Chapter-1

COMPANY PROFILE

Infidata Technologies is a dynamic start-up company. The foundation to Infidata technologies is built on the key focus area of domain expertise in industry verticals, customer focus by quickly aligning to client needs and company's unique delivery model is well executed through collaborative network of partners.

1.1 Introduction

Infidata technologies an ISO 9001: 2015 Certified Company (Accredited by International Accreditation Service: IAS) headquartered in "silicon valley" of India Bengaluru started in the year 2015. Company is highly specialized in the design and development of website, software application development, Mobile application development, E- Commerce solution and more. Company has a team of expert's professional works on the latest software tools and technologies to give the best and promising services to the customers. Company is equipped with the state-of-art work station in the software application development.

Vision

To become a leading performer and grow as a major IT service provider, in providing quality Web application, Software Development solutions and corporate training in the competitive global marketplace.

Mission

To ensure strategic planning with quality products and Profitable growth through customer service, innovation, quality and commitment.

1.2 Team

Infidata Technologies is a team of experienced professionals providing a wide range of complex software and web application development services. The energetic professionals with vast experience who are working in almost all technologies & spheres of IT.

1.3 Services

Infidata Technologies offer the services in the following areas:

- Enterprise Application Services

- Web designing and Development
- Mobile Application Development
- Internet of Things
- Training Services

1.3.1 Enterprise Application Services

An enterprise application is a software system platform designed to operate in a corporate environment such as business or government. EA software services include online shopping and payment processing, interactive product catalogs, computerized billing systems, security, content management, IT service management, content switching modules, resource planning, business intelligence, human resource management, manufacturing, application integration, forms automation, sales force automation, enterprise resource planning and business process management. EA also includes news relevant to IT compliance, business intelligence, office productivity suites, enterprise resource planning etc.

1.3.2 Web designing and Development

The web development process includes web design, web content development, client-side/server-side scripting and network security configuration, among other tasks.

Web development ranges from creating plain text pages to complex web-based applications, social network applications and electronic business applications.

The web development hierarchy is as follows:

- Client-side coding
- Server-side coding
- Database technology

1.3.3 Mobile Application Development

A mobile application, most commonly referred to as an app, is a type of application software designed to run on a mobile device, such as a smartphone or tablet computer.

Mobile applications frequently serve to provide users with similar services to those accessed on PCs. Apps are generally small, individual software units with limited function.

Vendors usually offer a mobile application platform to clients that want to go mobile or enter the mobile market. The platform includes migration tools and resources that support a mobile interface, or a development environment that allows creating new apps aimed at the Apple and Android markets. A platform approach to mobile application development helps provide a comprehensive model with tool suites that are accessible to developers and other users.

1.3.4 Internet of Things

Internet of things (IoT) is a computing concept that describes the idea of everyday physical objects being connected to the internet and being able to identify themselves to other devices. Most people think about being connected in terms of computers, tablets and smartphones. IoT describes a world where just about anything can be connected and communicate in an intelligent fashion. In other words, with the internet of things, the physical world is becoming one big information system.

1.3.5 Training Services

The training programs that are organized by us are of the best standards. We have some of the best trainers who conduct these programs. All these trainers have been in the industry for a long period. They know the problems faced by the industries. They have the total firsthand knowledge. Their rich experience is something that you would gain from. They can train students/ employee step by step. We offers training on Java, Web, C# & .Net, Python, Machine Learning, Mobile Application Development, Software Testing and More.

1.4 Company Products

- ECAMS
- Billing Software
- ERP Solution
- Retail Store

1.5 Our Clients

- Streams Inc
- Vishnu Enterprises
- Deepthi Engineering
- JB Transport India

- BreakDQ
- SJCPU
- Rbits Technologies

ABOUT THE DEPARTMENT

Infidata Technologies provided an internship in the department of Machine Learning division. The Objective is the development of new model and technology ensuring that the development of applications with the highest quality.

Machine Learning Department

Machine Learning organizes a practical procedure and approach in application development. Infidata technologies want to streamline their internal departments and functions, operations, sales and project management, etc. and want to take advantage of a ML based applications flexibility and versatility, by moving away from the traditional web application platform to the smart application platform and want to gain more clients for better service their current clients by offering convenient services and solutions online to build more smart applications to offer innovative services or solutions to online users and businesses.

The structure helps produce best practice coding with consistent logic and coding standards, and provides other developers the ability to become familiar with the code in a short time. Builds are based on the module, libraries and tools, allowing programmers to easily share libraries and implement complex functionalities and features in a fast and efficient manner.

Tips for ML application development

The following list of procedures and documents provide a good outline for a ML Application Lifecycle and Process:

1. Product Search

To give users relevant information according to their pursuit of the e-commerce app, our developers implement the whole set of ML tools such as ranking, query understanding and expansion related questions and so on.

For instance, for product ranking, we use customer information about the click-through rate or product sell-through rate. Additionally, we analyse behavioural data during

searching and the purchase process. Drawing on this, we create graphs between different goods and queries.

Another interesting tool is query intent detection. It comes from understanding the user`s portrait, his search history, and semantics outcome.

2. Product Recommendation and Promotions

The recommendation system is built on the collaborating filtering method. The App Solutions team together with our partner Soft cube provide clients with significant data service for smart recommendations and digital merchandising ("this item fits…").

The system is built upon the site content analysis, user behaviour or purchase patterns, and even upon the business logic of the enterprise. Predictive analytics makes the challenge easier, and recommendations become even more relevant with time. Such technology gives up to 7-12% from the same traffic.

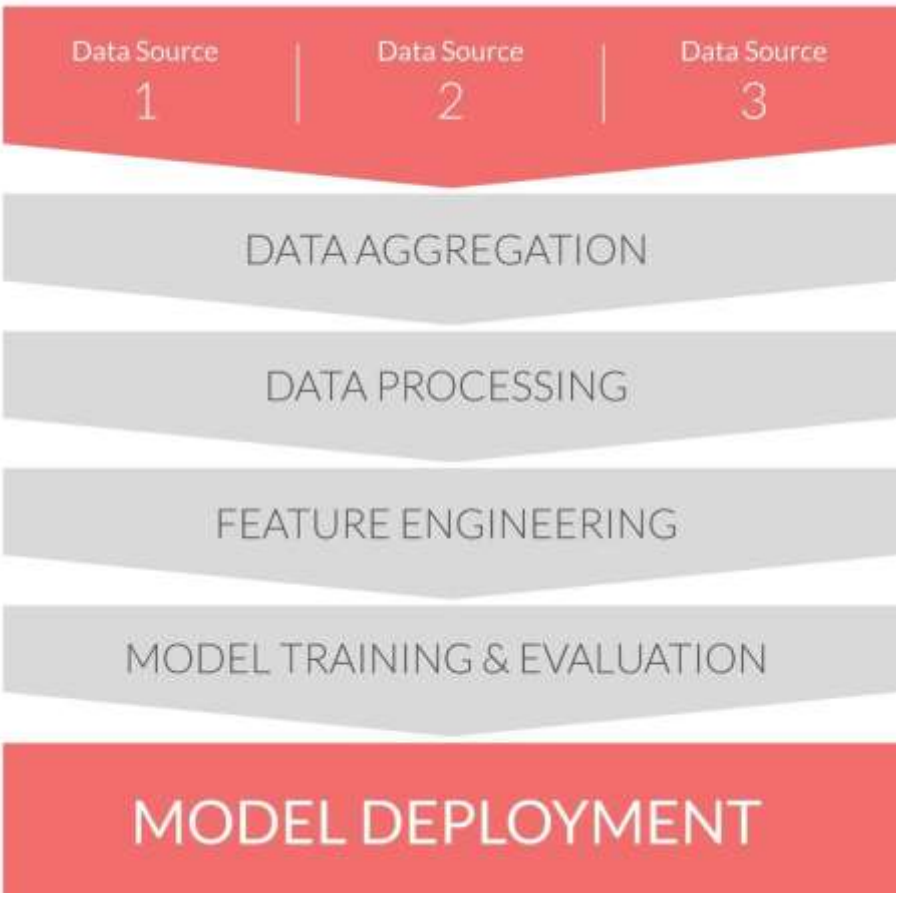
3. Trend forecasting and analytics

The e-Commerce enterprises, especially those, working in the fashion industry, always have a lack of information to understand and quickly respond to the latest trends. They have information about past season sales and upcoming tendencies. But between these two sources, there is a huge gap of missing opportunities.

Big data ML allows aggregating the trends and sales information from different open sources (inspirational blogs, social media, designer reports) and give predictions in real time. The same issue could be implemented in price management.

4. Fraud detection and prevention

One way or another, every e-Commerce company has faced this challenge. The annual fraud costs reached the point of $32 billion which is 38% more than the year before. Machine learning plays a critical role in building a defense system. It involves the ongoing monitoring of online activities and triggering of alarms. Here is the general workflow of "abnormal" behaviour patterns detection:


Data Source 1
Data Source 2
Data Source 3
DATA AGGREGATION
DATA PROCESSING
FEATURE ENGINEERING
MODEL TRAINING & EVALUATION
MODEL DEPLOYMENT

Chapter-2

INTERNSHIP DOMAIN

2.1 Introduction

The technology that promises to bring massive changes to the world next years is ML. Machine learning is a subfield of the Artificial Intelligence research and got the highest spotlight in business.

ML represents a new era in software development where computers, gadgets, and other devices do not require special programming to complete tasks anymore. Instead, they can collect and analyze information that is needed to draw appropriate conclusions and learn during program performance. Now machines can accumulate previous experience in order to make decisions as it occurs among human beings. Of course, the process of learning requires special algorithms that would “teach” machines. That is why, at The App Solutions, we use machine learning in mobile app development.

To understand the scale of ML industry, let's take a general outlook on the Artificial Intelligence market. According to Bank of America Merrill Lynch, over the next five years, the market will extend to $153 billion compared to $58 billion in 2014.

Venture Scanning gives an info graphic that summarizes the Artificial Intelligence market and shows funding of every category. The chart shows that ML applications category is leading with over $2 billion market share. This is three times more than the total funding of the next Natural Learning Processing group.

Basic Difference in ML and Traditional Programming?

- **Traditional Programming :**We feed in DATA (Input) + PROGRAM (logic), run it on machine and get output.
- **Machine Learning :**We feed in DATA(Input) + Output, run it on machine during training and the machine creates its own program(logic), which can be evaluated while testing.

What does exactly learning means for a computer?

A computer is said to be learning from Experiences with respect to some class of tasks, if its performance in a given Task improves with the Experience.

A computer program is said to learn from experience E 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

Example: playing checkers.

E = the experience of playing many games of checkers

T = the task of playing checkers.

P = the probability that the program will win the next game

In general, any machine learning problem can be assigned to one of two broad classifications:

Supervised learning and Unsupervised learning.

How things work in reality:-

Talking about online shopping, there are millions of users with unlimited range of interests with respect to brands, colors, price range and many more. While online shopping, buyers tend to search for a number of products. Now, searching a product frequently will make buyer's Facebook, web pages, search engine or that online store start recommending or showing offers on that particular product. There is no one sitting over there to code such task for each and every user, all this task is completely automatic.

Here, ML plays its role. Researchers, data scientists, machine learners build models on machine using good quality and huge amount of data and now their machine is automatically performing and even improving with more and more experience and time.

Traditionally, advertisement was only done using newspapers, magazines and radio but now technology has made us smart enough to do Targeted advertisement (online ad system) which is a way more efficient method to target most receptive audience.

Even in health care also, ML is doing a fabulous job. Researchers and scientists have prepared models to train machines for detecting cancer just by looking at slide – cell images. For humans to perform this task it would have taken a lot of time. But now, no more delay, machines predict the chances of having or not having cancer with some accuracy and doctors just have to give a assurance call, that's it. The answer to – how is this possible is very simple -all that is required, is, high computation machine, large amount of good quality image data, ML model with good algorithms to achieve state-of- the-art results.

Doctors are using ML even to diagnose patients based on different parameters under consideration.

You all might have use IMDB ratings, Google Photos where it recognizes faces, Google Lens where the ML image-text recognition model can extract text from the images you feed in, Gmail which categories Email as social, promotion, updates or forum using text classification, which is a part of ML.

How ML works?

Gathering past data in the form of text file, excel file, images or audio data. The more better the quality of data, the better will be the model learning Data Processing – Sometimes, the data collected is in the raw form and it needs to be rectified.

Example: if data has some missing values, then it has to be rectified. If data is in the form of text or images then converting it to numerical form will be required, be it list or array or matrix. Simply, Data is to be made relevant and understandable by the machine

Building up models with suitable algorithms and techniques and then training it.

Testing our prepared model with data which was not feed in at the time of training and so evaluating the performance – score, accuracy with high level of precision.

Prerequisites to learn ML:

- Linear Algebra
- Statistics and Probability
- Calculus
- Graph theory
- Programming Skills – Language such as Python, R, MATLAB, C++ or Octave

2.2 Tools Used

List of tools used during internship

- Pycharm-community-2021.3.2
- Python

Pycharm-community

PyCharm Community 2021.3.2 is an Integrated Development Environment (IDE) for Python programming.

It is a free and open-source version of PyCharm, which is developed by JetBrains.

PyCharm Community is designed for beginners and advanced Python developers alike, providing them with powerful tools to develop, debug, and maintain their Python projects.

Some of the key features of PyCharm Community 2021.3.2 include:

1. Code analysis and error highlighting: PyCharm Community 2021.3.2 has a powerful code analysis engine that can detect errors and highlight them in real-time. This helps developers to catch errors early and fix them before they become a problem.

2. Code completion and refactoring: PyCharm Community 2021.3.2 provides intelligent code completion and refactoring tools that help developers write clean, efficient, and maintainable code. These tools can automatically suggest code improvements and refactoring, saving developers time and effort.

3. Debugging and testing: PyCharm Community 2021.3.2 provides a powerful debugger that can be used to step through code, set breakpoints, and inspect variables. It also has built-in support for testing frameworks like pytest, unittest, and doctest.

4. Version control integration: PyCharm Community 2021.3.2 has built-in support for version control systems like Git, Mercurial, and Subversion. This allows developers to easily manage and track changes to their code.

5. Cross-platform support: PyCharm Community 2021.3.2 is available for Windows, macOS, and Linux, making it accessible to developers on different platforms.

6. Customizable interface: PyCharm Community 2021.3.2 has a customizable interface that allows developers to customize the look and feel of the IDE to suit their preferences.

Overall, PyCharm Community 2021.3.2 is a powerful IDE for Python development, and it provides a rich set of features that make Python development more efficient and enjoyable**.**

Python

Python is an interpreted, high-level, general-purpose programming language. Created by Guido van Rossum and first released in 1991, Python has a design philosophy that emphasizes code readability, notably using significant whitespace. It provides constructs that enable clear programming on both small and large scales. Van Rossum led the language community until stepping down as leader in July 2018.

Python features a dynamic type system and automatic memory management. It supports multiple programming paradigms, including object-oriented, imperative, functional and procedural, and has a large and comprehensive standard library.

Python interpreters are available for many operating systems. CPython, the reference implementation of Python, is open source software and has a community-based development model, as do nearly all of Python's other implementations. Python and CPython are managed by the non-profit Python Software Foundation.

It is used for:

- Web development (server-side),
- Software development,
- Mathematics,
- System scripting.

What can Python do?

- Python can be used on a server to create web applications.
- Python can be used alongside software to create workflows.
- Python can connect to database systems. It can also read and modify files.
- Python can be used to handle big data and perform complex mathematics.
- Python can be used for rapid prototyping, or for production-ready software development.

Why Python?

- Python works on different platforms (Windows, Mac, Linux, Raspberry Pi, etc).
- Python has a simple syntax similar to the English language.
- Python has syntax that allows developers to write programs with fewer lines than some other programming languages.
- Python runs on an interpreter system, meaning that code can be executed as soon as it is written. This means that prototyping can be very quick.
- Python can be treated in a procedural way, an object-orientated way or a functional way.

Good to know

- The most recent major version of Python is Python 3, which we shall be using in this tutorial. However, Python 2, although not being updated with anything other than security updates, is still quite popular.
- In this tutorial Python will be written in a text editor. It is possible to write Python in an Integrated Development Environment, such as Thonny, Pycharm, Netbeans or Eclipse which are particularly useful when managing larger collections of Python files.

Python Syntax compared to other programming languages

- Python was designed to for readability, and has some similarities to the English language with influence from mathematics.
- Python uses new lines to complete a command, as opposed to other programming languages which often use semicolons or parentheses.
- Python relies on indentation, using whitespace, to define scope; such as the scope of loops, functions and classes. Other programming languages often use curly-brackets for this purpose.

Chapter-3

TASK PERFORMED

PROJECT TITLE:

3.1 Introduction to Project

"Loan Prediction" is very helpful for employee of banks as well as for the applicant also. The aim of this Project is to provide quick, immediate and easy way to choose the deserving applicants. Loan Prediction System is a software which checks the eligibility of a particular customer who is capable of paying loan or not.

They have presence across all urban, semi urban and rural areas. Customer first apply for loan after that company or bank validates the customer eligibility for loan. Company or bank wants to automate the loan eligibility process (real time) based on customer details provided while filling application form. These details are Gender, Marital Status, Education, Number of Dependents, Income, Loan Amount, Credit History and other.

This project has taken the data of previous customers of various banks to whom on a set of parameters loan were approved. So the machine learning model is trained on that record to get accurate results. Our main objective of this project is to predict the safety of loan. This model is applied on the test data set for getting required output. The output generated will be in the form of yes or no. Yes indicates that a particular customer is capable of paying loan and No indicates that the particular customer is not capable of paying loan. Based on these factors we can approve loans for customers.

The Loan Prediction project using Decision Tree Algorithm with Loan dataset can be broken down into the following steps:

1. Data Collection: The first step is to collect data on applicant who have been applied for loan. The data should include the objective of the data is to use Gender, Marital Status, Education, Number of Dependents, Income, Loan Amount, Credit History, and other factors and predict the approval probability of each application. There are 614 values in this dataset. The output generated will be in the form of yes or no.

2. Data Preprocessing: The collected data needs to be preprocessed to clean and transform it into a usable format. This step involves removing duplicates, filling missing values, and

encoding categorical variables.

3. Data Splitting: The preprocessed data is split into training and testing datasets. The training dataset is used to train the Decision Tree model, and the testing dataset is used to evaluate the model's performance.

4. Decision Tree Model Training: The Decsion Tree model is trained on the training dataset using the available features (Gender, Marital Status, Education, Income, Loan Amount, Credit History) and the target variable (whether the applicant having a Loan or not).

5. Model Evaluation: The trained model is evaluated on the testing dataset to measure its accuracy and performance.

6. Prediction: Once the model is trained and evaluated, it can be used to predict whether a new female patient is likely to have diabetes or not based on their age, BMI, blood pressure, and glucose levels.

Overall, Loan approval is a very important process for banking organizations. The system approved or reject the loan applications. Recovery of loans is a major contributing parameter in the financial statements of a bank. It is very difficult to predict the possibility of payment of loan by the customer. Using Machine learning we predict the loan approval. This project can be further improved by collecting more data from diverse sources and using more advanced Machine Learning algorithms.

3.2 Software/Hardware Requirements

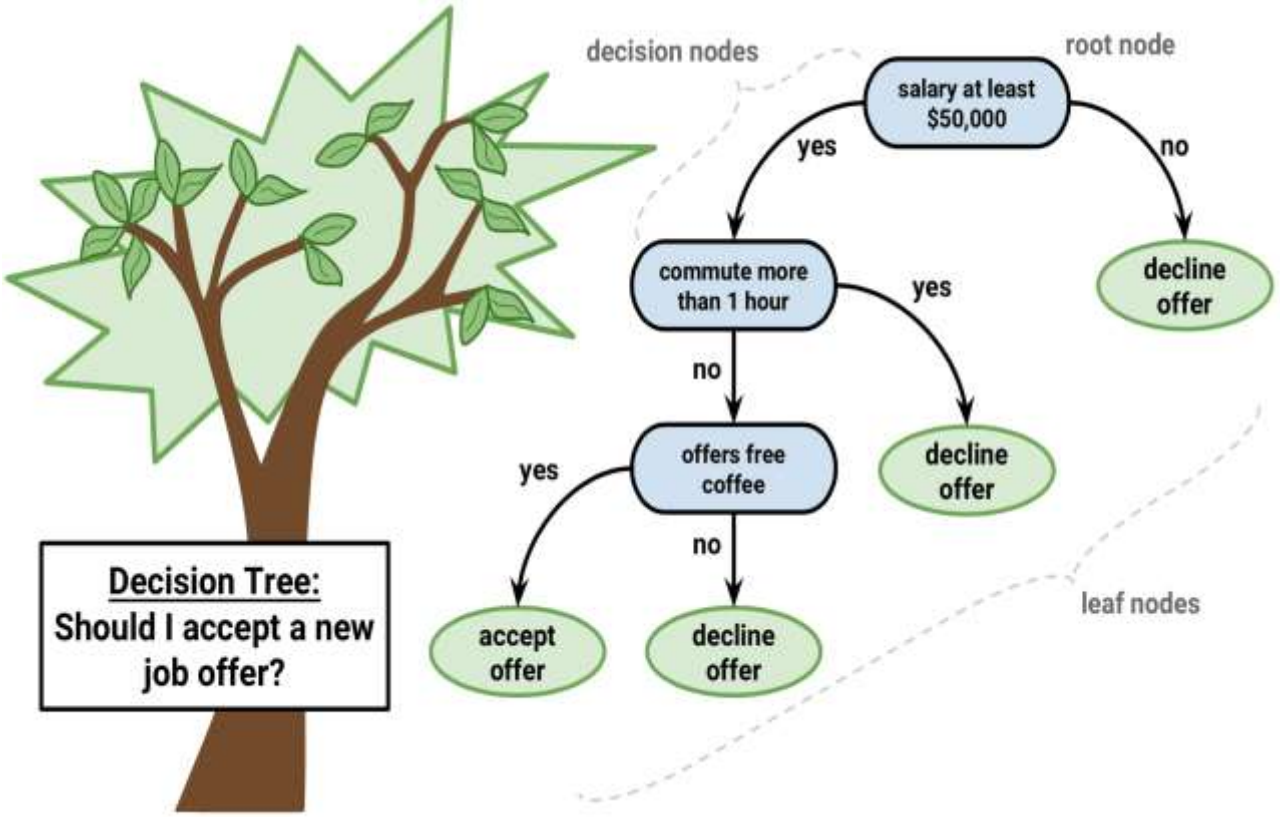
Software Requirement:

- Python 3.6/3.7
- Pycharm-community-2021.3.2

Hardware Requirement:

- Minimum PC Requirements

3.3. Modules Description



A decision tree is a type of supervised learning algorithm used in machine learning and data mining. It is a tree-like model where each node represents a feature or attribute, each branch represents a decision rule, and each leaf node represents a class label or a decision.

What is Decision Tree Algorithm?

The decision tree algorithm is useful for both classification and regression problems. In a classification problem, the tree predicts the class label of the input data, whereas in a regression problem, the tree predicts the numerical value of the output variable.

The decision tree algorithm builds the tree by recursively splitting the data based on the most significant attribute or feature that best separates the data into different classes or categories. The splitting process continues until all the data in a particular branch belongs to the same class or there are no more features to split on.

Decision Tree pseudocode:

```
function DecisionTree(trainingData, targetAttribute):

    if all instances in trainingData belong to the same class:
        return a leaf node with the class label

    if trainingData is empty:
        return a leaf node with the most common class label in the parent node

    if targetAttribute is empty:
        return a leaf node with the most common class label in trainingData

    else:
        select the best attribute to split on based on information gain
        create a new internal node with the selected attribute
        split the trainingData into subsets based on the values of the selected attribute
        for each subset:
            create a child node by recursively calling DecisionTree with the subset of data and
targetAttribute - {selected attribute}
        return the new internal node
```

This pseudocode assumes that the input trainingData is a set of labeled examples, where each example has a set of attributes and a class label. The targetAttribute is the attribute we want to predict. The algorithm recursively constructs a decision tree that partitions the data based on the values of the attributes and their relationship to the target attribute.

The information gain heuristic is used to select the best attribute to split on at each step

Decision Tree prediction pseudocode:

```
function predict(decision_tree, example):
    node = decision_tree.root
    while node is not a leaf node:
        feature = node.feature
        if example[feature] < node.threshold:
            node = node.left_child
        else:
            node = node.right_child
    return node.value
```

Here, **decision_tree** is the decision tree model, and **example** is a new data point that we want to predict. We start at the root node of the tree and traverse down the tree until we reach a leaf node. At each node, we check the value of the corresponding feature in the example data point against the threshold value of the node.

If the feature value is less than the threshold, we move to the left child node, and if it is greater than or equal to the threshold, we move to the right child node. We repeat this process until we reach a leaf node, which contains the predicted value for the example.

Note that this is a basic pseudocode, and the implementation details may vary depending on the specific decision tree algorithm and programming language used

Chapter-4

REFLECTION NOTES

I had been interned in the course and I have a very nice experience to share. The team was amazing and flexible. They are ready to help us at our convenience. There was lot to know about the subject, sensors in specific. To me from CS background they made it as a very easy task. I want to thank everyone who helped me to make it possible.”

"I thoroughly enjoyed my internship this summer and now have very valuable experience under my belt. I know this will help when looking for jobs and needing references. I was dreading it in the beginning, but now I am so happy it was required. As much as the curriculum changes, I hope that class remains constant."

"We all know that practical experience is the best, and internships give students that hands-on experience they need. I feel that quality internships are essential to develop key skills that you can't get in a classroom. Skills such as multitasking, communicating, learning to deal with diversity, and dealing with deadlines are different when you are working for someone else, not yourself like you do in college. Internships are also a great way to network with people in the industry. My boss and co-workers were great about giving me contacts and referring me to open positions in the industry."

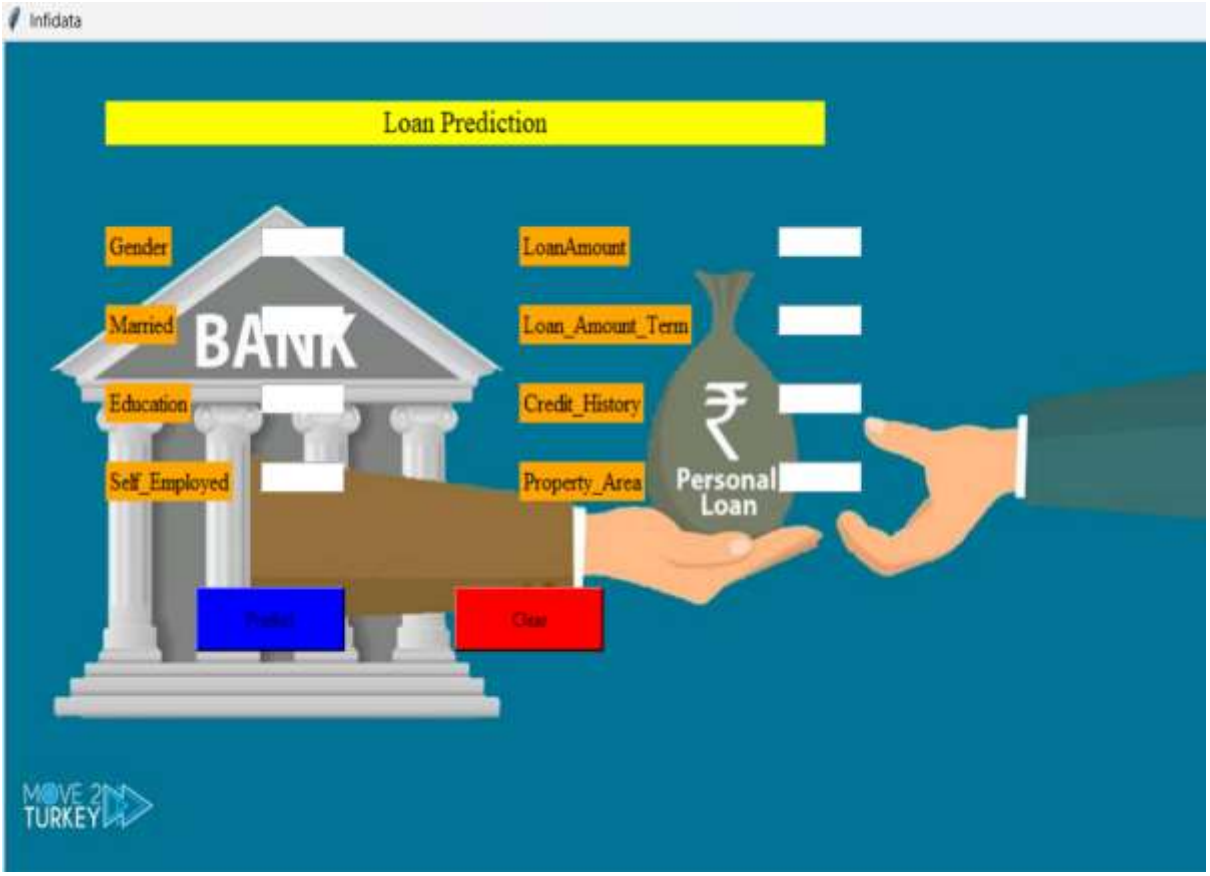
"I learned that customer service is hard. You can not please everyone. There are people in this world that are simply difficult. I have learned that stressing over little things will not get me anywhere. I have learned to work well as a team and that without my counterparts the work would not get done. Another aspect that I learned throughout my internship is to never be afraid to ask lots of questions. By asking questions I got answers."

"I learned more than I ever expected. I felt in the beginning that I would really have a head start being that I am a senior and have taken all my classes already but I realized that I don’t think any amount of classes will ever prepare you for all the world has to offer. Don’t get me wrong, I feel that Auburn did a wonderful job teaching me but I feel that it is most valuable if combined with first hand work .

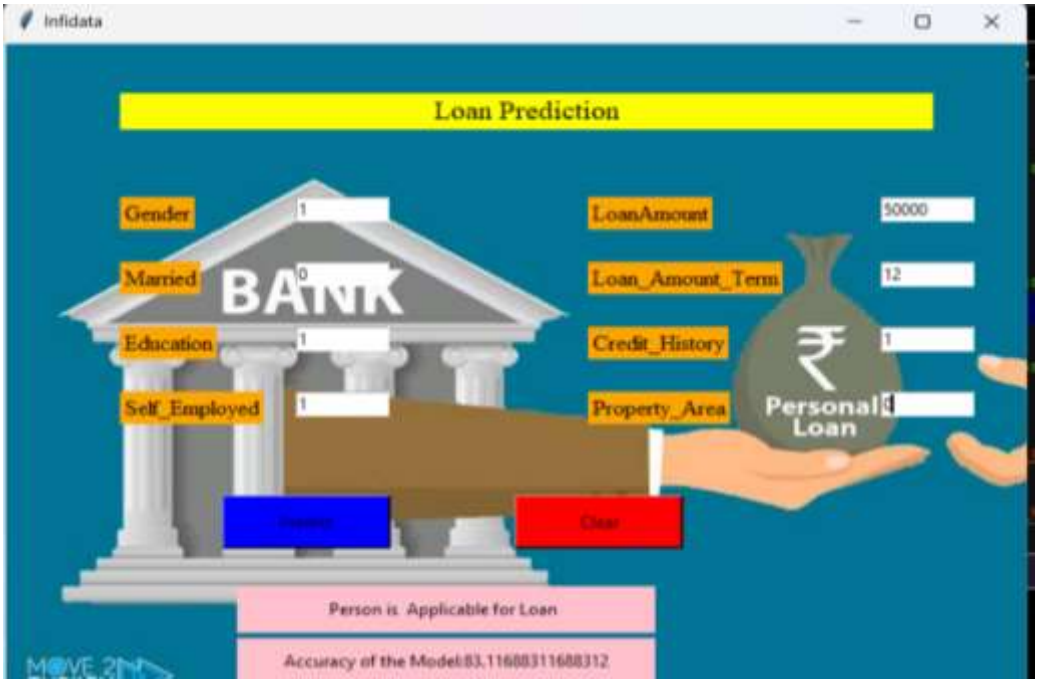
I feel that during my time interning my most challenging part of my work was having to share my tasks and coordinate my work with others. I have always been kind of a perfectionist and liked to have things done my way so working with others and sharing duties was a slight change I must say. It took me a couple of weeks to really feel comfortable allowing someone else to share my work but eventually I learned it was ok and that two heads were better than one.

Chapter-5

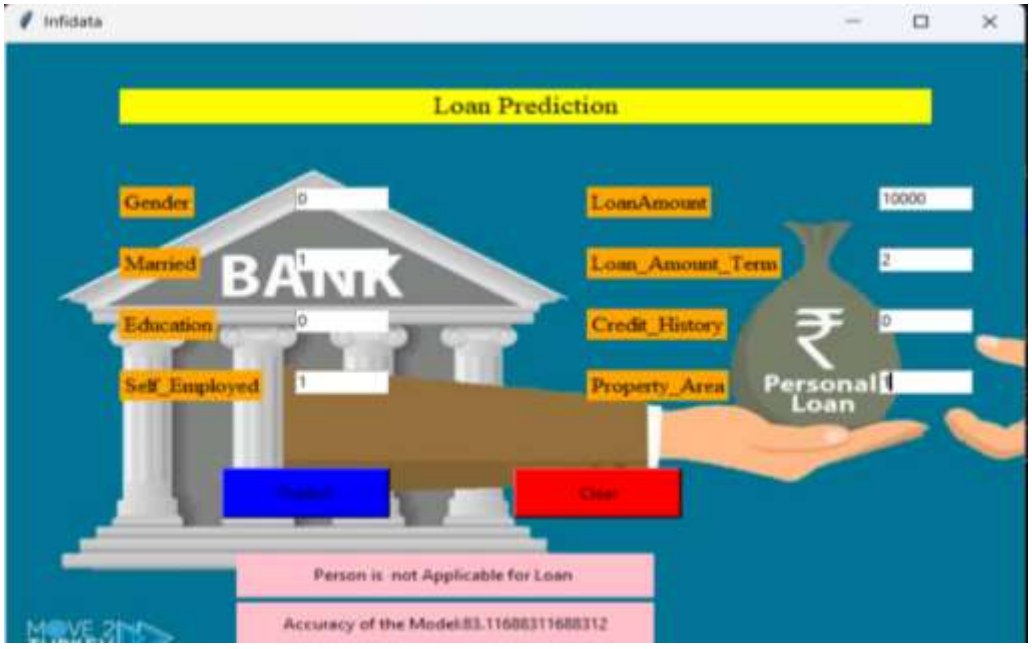
RESULTS AND INTERNSHIP OUTCOMES



Interface of the GUI



The GUI when applicant applicable for loan



The GUI when applicant is Not applicable for loan.

Chapter-6

CONCLUSION

Machine learning and artificial intelligence essentially empower the mobile application with enough customization features to make it more useful, efficient and effective. These technologies in accordance with the latest style are fast and safe. So it is a good idea to employ these technologies for your next mobile app development project. It is one of those defined limits that differentiate your application from that of your competitors. Now if you are looking to develop a mobile application with the necessary impulse of AI and ML for your business and make it jump.

There is likely room for further improvement, but this is a big improvement over the best decision tree error of 250,000. There are parameters which allow you to change the performance of the Random Forest much as we changed the maximum depth of the single decision tree. But one of the best features of Random Forest models is that they generally work reasonably even without this tuning.