**Comparative Analysis of Loan Status Prediction using different Machine Learning Algorithms**

A MINI PROJECT

*Submitted by*

**AKSHIT CHAUDHARY (RA2011027010106)**

**SAI SANJANA (RA2011027010109)**

**SIDDHANT SHEKHAR (RA2011027010130)**

*Under the guidance of*

**Dr. E. Sasikala**

**Professor**

**Department of Data Science and Business Systems**

In partial fulfilment for the

Course of

**18CSE392T- Machine Learning-I**

in

**Department of Data Science and Business Systems**



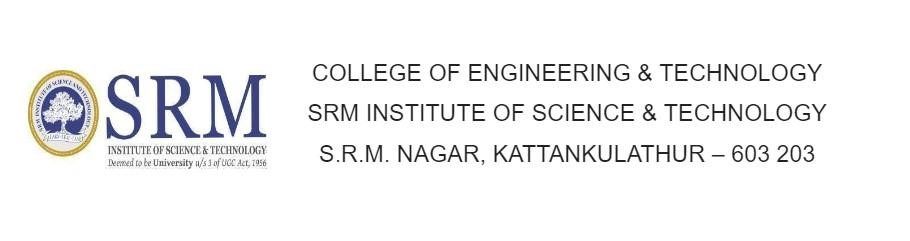
**SCHOOL OF COMPUTING**

**COLLEGE OF ENGINEERING AND TECHNOLOGY**

**SRM INSTITUTE OF SCIENCE AND TECHNOLOGY**

**KATTANKULATHUR – 603203**

**October 2023**



# BONAFIDE CERTIFICATE

Certified that this mini project report "**Online Flight Reservation System**" is the Bonafide work of **Akshit Chaudhary (RA2011027010106), Sai Sanjana (RA2011027010109)** and **Siddhant Shekhar (RA2011027010130)** who carried out the project work under my supervision.

|  |  |
| --- | --- |
| Dr. E. Sasikala | Dr. M Lakshmi |
| Professor | Professor & HOD |
| Department of Data Science and Business Systems | Department of DSBS |
| SRM institute of science and technology | SRM institute of science and technology |

# ABSTRACT

Technology has boosted the existence of humankind and the quality of life they live. Every day we are planning to create something new and different. We have a solution for every other problem. We have machines to support our lives and make us somewhat complete in the banking sector candidate gets proofs/ backup before approval of the loan amount. The application approved or not approved depends upon the historical data of the candidate by the system. Every day lots of people apply for the loan in the banking sector but Bank would have limited funds. In this case, the right prediction would be very beneficial using some class-function algorithm. An example the logistic regression, random forest classifier, support vector machine classifier, etc. A Bank's profit and loss depend on the amount of the loans, that is whether the Client or customer is paying back the loan. Recovery of loans is the most important for the banking sector. The improvement process plays an important role in the banking sector. The historical data of candidates was used to build a machine learning model using different classification algorithms. The main objective of this paper is to predict whether a new applicant granted the loan or not using machine learning models trained on the historical data set.

**TABLE OF CONTENTS**

|  |  |  |  |
| --- | --- | --- | --- |
| **CHAPTER NO.** | | **TITLE** | **PAGE NO.** |
|  | | **ABSTRACT** | 3 |
| **TABLE OF CONTENTS** | 4 |
| **LIST OF FIGURES** | 5 |
| **ABBREVIATIONS** | 6 |
| 1. |  | **INTRODUCTION** |  |
|  | 1.1 | Aim, Synopsis | 7 |
|  | 1.2 | Software Requirements Specification | 8 |
| 2. |  | **LITERATURE SURVEY** |  |
|  | 2.1 | Literature Review | 9 |
| 3. |  | **SYSTEM ARCHITECTURE AND DESIGN** |  |
|  | 3.1 | Architecture Diagram | 10 |
|  | 3.2 | ER Diagram | 11 |
|  | 3.3 | Use case Diagram | 12 |
|  | 3.4 | Front end (UI)design | 13 |
|  | 3.5 | Back end (Database) design | 14 |
| 4. |  | **MODULES AND FUNCTIONALITIES** | 17 |
|  | 4.1 | Functional Requirements | 18 |
|  | 4.2 | Non-Functional Requirements | 19 |
| 5. |  | **CODING AND TESTING** | 22 |
| 6. |  | **RESULTS AND DISCUSSION** | 25 |
|  | 6.1 | Front End Screenshots | 26 |
|  | 6.2 | Back End Screenshots | 35 |
| 7. |  | **CONCLUSION AND FUTURE ENHANCEMENT** | 39 |
| 8. |  | **REFERENCES** | 40 |

LIST OF FIGURES

|  |  |  |
| --- | --- | --- |
| **Figure No.** | **Figure Name** | **Page No** |
| 3.1 | Architecture Diagram | **10** |
| 3.2 | Use case Diagram | **12** |
| 3.3 | ER Diagram | **13** |

**ABBREVIATIONS**

**CSS** Cascading Style Sheet

**DB** Data Base

**ER** Entity Relationship

**SQL** Structured Query Language

**HTML** Hyper Text Markup Language

**UI** User Interface

# OBJECTIVE

**Aim:**

To determine the loan approval system using machine learning algorithms.

**Synopsis:**

Loan approval is a very important process for banking organizations. The systems approved or rejected the loan applications. Recovery of loans is a major contributing parameter in the financial statements of a bank. It is very difficult to predict the possibility of payment of loan by the customer. In recent years many researchers have worked on loan approval prediction systems. Machine Learning (ML) techniques are very useful in predicting outcomes for large amount of data. In this paper different machine learning algorithms are applied to predict the loan approval of customers. In this paper, various machine learning algorithms that have been used in past are discussed and their accuracy is evaluated. The focus of this paper is to determine whether the loan given to a particular person, or an organization shall be approved or not.

**HARDWARE AND SOFTWARE SPECIFICATIONS**

**HARDWARE REQUIREMENTS:**

* LA device (Compotes laptop)
* Memory (RAM) Mini 2GB RAM
* Processor M 1GHZ Recommended 2G1Z or more.
* Hard disk-40 GB Recommended 64 GB or more.
* Ethernet connection (LAN) or a wireless adapter (Wi-Fi)

**SOFTWARE REQUIREMENTS:**

* A database like DBMS to store the list of authors and articles.
* A web browser like Chrome, Mozilla Firefox etc.
* Operating System - Windows, Linux, macOS 32 bit and 64 bits

**LITERATURE REVIEW**

In the domain of loan prediction using machine learning models, a comprehensive literature survey revealed several noteworthy contributions. In the paper titled "Loan Prediction by using Machine Learning Models" (2017) by P Supriya and M Pavani, the authors focused on data collection and preprocessing, leveraging machine learning models, and implementing training and testing modules. During preprocessing, they paid particular attention to outlier detection and removal, as well as imputation processes. Their model, employing gradient boosting techniques, aimed to predict loan approval outcomes, following an 80:20 dataset split. The Decision Tree model stood out with an impressive accuracy of 81.1%.

Another significant study, "Credit Risk Analysis and Prediction Modelling of Bank Loans Using R" (2019) by Sudhamathy G., introduced a risk analysis approach for loan sanctioning. The research encompassed data selection, preprocessing, feature extraction and selection, model construction, prediction, and evaluation. The dataset from the USI repository was employed, and the Logistic Regression classifier was built after meticulous preprocessing. This method achieved a precision of 83.3%.

In the paper "Developing Prediction Model of Loan Risk in Banks using Data Mining" (2020) by Jafar Hamid and Tarig Mohammed Ahmed, three algorithms, namely j48, Bayes Net, and Naive Bayes, were employed to construct predictive models for classifying loan applications as good or bad based on customer behavior and past credit repayment history. The Weka application was used to develop the model, and it was observed that the j48 algorithm outperformed others in terms of accuracy and mean absolute error, making it the preferred choice.

Furthermore, "Loan Prediction Using Ensemble Technique" (2020) by Anchal Goyal and Ranpreet Kaur introduced an ensemble model incorporating eleven machine learning models with nine distinct properties for predicting credit risk in loan applications. This work aimed to evaluate model accuracy, employing parameters like Accuracy, Gini, AUC, Roc, and others while assessing different training algorithms. Real Coded Genetic Algorithms were used to calculate feature importance, facilitating credit risk prediction. The K-fold validation method was employed to ensure the robustness of the predictive model, achieving a maximum accuracy of 81.25% with the Tree model for genetic algorithm. These papers collectively contribute valuable insights into the field of loan prediction and offer diverse methods for improving loan approval and risk assessment processes.

**ARCHITECTURE DIAGRAM**

**A diagram of a computer

Description automatically generated**

**USE CASE DIAGRAM**

**A diagram of a machine learning

Description automatically generated**

**ER DIAGRAM**

**A diagram of a machine learning

Description automatically generated**

**FRONT-END LANGUAGES USED**

● HTML Hypertext Mark-up Language (HTML) is the standard mark-up language for creating web pages and web applications. With Cascading Style Sheets (CSS) and JavaScript it forms a triad of cornerstone technologies for the World Wide Web. Web browsers receive HTML documents from a web server or from local storage and render them into multimedia web pages. HTML describes the structure of a web page semantically and originally included cues for the appearance of the document. HTML elements are the building blocks of HTML pages. With HTML constructs, images and other objects, such as forms, may be embedded into the rendered page. It provides a means to create structured documents by denoting structural semantics for text such as headings, paragraphs, lists, links, quotes a nd other elements are delineated by tags, written using angle brackets. Tags such as surround and provide information about document text and may include other tags as sub-elements. Browsers do not display the HTML tags, but use them to interpret the content of the page.

● HTML is used to define the structure of web pages, including text, images, videos, and links. It uses a set of markup tags, also known as elements, to define the structure and content of the page. HTML tags are enclosed in angle brackets and are typically used in pairs, with the opening tag indicating the start of an element and the closing tag indicating the end.

● HTML includes a range of elements, such as headings, paragraphs, lists, tables, and forms. It also provides support for multimedia elements, such as images, audio, and video, and allows for the inclusion of scripts and stylesheets for enhanced functionality and styling.

● HTML is widely used by web developers to create static and dynamic web pages. It is constantly evolving, with new versions and features being released regularly. HTML5 is the latest version of HTML, which provides a range of new features and functionality, including support for multimedia elements and enhanced accessibility.

• CSS Cascading Style Sheets (CSS) is a style sheet language used for describing the presentation of a document written in a mark-up language. Although most often used to set the visual style of web pages and user interfaces written in HTML and XHTML, the language can be applied to any XML document, including plain XML, SVG and XUL, and is applicable to rendering in speech, or on other media. Along with HTML and JavaScript, CSS is a cornerstone technology used by most websites to create visually engaging webpages, user interfaces for web applications, and user interfaces for many mobile applications

● CSS is designed primarily to enable the separation of presentation and content, including aspects such as the layout, colors, and fonts. This separation can improve content accessibility, provide more flexibility and control in the specification of presentation characteristics, enable multiple HTML pages to share formatting by specifying the relevant CSS in a separate .CSS file, and reduce complexity and repetition in the structural content.

● CSS, which stands for Cascading Style Sheets, is a style sheet language used to describe the presentation of web pages. It is used to control the layout, formatting, and appearance of web pages, including fonts, colors, spacing, and positioning.

● CSS is used in conjunction with HTML and JavaScript to create responsive and visually appealing web pages. CSS styles are typically defined in separate files or within the HTML file using the "style" tag. They are applied to HTML elements using selectors, which target specific elements or groups of elements on the page.

● CSS includes a wide range of features, such as layout and positioning of elements, typography, and color management. It also provides support for responsive design, allowing web pages to adapt to different screen sizes and devices.

● CSS is widely used by web developers to create visually appealing and user-friendly web pages. It is supported by all modern web browsers and is constantly evolving to include new features and functionality.

**BACK-END LANGUAGES USED**

• PHP MY ADMIN PhpMyAdmin is a free and open-source web-based tool written in PHP used to manage MySQL and MariaDB databases. It provides an intuitive graphical user interface that allows users to perform various tasks related to database management, such as creating and deleting databases, tables, and fields, inserting, updating, and deleting data, running SQL queries, and managing user accounts and permissions. PhpMyAdmin offers a range of features, including support for multiple languages, import and export of data in various formats, visual representation of database relationships, and the ability to create and execute SQL queries directly from the interface. It also includes a set of security features, such as the ability to restrict access to specific databases and tables, and an audit log that records all actions performed by users. PhpMyAdmin is widely used by web developers, system administrators, and database administrators for managing and maintaining MySQL and MariaDB databases. It is available for download from the official website and can be installed on various web servers, including Apache and Nginx.

• MY SQL We have used, in this project, MySQL which is an open-source relational database management system. MySQL is a central component of the LAMP open-source web application software stack (and other "AMP" stacks). LAMP is an acronym for "Linux, Apache, MySQL, Perl/PHP/Python". Applications that use the MySQL database include: TYPO3, MoDx, Joomla, WordPress, phpBB, MyBB, and Drupal. MySQL is also used in many high-profile, large-scale websites, including Google (though not for searches), Facebook, Twitter, Flickr, and YouTube.

● MySQL is a popular open-source relational database management system (RDBMS) used to store, manage and retrieve data. It is widely used for web-based applications, including content management systems, e-commerce platforms, and online forums.

● MySQL supports various platforms, including Windows, Linux, and macOS, and provides a range of features, such as support for multiple databases, multi-threading, and indexing for efficient data retrieval. It also includes a set of programming interfaces and tools, such as JDBC, ODBC, and MySQL Workbench, to facilitate application development and database management.

● MySQL uses a structured query language (SQL) for managing data, which allows users to create, modify, and retrieve data in a structured and efficient manner. It also supports transactions and provides various security features, such as encryption and authentication, to ensure the security of data stored in the database.

**MODULES**

⮚ Dataset collection

⮚ Machine Learning Algorithm

⮚ Prediction

**MODULE EXPLANATION:**

*Dataset collection:*

Dataset is collected from kaggle.com. That dataset has some value like gender, marital status, self-employed or not, monthly income, etc. Dataset has the information, whether the previous loan is approved or not depends on the customer information. That data will be preprocessed and proceed to the next step.

*Machine learning Algorithm:*

In this stage, the collected data will be given to the machine algorithm for the training process. We use multiple algorithms to get a high accuracy range of prediction. A preprocessed data set is processed in different machine learning algorithms. Each algorithm gives some accuracy level. Each one is undergoing for the comparison.

✔ Logistic Regression

✔ Random Forest Classifier

✔ Decision Tree Classifier

✔ SVM

*Prediction:*

Preprocessed data are trained, and input given by the user goes to the trained dataset. The Logistic Regression trained model is used to predict and determine whether the loan given to a particular person shall be approved or not.

**Design and Implementation Constraints**

**Constraints in Analysis**

♦ Constraints as Informal Text

♦ Constraints as Operational Restrictions

♦ Constraints Integrated in Existing Model Concepts

♦ Constraints as a Separate Concept

♦ Constraints Implied by the Model Structure

**Constraints in Design**

♦ Determination of the Involved Classes

♦ Determination of the Involved Objects

♦ Determination of the Involved Actions

♦ Determination of the Require Clauses

♦ Global actions and Constraint Realization

**Constraints in Implementation**

A hierarchical structuring of relations may result in more classes and a more complicated structure to implement. Therefore, it is advisable to transform the hierarchical relation structure to a simpler structure such as a classical flat one. It is rather straightforward to transform the developed hierarchical model into a bipartite, flat model, consisting of classes on the one hand and flat relations on the other. Flat relations are preferred at the design level for reasons of simplicity and implementation ease. There is no identity or functionality associated with a flat relation. A flat relation corresponds with the relation concept of entity-relationship modeling and many object-oriented methods.

**Other Nonfunctional Requirements**

**Performance Requirements**

The application at this side controls and communicates with the following two main general components.

⮚ embedded browser in charge of the navigation and accessing to the web service.

⮚ Server Tier: The server side contains the main parts of the functionality of the proposed. architecture. The components at this tier are the following.

Web Server, Security Module, Server-Side Capturing Engine, Preprocessing Engine, Database System, Verification Engine, Output Module.

**Safety Requirements**

1. The software may be safety critical. If so, there are issues associated with its integrity level.

2. The software may not be safety-critical although it forms part of a safety-critical system. For example, software may simply log transactions.

3. If a system must be of a high integrity level and if the software is shown to be of that integrity level, then the hardware must be at least of the same integrity level.

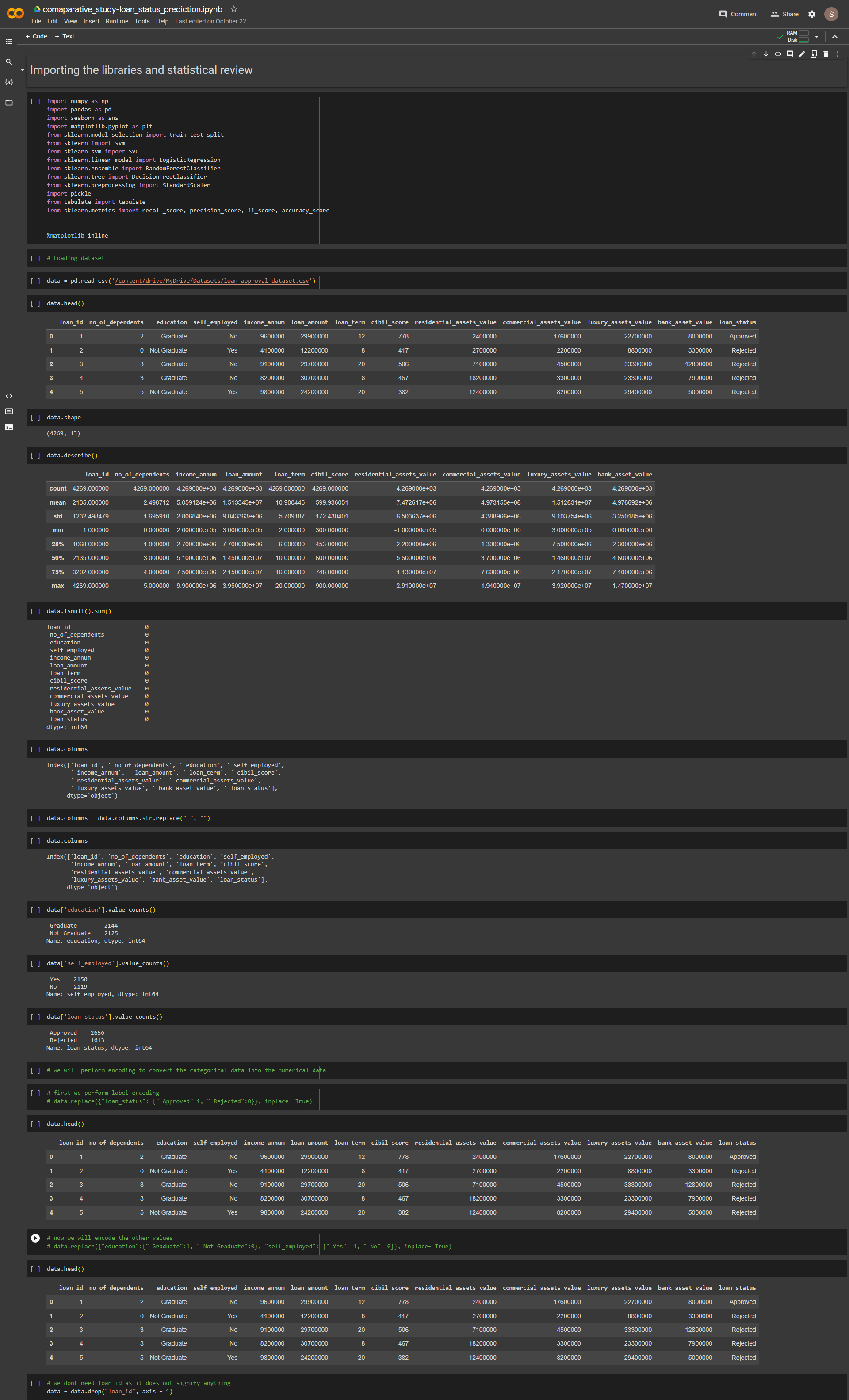
4. There is little point in producing 'perfect' code in some language if hardware and system software (in the widest sense) are not reliable.

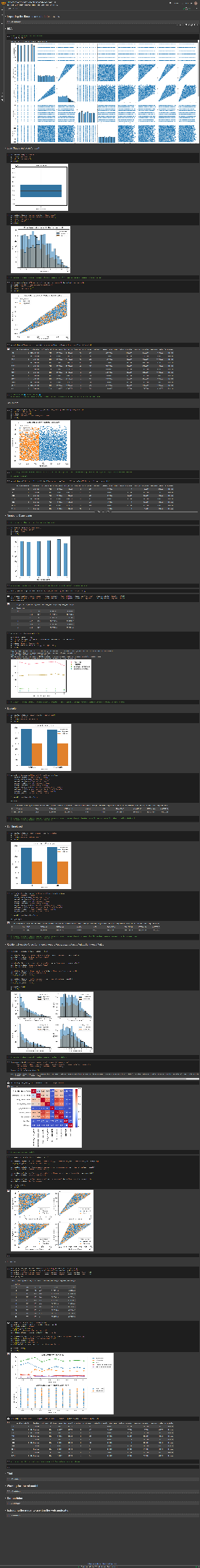
5. If a computer system is to run software of a high integrity level, then that system should not at the same time accommodate software of a lower integrity level.

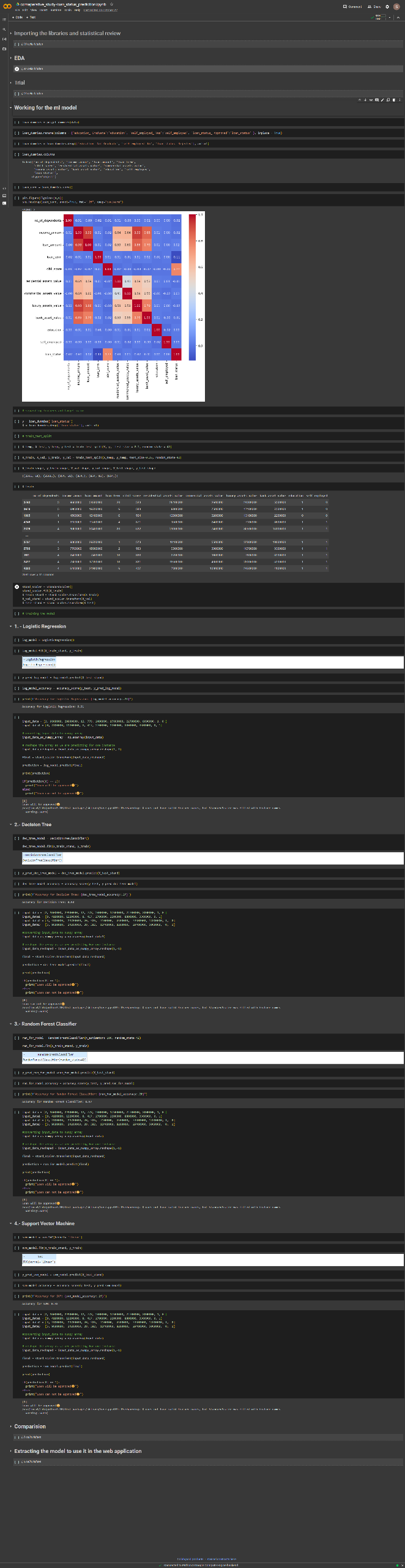
6. Systems with different requirements for safety levels must be separated.

7. Otherwise, the highest level of integrity required must be applied to all systems in the same environment.

**SOURCE CODE**

****

****

****

**Front-End**

**RESULT AND DISCUSSION**

* + This project has something like cancelation of tickets are missing but it can surely help with the airline reservation system.
  + In near future there can be upgrades for flight cancelation also which make this project more beneficial
  + We hope this project will work fine & in near future we will do more work to improve this project. Major work on project UI will also be done.

**CONCLUSION AND FUTURE WORK**

* + This project has something like cancelation of tickets are missing but it can surely help with the airline reservation system.
  + In near future there can be upgrades for flight cancelation also which make this project more beneficial
  + We hope this project will work fine & in near future we will do more work to improve this project. Major work on project UI will also be done.

**REFERENCES**

1. <https://ijisrt.com/wp-content/uploads/2018/04/Implementation-of-mysql-database>

2. <https://ijiet.com/wp-content/uploads/2014/12/4.pdf>

3. <https://www.researchgate.net/figure/The-architecture-of-proposed-patient-appointmentscheduling-system_fig1_249643353>

4. <https://app.creately.com/d/IIB3y7rSN7n/edit>

5. <https://www.javatpoint.com/phpmyadmin>