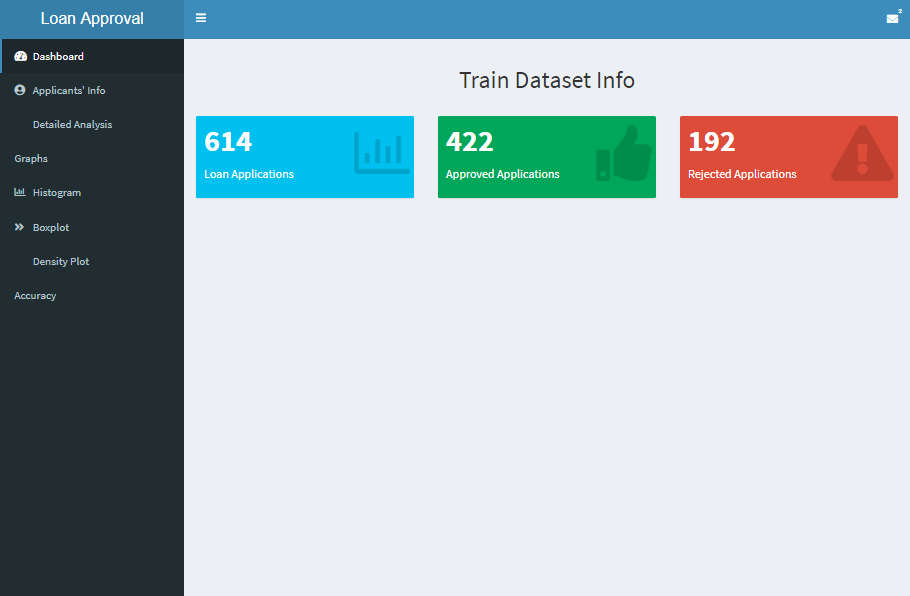
## User Interfaces

The login page where users enter their credentials so as to access the system components.



Users of the system can access the system functionality after verification of their credentials via the login page

The system dashboard



The different system functionalities of the system module can be accessed via the dashboard

## Hardware Interfaces

The minimum hardware requirements of the system are 1GB – 2GB RAM, 1-2 Cores (1.6 GHz), 1-2 GB SATA (hard disk) and configuration for better performance 2GB – 4GB RAM. A compatible graphics card is required for efficient visualization. A system with these specifications can handle a Network of approximately 1000 edges and nodes. For bigger networks, additional memory is required

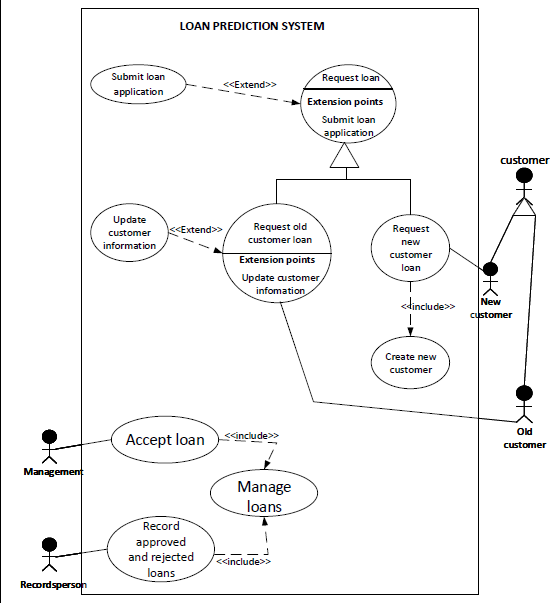
## Software Interfaces

The system requires R and RStudio to be installed on the system, more specifically R version 3.4.0 and above for its latest release. Additional information can be found in Section 2.7 of this document.

## Communications Interfaces

The system requires an internet connection to install new packages, update already installed ones and update some of its components (APIs, modules etc.), and verification of users. Additional information can be found in Section 2.5 of this document.

# System Features



|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Name | Goal | Input | Output | Main scenario | Pre-condition |
| Login | User verification | Account credentials | System access | Validation |  |
| View analyzed data | Visualization | Developer Survey data file | Analyzed results | Analysis | Knowledge about visualization tools |
| Input Application Data | For computation | Loan Application | Eligibility Predictions | Computations |  |

# Other Requirements

# Appendix A: Glossary

1. Graph: A system of nodes connected in pairs by edges. Often sub-divided into directed graphs or undirected graphs according to whether the edges have an orientation or not. Mixed graphs include both types of edges.
2. Node: A node is (together with edges) one of the two basic units out of which graphs are constructed. Nodes of graphs are often considered to be atomic objects, with no internal structure.
3. Edge: An edge is (together with vertices) one of the two basic units out of which graphs are constructed. Each edge has two (or in hyper graphs, more) vertices to which it is attached, called its endpoints.

## Appendix B: Analysis Models

1. Logistic Regression is a predictive modelling algorithm that is used when the Y variable is binary categorical. That is, it can take only two values like 1 or 0. The goal is to determine a mathematical equation that can be used to predict the probability of event 1. Once the equation is established, it can be used to predict the Y when only the X’s are known.