CREDIT CARD FRAUD DETECTION USING MACHINE LEARNING ALGORITHMS WITH LSTM

PROJECT REPORT

PHASE II

Submitted by

AMANULLA M	**[19CS002]**
ELAVARASAN S	**[19CS015]**
HARIPRASATH M	**[19CS022]**
VIMALKUMAR M	**[19CS059]**

In partial fulfillment for the award of the degree
of

BACHELOR OF ENGINEERING

in

COMPUTER SCIENCE AND ENGINEERING

MUTHAYAMMAL ENGINEERING COLLEGE (AUTONOMOUS)

RASIPURAM - 637 408

ANNA UNIVERSITY :: CHENNAI – 600 625

MAY 2023

MUTHAYAMMAL ENGINEERING COLLEGE

(AUTONOMOUS)

RASIPURAM

BONAFIDE CERTIFICATE

Certified that this Report **“CREDIT CARD FRAUD DETECTION USING MACHINE LEARNING ALGORITHMS WITH LSTM”** is the confide work of **“AMANULLA M [19CS002], ELAVARASAN S [19CS015] HARIPRASANTH M [19CS022], VIMALKUMAR M [19CS059]”** who
Carried out the work under my supervision.

Signature of the HOD

Dr .G. KAVITHA ,
M.S(By Research),Ph.D.,
PROFESSOR
HEAD OF THE DEPARTMENT
Department of Computer Science and
Engineering ,
Muthayammal Engineering
College (Autonomous),
Rasipuram-637408.

Signature of the Supervisor

MRS.S.NAZEEMA M.E.,
ASSISTANT PROFESSOR
SUPERVISOR
Department of Computer Science
and Engineering,
Muthayammal Engineering
College (Autonomous),
Rasipuram- 637408.

Submitted for the PROJECT PHASE- II examination held on ______________

INTERNAL EXAMINER **EXTERNAL EXAMINER**

ACKNOWLEDGEMENT

We would like to thank our College Chairman **Shri.R.KANDASAMY** and our Secretary **Dr.K.GUNASEKARAN, M.E., Ph.D.,** who encourages us in all activities.

We here like to record our deep sense of gratitude to our belovedPrincipal **Dr.M.MADHESWARAN, M.E., MBA., Ph.D.,** for providing us the required facility to complete our project successfully.

We extend our sincere thanks and gratitude to our Head of the Department **Dr.G.KAVITHA, M.S., Ph.D.,** Department of Computer Science and Engineering for her valuable suggestions throughout the project.

It is pleasure to acknowledge the contribution made by our Project Coordinator **Dr.N.NAVEENKUMAR., M.E., Ph.D.,** Associate Professor, Department of Computer Science and Engineering for his efforts to complete our project successfully.

It is grateful to acknowledge the support provided by our Project Guide **Mrs.S.NAZEEMA, M.E.,** Assistant Professor, Department of Computer Science and Engineering for her guidance to complete our project successfully.

We are very much thankful to our Parents, Friends and all Faculty Members of the Department of Computer Science and Engineering, who helped us in the successfulcompletion of the project.

TABLE OF CONTENTS

ABSTRACT

Banking industry has the major activity of lending money to those who are in need of money. In order to payback the principle borrowed from the depositor bank collects the interest made by the principle borrowers. Credit risk analysis is becoming an important field in financial risk management. Many credit risk analysis techniques are used for the evaluation of credit risk of the customer dataset. The evaluation of the credit risk datasets leads to the decision to issue the loan of the customer or reject the application of the customer is the difficult task which involves the deep analysis of the customer credit dataset or the data provided by the customer. In this paper we are surveying different techniques for the credit risk analysis which are used for the evaluation for the credit risk datasets. Credit card fraud is a serious problem in financial services Machine learning algorithm based fraud detection schemeis implemented for detect the fraud card. The methods which use long short-term memory (LSTM) networks and majority voting methods are applied. Toevaluate the model efficacy, a publicly available credit card data set is used. Then, a real-world credit card data set from a financial institution is analyzed.

LIST OF FIGURES

LIST OF ABBREVIATION

TERM	**ABBREVIATION**
ML	Machine Leaning
LSTM	Long Short-Team Memory
LR	Logistic Regression
DL	Deep Leaning
CMI	Computer Machinery and Intelligence
CVV	Card Verification Value
NLP	Natural Language Processing
SVM	Support Vector Machine
MLP	Multi-Layer Perceptron
GMDH	Group Method of Data Handling

CHAPTER 1

INTRODUCTION

1.1 OVERVIEW

Credit card fraud is a wide-ranging term for theft and fraud committed using or involving a payment card, such as a credit card or debit card, as a fraudulent source of funds in a transaction. The purpose may be to obtain goods without paying, or to obtain unauthorized funds from an account. Credit card fraud is also an adjunct to identity theft. According to the United States Federal Trade Commission, while the rate of identity theft had been holding steady during the mid-2000s, it increased by 21 percent in 2008. However, credit card fraud, that crime which most people associate with ID theft, decreased as a percentage of all ID theft complaints for the sixth year in a row.

Machine learning (ML) is the scientific study of algorithms and statistical models that computer systems use to effectively perform a specific task without using explicit instructions, relying on models and inference instead. It is seen as a subset of artificial intelligence. Machine learning algorithms build a mathematical model of sample data, known as "training data", in order to make predictions or decisions without being explicitly programmed to perform the task. Machine learning algorithms are used in the applications of email filtering, detection of network intruders, and computer vision, where it is infeasible to develop an algorithm of specific instructions for performing the task. Machine learning is closely related to computational statistics, which focuses on making predictions using computers. The study of mathematical optimization delivers methods, theory and application domains to the field of machine learning. Data mining is a field of study within machine learning, and focuses on through unsupervised learning. In its application across business problems, machine learning is also referred to as predictive analytics.

The name machine learning was coined in 1959 by Arthur Samuel. Tom M. Mitchell provided a widely quoted, more formal definition of the algorithms studied in the machine learning field: "A computer program is said to learn from experienceE with respect to some class of tasks T and performance measure P if its performance at tasks in T, as measured by P, improves with experience E."[6] This definition of the tasks in which machine learning is concerned offers a fundamentally operational definition rather than defining the field in cognitive terms. This follows Alan Turing's proposal in his paper "Computing Machinery andIntelligence", in which the question "Can machines think?" is replaced with the question "Can machines do what we (as thinking entities) can do?”. In Turing's proposal the various characteristics that could be possessed by a thinking machine and the various implications in constructing one are exposed.

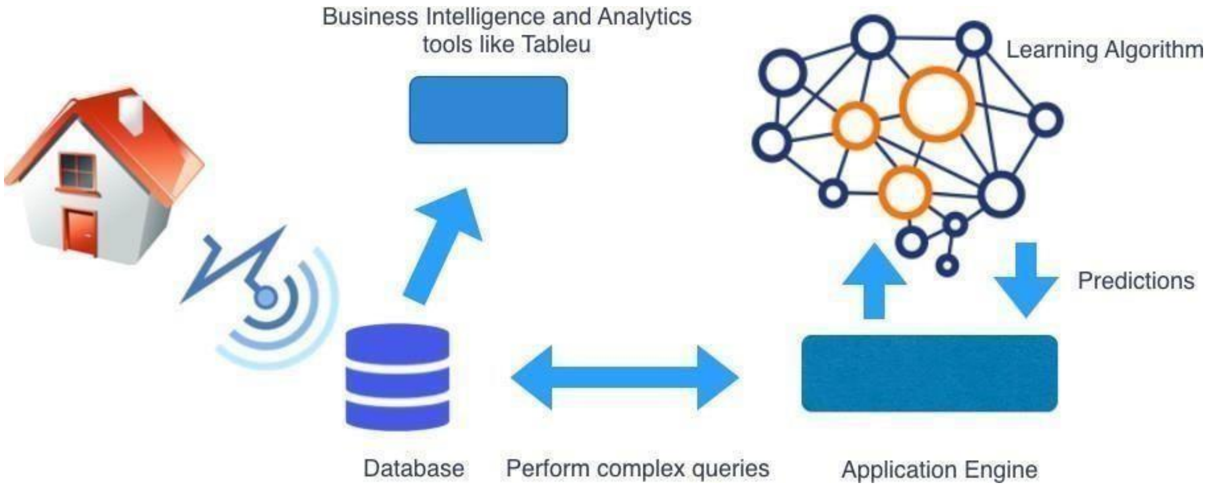


Fig 1.1 Computer Machinery and Intelligence.

1.2 TASKS

Machine learning tasks are classified into several broad categories. In supervised learning, the algorithm builds a mathematical model of a set of data that contains both the inputs and the desired outputs. For example, if the task were determining whether an image contained a certain object, the training data for a supervised learning algorithm would include images with and without that object (the input), and each image would have a label (the output) designating whether it contained the object. In special cases, the input may be only partially available, or restricted to special feedback, Semi-supervised learning algorithms develop mathematical models from incomplete training data, where a portion of the sample inputs are missing the desired output.

Data mining, the extraction of hidden predictive information from large databases, is a powerful new technology with great potential to help companies focus on the most important information in their data warehouses. Data mining tools predict future trends and behaviors, allowing businesses to make proactive, knowledge-driven decisions. The automated, prospective analyses offered by data mining move beyond the analyses of past events provided by retrospective tools typical of decision support systems. Data mining tools can answer business questions that traditionally were too time consuming to resolve. They scour databases for hidden patterns, finding predictive information that experts may miss because it lies outside their expectations.

The Foundations of Data Mining techniques are the result of a long process of research and product development. This evolution began when business data was first stored on computers, continued with improvements in data access, and more recently, generated technologies that allow users to navigate through their data in real time. Data mining takes this evolutionary process beyond retrospective data access and navigation to prospective and proactive information delivery. Data mining is ready for application in the business community because it is supported by three technologies that are now sufficiently mature: Massive data collection Powerful multiprocessor computers Data mining algorithms Commercial databases are growing at unprecedented rates. A recent META Group survey of data warehouse projects found that 19% of respondents are beyond the 50 gigabyte level, while 59% expect to be there by second quarter of 1996.1. In some industries, such as retail, these numbers can be much larger. The accompanying need for improved

Computational engines can now be met in a cost-effective manner with parallel multiprocessor computer technology. Data mining algorithms embody techniques that have existed for at least 10 years, but have only recently been implemented as mature, reliable, understandable tools that consistently outperform older statistical methods.

The most commonly used techniques in data mining are:

Artificial neural networks:

Non-linear predictive models that learn through training and resemble biological neural networks in structure.

Decision trees:

Tree-shaped structures that represent sets of decisions. These decisions generate rules for the classification of a dataset. Specific decision tree methods include Classification and Regression Trees (CART) and Chi Square Automatic Interaction Detection (CHAID).

Genetic algorithms:

Optimization techniques that use processes such as genetic combination, mutation, and natural selection in a design based on the concepts of evolution.

Nearest neighbor method:

A technique that classifies each record in a dataset based on a combination of the classes of the k record(s) most similar to it in a historical dataset (where k ³ 1). Sometimes called the k-nearest neighbor technique. Rule induction: The extraction of useful if-then rules from data based on statistical significance.

Pre-processing Before data mining algorithms can be used, a target data set must be assembled. As data mining can only uncover patterns actually present in the data, the target data set must be large enough to contain these patterns while remaining concise enough to be mined within an acceptable time limit. A common source for data is a data mart or data warehouse. Pre-processing is essential to analyze the multivariate data sets before data mining. The target set is then cleaned. Data cleaning removes the observations containing noise and those with missing data.

1.3 LONG SHORT-TERM MEMORY

LSTM stands for Long Short-Term Memory, and it is a type of recurrent neural network (RNN) architecture used in deep learning for sequential data processing. LSTMs are designed to address the limitations of traditional RNNs, which can suffer from the "vanishing gradient" problem when processing long sequences of data.

LSTMs are capable of capturing long-term dependencies in sequential data by using specialized memory cells that can retain information over extended time steps. These memory cells are equipped with gates that control the flow of information, allowing LSTMs to selectively remember, forget, or update information based on the input data and the network's learned weights.

The key components of an LSTM cell include:

1. Input Gate: This gate controls the flow of information from the input data into the memory cell. It determines which information is relevant to be stored in the memory cell and which can be ignored.
2. Forget Gate: This gate controls the flow of information from the previous memory cell to the current time step. It determines which information should be forgotten or erased from the memory cell based on the current input.
3. Output Gate: This gate controls the flow of information from the memory cell to the output at the current time step. It determines how much of the memory cell's content should be exposed as the output.
4. Memory Cell: This is the memory component of the LSTM, which retains information over time steps. It can store and update information from the input gate and the forget gate, and pass it to the output gate.

LSTMs are widely used in various applications that involve sequential data, such as natural language processing (NLP), speech recognition, time series analysis, and many more. Their ability to capture long-term dependencies makes them effective in modeling complex patterns in sequential data, and they have been widely adoptedin both research and industry for a wide range of tasks.

CHAPTER 2

LITERATURE SURVEY

2.1 TITLE: Use Of Optimized Fuzzy C-Means Clustering And Supervised Classifiers For Automobile Insurance Fraud Detection

AUTHOR : Sharmila Subudhi, Suvasini Panigrahi

This paper presents a novel hybrid approach for detecting frauds in automobile insurance claims by applying Genetic Algorithm (GA) based Fuzzy C -Means(FCM) clustering and various supervised classifier models. Initially, a test set is extracted from the original insurance dataset. The remaining train set is subjected to the clustering technique for under sampling after generating some meaningful clusters. The test instances are then segregated into genuine, malicious or suspicious classes after subjecting to the clusters. The genuine and fraudulent records are discarded, while the suspicious cases are further analyzed by four classifiers – Decision Tree (DT), Support Vector Machine (SVM), Group Method of Data Handling (GMDH) and Multi-Layer Perceptron (MLP) individually. The 10-fold cross validation method is used throughout the work for training andvalidation of the models. The efficacy of the proposed sys- tem is illustrated by conducting several experiments ona real world automobile insurance dataset.

2.2.TITLE: Credit Card Fraud Detection Using Self- Organizing Maps.

AUTHOR: Vladimir Zaslavsky And Anna Strizhak

Nowadays, credit card fraud detection is of great importance to financial institutions. This article presents an automated credit card fraud detection system based on the neural network technology. The authors apply the Self-Organizing Map algorithm to create a model of typical cardholder's behavior and to analyzethe deviation of transactions. Thus finding suspicious transactions.

2.3 TITLE: Evaluating Credit Card Transactions In The Frequency Domain For A Proactive Fraud Detection Approach

AUTHOR : Roberto Saia and Salvatore Carta

The massive increase in financial transactions made in the e-commerce field has led to an equally massive increase in the risks related to fraudulent activities. It is a problem directly correlated with the use of credit cards, considering that almost all the operators that offer goods or services in t he e-commerce space allow their customers to use them for making payments. The main disadvantage of these powerful methods of payment concerns the fact that they can be used not only by the legitimate users (cardholders) but also by fraudsters. Literature reports a considerable number of techniques designed to face this problem, although their effectiveness is jeopardized by a series of common problems, such as the imbalanced distribution and the heterogeneity of the involved data. The approach presented in this paper takes advantage of a novel evaluation criterion based on the analysis, in the frequency domain, of the spectral pattern of the data. Such strategy allows us to obtain a more stable model for representing information, with respect to the canonical ones, reducing both the problems of imbalance and heterogeneity of data. Experiments show that the performance of the proposed approach is comparable to that of its state-of-the-art competitor, although the model definition does not use any fraudulent previous case, adopting a proactive strategy.As a result, fraud detection becomes the essential tool and probably the best way to stop such fraud types. In this study, a new cost-sensitive decision tree approach which minimizes the sum of misclassification costs while selecting the splitting attribute

2.4 TITLE: Real Time Credit Card Fraud Detection Using Computational Intelligence
AUTHOR : Jon T. S. Quah and M. Sriganesh

Online banking and e-commerce have been experiencing rapid growth over the pastfew years and show tremendous promise of growth even in the future. This has made it easier for fraudsters to indulge in new and abstruse ways of committing credit card fraud over the Internet. This paper focuses on real-time fraud detection and presents a new and innovative approach in understanding spending patterns to decipher potential fraud cases. It makes use of Self Organization Map to decipher, filter and analyze customer behavior for detection of fraud.

2.5 TITLE: A Cost-Sensitive Decision Tree Approach For Fraud Detection.
AUTHOR: Yusuf Sahin , Serol Bulkan , Ekrem Duman

With the developments in the information technology, fraud is spreading all over the world, resulting in huge financial losses. Though fraud prevention mechanisms suchas CHIP&PIN are developed for credit card systems, these mechanisms do not prevent the most common fraud types such as fraudulent credit card usages

over virtual POS (Point Of Sale) terminals or mail orders so called online credit card fraud. As a result, fraud detection becomes the essential tool and probably the best way to stop such fraud types. In this study, a new cost-sensitive decision tree approach which minimizes the sum of misclassification costs while selecting the splitting attribute at each non-terminal node is developed and the performance of this approach is compared with the well-known traditional classification models ona real world credit card data set. In this approach, misclassification costs are taken as varying. The results show that this cost-sensitive decision tree algorithm outperforms the existing well-known methods on the given problem set with respect to the well-known performance metrics such as accuracy and true positive rate, but also a newly defined cost-sensitive metric specific to credit card fraud detection domain. Accordingly, financial losses due to fraudulent transactions can be decreased more by the implementation of this approach in fraud detectionsystems.

2.6 TITLE: A Novel Model For Credit Card Fraud Detection Using Artificial Immune Systems

AUTHOR: Neda Soltani Halvaiee, , Mohammad Kazem Akbari

The amount of online transactions is growing these days to a large number. A big portion of these trans- actions contains credit card transactions. The growth of online fraud, on the other hand, is notable, which is generally a result of ease of access to edge technology for everyone. There has been research done on many models and methods for credit card fraud prevention and detection. Artificial Immune Systems is one of them. However, organizations need accuracy along with speed in the fraud detection systems, which is not completely gained yet. In this paper we address credit card fraud detection using Artificial Immune Systems (AIS), and introduce a new model called AIS- based Fraud Detection Model (AFDM). We will use an immune system inspired algorithm (AIRS) and improve it for fraud detection. We increase the accuracy up to 25%, reduce the cost up to 85%, and decrease system response time up to 40% compared to the base algorithm.

2.7 TITLE: Detecting Credit Card Fraud By Genetic Algorithm And Scatter Search

AUTHOR: Ekrem Duman , M. Hamdi Ozcelik

In this study we develop a method which improves a credit card fraud detection solution currently being used in a bank. With this solution each transaction is scoredand based on these scores the transactions are classified as fraudulent or legitimate. In fraud detection solutions the typical objective is to minimize the wrongly classified number of transactions. However, in reality, each transaction do not have the same effect in that if a card is in the hand of fraudsters its whole available limit is used up. Thus, the misclassification cost should be taken as the available limit of the card. This is what we aim at minimizing in this study. As for the solution method, we suggest a novel combination of the two well known meta-heuristic approaches, namely the genetic algorithms and the scatter search. The method is applied to real data and very successful results are obtained compared to current practice.

2.8 TITLE: Detecting Fraud In Online Games Of Chance And Lotteries

AUTHOR: I.T. Christou, M. Bakopoulos

Fraud detection has been an important topic of research in the data mining community for the past two decades. Supervised, semi-supervised, and unsupervised approaches to fraud detection have been pro- posed for the telecommunications, credit, insurance and health-care industries. We describe a novel hybrid system for detecting fraud in the highly growing lotteries and online games of chance sector. While the objectives of fraudsters in this sector are not unique, money laundering and insider attack scenarios are much more prevalent in lotteries than in the previously studied sectors. The lack of labeled data for supervised classifier design, user anonymity, and the size of the data-sets are the other key factors differentiating the problem from previous studies, and are the keydrivers behind the design and implementation decisions for the system described.

2.9 TITLE: Detecting Corporate Tax Evasion Using A Hybrid Intelligent System: A Case Study Of Iran.

AUTHOR: Teymur Rahmani, Mehdi Ghazanfari.

This paper concentrates on the effectiveness of using a hybrid intelligent system that com- bines multilayer perceptron neural network, support vector machine (SVM), and logistic regression (LR) classification models with harmony search (HS) optimization algorithm to detect corporate tax evasion for the Iranian National Tax

Administration (INTA). In this research, the role of optimization algorithm is to search and find the optimal classification model parameters and financial variables combination. Our proposed system finds optimal structure of the

classification model based on the characteristics of the imported dataset. This system has been tested on the data from the food and textile sectors using an iterative structure of 10-fold cross-validation involving 2451 and 2053 test set samples from the tax returns of a two-year period and 1118 and 906 samples as out-of-sample using the tax returns of the consequent year. The results from out-of-sample data show that MLP neural network in combination with HS optimization algorithm outperforms other combinations with 90.07% and 82.45% accuracy, 85.48% and 84.85% sensitivity, and 90.34% and 82.26% specificity, respectivelyin the food and textile sectors. In addition, there is also a difference between the selected models and obtained accuracies based on the test data and out-of-sample data in both sectors and selected financial variables of every sector.

2.10 TITLE: Credit Card Fraud Detection Using Self Organised**.**

AUTHOR: Mitali Bansal, Suman

Self organizing Maps (SOMs) are most well-known, unsupervised approach of neural network that is used for clustering and are very efficient in handling large and high dimensional dataset. As SOMs can be applied on large complex. Set, so it can be implemented to detect credit card fraud. Online banking and e- commerce has been experiencing rapid growth over past years and willshowtremendous growth even in future. So, it is very necessary to keep an eye on fraudsters and find out some ways to depreciate the rate of frauds. This paper focuses on Real Time Credit Card Fraud Detection and presents a new and innovative approach to detect the fraud by the help of SOM.

CHAPTER 3

SYSTEM ANALYSIS

3.1 EXISTING SYSTEM

Our existing has made a detail study on fraud detection using the method of natural observation of the events happened from the customer side. The existing has worked on the methods of collecting the data from the social media and framing them in terms of big data models and working on the challenges existed the field. Implemented a system which supports in the detection of the scams or frauds in the field of the business by recording the transactions and there by building a model using data mining models. In existing system AdaBoost algorithm has been implemented. The Random Forest (RF) algorithm has been implemented. Loss from credit card fraud affects the merchants, where they bear all costs, including card issuer fees, charges, and administrative charges.

3.1.1 DRAWBACKS

- Difficult to identify a fraud credit card.
- Less Accuracy.

3.2 PROPOSED SYSTEM

The proposed system first step is data collection in this step data collection from Kaggle. After collecting data to processing using machine learning algorithm. The processing data is converted into principle component analysis. In this step using 10 different component. Data splitting process to split a data from principle component analysis and k-told cross validation used. Model development process using machine learning algorithm to train a data to send next step of the process. Performance evaluation implemented this project using formulas. Finally we deployed the project using model selection. Majority voting is frequently used in data classification, which involves a combined model with at least two algorithms.

Machine learning based algorithm is implemented for fraud detection. Each algorithm makes its own prediction for every test sample. It will be extended to online learning models. In addition, other online learning models will be investigated. The use of online learning will enable rapid detection of fraud cases, potentially in real-time. In the proposed system the KNN algorithm is used. The k-nearest neighbor algorithm is a non-parametric method used for classification and regression. By using KNN the processing time is reduced and also process the larger datasets.

3.2.1 ADVANTAGES

- Stable system
- More accuracy
- Less time to predict the fraud

3.2 SYSTEM ARCHITECTURE

A system architecture is the conceptual model that defines the structure, behavior, and more views of a system. A system architecture can consist of system components and the sub-systems developed, that will work together to implement the overall system**.**

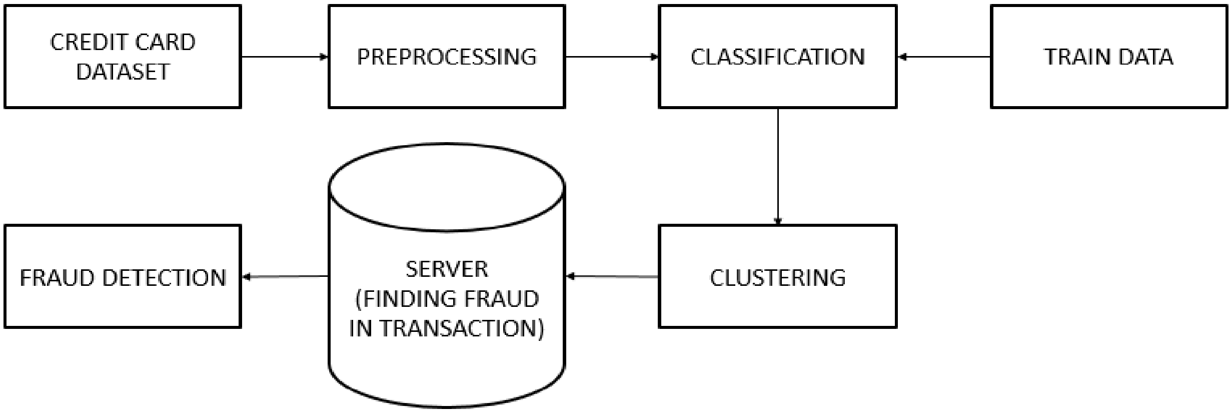


Fig. 3.3 System Architecture.

CHAPTER 4

SYSTEM REQUIREMENT

4.1 HARDWARE REQUIREMENT

- Processor - Pentium – IV

  RAM - 4 GB (min)

The compiler.ast module defines an abstract syntax for Python. In the abstract syntax tree, each node represents a syntactic construct. The root of the treeis Module object.

The abstract syntax offers a higher level interface to parsed Python source code. The parser module and the compiler written in C for the Python interpreter use a concrete syntax tree. The concrete syntax is tied closely to the grammar description used for the Python parser. Instead of a single node for a construct, there are often several levels of nested nodes that are introduced by Python's precedence rules

The abstract syntax tree is created by the compiler.transformer module. The transformer relies on the built-in Python parser to generate a concrete syntax tree. It generates an abstract syntax tree from the concrete tree.

The transformer module was created by Greg Stein and Bill Tutt for an experimental Python-to-C compiler. The current version contains a number of modifications and improvements, but the basic form of the abstract syntax and of the transformer are due to Stein and Tutt. The compiler.ast module defines an abstract syntax for Python. In the abstract syntax tree, each node represents a syntactic construct. The root of the tree is Module object.

The abstract syntax offers a higher level interface to parsed Python source code. The parser module and the compiler written in C for the Python interpreter use a concrete syntax tree. The concrete syntax is tied closely to the grammar description used for the Python parser. Instead of a single node for a construct, there are often several levels of nested nodes that are introduced by Python's precedence rules

The abstract syntax tree is created by the compiler.transformer module. The transformer relies on the built-in Python parser to generate a concrete syntax tree. It generates an abstract syntax tree from the concrete tree.

The transformer module was created by Greg Stein and Bill Tutt for an experimental Python-to-C compiler. The current version contains a number of modifications and improvements, but the basic form of the abstract syntax and of the transformer are due to Stein and Tutt.

4.2 SOFTWARE REQUIREMENT

Operating System	: Windows 7 or 8
Front Programming	: HTML and CSS.
Back Programming	: Python Idle

Python Technology:

Python is an interpreted, high-level, general-purpose programming language. It supports multiple programming paradigms, including procedural, object-oriented, and functional programming. Python is often described as a "batteries included" language due to its comprehensive standard library.

Python Programing Language:

Python is a multi-paradigm programming language. Object-oriented programming and structured programming are fully supported, and many of its features support functional programming and aspect-oriented programming (including by metaprogramming and met objects (magic methods)). Many other paradigms are supported via extensions, including design by contract and logic programming.

Python uses dynamic typing and a combination of reference counting and a cycle - detecting garbage collector for memory management. It also features dynamic name resolution (late binding), which binds method and variable names during program execution.

Python is meant to be an easily readable language. Its formatting is visually uncluttered, and it often uses English keywords where other languages use punctuation. Unlike many other languages, it does not use curly brackets to delimit blocks, and semicolons after statements are optional. It has fewer syntactic exceptions and special cases than C or Pascal.

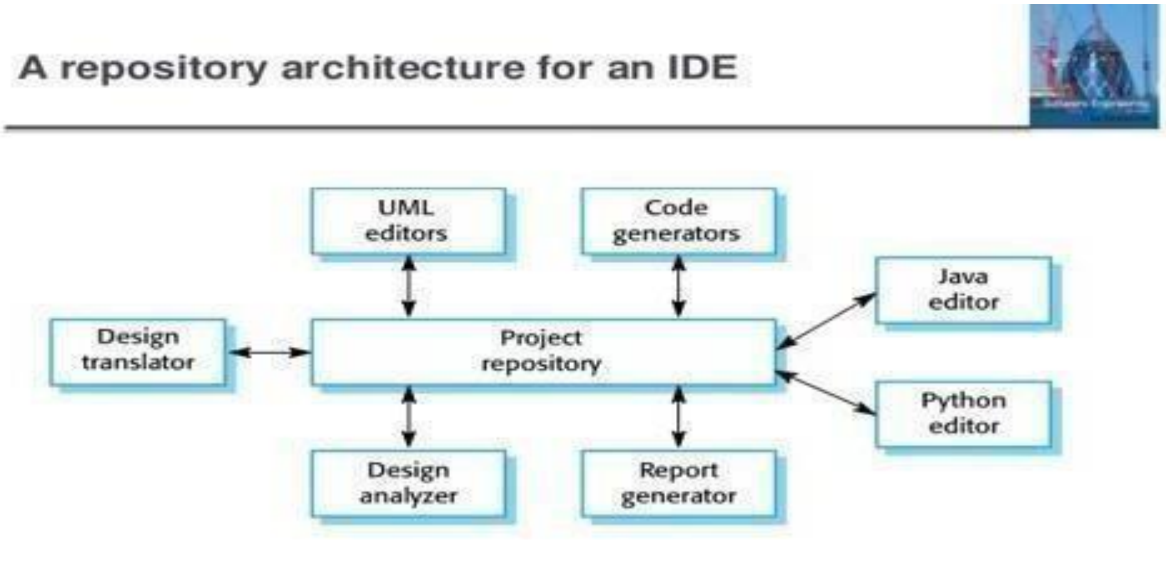


Fig 4.2 Architecture for an IDE

Python strives for a simpler, less-cluttered syntax and grammar while giving developers a choice in their coding methodology. In contrast to Perl's "there is more than one way to do it" motto, Python embraces a "there should be one and preferably only one obvious way to do it" design philosophy. Alex Martelli, a Fellow at the Python Software Foundation and Python book author, writes that "To describe something as 'clever' is not considered a compliment in the Python culture."

Python's developers strive to avoid premature optimization, and reject patches to non-critical parts of the Python reference implementation that would offer marginalincreases in speed at the cost of clarity. When speed is important, a Python programmer can move time-critical functions to extension modules written in languages such as C, or use PyPy, a just-in-time compiler. Python is alsoavailable, which translates a Python script into C and makes direct C-level API calls into the Python interpreter.

An important goal of Python's developers is keeping it fun to use. This is reflected in the language's name a tribute to the British comedy group Monty Python and in occasionally playful approaches to tutorials and reference materials, such as examples that refer to spam and eggs (from a famous Monty Python sketch) instead of the standard foo and bar. Python strives for a simpler, less-cluttered syntax and grammar while giving developers a choice in their coding methodology. In contrastto Perl's "there is more than one way to do it" motto, Python embraces a "there should be one and preferably only one obvious way to doit" design philosophy. Alex Martelli, a Fellow at the Python Software Foundation and Python book author, writes that "To describe something as 'clever' is not considered a compliment in the Python culture."

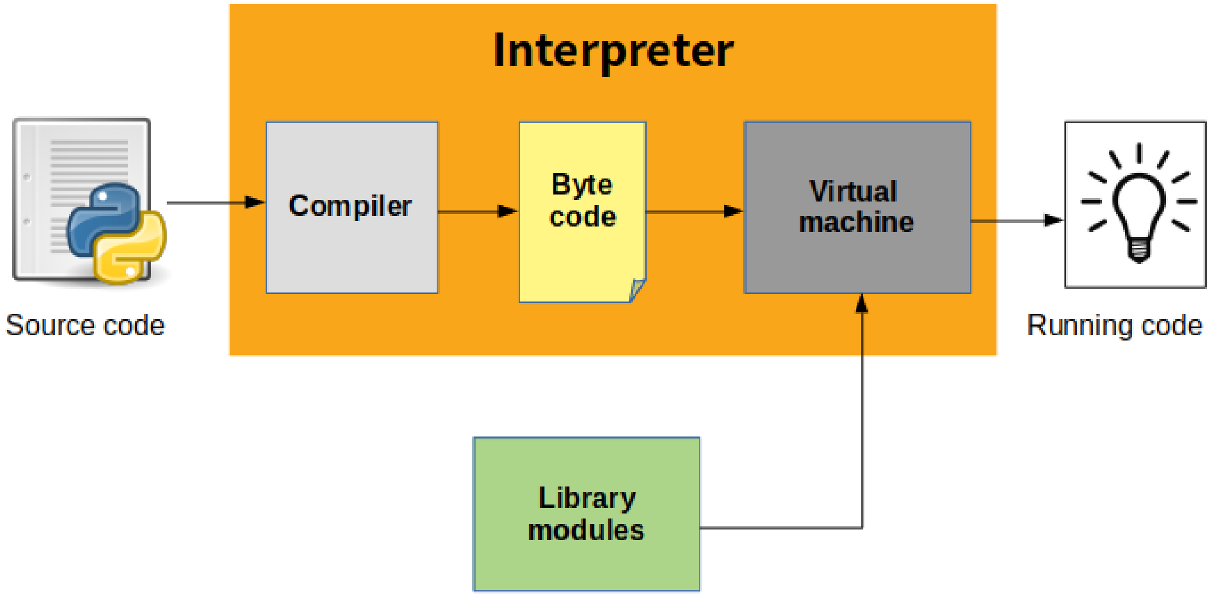


Fig 4.2.1 Python Running code

Python uses duck typing and has typed objects but untyped variable names. Type constraints are not checked at compile time; rather, operations on an object may fail, signifying that the given object is not of a suitable type. Despite being dynamically typed, Python is strongly typed, forbidding operations that are not well - defined (for example, adding a number to a string) rather than silently attempting tomake sense of them.

The Python Platform:

The platform module in Python is used to access the underlying platform's data, such as, hardware, operating system, and interpreter version information. The platform module includes tools to see the platform's hardware, operating system, and interpreter version information where the program is running.

There are four functions for getting information about the current Python interpreter. python_version() and python_version_tuple() return different forms of the interpreter version with major, minor, and patch level components.

python_compiler() reports on the compiler used to build the interpreter. And python_build() gives a version string for the build of the interpreter.

Platform () returns string containing a general purpose platform identifier. The function accepts two optional Boolean arguments. If aliased is true, the names in the return value are converted from a formal name to their more common form. When terse is true, returns a minimal value with some parts dropped.

Python compiler

The Python compiler package is a tool for analyzing Python source code and generating Python bytecode. The compiler contains libraries to generate an abstract syntax tree from Python source code and to generate Python bytecode from the tree.

The compiler package is a Python source to bytecode translator written in Python. It uses the built-in parser and standard parser module to generate a concrete syntax tree. This tree is used to generate an abstract syntax tree (AST) andthen Python bytecode.

The full functionality of the package duplicates the built-in compiler provided with the Python interpreter. It is intended to match its behavior almost exactly. Whyimplement another compiler that does the same thing? The package is useful for a variety of purposes. It can be modified more easily than the built-in compiler. The AST it generates is useful for analyzing Python source code.

CHAPTER 5
SYSTEM DIAGRAM

UML DIAGRAMS

UML stands for Unified Modeling Language. UML is a standardized general - purpose modeling language in the field of object-oriented software engineering. The standard is managed, and was created by, the Object Management Group.

The goal is for UML to become a common language for creating models of object oriented computer software. In its current form UML is comprised of two major components: a Meta-model and a notation. In the future, some form of method or process may also be added to; or associated with, UML.

The Unified Modeling Language is a standard language for specifying, Visualization, Constructing and documenting the artifacts of software system, as well as for business modeling and other non-software systems.

The UML represents a collection of best engineering practices that have proven successful in the modeling of large and complex systems.

The UML is a very important part of developing objects oriented software and the software development process. The UML uses mostly graphical notations to express the design of software projects.

GOALS

The Primary goals in the design of the UML are as follows:

1. Provide users a ready-to-use, expressive visual modeling Language so that they can develop and exchange meaningful models.
2. Provide extendibility and specialization mechanisms to extend the core concepts.
3. Be independent of particular programming languages and development process.
4. Provide a formal basis for understanding the modeling language.
5. Encourage the growth of OO tools market.
6. Support higher level development concepts such as collaborations, frameworks, patterns and components.
7. Integrate best practices.

5.1 USE CASE DIAGRAM

A use case diagram in the Unified Modeling Language (UML) is a type of behavioral diagram defined by and created from a Use-case analysis. Its purpose is to present a graphical overview of the functionality provided by a system in terms of actors, their goals (represented as use cases), and any dependencies between those use cases. The main purpose of a use case diagram is to show what system functions are performed for which actor. Roles of the actors in the system can be depicted.

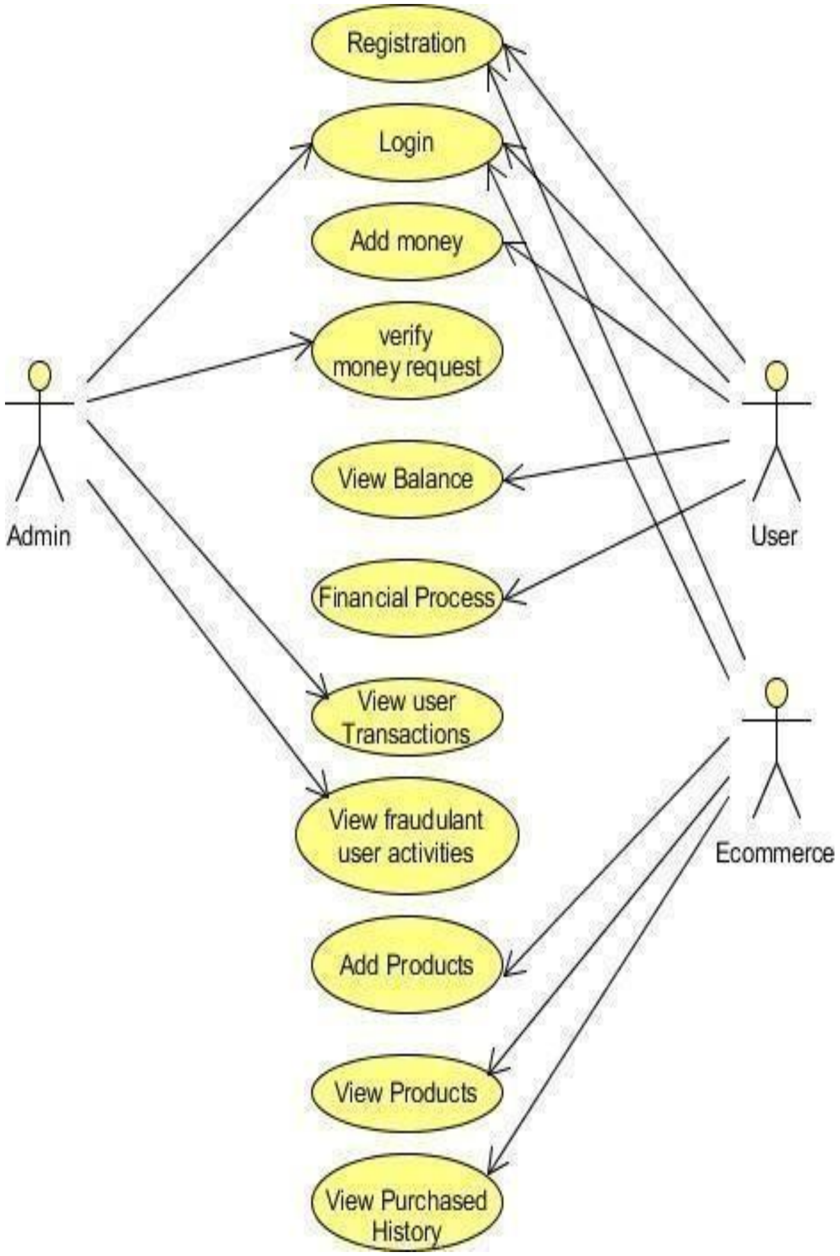


Fig 5.1 User case Diagram

5.2 CLASS DIAGRAM

In software engineering, a class diagram in the Unified Modeling Language (UML)is a type of static structure diagram that describes the structure of a system by showing the system's classes, their attributes, operations (or methods), and the relationships among the classes. It explains which class contains information.

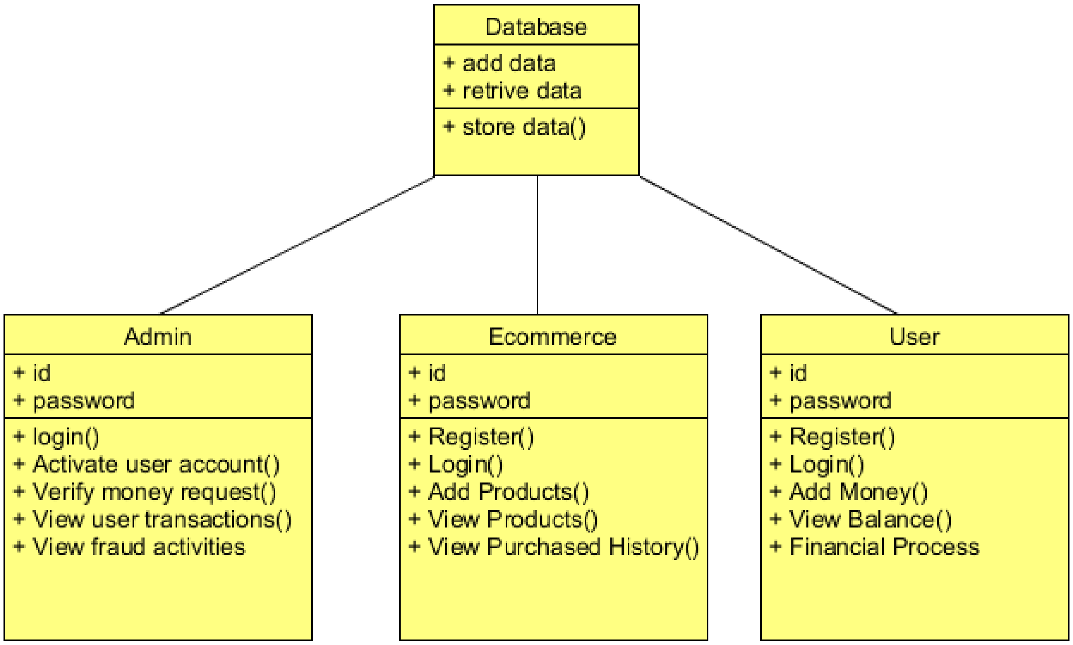


Fig 5.2 Class Diagram

5.3 SEQUENCE DIAGRAM

A sequence diagram in Unified Modeling Language (UML) is a kind of interactiondiagram that shows how processes operate with one another and in what order. It is a construct of a Message Sequence Chart. Sequence diagrams are sometimes calledevent diagrams, event scenarios, and timing diagrams.

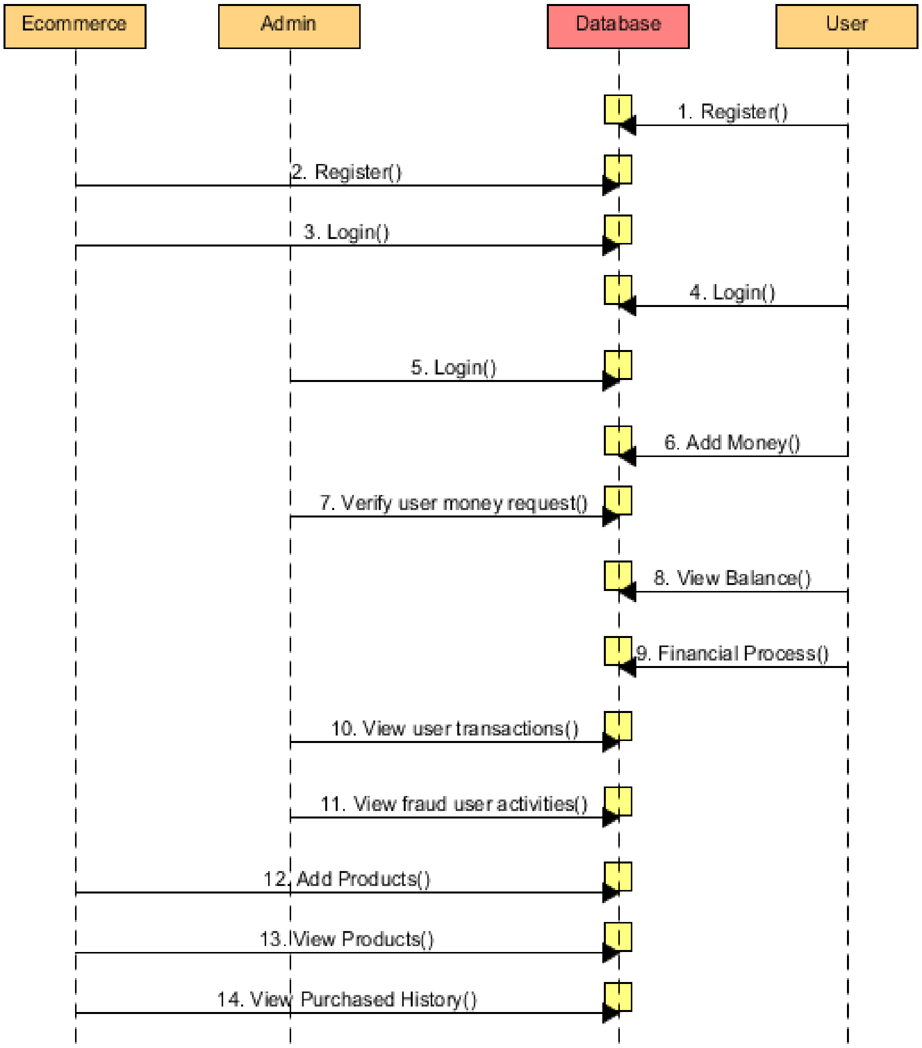


Fig 5.3 Sequence Diagram

5.4 DEPLOYMENT

Component diagrams are used to describe the components and deployment diagrams shows how they are deployed in hardware. UML is mainly designed to focus on the software artifacts of a system. However, these two diagrams are special diagrams used to focus on software and hardware components.

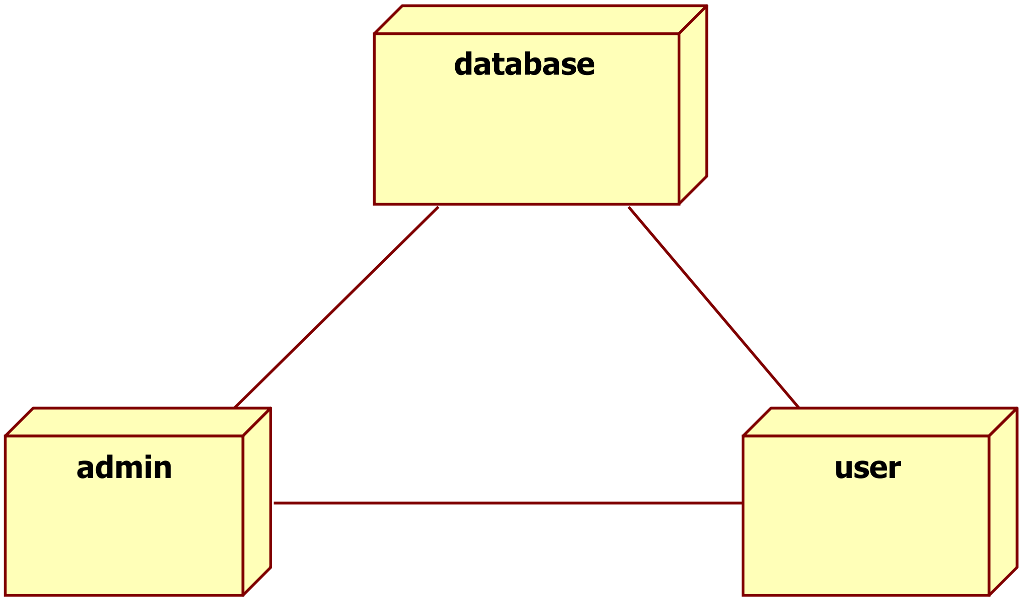


Fig 5.4 Deployment.

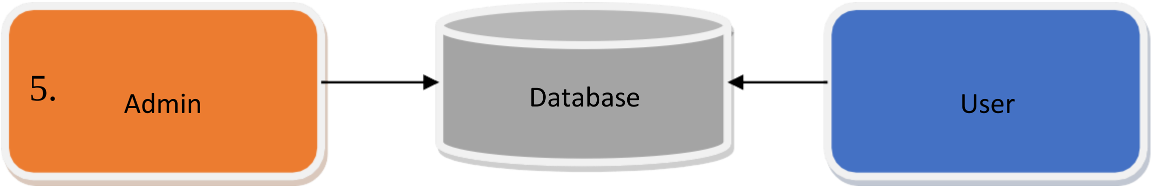
5.5 DATA FLOW DIAGRAM

1. The DFD is also called as bubble chart. It is a simple graphical formalism that can be used to represent a system in terms of input data to the system, various processing carried out on this data, and the output data is generated by this system.
2. The data flow diagram (DFD) is one of the most important modeling tools. It is used to model the system components. These components are the system

process, the data used by the process, an external entity that interacts with the system and the information flows in the system.

3. DFD shows how the information moves through the system and how it is modified by a series of transformations. It is a graphical technique that depicts information flow and the transformations that are applied as data moves from input to output.
4. DFD is also known as bubble chart. A DFD may be used to represent a system at any level of abstraction. DFD may be partitioned into levels that represent increasing information flow and functional detail.

DFD DIAGRAM



5.5 Data Flow Diagram.

CHAPTER 6

SYSTEM DESIGN

6.1 INPUT DESIGN

The input design is the link between the information system and the user. It comprises the developing specification and procedures for data preparation and those steps are necessary to put transaction data in to a usable form for processing can be achieved by inspecting the computer to read data from a written or printed document or it can occur by having people keying the data directly into the system. The design of input focuses on controlling the amount of input required, controlling the errors, avoiding delay, avoiding extra steps and keeping the processsimple. The input is designed in such a way so that it provides security and ease of use with retaining the privacy. Input Design considered the following things:

- What data should be given as input?
- How the data should be arranged or coded?
- The dialog to guide the operating personnel in providing input.
- Methods for preparing input validations and steps to follow when error occur.

6.2 OBJECTIVES

1. Input Design is the process of converting a user-oriented description of the input into a computer-based system. This design is important to avoid errors in the data input process and show the correct direction to the management for getting correct information from the computerized system.

2. It is achieved by creating user-friendly screens for the data entry to handle large volume of data. The goal of designing input is to make data entry easier and to be free from errors. The data entry screen is designed in such a way that all the data manipulates can be performed. It also provides record viewing facilities.

3. When the data is entered it will check for its validity. Data can be entered with the help of screens. Appropriate messages are provided as when needed so that the userwill not be in maize of instant. Thus the objective of input design is to create an input layout that is easy to follow.

6.3 OUTPUT DESIGN

A quality output is one, which meets the requirements of the end user and presents the information clearly. In any system results of processing are communicated to the users and to other system through outputs. In output design it is determined how theinformation is to be displaced for immediate need and also the hard copy output. It is the most important and direct source information to the user. Efficient and intelligent output design improves the system’s relationship to help user decision-making.

1. Designing computer output should proceed in an organized, well thought out manner; the right output must be developed while ensuring that each output element is designed so that people will find the system can use easily and effectively. When analysis design computer output, they should Identify the specific output that is needed to meet the requirements.

2. Select methods for presenting information.

3. Create document, report, or other formats that contain information produced by the system.

CHAPTER 7

MODULE IMPLEMENTATION

7.1. MODULE

- ADMIN MODULE
- BANK ADMIN
- USER MODULE
- PRODUCT SEARCH MODULE
- PAYMENT MODULE

7.2 MODULE DESCRIPTION

7.2.1 ADMIN MODULE

- In this module admin verify and authorize the both registered bank admin and registered user account.
- Admin add and view Products. User purchased history also admin view in this module.
- The Random Forest (RF) creates an ensemble of random trees it will classify the data’s into CVV fraud and Expiry Date problems.
- Depends upon RF algorithm Admin view the user fraudulent activities graphically in this module.

7.2.2 BANK ADMIN

- In this module bank admin manage the user requirements and fraudulent activities.
- Accept the user credit card request and show the usage details also.
- Send the card CVV and expiry date for newly approved credit card user in the module.

7.2.3 USER MODULE

- In this module user needs registration and login.
- Newly registered user send credit card request for bank admin.
- After approval of your card you will use that card for purchasing and transactions.
- User check his card usage for his account.

7.2.4 PRODUCT SEARCH MODULE

- They can find products easily word search.
- After viewing the descriptions selecting their required brands and on confirmation they can bought products.
- The user can easily go through the site by just having the minimum knowledge of computer is sufficient to use this site.

7.2.5 PAYMENT MODULE

- The module takes care of the all the secured payments that should happen for the purchases that happens online, so to implement a security algorithm is one of the major concepts of the payment.
- Verify the CVV details as well as balance of your card also.
- If user enter wrong CVV or low balance account means it will terminate the transaction process.
- Here we are calculate the expiry date of our credit card also.
- If fraud user try to used expired card means it will analyze and terminate the payment process.

CHAPTER 8

SYSTEM STUDY

8.1 FEASIBILITY STUDY

The feasibility of the project is analyzed in this phase and business proposal is put forth with a very general plan for the project and some cost estimates. During system analysis the feasibility study of the proposed system is to be carried out. This is to ensure that the proposed system is not a burden to the company. For feasibility analysis, some understanding of the major requirements for the systemis essential. Three key considerations involved in the feasibility analysis are

- ♦ ECONOMICAL FEASIBILITY
- ♦ TECHNICAL FEASIBILITY
- ♦ SOCIAL FEASIBILITY

8.2 ECONOMICAL FEASIBILITY

This study is carried out to check the economic impact that the system will have on the organization. The amount of fund that the company can pour into the research anddevelopment of the system is limited. The expenditures must be justified. Thus the developed system as well within achieved because most of the technologies used are freely available. Only thecustomized products had to be purchased achieved because most of the technologies used are freely available. Only thecustomized products had to be purchased.

8.3 TECHNICAL FEASIBILITY

This study is carried out to check the technical feasibility, that is, the technical requirements of the system. Any system developed must not have a high demand on the available technical resources. This will lead to high demands on theavailable technical resources. This will lead to high demands being placed on the client. The developed system must have a modest requirement, as only minimal or null changes are required for implementing this system.

8.4 SOCIAL FEASIBILITY

The aspect of study is to check the level of acceptance of the system by the user. This includes the process of training the user to use the system efficiently. The user must not feel threatened by the system, instead must accept it as a necessity. The level of acceptance by the users solely depends on the methods that are employed to educate the user about the system and to make him familiar withit. His level of confidence must be raised so that he is also able to make some constructive criticism, which is welcomed, as he is the final user of the system.

CHAPTER 9

SYSTEM TESTING

The purpose of testing is to discover errors. Testing is the process of trying to discover every conceivable fault or weakness in a work product. It provides a way to check the functionality of components, sub-assemblies, assemblies and/or a finished product It is the process of exercising software with the intent of ensuring that the Software system meets its requirements and user expectations and does not fail in an unacceptable manner. There are various types of test. Each test type addresses a specific testing requirement.

TYPES OF TESTS

9.1 UNIT TESTING

Unit testing involves the design of test cases that validate that the internal program logic is functioning properly, and that program inputs produce validoutputs. All decision branches and internal code flow should be validated. It is the testing of individual software units of the application .it is done after the completion of an individual unit before integration. This is a structural testing, that relies on knowledge of its construction and is invasive. Unit tests perform basic tests at component level and test a specific business process, application, and/or system configuration. Unit tests ensure that each unique path of a business process performs accurately to the documented specifications and contains clearly defined inputs and expected results.

9.2 INTEGRATION TESTING

Integration tests are designed to test integrated software components to determine if they actually run as one program. Testing is event driven and is more concerned with the basic outcome of screens or fields. Integration tests demonstrate that although the components were individually satisfaction, as shown by successfully unit testing, the combination of components is correct and consistent. Integration testing is specifically aimed at exposing the problems that arise from the combination of components.

9.3 FUNCTIONAL TESTING

Functional tests provide systematic demonstrations that functions tested are available as specified by the business and technical requirements, system documentation, and user manuals.

Functional testing is centered on the following items:

Valid Input : identified classes of valid input must be accepted.

Invalid Input : identified classes of invalid input must be rejected.

Functions : identified functions must be exercised.

Output : identified classes of application outputs must be exercised.

Systems/ Procedures : interfacing systems or procedures must be invoked.

Organization and preparation of functional tests is focused on requirements, key functions, or special test cases. In addition, systematic coverage pertaining toidentify Business process flows; data fields, predefined processes, and successive processes must be considered for testing. Before functional testing is complete, additional tests are identified and the effective value of current tests is determined.

System Test

System testing ensures that the entire integrated software system meets requirements. It tests a configuration to ensure known and predictable results. An example of system testing is the configuration oriented system integration test. System testing is based on process descriptions and flows, emphasizing pre-driven process links and integration points.

White Box Testing

White Box Testing is a testing in which in which the software tester has knowledge of the inner workings, structure and language of the software, or at least its purpose. It is purpose. It is used to test areas that cannot be reached from a black box level.

Black Box Testing

Black Box Testing is testing the software without any knowledge of the inner workings, structure or language of the module being tested. Black box tests, as most other kinds of tests, must be written from a definitive source document, such as specification or requirements document, such as specification or requirements document. It is a testing in which the software under test is treated, as a black box you cannot "see" into it. The test provides inputs and responds to outputs without considering how the software works. components or software applications, e.g. components in a software system or –one step up – software applications at the company level – interact without error.

Acceptance Testing

User Acceptance Testing is a critical phase of any project and requires significant participation by the end user. It also ensures that the system meets the functional requirements.

Test Results:

All the test cases mentioned above passed successfully. No defects encountered.

CHAPTER 10
CONCLUTION AND FUTURE WORK

10.1 CONCLUSION

It is worth keep in mind that objective of the paper is to surveying on the different classifier which are used in the credit risk evaluation. In this paper different types of classifiers are discussed and also different types of ensemble classifiers are briefed. The dataset which are used in the classifier is discussed in the paper. We have analyzed and compare their accuracies using different types classifiers and from comparison table we found that the LSTM classifier gives better accuracies compare to other classifiers that is LSTM gives 96.33(%) in German dataset and 96.32(%) in Australian dataset.

10.2 FUTURE WORK

The methods studied in this paper will be extended to online learning models. In addition, other online learning models will be investigated. The use of online learning will enable rapid detection of fraud cases, potentially in real-time. This in turn will help detect and prevent fraudulent transactions before they take place, which will reduce the number of losses incurred every day in the financial sector.

APPENDIX

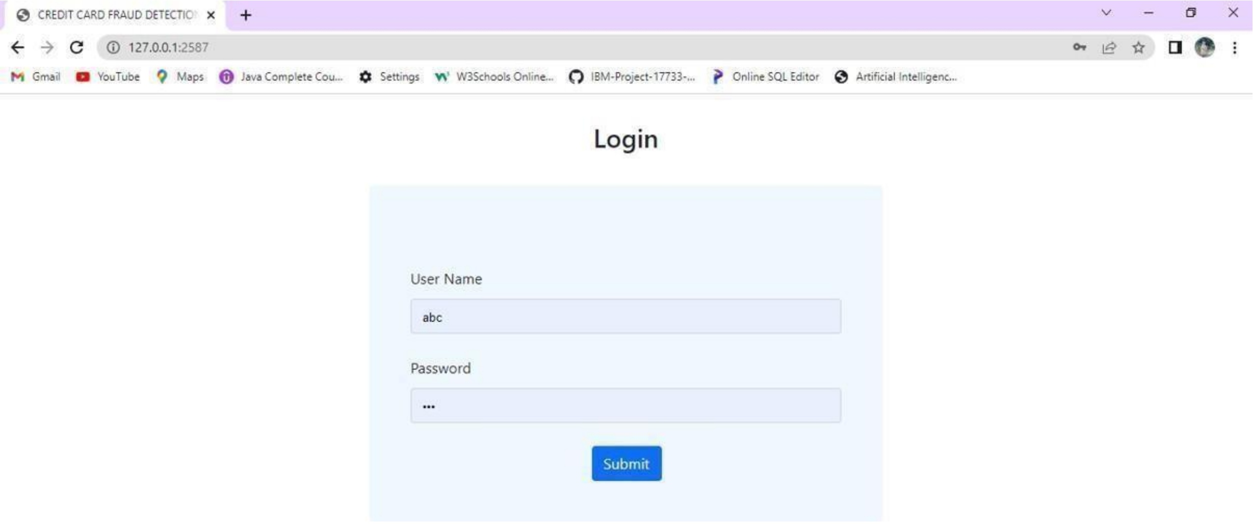
SOURCE CODE

```
from flask import Flask, render_template, request
importjoblib
import pandas as pd
import random
app = Flask(__name__)
df = pd.read_csv('final.csv')
df2 = pd.read_csv('creditcard.csv')

@app.route('/')
def index():
returnrender_template('login.html')


@app.route('/validate', methods = ['POST','GET'])
def validate():
ifrequest.method == 'POST':
ifrequest.form.get('uname') == 'abc' and request.form.get('upass') == '123':
```

```
returnrender_template('index.html')

else:

returnrender_template('login.html',  msg =  'Invalid')



@app.route('/model')

def model():

model =  joblib.load('model.pkl')




@app.route('/predict', methods = ['POST','GET'])

def predict():

ifrequest.method == 'POST':

card_no = int(request.form.get('card_no'))

        df1 = df.loc[df['Card  No'] == card_no]

location  =  df1['Location']

fori in location:
```

```
location = i

classes = df1['Class']

value = ['Fraud','Non Fraud']

fori in classes:

        c = i

     c = value[c]


returnrender_template('result.html',card = card_no, cla = c, las =
(random.randint(20,50))*100, loc = location)


if__name__== '__main__':

app.run(debug = True)
```

SCREENSHOTS



Login

User Name

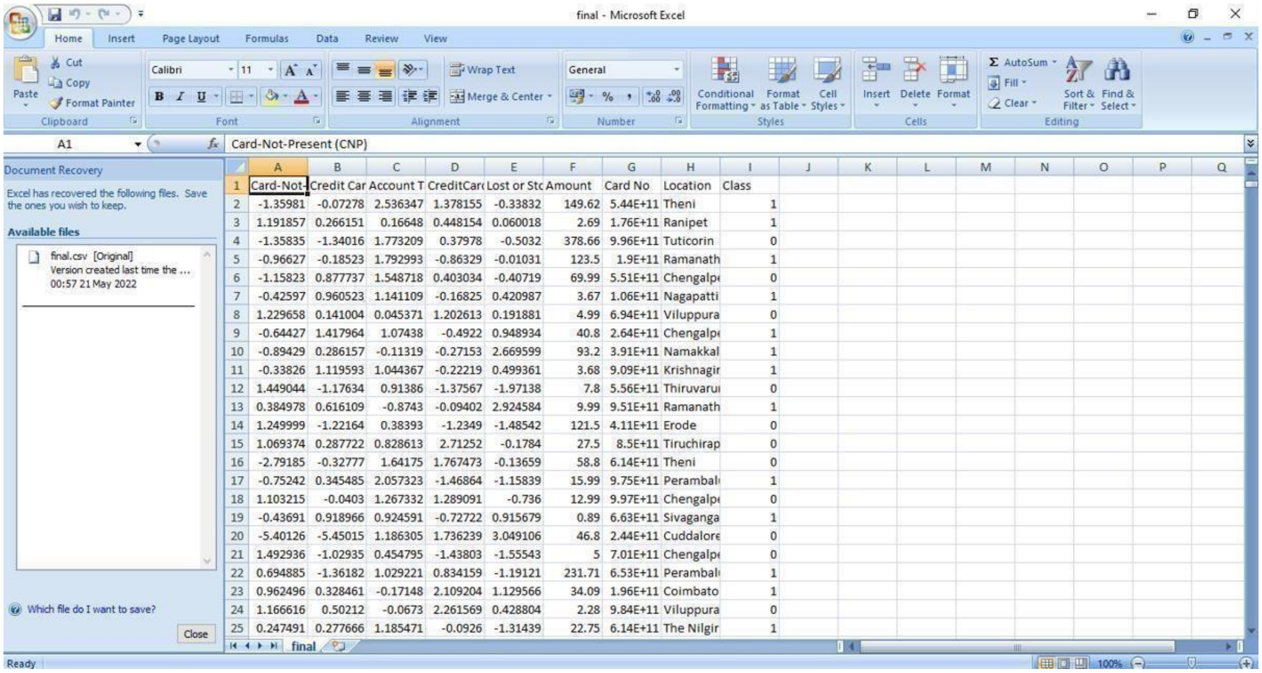
abc

Password

•••

Submit

A.2.1Credit card fraud Login Page .



Document Recovery

Excel has recovered the following files. Save the ones you wish to keep.

Available files

final.csv [Original]
Version created last time the ...
00:57 21 May 2022

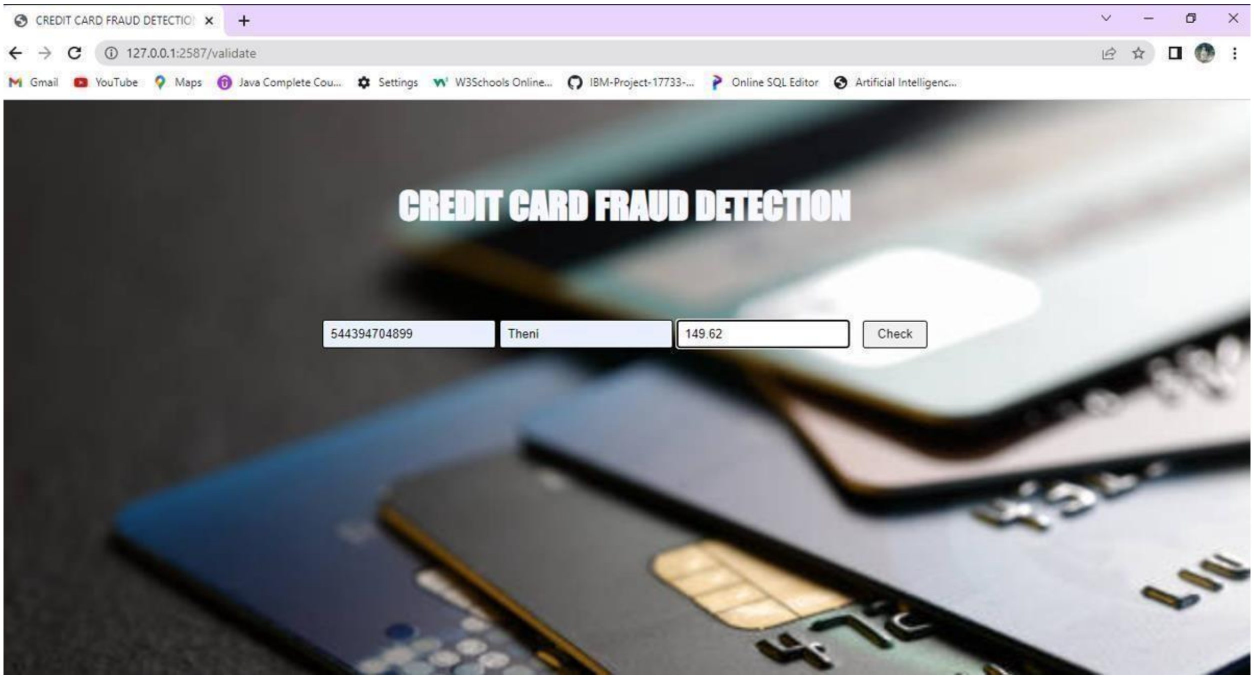
Which file do I want to save?

Close

Card-Not-Present (CNP)

	A	B	C	D	E	F	G	H	I
1	Card-Not-	Credit Car	Account T	CreditCar	Lost or Stc	Amount	Card No	Location	Class
2	-1.35981	-0.07278	2.536347	1.378155	-0.33832	149.62	5.44E+11	Theni	1
3	1.191857	0.266151	0.16648	0.448154	0.060018	2.69	1.76E+11	Ranipet	1
4	-1.35835	-1.34016	1.773209	0.37978	-0.5032	378.66	9.96E+11	Tuticorin	0
5	-0.96627	-0.18523	1.792993	-0.86329	-0.01031	123.5	1.9E+11	Ramanath	1
6	-1.15823	0.877737	1.548718	0.403034	-0.40719	69.99	5.51E+11	Chengalp	0
7	-0.42597	0.960523	1.141109	-0.16825	0.420987	3.67	1.06E+11	Nagapatti	1
8	1.229658	0.141004	0.045371	1.202613	0.191881	4.99	6.94E+11	Viluppura	0
9	-0.64427	1.417964	1.07438	-0.4922	0.948934	40.8	2.64E+11	Chengalp	1
10	-0.89429	0.286157	-0.11319	-0.27153	2.669599	93.2	3.91E+11	Namakkal	1
11	-0.33826	1.119593	1.044367	-0.22219	0.499361	3.68	9.09E+11	Krishnagir	1
12	1.449044	-1.17634	0.91386	-1.37567	-1.97138	7.8	5.56E+11	Thiruvaru	0
13	0.384978	0.616109	-0.8743	-0.09402	2.924584	9.99	9.51E+11	Ramanath	1
14	1.249999	-1.22164	0.38393	-1.2349	-1.48542	121.5	4.11E+11	Erode	0
15	1.069374	0.287722	0.828613	2.71252	-0.1784	27.5	8.5E+11	Tiruchirap	0
16	-2.79185	-0.32777	1.64175	1.767473	-0.13659	58.8	6.14E+11	Theni	0
17	-0.75242	0.345485	2.057323	-1.46864	-1.15839	15.99	9.75E+11	Perambal	1
18	1.103215	-0.0403	1.267332	1.289091	-0.736	12.99	9.97E+11	Chengalp	0
19	-0.43691	0.918966	0.924591	-0.72722	0.915679	0.89	6.63E+11	Sivaganga	1
20	-5.40126	-5.45015	1.186305	1.736239	3.049106	46.8	2.44E+11	Cuddalore	0
21	1.492936	-1.02935	0.454795	-1.43803	-1.55543	5	7.01E+11	Chengalp	0
22	0.694885	-1.36182	1.029221	0.834159	-1.19121	231.71	6.53E+11	Perambal	1
23	0.962496	0.328461	-0.17148	2.109204	1.129566	34.09	1.96E+11	Coimbato	1
24	1.166616	0.50212	-0.0673	2.261569	0.428804	2.28	9.84E+11	Viluppura	0
25	0.247491	0.277666	1.185471	-0.0926	-1.31439	22.75	6.14E+11	The Nilgir	1

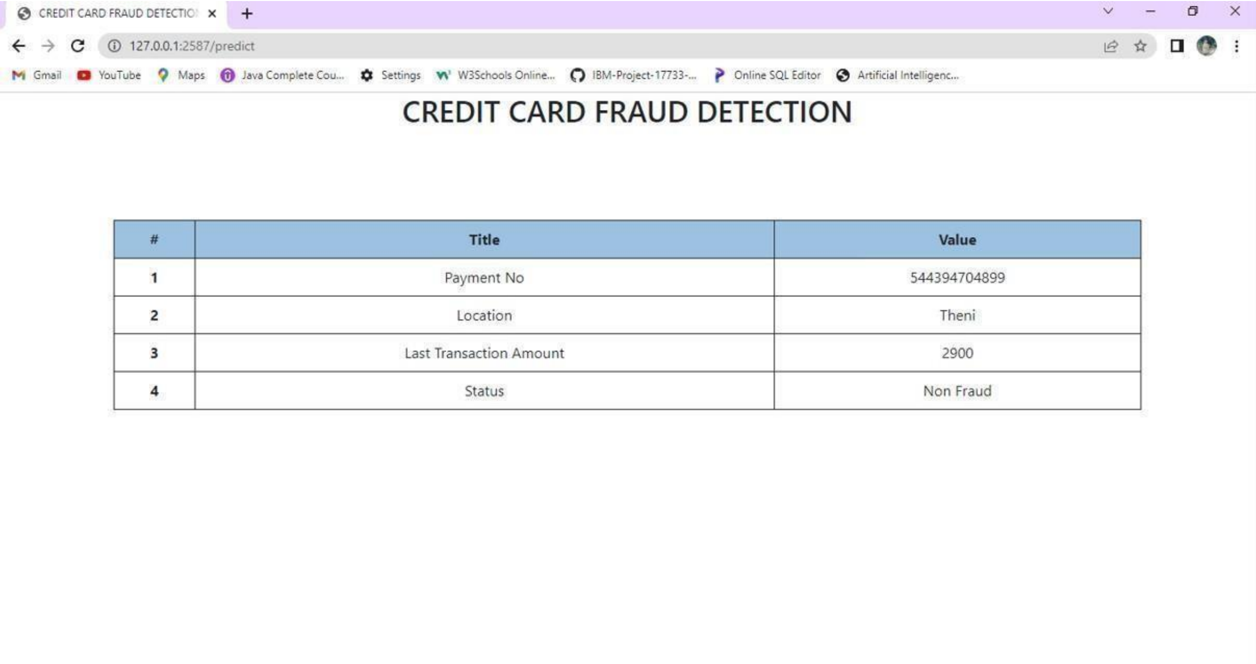
A.2.2 Data Set.



CREDIT CARD FRAUD DETECTION

544394704899 | Theni | 149.62

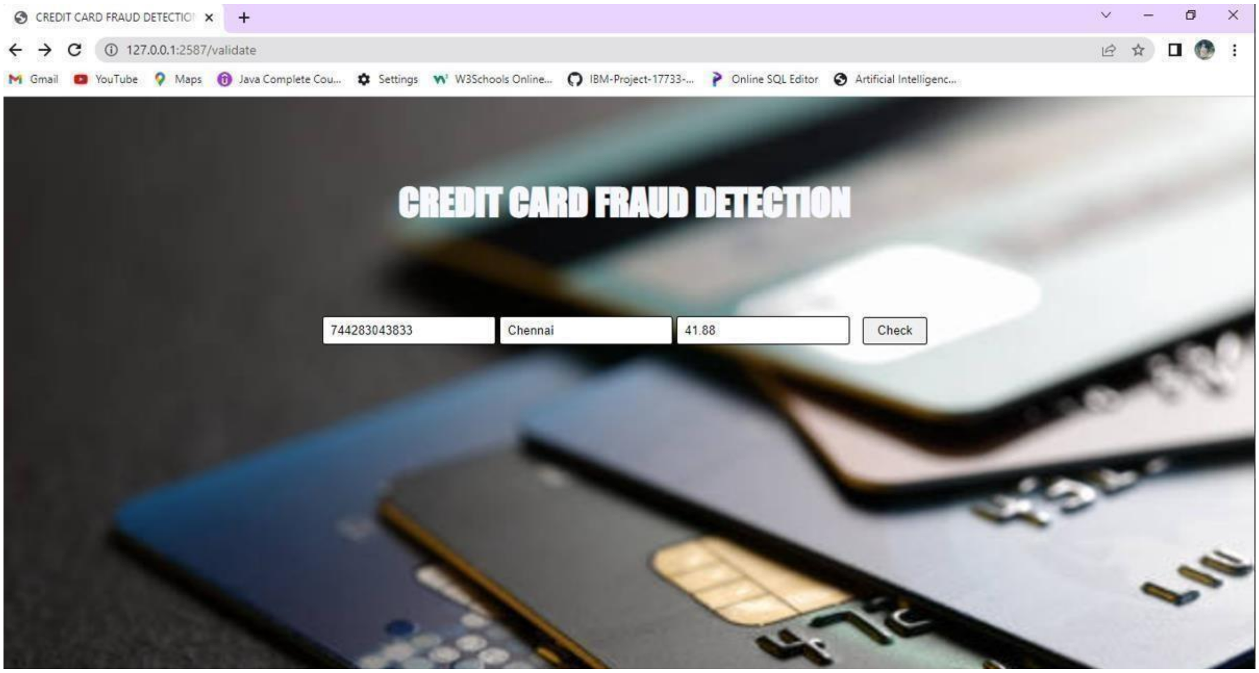
A.2.3 Enter the Credit card no , Location, Amount.



CREDIT CARD FRAUD DETECTION

#	Title	Value
1	Payment No	544394704899
2	Location	Theni
3	Last Transaction Amount	2900
4	Status	Non Fraud

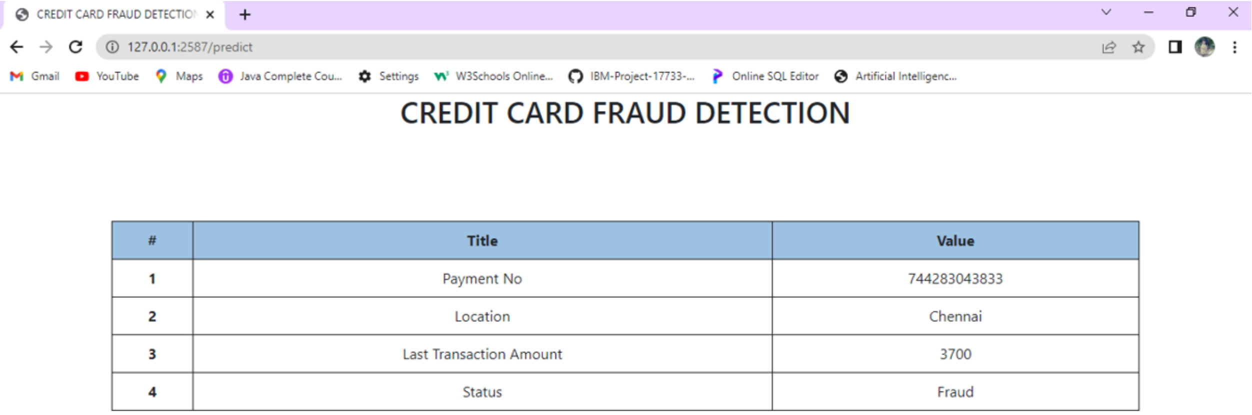
A.2.4 Find the Not Fraud**.**



CREDIT CARD FRAUD DETECTION

744283043833 | Chennai | 41.88 | Check

A.2.5 Enter the New Data Set , Credit card , location,Amount.



CREDIT CARD FRAUD DETECTION

#	Title	Value
1	Payment No	744283043833
2	Location	Chennai
3	Last Transaction Amount	3700
4	Status	Fraud

A.2.6 Find the Fraud .

REFERENCES

[1] A. O. Adewumi and A. A. Akinyelu, "A survey of machine-learning and nature-inspired based credit card fraud detection techniques," International Journal of System Assurance Engineering and Management, vol. 8, pp. 937– 953, 2017.

[2] C. F. Tsai, "Combining cluster analysis with classifier ensembles to predict financial distress" Information Fusion, vol. 16, pp. 46–58, 2014.

[3] E. Duman, A. Buyukkaya, and I. Elikucuk, "A novel and successful credit card fraud detection system Implemented in a Turkish Bank," In IEEE 13thInternationalConference on Data Mining Workshops (ICDMW), pp. 162–171, 2013.

[4] F. H. Chen, D. J. Chi, and J. Y. Zhu, "Application of Random Forest, Rough Set Theory, Decision Tree and Neural Network to Detect Financial Statement Fraud–Taking Corporate Governance into Consideration," In International Conference on Intelligent Computing, pp. 221–234, Springer, 2014.

[5] M. Seera, C. P. Lim, K. S. Tan, and W. S. Liew, "Classification of transcranial Doppler signals using individual and ensemble recurrent neural networks," Neurocomputing, vol. 249, pp. 337-344, 2017.

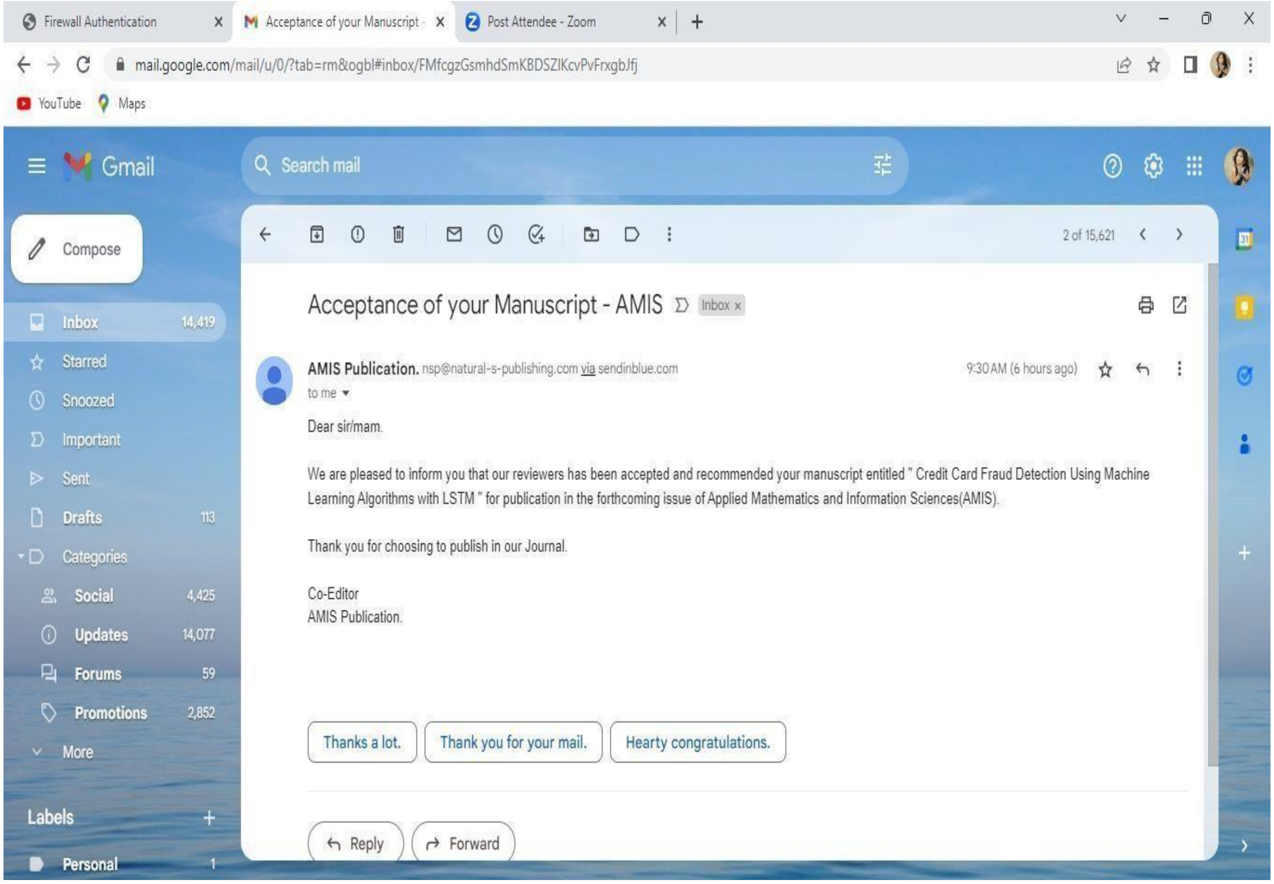
[6] N. Mahmoudi and E. Duman, "Detecting credit card fraud by modified Fisher discriminant analysis," Expert Systems with Applications, vol. 42, no. 5, pp. 2510–2516, 2015.

[7] R. Saia and S. Carta, "Evaluating Credit Card Transactions in the Frequency Domain for a Proactive Fraud Detection Approach," In Proceedings of the 14th International Joint Conference on e- Business and Telecommunications, vol. 4, pp. 335–342, 2017.

[8] Y. Li, C. Yan, W. Liu, and M. Li, "A principle component analysis- based random forest with the potential nearest neighbor method for automobile insurance fraud identification," Applied Soft Computing ,to be published. DOI: 10.1016/j.asoc.2017.07.027.

[9] Y. Sahin, S. Bulkan, and E. Duman, "A cost-sensitive decision tree approach for frauddetection," Expert Systems with Applications, vol. 40, no. 15, pp. 595923,2013.

JOURNAL PUBLISH STATUS



Acceptance of your Manuscript - AMIS

AMIS Publication. nsp@natural-s-publishing.com via sendinblue.com

to me

Dear sir/mam.

We are pleased to inform you that our reviewers has been accepted and recommended your manuscript entitled " Credit Card Fraud Detection Using Machine Learning Algorithms with LSTM " for publication in the forthcoming issue of Applied Mathematics and Information Sciences(AMIS).

Thank you for choosing to publish in our Journal.

Co-Editor
AMIS Publication.