

CREDIT-WORTHINESS DETERMINANT MODEL (USING DECISION TREE ALGORITHM)

TABLE OF CONTENTS

Title Page

Certification

Dedication

Acknowledgments

Abstract

Table of Contents

List of Tables

List of Figures

CHAPTER ONE: GENERAL INTRODUCTION

- 1.1 Background to the Study
- 1.2 Statement of the Problem
- 1.3 Aim and Objectives of the Study
- 1.4 Significance of the Study
- 1.5 Scope of the Study
- 1.6 Organization of the Study

CHAPTER TWO: REVIEW OF RELATED PAST WORK

- 2.1 Review of Related Works
- 2.2 Loan System Credibility Prediction System using Decision Tree Algorithm

CHAPTER THREE: RESEARCH METHODOLOGY AND ANALYSIS

- 3.1 Research Methodology
- 3.2 Analysis of the Existing System

- 3.3 Problems of the Existing System
- 3.4 Analysis of the Proposed System
- 3.5 Advantages of the New System Over the Existing System

CHAPTER FOUR: DESIGN IMPLEMENTATION AND DOCUMENTATION OF THE SYSTEM

- 4.1 Design of the System
 - 4.1.1 Output Design
 - 4.1.2 Input Design
 - 4.1.3 Database Design
 - 4.1.4 Procedure Design
- 4.2 System Implementation
 - 4.2.1 Choice of Programming Language
 - 4.2.2 Hardware Support
 - 4.2.3 Software Support
 - 4.2.4 Implementation Techniques
- 4.3 System Documentation
 - 4.3.1 Program Documentation
 - 4.3.2 Operation of the System
 - 4.3.3 Maintaining the System

CHAPTER FIVE: SUMMARY, CONCLUSION AND RECOMMENDATION

- 5.1 Summary
 - 5.2 Experience Gained and Problems Encountered
 - 5.3 Conclusions
 - 5.4 Recommendations
- References

APPENDICES

Program Flowchart of the Proposed System

Source Program Listing

ABSTRACT

With the enhancement in the banking sector, very many 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 in this paper, we try to reduce the 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 gives the most accurate result. The main objective of this paper is to predict whether assigning the loan to a particular person will be safe or not. This paper is divided into four sections (i) Data collection (ii) Comparison of machine learning models on collected data (iii) Training of systems on most promising model and (iv) Testing.

Keywords: Credit-worthiness, Decision Tree Algorithm, Machine Learning, Predictive Model, Credit Risk, Classification, Financial Assessment, Loan Management

CHAPTER ONE

INTRODUCTION

1.1 BACKGROUND TO THE STUDY

Nowadays, banks struggle a lot to get an upper hand over each other to enhance overall business due to tight schedules. Banks have recognized that customer retention and fraud prevention are critical strategies in today's competitive environment (Hasheminejad & Khorrami, 2018). Availability of the huge quantity of data, creation of knowledge base and efficient utilization of the same have helped banks to open up efficient delivery channels. Business decisions can be optimized through data mining (Islam & Habib, 2015). Customer segmentation, banking profitability, credit scoring and approval, predicting payment from customers, marketing, detecting fraud transactions, cash management and forecasting operations, optimizing stock portfolios and ranking investments are some of the areas where data mining techniques can be used in the banking industry (Al-Mashaqbeh et al., 2022).

Credit risks, which account for the risk of loss and loan defaults, are the major sources of risk encountered by the banking industry (Gerritsen, 2018). Data mining techniques like classification and prediction can be applied to overcome this to a great extent. There are mainly two objectives that are to be achieved through these techniques. They are: (i) Identification of the relevant attributes that signal the capacity of borrowers to pay back the loan (ii) Determining the best model(s) to evaluate credit risk. Decision Tree Algorithm is one of the best techniques to achieve this objective (Madan, 2016). The model being @! developed will provide a better credit risk assessment, which will potentially lead to a better allocation of the bank's capital.

In this regard, a study is conducted and an efficient prediction model, which helps to reduce the proportion of unsafe borrowers, is introduced herewith. Due to the significance of credit risk analysis, this study helps the banking industry by providing additional information to the loan decision-making process, potentially decreases the cost and time of loan

applications appraisal, and decreases the level of uncertainty for loan officers by providing knowledge extracted from previous loans. The Decision Tree Algorithm used in this model is the data mining technique for predicting credible customers.

1.2 STATEMENT OF THE PROBLEM

There are numerous areas in which data mining can be used in the banking industry to support customer relationships management. Data mining techniques will help distinguish borrowers who repay loans promptly from those who do not. It also helps to predict the credit worthiness of the borrower by analyzing the behaviour and reliability of the customers. With data mining techniques, banks can do a thorough profiling and ranking of their branches with respect to loan fraud risk. In developing countries like India, Bankers face more problems with the fraudsters. Using data mining techniques, it is simple to build a successful predictive model and visualize the report into meaningful information to the user. Data mining can be applied to reduce the risk associated with lending due to fraud as well as find an appropriate solution to the borrower's need for funds, with proper assessment of risk and the inclusion of sufficient control systems to ensure repayment. Loan officers are tasked with entering borrower's credit data while the system does risk computation and sends the result to the database, which is then assessed by the loan committee at head office. The decision is sent to the loan database which is then assessed by the loan officer. The officer then informs the customer about the decision. Those banks and retailers that have realized the utility of data mining and are in the process of building a data mining environment for their decision making process will get immense benefit and advantages in future. The problem statement is "Propose a data mining methodology to analyse, design and test efficient data mining framework for customer loan credibility prediction".

1.3 AIM AND OBJECTIVES

The aim of this research work is Credit-worthiness Determinant model using Decision Rule Algorithm. The objectives are to:

- i. develop a system that identifies relevant attributes that signal the capacity of borrowers to pay back the loan;
- ii. determine the best model(s) to evaluate credit risk; and

- iii. implement the system in C# as the programming language.

1.4 SIGNIFICANCE OF THE STUDY

The main importance of this system is that it helps financial institutions to recognize the importance of the information it has about its customers. Also, it helps to find patterns that distinguish borrowers who repay promptly from those who don't. This system helps to get rid of credit risks, which account for the risk of loss and loan defaults which are the major source of risk encountered by the banking industry.

1.5 SCOPE AND LIMITATION OF THE STUDY

The scope of this study focuses mainly on the development of a Credit-worthiness determinant model using Decision Tree Algorithm. It does not go beyond this.

Limitation of the Study

The limitations of this study include:

Data Availability: The accuracy of the Credit-worthiness determinant model depends heavily on the quality and quantity of the data used for training the decision tree algorithm. Limited or incomplete data may affect the model's performance.

Algorithm Constraints: The study is limited to using only the decision tree algorithm and does not compare its results with other machine learning algorithms like Random Forest or Neural Networks, which might offer different accuracy levels.

Feature Selection: The study uses selected customer data attributes for prediction, and may not capture all possible factors influencing credit-worthiness.

Scope of Application: The model developed is for academic demonstration and may not directly fit all real-world banking environments without further validation and testing.

Computational Resources: Due to the constraints of computational resources and time, the model's testing and validation are limited to a smaller dataset.

1.6 ORGANISATION OF THE REPORT

This report is divided into five chapters as stated below. Chapter One deals with the general introduction and overview of the whole research. It also discusses the statement of the problem, aim and objectives of the proposed system, significance of the study, scope and limitations of the study as well as organization of the report.

Chapter two deals with the literature review. It reviews the related topics to the project. It also contains a discussion of related aspects of the project topic relative to computer technology.

Chapter three deals with analysis of the system, which includes the data collection method employed, the description of the existing system and its problems and the description of the proposed system and possible advantages it will provide which will solve problems encountered in the existing manual system.

While chapter four deals with the design, implementation and the description of the proposed system, it also covers description of the output design, input design, database design and procedure design. The implementation techniques used, the programming language used in developing the new system and system requirements for running the system. And it talks about the program documentation as well as user documentation.

Finally, chapter five presents a brief summary of the work done, experience gained and problems encountered in the course of the project, conclusion and recommendation.

1.7 DEFINITION OF TERMS

Loan: A thing that is borrowed, especially a sum of money that is expected to be paid back with interest.

Creditability: The quality of being credible.

Prediction: A thing predicted; a forecast.

Algorithm: A process or set of rules to be followed in calculations or other problem-solving operations, especially by a computer.

Decision Tree: A decision tree is a flowchart-like structure in which each internal node represents a “test” on an attribute (e.g. whether a coin flip comes up heads or tails), each branch

represents the outcome of the test, and each leaf node represent a class label (**decision** taken after computing all attributes).

CHAPTER TWO

LITERATURE REVIEW

2.1 REVIEW OF RELATED WORKS

In Sudhamathy and Jothi, (2016), the authors introduce a framework to effectively identify the probability of Default of a Bank Loan Applicant. The metrics derived from the predictions reveal the high accuracy and precision of the built model. The model proposed in Sudhaka and Reddy, (2016) provides an effective prediction model for predicting the credible customers who have applied for bank loans. Decision Tree is applied to predict the attributes relevant for credibility. This prototype model can be used to sanction the loan request of the customers or not. The model proposed in (Aboobyda and Tarig, 2016) has been built using data from the banking sector to predict the status of loans. This model uses three classification algorithms namely j48, bayes Net and naïve Bayes. The model is implemented and verified using Weka. The best algorithm j48 was selected based on accuracy. An improved Risk prediction clustering Algorithm that is Multi-dimensional is implemented in Kavitha, (2016) to determine bad loan applicants. In this work, the Primary and Secondary Levels of Risk assessments are used and to avoid redundancy, Association Rule is integrated. In (Somayyeh and Abdolkarim, 2015) a decision tree model was used as a classifier and for feature selection a genetic algorithm was used. The model was tested using Weka. The work in (Hussain and Shorouq, 2014) developed two data mining models for credit scoring that helps in decision making of giving loans for the banks in Jordan. Considering the rate of accuracy, the regression model is found to perform better than the radial function model.

The work in Blanco, Mejias, Lara and Rayo. (2023) develops many credit scoring models that are based on the multilayer approach. The work proves its performance more than the other models that use logistic regression techniques. The results show that the neural network model performs better than the other three techniques. The work in (Harris, 2017) compares support vector machine based models for credit-scoring developed using various default definitions. The work concluded that the broad definition models are better than the

narrow definition models in their performance. Financial data analysis is done in (Dileep, Desai and Kulkarni, 2018) using techniques such as Decision Tree, Random Forest, Boosting, Bayes classification, Bagging algorithm and others. Support Vector Machine, Decision Tree, Logistic Regression, Neural Network, Perception model, all these techniques are combined in this model.

The effectiveness of applying the above technique on credit scoring is studied. The analysis results show the performance is outstanding based on accuracy. The aim of the study in Francesca, (2012) is to introduce a discrete survival model to study the risk of default and to provide experimental evidence using the Italian banking system.

Soni and Varghese (2019) wrote an article on “Algorithm for the Credit-worthiness determinant model” They discovered that Nowadays, people approach or select bank loans to fulfill their needs, which are very common. This practice has been increasing day by day especially for business, education, marriage, agriculture as well. But several people take advantage and misuse the facilities given by the bank. With technology developing at such a peak stage these days, data mining plays a key role in computer science to solve such issues. Classification is the most suitable predictive modeling technique in data mining to predict the loan repayment capability of a customer in a banking industry. There are various methods to improve the accuracy of a classification algorithm. The accuracy of random forest classification algorithms can be improved using Ensemble methods, Optimization techniques and Feature selection. Various feature selection methods are available. In this research work a novel hybrid feature selection algorithm using wrapper model and fisher score is introduced. The main objective of this paper is to prove that the new hybrid model produces better accuracy than the traditional random forest algorithm. This paper also compares the results obtained from other classification methods and feature selection methods to prove that the proposed algorithm produces better classification accuracy. The experiments were being done using tools such as weka, R, and python programming. This research aims at introducing a new technique which can increase the progress of the banking sector. The accuracy level of this new algorithm in finding the potential of the customer is much higher than the data mining classification algorithm and thus it proves to be very helpful for bank officers.

Semiu and Akanmu (2019) wrote an article “A Boosted Decision Tree Model for Predicting Loan Default in P2P Lending Communities” The study discovers that Loan Default Prediction For Social Lending is an Emerging Area of Research in Predictive Analytics. The

Need For Large Amount Of Data And Few Available Studies In The Current Loan Default Prediction Models For Social Lending Suggest That Other Viable And Easily Implementable Models Should Be Investigated And Developed. In View Of This, This Study Developed A Data Mining Model For Predicting Loan Default Among Social Lending Patrons, Specifically The Small Business Owners, Using Boosted Decision Tree Model. The United States Small Business Administration (Usba) Publicly Available Loan Administration Dataset Of 27 Features And 899164 Data Instances Was Used In 80:20 Ratios For The Training and Testing of the Model. 16 Data Features Were Finally Used As Predictors After Data Cleaning And Feature Engineering. The Gradient Boosting Decision Tree Classifier Recorded 99% Accuracy Compared To The Basic Decision Tree Classifier Of 98%. The Model is Further Evaluated With (A) Receiver Operating Characteristics (Roc) and Area Under Curve (Auc), (B) Cumulative Accuracy Profile (Cap). And (C) Cumulative Accuracy Profile (Cap) Under Auc. Each Of These Model Performance Evaluation Metrics, Especially Roc-Auc, Showed The Relationship Between the True Positives And False Positives that Implies the Model is A Good Fit.

Sivasree and Rekha (2015) wrote an article on “Loan Credibility Prediction System Based on Decision Tree Algorithm” Their study revealed that Data mining techniques are becoming very popular nowadays because of the wide availability of huge quantities of data and the need for transforming such data into knowledge. Techniques of data mining are implemented in various domains such as retail industry, telecommunication industry, biological data analysis, intrusion detection and other scientific applications. Data mining techniques can also be used in the banking industry which help them compete in the market well equipped. In this paper we introduce an effective prediction model for the bankers that helps them predict the credible customers who have applied for a loan. Decision Tree Induction Data Mining Algorithm is applied to predict the attributes relevant for credibility. A prototype of the model is described in this paper which can be used by the organizations in making the right decision to approve or reject the loan request of the customers.

2.2 LOAN SYSTEM CREDITABILITY PREDICTION SYSTEM USING DECISION TREE ALGORITHM

A. Data Mining

Data mining is the process of analyzing data from different perspectives and extracting useful knowledge from it. It is the core of the knowledge discovery process. The various steps

involved in extracting knowledge from raw data as depicted in figure-1. Different data mining techniques include classification, clustering, association rule mining, prediction and sequential patterns, neural networks, regression etc. (Frawley, Piatetsky-Shapiro & Matheus, 2012). Classification is the most commonly applied data mining technique, which employs a set of pre-classified examples to develop a model that can classify the population of records at large (Bharati, 2018). Fraud detection and credit risk applications are particularly well suited to classification techniques. This approach frequently employs Decision Tree based Classification Algorithms. In classification, a training set is used to build the model as the classifier which can classify the data items into its appropriate classes. A test set is used to validate the model.

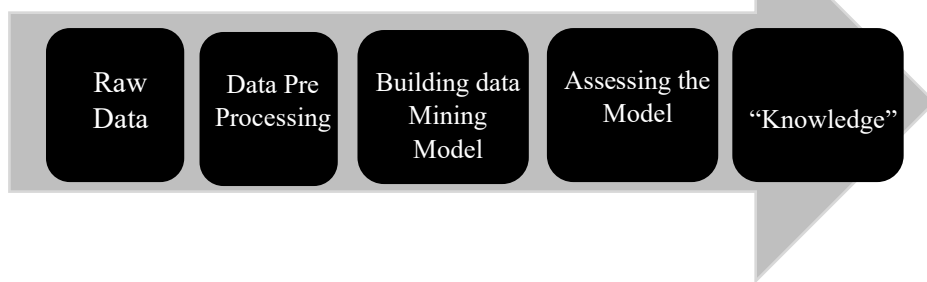


Figure 2.1: Steps in Knowledge Extraction

Source:

A Decision Tree is a structure that includes a root node, branches, and leaf nodes. Each internal node denotes a test on an attribute, each branch denotes the outcome of a test, and each leaf node holds a class label. The topmost node in the tree is the root node. An example of a Decision Tree is depicted in figure2.

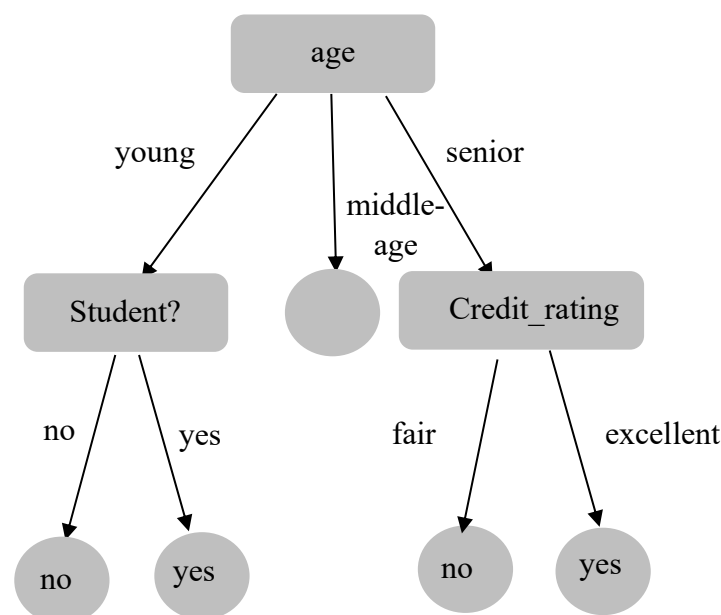


Figure. 2.2. Decision Tree Induction

Source:

B. Data Mining in Banking

Due to the tremendous growth in data the banking industry deals with, analysis and transformation of the data into useful knowledge has become a task beyond human ability (Kumar & Singh, 2023). Data mining techniques can be adopted in solving business problems by finding patterns, associations and correlations which are hidden in the business information stored in the data bases (Bharati, 2018). By using data mining techniques to analyze patterns and trends, bank executives can predict, with increased accuracy, how customers will react to adjustments in interest rates, which customers are likely to accept new product offers, which customers will be at a higher risk for defaulting on a loan, and how to make customer relationships more profitable (Madan, 2016). Globalization and the stiff competition had led the banks focus towards customer retention and fraud prevention. To help them for the same, data mining is used. By analyzing the past data, data mining can help banks to predict credible customers. Thus they can prevent frauds, they can also plan for launching different special offers to retain those customers who are credible. Certain areas that effectively utilize data mining in the banking industry are marketing, risk management and customer relationship management.

Marketing: It is one of the most widely used areas of data mining in the banking industry. The consumer behavior with reference to product, price and distribution channel can be analyzed by the marketing department. The reaction of the customers to the existing and new products can also be known. This information can be used by the banks to promote the products, improve quality of products and services, and gain competitive advantages. Bank analysts can also analyze past trends, determine present demands, and forecast customer behavior of various products and services in order to seize more business opportunities (Gupta & Verma, 2023).

Risk Management: It is widely used for managing risks in the banking industry. Bank executives need to know the credibility of customers they are dealing with. Offering new customers credit cards, extending existing customers' lines of credit, and approving loans can

be risky decisions for banks if they do not know anything about their customers (Patel & Sharma, 2023). Banks provide loans to their customers by verifying the various details relating to the loan, such as amount of loan, lending rate, repayment period etc. Even though banks are cautious while providing loans, there are chances of loan repaying defaults by customers. Data mining techniques help to distinguish borrowers who repay loans promptly from those who default.

Customer Relationship Management: Data mining can be useful in all the three phases of a customer relationship cycle such as customer acquisition, increasing value of the customer and customer retention (Rajanish, 2018). Customer acquisition and retention are very important concerns of any industry, especially the banking industry (Singh, Kumar, & Rao, 2023). Banks have to cater the needs of the customers by providing the services they prefer. This will ultimately lead to customer loyalty and customer retention. Data mining techniques help to analyze the customers who are loyal from those who shift to other banks for better services. If the customer is shifting from his bank to another, reasons for such shifting and the last transaction performed before shifting can be known, and this will help the banks to perform better and retain their customers (Sharma & Patel, 2023).

CHAPTER THREE

3.1 RESEARCH METHODOLOGY AND ANALYSIS

The reference model for our work is cross-industry standard process for data mining (CRISP-DM) Figure 3.1 which is well-known to develop Data Mining projects.

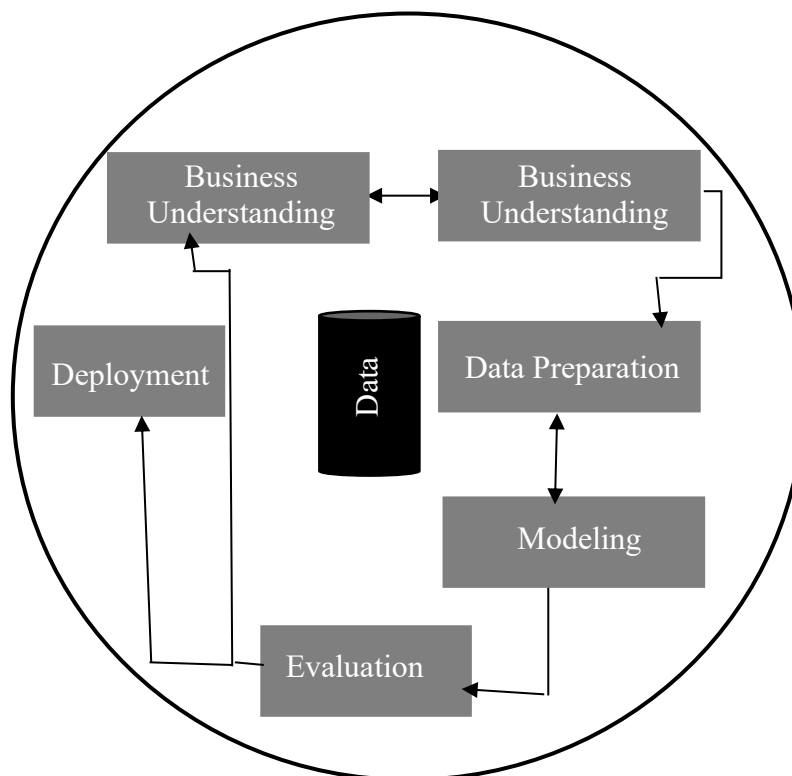


Figure 3.1. CRISP-DM framework

Source:

According to this methodology, the steps of the research can be described as follows:

A. Business understanding

Business understanding: It is the initial phase which focuses on understanding the project objectives and requirements from a business perspective, then converting this knowledge into a data mining problem definition and a preliminary plan designed to achieve the objectives.

B. Data understanding.

The data understanding phase focuses on initial data collection, familiarization of data, identification of data quality problems, and interesting subsets to form hypotheses for hidden information etc.

C. Data preparation

The data preparation phase covers all activities to construct the final dataset (data that will be fed into the modelling tool from the initial raw data). There is no prescribed order for data preparation tasks. Sometimes tasks are to be repeatedly performed, like selection of table, record and attribute as well as transformation and cleaning of data for modelling tools.

D). Modeling

Various modelling techniques are selected and applied in this phase. Typically, there are several techniques for the same data mining problem type. Since some techniques have specific requirements on the form of data, sometimes it needs to go back to the data preparation phase.

E. Evaluation

This phase is to be covered before proceeding to the final deployment of the model, to be certain that business objectives are properly achieved. Consideration and successful implementation of all important business issues are to be confirmed. At the end of this phase, a decision on the use of the data mining results should be reached.

F. Deployment

Creation of the model is generally not the end of the project. The knowledge gained will have to be organized and presented in such a way that the customer can use it.

3.2 ANALYSIS OF THE EXISTING SYSTEM

The existing system used in the banking or financial institutions is a manual method where people were given loans while the banks will be giving them the deadline of the loan given to them. This system causes some banks to experience debt because some people have a default loan. The prediction in this system is done manually where the person leasing the loan will need to forecast by himself who loan should be given next.

3.3 PROBLEMS OF EXISTING SYSTEM

Some problems of the existing system are listed below:

- i. **Necessity for a borrower to visit the branch:** For many lenders and their customers, the loan process begins with a hurdle: the necessity for the customer to visit a physical branch. Filling out application forms delays the processing of the loan and eats away at loan officers' time.
- ii. **Labor intensity for a lender:** Although labor-intensive, the manual application process is not entirely without value. It does allow the loan origination officer to conduct an initial check of whether all the required information has been supplied. Furthermore, the lender can ensure that regulatory requirements for documentation and KYC standards are being met. Finally, an in-person meeting allows the customer to discuss any unusual circumstances that might not be properly addressed by the application forms. On the other hand, if the customer forgets any required documents, he or she must visit the branch again on a later date. In addition, if the loan officer discovers that the customer is wholly ineligible for a loan, both the loan officer and the customer's time has been wasted.
- iii. **Wasted time, analyzing fairly simple cases:** Finally, the underwriting process or the analysis of creditworthiness as well as the loan decision takes additional time if done manually. Again, the manual consumer lending process is not without some advantages. Problems with applications that require human judgment to determine whether a loan should be approved can be addressed on the spot. However, the manual process means that underwriters spend nearly as much time as normal. unproblematic loan files - which for most lenders account for a majority of underwriting volume - as they do on those that require extra attention..
- iv. **Data entry errors:** Assuming, however, that the customer completes an application and meets all the necessary requirements for moving on, the next hurdle to overcome is data management. The loan origination officer will need to follow at least these steps:
 - Enter the customer's information into the bank's system
 - Upload or enter information from supporting documents
 - Alert the personnel conducting the verification and underwriting processes that a new application has been received

These steps take time and increase the risk of error or mis-entry. As the departments that handle verification and underwriting may be in different locations, and may even use a different local system, the number of handling and communication steps between the bank and the customer increases, Each added step increases the time and potential for error in the overall process.

3.4 DESCRIPTION OF THE PROPOSED SYSTEM

3.4.1 PROPOSED MODEL

The proposed model focuses on predicting the credibility of customers for loan repayment by analyzing their behavior. The input to the model is the customer behavior collected. Based on the output from the classifier, a decision on whether to approve or reject the customer request can be made. Decision Tree Induction data mining technique is used to generate the relevant attributes and also make the decision in the model. Data mining model of the proposed system is as depicted in figure4.

A. Problem Understanding

The data mining model is initiated with a collection of details regarding the banking sector and the existing loan processing procedures. The challenges and the main risks associated with the loan approval/rejection in the banking sector are thus better understood.

B. Data Understanding

In the Data Understanding phase, the bank dataset of customer details, which is required for data mining, is collected and familiarized with. Various attributes needed are also studied.

C. Data Filtering

The attributes in the bank data set are filtered and the relevant attributes needed for prediction are selected. After that the incomplete and noisy records in the dataset are removed and prepared for mining.

D. System Modelling

In this stage the system is developed in an efficient and user-friendly manner so that even those users with less technical knowledge can also use it comfortably. The system provides the most relevant attributes that help in determining whether to approve or reject the loan application. This aids in predicting the credibility of future customers.

E. System Evaluation

In the final stage, the designed system is tested with a test set and the performance is assured.

3.4.2 ARCHITECTURE OF PROPOSED MODEL

Architecture of the proposed model is as shown in the figures.

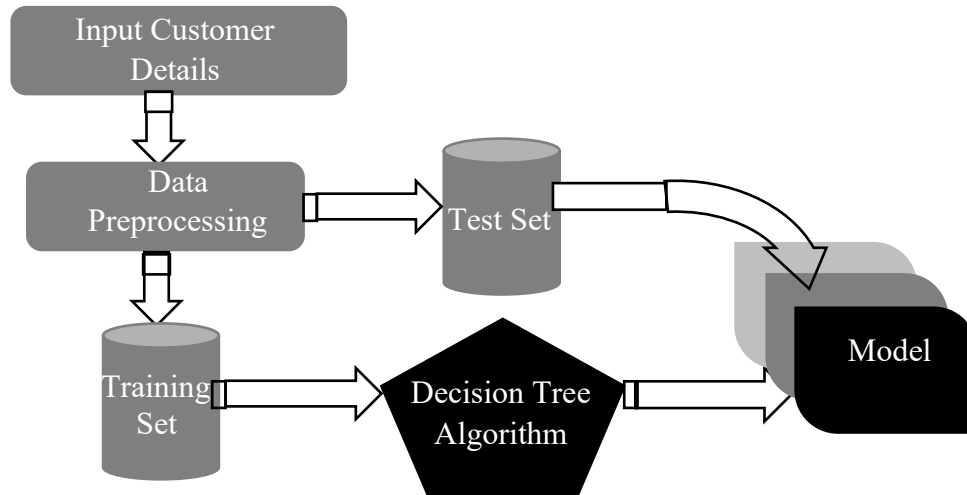


Figure 3.2. Architecture of Proposed Model

A. Input

The main highlight of this Loan Credibility Prediction System is that it uses Decision Tree Induction Data Mining Algorithm to screen/filter out the loan requests. A Decision Tree is developed by performing data mining on an existing bank dataset containing 4520 records and 17 attributes.

B. Data Pre-processing

Initially the Attributes which are critical to make a Loan Credibility Prediction is identified with Information Gain as the attribute-evaluator and Ranker as the search-method. Manual preprocessing is also performed.

C. Data Filtering

Final dataset after preprocessing is divided in such a way that there is 66 % training set and 34 % test set. Test set is used to validate the final result of the classifier.

D. Decision Tree Algorithm

An efficient Decision Tree is formulated with Decision Tree Induction Algorithm. It produces a model with the most relevant 6 attributes. Attribute with rank-1 is placed as the root node of the Decision tree, other attributes from Rank-2 to Rank-6 constitute the intermediate nodes. A

decision is made at each node and the leaf node gives us the final result. That is, if the customer possesses the minimum loan repayment capacity, then the future risks can be avoided. The main benefit of applying Data Mining is that we can always rely on the result of the algorithm to accept or reject the loan application.

3.5 ADVANTAGES OF THE PROPOSED SYSTEM

The proposed system has the following advantages over the existing system:

- i. It is reliable when it comes to Loan Credibility Prediction: This proposed system is very reliable as customers don't need to be afraid of data loss or loan dismissal.
- ii. Its accuracy is of highest order: It is very accurate in prediction than the existing system
- iii. It saves time and energy: It is very fast in prediction as it is automatic.

CHAPTER FOUR

IMPLEMENTATION AND DOCUMENTATION OF THE SYSTEM

4.1 DESIGN OF THE SYSTEM

Generally, all efforts are geared toward designing a system program that eliminates all the setbacks of the existing manual method. This new approach is simple, efficient and guaranteed.

The proposal is designed in order to meet the goal and objective of the organization. The design of the system is intended to solve the various problems of online goods ordering

4.1.1 OUTPUT DESIGN

It refers to the report expected to be generated from files in the database. The generated reports are lists of view, index page, etc. Below are the snapshot of the output designs

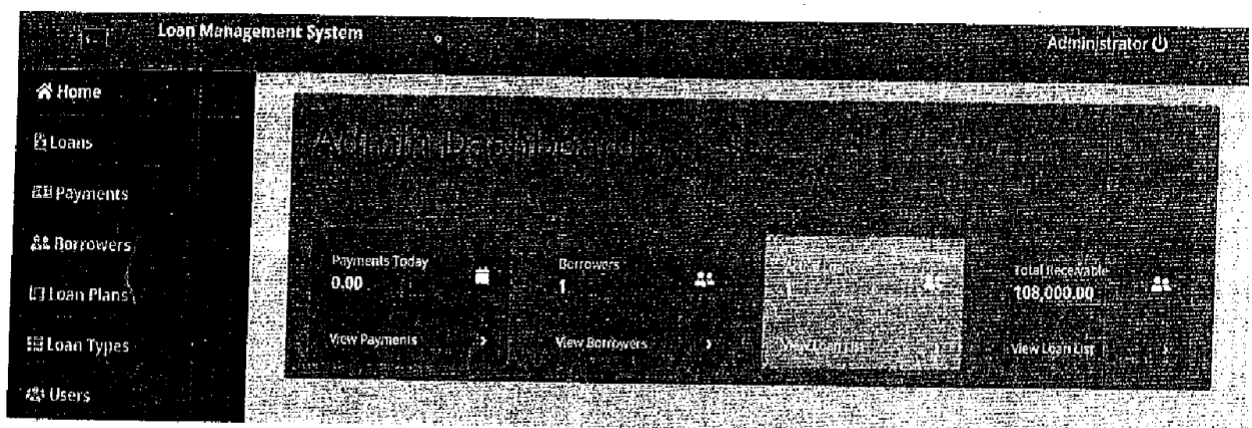


Figure 4.1 Admin Dashboard: This is the main menu of the system where all functions can be carried out through navigation.

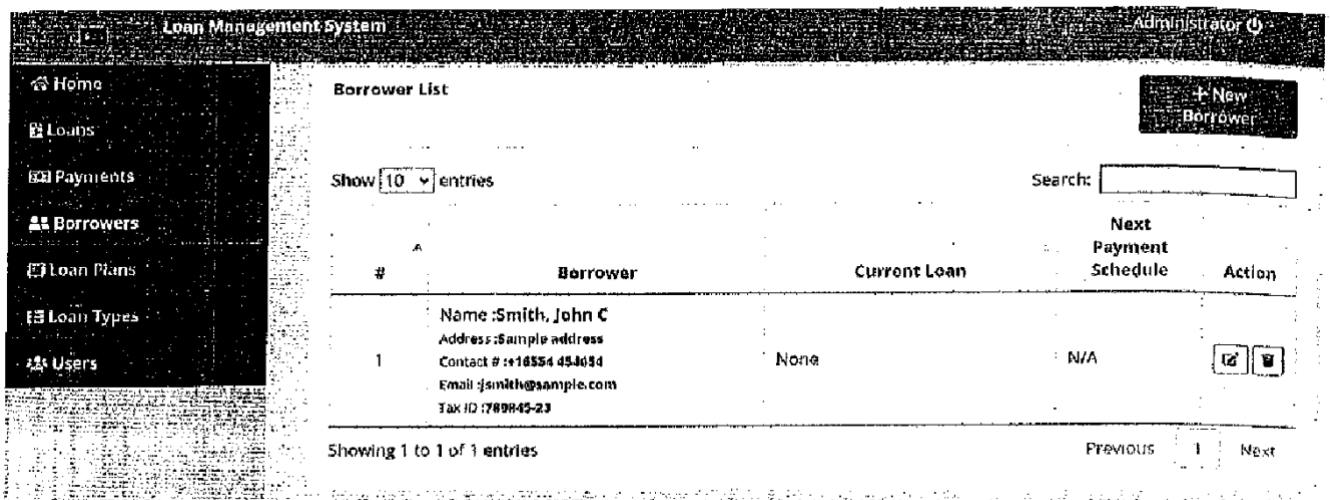


Figure 4.2: Borrower List: The display list of all members who are on loan to make payment.

#	Plan	Action
1	Years/Month: 3 yrs. Interest: 8% Over due Penalty: 3%	Edit Delete
2	Years/Month: 2 yrs. Interest: 5% Over due Penalty: 2%	Edit Delete
3	Years/Month: 2 yrs. and 3month/s Interest: 6% Over due Penalty: 2%	Edit Delete

Figure 4.3 Loan Plan: This allow admin to set the type of loan user can apply

#	Loan Reference No	Payee	Amount	Penalty	Action
1	81409630	Smith, John C	3,000.00	90.00	Edit Delete

Figure 4.4 Payment List: This display the list of all payments made in the system

4.1.2 INPUT DESIGN

Design is a way of mapping and arranging parts into a whole, which satisfies the objective involved. The input design refers to the type of all data used to feed information into the database via the program and the format and the input used for the system. Below are the snapshots of the input designs

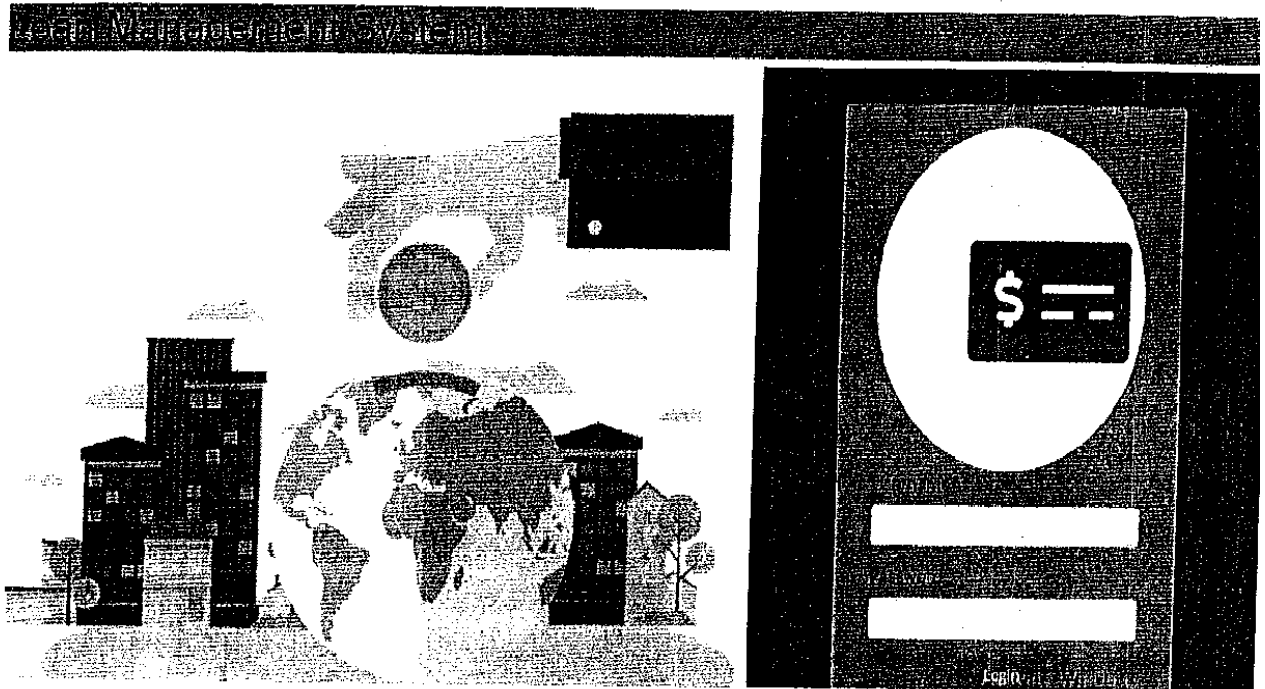


Figure 4.5 Login Page: This allow the admin get access into the system

New Loan Application

Borrower <input type="text" value="Please select here"/>	Loan Type <input type="text" value="Please select here"/>
Loan Plan <input type="text" value="Please select here"/>	Loan Amount <input type="text"/>
months [interest%,penalty%] <input type="text"/>	
Purpose <input type="text"/>	Calculate
Save	Cancel

New borrower

Last Name <input type="text"/>	First Name <input type="text"/>	Middle Name <input type="text"/>
Address <input type="text"/>	Contact # <input type="text"/>	
Email <input type="text"/>	Tax ID <input type="text"/>	
Save		Cancel

Figure 4.7 New Borrower: This allow user to borrow money

4.1.3 DATABASE DESIGN

Database entails the organization and access mode that are implemented in carrying out our specified operations upon the records of the file.

The files used in this system have records that have been organized randomly (random organization) and for this purpose, it will be more appropriate to access the records of the files randomly (random access mode).

Table 4.1 Loan Database

Table	Action	Rows	Type	Collation	Size	Overhead
<input type="checkbox"/> borrowers		1	InnoDB	utf8mb4_general_ci	16 K18	-
<input type="checkbox"/> loan_list		1	InnoDB	utf8mb4_general_ci	16 K18	-
<input type="checkbox"/> loan_plan		3	InnoDB	utf8mb4_general_ci	16 K18	-
<input type="checkbox"/> loan_schedules		36	InnoDB	utf8mb4_general_ci	16 K18	-
<input type="checkbox"/> loan_types		3	InnoDB	utf8mb4_general_ci	16 K18	-
<input type="checkbox"/> payments		1	InnoDB	utf8mb4_general_ci	16 K18	-
<input type="checkbox"/> users		1	InnoDB	utf8mb4_general_ci	16 K18	-
7 tables	Sum	46	InnoDB	utf8mb4_general_ci	112 K18	9.8

This displays the list of all borrowers in the database.

4.1.4 PROCEDURE DESIGN

Procedures are steps which verify the whole process i.e which are everything put together to produce the desired output. This involves the organization of the source document and ends with the output result. Documents are sent to various departments to be filled by the employees and later returned to the personnel department which are analysed to determine which record goes into the computer. After selecting the necessary data, this serves as input to the computer system.

4.2 IMPLEMENTATION OF THE SYSTEM

It is always good to develop new ideas, to implement them on a computer and eventually to relish the satisfaction of achieving a successful result. The implementation process involves converting the system design into a complete and tested EDP that is fully operational and that

can be used by the system users to meet their business needs. During the implementation phase, the hardware and the software must be implemented.

Implementation of a system can be explained in six steps:-

1. Review design specification
2. Code, test and document programs
3. Train users
4. Perform system test
5. Convert to new system
6. Evaluate and maintain the new system

4.2.1 CHOICE OF PROGRAMMING LANGUAGE

The application is designed in Adobe CS4 Dreamweaver web development package which involves the use of PHP server-side scripting language, MYSQL for database management and HTML (with other embedded functionalities) for the page design and layout settings. Hence, the program testing simply involves running it directly from a Mozilla Firefox web browser on a local host server provided by Apache 2.0 in WampServer 2.0 application.

In preparation for the installation of the new system, the method of changeover is given serious consideration to determine the success of the new system. Suitable changeover technique for this system is pilot changeover. The pilot changeover operates by applying the new system bit-by-bit until it covers the whole of the operations. The result obtained from using the pilot method on a small portion of the operations would be used in determining the suitability of the new system for the rest of the operations. This method is similar to testing a small sample of a distribution. If the test yields a good result then the whole system becomes fully operational and the manual/existing system is eliminated.

4.2.2 HARDWARE SUPPORT

The computer configuration required to run the software is;

Computer/memory processor PC with a 486dx, MHZ or Pentium, Intel or higher processor required.

Memory:	2GB of RAM
Cache memory:	512KB
Hard disk Minimum size	500GB
Recommended	200GB
Virtual Memory	32 Bits
Cache memory	512KB

4.2.3 SOFTWARE

- i. Interface Design Language, windows Notepad for help interface design Hypertext Markup Language (HTML)
- ii. MY SQL Database Management Software
- iii. Programming PHP (Hypertext Preprocessor)
- iv. Operating system window 07 professional
- v. Graphic software paint shop and choosing these two formats GIF (Graphic Image Format)
- vi. Scanner software, Mira scans
- vii. Web browser software MOZILLA

4.3 PROGRAM DOCUMENTATION

4.3.1 OPERATING THE SYSTEM

Step 1: Boot your computer and click on start button on taskbar
Step 2: Launch WampServer
Step 3: Locate Online _ goods
Step 4: Click on Options

- 4.1 Click on Product (to view available product)
- 4.2 Click on About Us (to view the about us page)
- 4.3 Click on Contact Us (to send message to the admin)
- 4.4 Click on Order Now (to order for your bags)

Step 5: Logout.

4.3.2 MAINTAINING THE SYSTEM

The use of the term maintenance for software is different from other references to maintenance. Unlike the tires on your car, software does not "wear out". If this is the case, then why does software maintenance account for such a high percentage of the Total Cost of Ownership for software?

The software maintenance definition refers to changes for defect correction, performance improvements, or adaptations to a changed environment (enhancements). According to this definition, if we build software that is defect-free, performs well, and contains user-controlled parameters to adjust processing rules in response to changing requirements then most maintenance would not be necessary.

Why does this happen? There are many reasons but the most common reasons are time constraints and lack of experience. Adding validation logic takes time. So, people make assumptions about the quality of in-bound data. Assumptions are also made about the volume of transactions and the impact on performance and the stability of the automated business processes. Finally, it is common for new software to be developed by younger developers who don't understand the maintenance impacts of their designs. The reality is that business requirements change and most of these assumptions are flawed. Transaction volumes increase, changing business processes require new transactions or new validation criteria, and software users will use the software incorrectly. The cost of software maintenance and the total cost of ownership can dramatically be reduced if developers build software that is adjusted to changes in transaction volumes; validate all inbound data and provide user-configurable options for decision logic and data validation.

CHAPTER FIVE

SUMMARY, CONCLUSION AND RECOMMENDATIONS

5.1 SUMMARY

Nowadays, banks struggle a lot to get an upper hand over each other to enhance overall business due to tight competition. Therefore, banks have realized that retaining the customers and preventing fraud must be the strategic tool for healthy competition. Availability of the huge quantity of data, creation of knowledge base and efficient utilization of the same has helped banks to open up efficient delivery channels. Business decisions can be optimized through data mining. Customer segmentation, banking profitability, credit scoring and approval, predicting payment from customers, marketing, detecting fraud transactions, cash management and forecasting operations, optimizing stock portfolios and ranking investments are some of the areas where data mining techniques can be used in the banking industry. There are numerous areas in which data mining can be used in the banking industry to support customer relationships management. Data mining techniques will help to distinguish borrowers who repay loans promptly from those who do not. It also helps to predict the credit worthiness of the borrower by analyzing the behavior and reliability of the customers. With data mining techniques, banks can do a thorough profiling and ranking of their branches with respect to loan fraud risk. In developing countries like India, Bankers face more problems with the fraudsters.

5.2 CONCLUSION

In this project, we have presented a loan credibility prediction system that helps the organizations i.e. banks in making the right decision to approve or reject the loan request of the customers. This will definitely help the banking industry to open up efficient delivery channels. Decision Tree Induction Algorithm is used for the prediction. Incorporation of other techniques that outperform the performance of popular data mining models have to be implemented and tested for the domain.

5.3 RECOMMENDATION

I recommend that this system should be adopted by banks or financial institutions for easy prediction of loans.

REFERENCES

- Aboobyda, J.H., & Tarig M.A., (2016). Developing Prediction Model of Loan Risk In Banks Using Data Mining, *In Machine Learning and Applications: An International Journal (MLAIJ)*, vol. 3, no.1, pp. 1-9.
- Blanco, A., Mejias, R., Lara, J., & Rayo, S., (2017). Credit scoring models for the microfinance industry using neural networks: evidence from Peru, *In Expert Systems with Applications*, vol. 40, pp. 356-364.
- Dileep, B., Desai & Kulkarni, R.V. (2018) A Review: Application of Data Mining Tools in CRM for Selected Banks, *In (IJCSIT) International Journal of Computer Science and Information Technologies*, Vol. 4 (2), 199-201.
- Francesca, G., (2015) A Discrete-Time Hazard Model for Loans: Some Evidence from Italian Banking System, *In American Journal of Applied Sciences*, vol. 9, no.9), pp. 1337-1346.
- Harris, T., (2023) Quantitative credit risk assessment using support vector machines: Broad versus Narrow default definitions, *In Expert Systems with Applications*, vol. 40, pp. 4404-4413.
- Hussain, A.B., & Shorouq, F.K.E., (2024). Credit risk assessment model for Jordanian commercial banks: Neurscoring approach, *In Review of Development Finance, Elsevier*, vol. 4, pp. 20-28.
- Kavitha, K., (2016). Clustering Loan Applicants based on Risk Percentage using K-Means Clustering Techniques, *In International Journal of Advanced Research in Computer Science and Software Engineering*, vol. 6(2), pp. 162-166.
- Semiu, A. & Akanmu, A.R.G. (2019). A Boosted Decision Tree Model for Predicting Loan Default in P2P Lending Communities, *In International Journal of Engineering and Advanced Technology (IJEAT) ISSN: 2249 - 8958, Volume-9 Issue-1, October 2019*.
- Sivasree, M.S. & Rekha, S.T. (2015). Loan Credibility Prediction System Based on Decision Tree Algorithm, *In International Journal of Engineering Research & Technology (IJERT)* ISSN: 2278-0181 UERTV4IS090708 www.ijert.org (This work is licensed under a Creative Commons Attribution 4.0 International License.) Vol. 4 Issue 09, September.

Somayyeh, Z., & Abdolkarim M., (2015). Natural Customer Ranking of Banks in Terms of Credit Risk by Using Data Mining A Case Study: Branches of Mellat Bank of Iran, *In Journal UMP Social Sciences and Technology Management*, vol. 3, no. 2, pp. 307-316.

Soni, P.M. & Varghese Paul (2019) wrote an article on Algorithm For the Loan Credibility Prediction System, *In International Journal of Recent Technology and Engineering (IJRTE)* ISSN: 2277-3878, Volume-8, Issue-IS4, June 2019

Sudhakar, M., & Reddy, C.V.K., (2016). Two Step Credit Risk Assessment Model For Retail Bank Loan Applications Using Decision Tree Data Mining Technique, *In International Journal of Advanced Research in Computer Engineering & Technology (IJARCET)*, vol. 5, no.3, pp. 705-718.

Sudhamathy, G. & Jothi, V. (2016). Analytics Using R for Predicting Credit Defaulters. IEEE international conference on advances in computer applications (1C AC A), 978-1-5090-3770-4.

APPENDICES

PROGRAM FLOWCHART

