**BANKING CUSTOMER AND LOAN PREDICTION SYSTEM**

A Dissertation Submitted for the Fulfilment of the Requirements for the project report of

**MASTER OF COMPUTER APPLICATION**

Submitted by

**Pranab Bora**

**Roll no-18MCA012**

Under the supervision and guidance of

**Prof. Utpal Bhattacharjee**

Head of the Department



**DEPARTMENT OF COMPUTER SCIENCE & ENGINEERING**

**Rajiv Gandhi University**

**(**A central University**)**

**Rono Hills, Doimukh-791112**

**ARUNACHAL PRADESH, INDIA**

**June-2021**



**DECLARATION**

I hereby declare that the dissertation work entitled “**BANKING CUSTOMER & LOAN PREDICTION SYSTEM”** submitted to the Department of computer science & engineering, Rajiv Gandhi University, is prepared by me for 6th semester Master of Computer Application in the dept. of computer science and engineering and was not submitted to any other institute for any other degree.

Pranab Bora

Roll no.- 18MCA012

Dept. of Computer Science & Engg.

Rajiv Gandhi University



**CERTIFICATE**

This is to certify that the project work entitled “**BANKING CUSTOMER & LOAN PREDICTION SYSTEM**”, is carried out by PRANAB BORA (18MCA012) under the guidance of Prof. UTPAL BHATTACHARJEE has been found satisfactory, and is hereby approved as a project work carried out and presented in a manner required for its acceptance in fulfilment of the requirements of MCA 6th semester Major Project work for the degree of Master of Computer Application Under Rajiv Gandhi University.

Signature of guide

Prof. Utpal Bhattacharjee



**CERTIFICATE**

This is to certify that the project work entitled “**BANKING CUSTOMER & LOAN PREDICTION SYSTEM**”, is carried out by PRANAB BORA (18MCA012) under the guidance of Prof. UTPAL BHATTACHARJEE has been found satisfactory, and is hereby approved as a project work carried out and presented in a manner required for its acceptance in fulfilment of the requirements of MCA 6th semester Major Project work for the degree of Master of Computer Application Under Rajiv Gandhi University.

Head of the Dept.

Prof. Utpal Bhattacharjee



**CERTIFICATE**

This project entitled “**BANKING CUSTOMER & LOAN PREDICTION SYSTEM**” submitted By “PRANAB BORA” (18MCA012) for fulfilment of the requirements of Master of Computer Application 6th Semester project report in computer science & engineering has been examined.

External Examiner

Date:

Place: Doimukh

**ACKNOWLEDGEMENT**

I would like to express my sincere thanks to my project guide, Prof. Utpal Bhattacharjee(Head of the Dept, CSE) for his constant support and making himself available whenever I sought him. Without his guidance and encouragement the progress of my project would have been suffered.

I would like to thank our project coordinator Dr. Firos A sir for giving us moral support and facilities. I would like to extend my gratitude to the entire faculty. Lastly but not the least I would like to extend my gratitude to all our friend of computer science and engineering department, Rajiv Gandhi University for their constant support and help.

Name: Pranab Bora

Roll no: 18MCA012

MCA 6th semester

Department of computer science and engineering

**ABSTRACT**

With the enhancement in the banking sector lots of people are applying for bank loans but the bank has its limited assets which it has to grant to limited people only, so finding out to whom the loan can be granted which will be a safer option for the bank is a typical process. So in this project we try to reduce this risk factor behind selecting the safe person so as to save lots of bank efforts and assets. This is done by mining the Big Data of the previous

records of the people to whom the loan was granted before and on the basis of these records/experiences the machine was trained using the machine learning model which give the most accurate result. The main objective of this project is to predict whether assigning the loan to particular person will be safe or not. This paper is divided into four sections

(i)Data Collection

(ii) Comparison of machine learning models on collected data

(iii) Training of system on most promising model

(iv) Testing.

In this paper we are predict the loan data by using some machine

learning algorithms they are classification, logic regression, Decision Tree and gradient boosting.

Contents

[Chapter 1: INTRODUCTION 1](#_Toc74943519)

[*1.1 Background* 1](#_Toc74943520)

[*1.2 Objective* 1](#_Toc74943521)

[*1.3 Purpose* 1](#_Toc74943522)

[*1.4 Data* 1](#_Toc74943523)

[*1.5 Literature review* 2](#_Toc74943524)

[*1.6 Scope* 2](#_Toc74943525)

[*1.7 Features* 2](#_Toc74943526)

[Chapter 2: SURVEY OF TECHNOLOGY 3](#_Toc74943527)

[*2.1 Open Source Software* 3](#_Toc74943528)

[*2.2 GNU License* 3](#_Toc74943529)

[*2.3 Used Technology* 3](#_Toc74943530)

[*2.3.1 Python* 3](#_Toc74943531)

[*2.3.2 Anaconda* 4](#_Toc74943532)

[*2.3.3 Numpy* 4](#_Toc74943533)

[*2.3.4 Pandas* 4](#_Toc74943534)

[*2.3.4 Matplotlib* 4](#_Toc74943535)

[*2.3.5 Open CV* 4](#_Toc74943536)

[*2.3.6 Scikit-learn* 5](#_Toc74943537)

[*2.3.7 Keras* 5](#_Toc74943538)

[*2.3.8 TensorFlow* 5](#_Toc74943539)

[*2.4 Computer Vision* 5](#_Toc74943540)

[Chapter 3: REQUIREMENT AND ANALYSIS 6](#_Toc74943541)

[*3.1 Requirement Specification* 6](#_Toc74943542)

[*3.2 Project planning and scheduling* 6](#_Toc74943543)

[*3.3 Software and hardware requirement* 6](#_Toc74943544)

[*3.3.1 Software requirement* 6](#_Toc74943545)

[*3.3.2Hardware requirement:* 6](#_Toc74943546)

[Chapter 4: ARTIFICIAL INTELLIGENCE 7](#_Toc74943547)

[*4.1 Introduction* 7](#_Toc74943548)

[*4.2 How AI used?* 7](#_Toc74943549)

[*4.3 Machine Learning & Deep Learning* 7](#_Toc74943550)

[*4.3 Neural Network* 8](#_Toc74943551)

[*4.4 Artificial Neural Network* 8](#_Toc74943552)

[Chapter 5: METHODOLOGY AND IMPLEMENTATION 10](#_Toc74943553)

[*5.1 Image Acquisition* 10](#_Toc74943554)

[*5.2 Image Preprocessing* 10](#_Toc74943555)

[*5.3 Shape Matching Based Detection* 10](#_Toc74943556)

[*5.4 Objects Features Analysing* 11](#_Toc74943557)

[*5.5 Architecture of the system* 11](#_Toc74943558)

[*5.6 Steps to Build the Project* 12](#_Toc74943559)

[*Step 1: Explore the dataset* 12](#_Toc74943560)

[*Step 2: Build a CNN model* 13](#_Toc74943561)

[*Step 3: Train and validate the model* 13](#_Toc74943562)

[*Step 4: Test our model with test dataset* 16](#_Toc74943563)

[Chapter 6: SNAPSHOTS 17](#_Toc74943564)

[Chapter 7: CONCLUSION 19](#_Toc74943565)

[*7.1 Conclusion* 19](#_Toc74943566)

[*7.2 Benefits* 19](#_Toc74943567)

[*7.3 Future scope* 19](#_Toc74943568)

[BIBLIOGRAPHY 20](#_Toc74943569)

# Chapter 1: INTRODUCTION

## *1.1 Background*

Banking Customer and Loan Prediction System is a system that can be used by bank employee to verify with the previous loan applicants data and whether the application was approved or not for the current customer. In this project , we will build a machine learning model to predict the loan approval probability.

## *1.2 Objective*

Following are the steps involved in creating a well-defined ML project:

* Understand and define the problem
* Analyse and prepare the data
* Apply the algorithms
* Reduce the errors
* Predict the result

## *1.3 Purpose*

A Company wants to automate the loan eligibility process (real time) based on customer detail provided while filling online application form. These details are Gender, Marital Status, Education, Number of Dependents, Income, Loan Amount, Credit History and others. To automate this process, they have given a problem to identify the customers segments, those are eligible for loan amount so that they can specifically target these customers. Here they have provided a data set.

## *Screenshot 2021-06-18 193011.png**1.4 Data*

## *1.5 Literature review*

Data mining is the process of analyzing data from different perspectives and extracting useful knowledge from it. It is the core of knowledge discovery process. There are various steps involved in extracting knowledge from raw data. Different data mining techniques include classification, clustering, association rule mining, prediction and sequential patterns, neural networks, regression etc. Classification is the most commonly applied data mining technique, which employs a set of pre-classified examples to develop a model that can classify the population of records at large. Fraud detection and credit risk applications are particularly well suited to classification technique. This approach frequently employs Decision tree based classification Algorithm. In classification, a training set is used to build the model as the classifier which can classify the data items into its appropriate classes. A test set is used to validate the model.

## *1.6 Scope*

It is widely used for managing risks in the banking industry. Bank executives need to know the credibility of customers they are dealing with. Offering new customers credit cards, extending existing customers’ lines of credit, and approving loans can be risky decisions for banks, if they do not know anything about their customers. Banks provide loans to their customers by verifying the various details relating to the loan, such as amount of loan, lending rate, repayment period etc. Even though, banks are cautious while providing loan, there are chances of loan repaying defaults by customers. Data mining technique helps to distinguish borrowers who repay loans promptly from those who default.

## *1.7 Features*

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

# Chapter 2: SURVEY OF TECHNOLOGY

## *2.1 Open Source Software*

**Open-source software** (**OSS**) is a type of computer software with its source code made available with a license in which the copyright holder provides the rights to study, change, and distribute the software to anyone and for any purpose. Open-source software may be developed in a collaborative public manner. According to scientists who studied it, open-source software is a prominent example of open collaboration. The term is often written without a hyphen as "open source software".

## *2.2 GNU License*

Here's a brief summary of what you need to do to release a program under one of our licenses:

* Get a copyright disclaimer from your employer or school.
* Give each file the proper copyright notices. Make sure to clearly identify which versions of the license users can use.
* Add a COPYING file with a copy of the GNU GPL or GNU AGPL.
* Also add a COPYING.LESSER file with a copy of the GNU LGPL, if you use that.
* Put a license notice in each file.
* (Optionally) make the program display a startup notice.
* (If using the AGPL) make the program offer copies of its source code.

## *2.3 Used Technology*

### *2.3.1 Python*

Python is a popular general-purpose programming language that can be used for a wide variety of applications. It includes high-level data structures, dynamic typing, dynamic binding, and many more features that make it as useful for complex application development as it is for scripting or "glue code" that connects components together. It can also be extended to make system calls to almost all operating systems and to run code written in C or C++. Python is a universal language found in a variety of different applications.

#### 2.3.1.1 About Python

First developed in the late 1980s by Guido van Rossum, Python has advanced as an open source programming language by managing public discussion through Python Enhancement Proposals (PEPs). In 2018, van Rossum stepped down as the language's Benevolent Dictator For Life (BDFL), and, as officially outlined in PEP 13, a steering council was put in place to serve as the leadership of the language.

The Python Software Foundation (PSF) is a 501(c)(3) non-profit corporation that holds the intellectual property rights behind the Python programming language. This includes Python version 2.1 and later, PyPI, the CPython reference implementation, and infrastructure to maintain the language. The PSF also provides grants for software craftship and runs multiple PyCon conferences a year.

Python is currently on its third major version and is regularly updated.

#### 2.3.1.2 Is Python open source?

Yes, all modern versions of Python are copyrighted under a GPL-compatible license certified by the Open Source Initiative.

### *2.3.2 Anaconda*

Anaconda Individual Edition is the world’s most popular Python distribution platform with over 20 million users worldwide. Anaconda has cloud-based repository to find and install over 7,500 data science and machine learning packages. With the conda-install command, we can start using thousands of open-source Conda, R, Python and many other packages.Individual Edition is an open source, flexible solution that provides the utilities to build, distribute, install, update, and manage software in a cross-platform manner. Conda makes it easy to manage multiple data environments that can be maintained and run separately without interference from each other.

Anaconda Navigator is a desktop GUI that comes with Anaconda Individual Edition. It makes it easy to launch applications and manage packages and environments without using command-line commands.

### *2.3.3 Numpy*

NumPy is a Python library used for working with arrays. It also has functions for working in domain of linear algebra, fourier transform, and matrices. NumPy was created in 2005 by Travis Oliphant. It is an open source project and you can use it freely. NumPy stands for Numerical Python.

### *2.3.4 Pandas*

Pandas is an open-source, BSD-licensed Python library providing high performance, easy-to-use data structures and data analysis tools for the Python programming language. Python with Pandas is used in a wide range of fields including academic and commercial domains including finance, economics, Statistics, analytics, etc. In this tutorial, we will learn the various features of Python Pandas and how to use them in practice.

### *2.3.4 Matplotlib*

Matplotlib is one of the most popular Python packages used for data visualization. It is a cross-platform library for making 2D plots from data in arrays. It provides an object-oriented API that helps in embedding plots in applications using Python GUI toolkits such as PyQt, WxPythonotTkinter. It can be used in Python and IPython shells, Jupyter notebook and web application servers also.

### *2.3.5 Open CV*

OpenCV is a cross-platform library using which we can develop real-time computer vision applications. It mainly focuses on image processing, video capture and analysis including features like face detection and object detection.

### *2.3.6 Scikit-learn*

Scikit-learn (Sklearn) is the most useful and robust library for machine learning in Python. It provides a selection of efficient tools for machine learning and statistical modelling including classification, regression, clustering and dimensionality reduction via a consistence interface in Python.

### *2.3.7 Keras*

Keras is the high-level API of TensorFlow 2. An approachable, highly-productive interface for solving machine learning problems, with a focus on modern deep learning. It provides essential abstractions and building blocks for developing and shipping machine learning solutions with high iteration velocity.

### *2.3.8 TensorFlow*

TensorFlow is an open-source library developed by Google primarily for deep learning applications. It also supports traditional machine learning. TensorFlow was originally developed for large numerical computations without keeping deep learning in mind. However, it proved to be very useful for deep learning development as well, and therefore Google open-sourced it.

TensorFlow accepts data in the form of multi-dimensional arrays of higher dimensions called tensors. Multi-dimensional arrays are very handy in handling large amounts of data.

TensorFlow works on the basis of data flow graphs that have nodes and edges. As the execution mechanism is in the form of graphs, it is much easier to execute TensorFlow code in a distributed manner across a cluster of computers while using GPUs.

## *2.4 Computer Vision*

Computer Vision can be defined as a discipline that explains how to reconstruct, interrupt, and understand a 3D scene from its 2D images, in terms of the properties of the structure present in the scene. It deals with modeling and replicating human vision using computer software and hardware. Computer Vision overlaps significantly with the following fields −

Image Processing − It focuses on image manipulation.

Pattern Recognition − It explains various techniques to classify patterns.

Photogrammetry − It is concerned with obtaining accurate measurements from images.

# Chapter 3: REQUIREMENT AND ANALYSIS

## *3.1 Requirement Specification*

This Problem is done by mining the Big Data of the previous records of the people to whom the loan was granted before and on the basis of these records/experiences the machine was trained using the machine learning model which give the most accurate result. The main objective of this paper is to predict whether assigning the loan to a particular person will be safe or not. We have implemented this loan prediction problem using Decision tree algorithm and data cleaning in Python as there are missing values in the dataset. We use map function for the missing values. The aim of this paper is to apply machine learning technique on dataset which has 1000 cases and 7 numerical and 6 categorical attributes. The creditability of a customer for sanctioning loan depend on several parameters, such as credit history, Installment etc.

## *3.2 Project planning and scheduling*

This is a six month time frame to implement a production system of an Banking Customer and Loan Prediction System from project commencement in time for fall 2021.

|  |  |  |
| --- | --- | --- |
| Phases | Date | No. of days |
| Analysis |  |  |
| Requirements |  |  |
| Coding |  |  |
| Testing |  |  |

## *3.3 Software and hardware requirement*

### *3.3.1 Software requirement*

In this requirements that involve in this project development are Python, Anaconda, Jupyter notebook, Spyder.

* Operating system : Windows 7 or Higher
* Programming Language : Python
* Dataset : Image Data

### *3.3.2Hardware requirement:*

* + Processor: Minimum i5 8th Gen
  + HDD/SDD: Minimum 256GB
  + RAM: Minimum 8GB or Higher
  + Graphics: Minimum 4GB

# Chapter 4: ARTIFICIAL INTELLIGENCE

## *4.1 Introduction*

Artificial intelligence (AI), the ability of a digital computer or computer-controlled robot to perform tasks commonly associated with intelligent beings. The term is frequently applied to the project of developing systems endowed with the intellectual processes characteristic of humans, such as the ability to reason, discover meaning, generalize, or learn from past experience. Since the development of the digital computer in the 1940s, it has been demonstrated that computers can be programmed to carry out very complex tasks-as, for example, discovering proofs for mathematical theorems or playing chess-with great proficiency. Still, despite continuing advances in computer processing speed and memory capacity, there are as yet no programs that can match human flexibility over wider domains or in tasks requiring much everyday knowledge.

## *4.2 How AI used?*

Artificial intelligence generally falls under two broad categories:

**Narrow AI**: Sometimes referred to as "Weak AI," this kind of artificial intelligence operates within a limited context and is a simulation of human intelligence. Narrow AI is often focused on performing a single task extremely well and while these machines may seem intelligent, they are operating under far more constraints and limitations than even the most basic human intelligence.

**Artificial General Intelligence (AGI):** AGI, sometimes referred to as "Strong AI," is the kind of artificial intelligence we see in the movies, like the robots from West world or Data from Star Trek: The Next Generation. AGI is a machine with general intelligence and, much like a human being, it can apply that intelligence to solve any problem.

## *4.3 Machine Learning & Deep Learning*

Much of Narrow AI is powered by breakthroughs in machine learning and deep learning. Understanding the difference between artificial intelligence, machine learning and deep learning can be confusing. Venture capitalist Frank Chen provides a good overview of how to distinguish between them, noting: Simply put, machine learning feeds a computer data and uses statistical techniques to help it "learn" how to get progressively better at a task, without having been specifically programmed for that task, eliminating the need for millions of lines of written code. Machine learning consists of both supervised learning (using labeled data sets) and unsupervised learning (using unlabeled data sets).

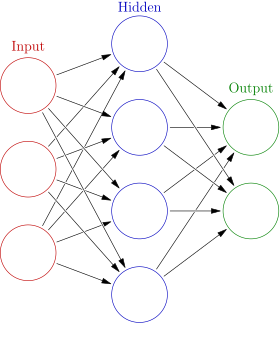
Deep learning is a type of machine learning that runs inputs through a biologically-inspired neural network architecture. The neural networks contain a number of hidden layers through which the data is processed, allowing the machine to go "deep" in its learning, making connections and weighting input for the best results.

**A few examples of Narrow AI include:**

* Google search
* Image recognition software
* Siri, Alexa and other personal assistants
* Self-driving cars
* IBM's Watson

## *4.3 Neural Network*

A neural network is a series of algorithms that ndeavours to recognize underlying relationships in a set of data through a process that mimics the way the human brain operates. A neural network contains layers of interconnected nodes.



## *4.4 Artificial Neural Network*

ANNs are biologically inspired computer programmes designed to simulate the way in which the human brain processes information (Dayfoff and DeLeo, 2001). ANNs gather their knowledge by detecting the patterns and relationships in data and learn (or are trained) through experience, not from programming, and there lies the basic difference between ANNs and other classical computer programmes. Another significant difference between ANNs software and other computer programmes is that the algorithms used for data analysis are flexible. They can be changed anytime during the progress of analysis. The distinctive feature of ANNs is their ability to deal effectively with multidimensional problems, including several thousands of features. An ANN is formed from hundreds of single units, i.e. artificial neurons or processing elements, connected with coefficients (weights), which constitute the neural structure and are organized in layers. The ability of neural computations comes from connecting neurons in a network. The better the neurons are connected in networks, the better is the prediction as output. The activity of a neural network is determined by transfer functions of its neurons, by the learning rule, and by the architecture itself. Achievement of successful result from ANNs studies depends on minimization of prediction error by optimization of interunit connections during training. By doing so as trial and error, the network reaches the specified level of accuracy. Once the network is trained with minimum prediction error and tested, it may be used with new input information to predict the output. The information in ANNs is encoded in the strength of the network's ‘synaptic’ connections (Zupan and Gasteiger, 1993; Kaliszan et al., 2003). Latest studies on ANNs are mainly centred on designing new network types by changing transfer connection of neurons, by changing learning rule, and by initiating new connection formula.

Artificial neural networks belong to a group of information-processing techniques which can be used to find knowledge, patterns or models from a large amount of data. Furthermore, intriguing advantages can be achieved by combining artificial neural networks with other computational models (FDM, FEM, FVM), which can provide the data to train the artificial neural network in order to create the model. Genetic algorithms can be used efficiently to find a suitable solution to a complex optimization problem. Genetic algorithms are an effective tool that is available for exploring large, complex search spaces based on Darwin's theory of survival of the fittest. In recent years, a lot of work has been presented combining genetic algorithms and artificial neural networks in the context of the development of hybrid methods. The main approaches that have been used in that sense are the following: (i) using genetic algorithms to improve the learning of artificial neural networks, and (ii) using genetic algorithms to perform a global search and to optimize the inputs of the neural network model versus the output results created by a network.

# Chapter 5: METHODOLOGY AND IMPLEMENTATION

## *5.1 Image Acquisition*

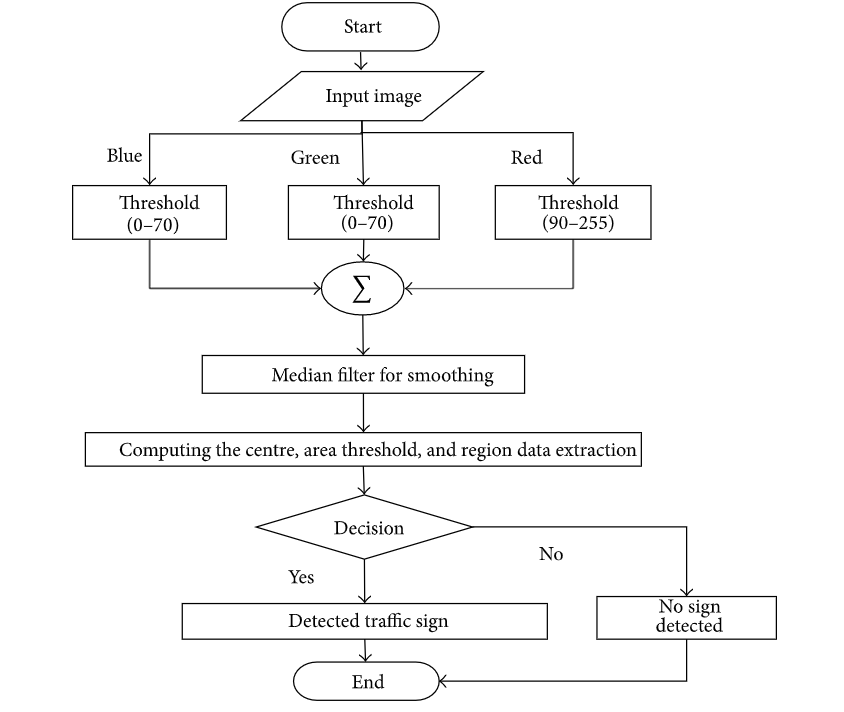
The dataset contains more than 50,000 images of different traffic signs. It is further classified into 43 different classes. The dataset is quite varying, some of the classes have many images while some classes have few images. The size of the dataset is around 300 MB. The dataset has a train folder which contains images inside each class and a test folder which we will use for testing our model.

## *5.2 Image Preprocessing*

Image preprocessing is an important part of the TSDR system whose main idea is to remove low-frequency background noise, normalising the intensity of the individual particles images, removing reflections, and masking portions of images. Below is a description of selected image preprocessing techniques. The input image is divided into channels R, G, and B separately. In the proposed approach, filters are applied on each channel threshold to select those regions of the image where the values of the pixels fall in the range of our target object. For example, for traffic signs with a red background (such as stop signs), the threshold for channel R is pixels with values in the range of 90–255 and for channels G and B the range is 0–70. The region of interest (ROI) is the logical sum of the three filtered channels of R, G, and B.

## *5.3 Shape Matching Based Detection*

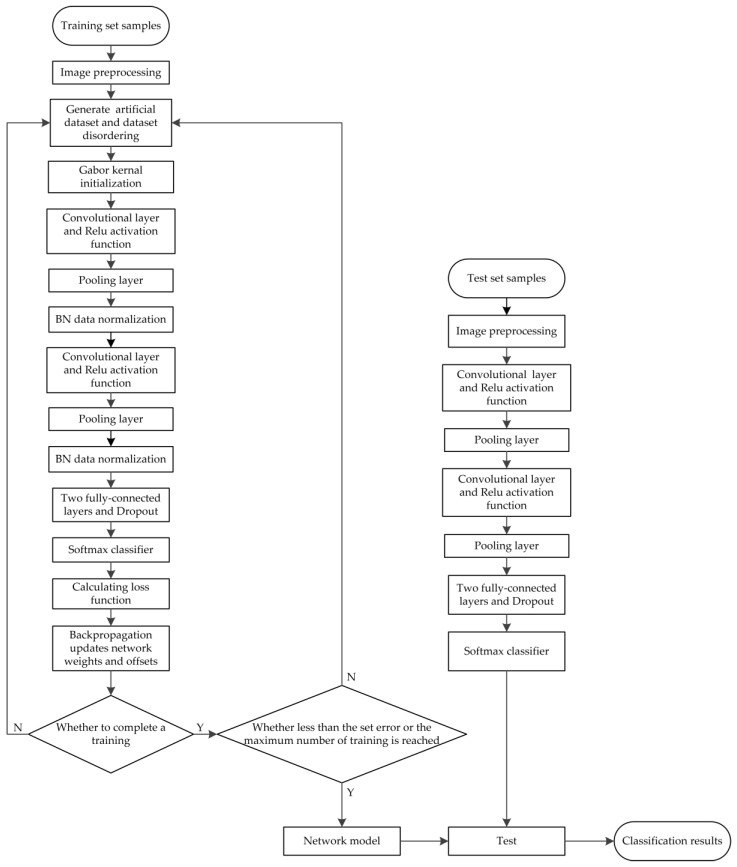
The idea is to use colour characteristic of the preferred object to accelerate the procedure without employing model-based classifiers which is a time consuming process. After filtering and analysing the features of the detected object, the candidates of the traffic sign are selected based on shape matching.



## *5.4 Objects Features Analysing*

One of the important steps is to eliminate noise from the image therefore to better deal with the ROI. Appropriate filters have an enormous effect on accuracy and speed of the procedure without deleting any useful information. In the proposed system, for image smoothing and filling up the smaller region to extract the region of interest, a median filter was used.

## *5.5 Architecture of the system*



## *5.6 Steps to Build the Project*

Our approach to building this traffic sign classification model is discussed in four steps:

* Explore the dataset
* Build a CNN model
* Train and validate the model
* Test the model with test dataset

### *Step 1: Explore the dataset*

Our ‘train’ folder contains 43 folders each representing a different class. The range of the folder is from 0 to 42. With the help of the OS module, we iterate over all the classes and append images and their respective labels in the data and labels list.

The PIL library is used to open image content into an array.



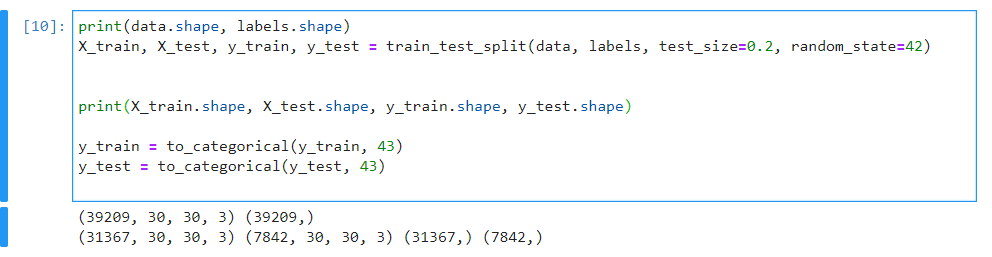
Finally, we have stored all the images and their labels into lists (data and labels).

We need to convert the list into numpy arrays for feeding to the model.

The shape of data is (39209, 30, 30, 3) which means that there are 39,209 images of size 30×30 pixels and the last 3 means the data contains colored images (RGB value).

With the sklearn package, we use the train\_test\_split() method to split training and testing data.

From the keras.utils package, we use to\_categorical method to convert the labels present in y\_train and t\_test into one-hot encoding.



### *Step 2: Build a CNN model*

To classify the images into their respective categories, we will build a CNN model (Convolutional Neural Network). CNN is best for image classification purposes.

The architecture of our model is:

2 Conv2D layer (filter=32, kernel\_size=(5,5), activation=”relu”)

MaxPool2D layer ( pool\_size=(2,2))

Dropout layer (rate=0.25)

2 Conv2D layer (filter=64, kernel\_size=(3,3), activation=”relu”)

MaxPool2D layer ( pool\_size=(2,2))

Dropout layer (rate=0.25)

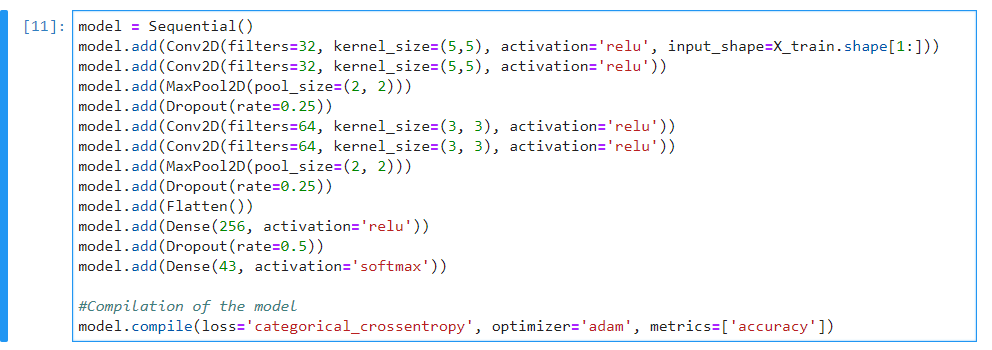
Flatten layer to squeeze the layers into 1 dimension

Dense Fully connected layer (256 nodes, activation=”relu”)

Dropout layer (rate=0.5)

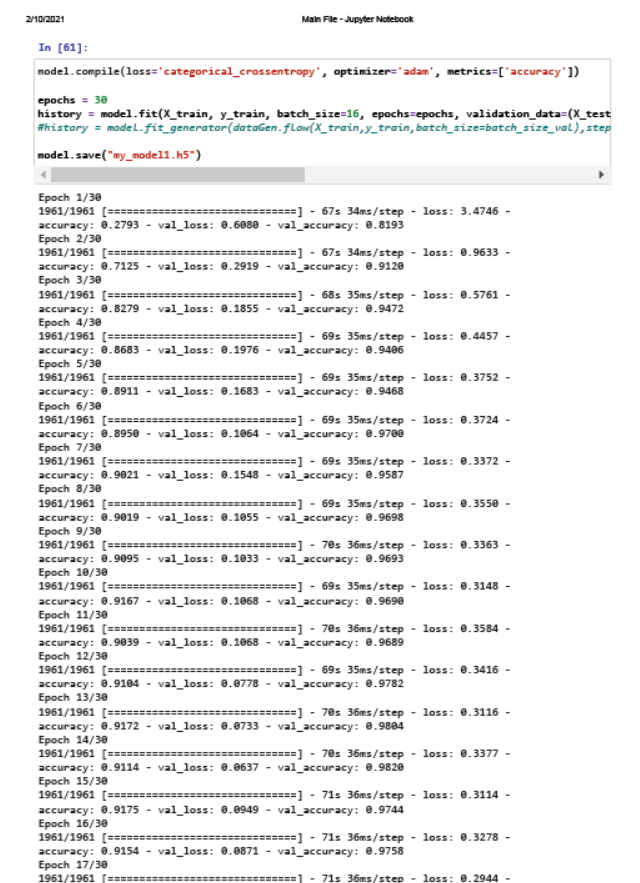
Dense layer (43 nodes, activation=”softmax”)

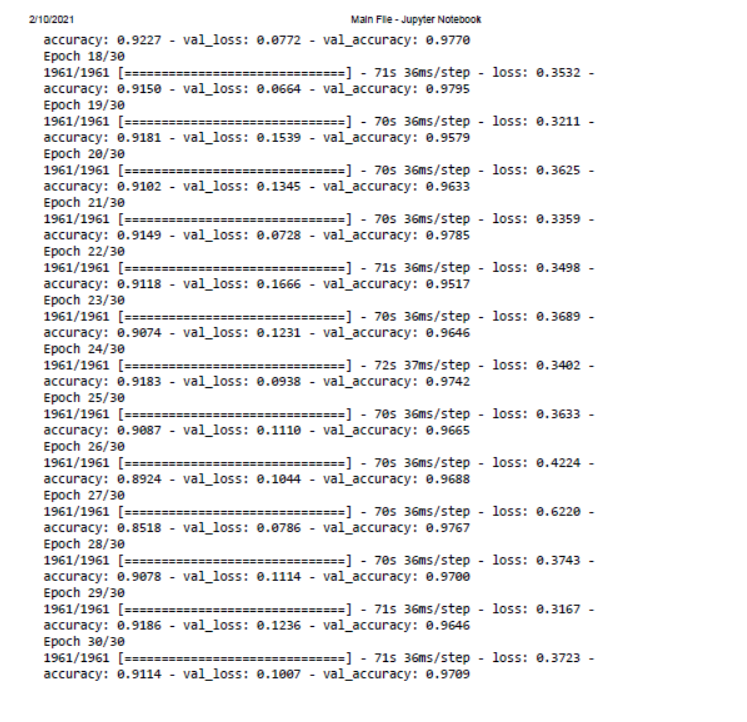
We compile the model with Adam optimizer which performs well and loss is “categorical\_crossentropy” because we have multiple classes to categorise.



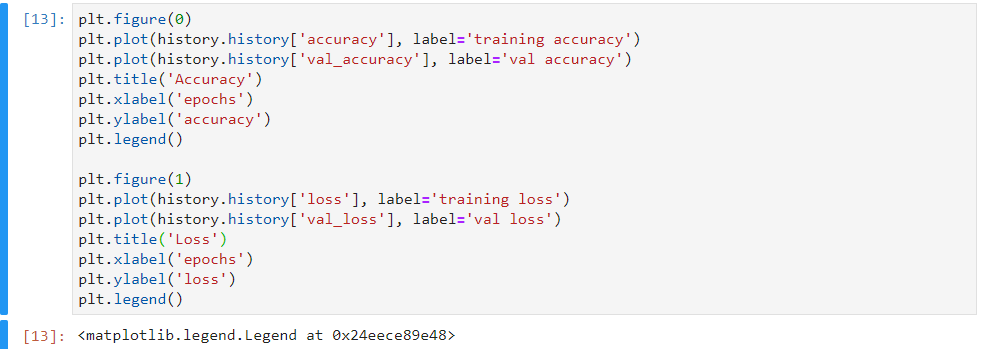
### *Step 3: Train and validate the model*

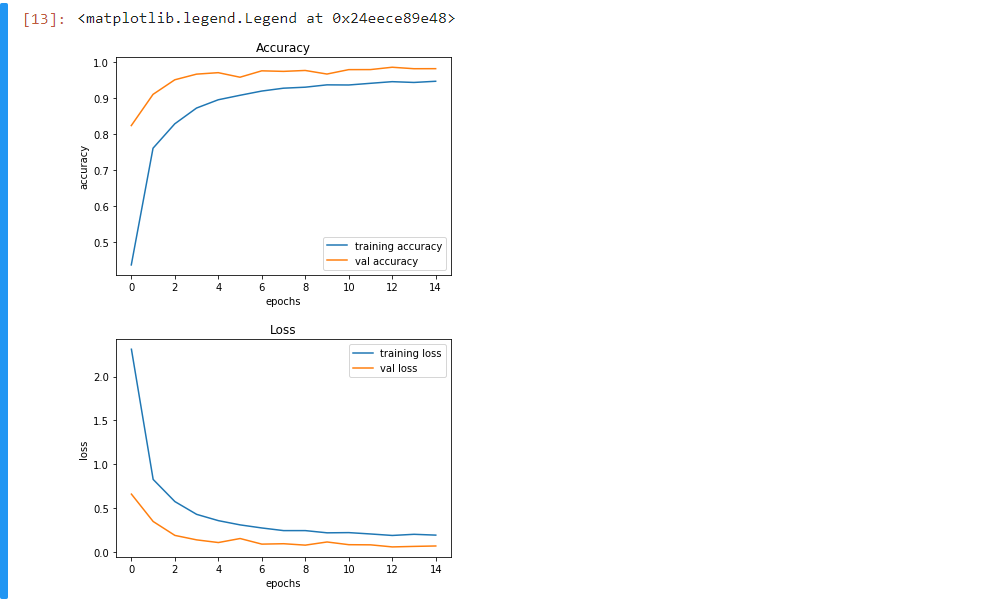
After building the model architecture, we then train the model using model.fit(). I tried with batch size 32 and 16. Our model performed better with 16 batch size. And after 30 epochs the accuracy was stable.





Our model got a 91% accuracy on the training dataset. With matplotlib, we plot the graph for accuracy and the loss.



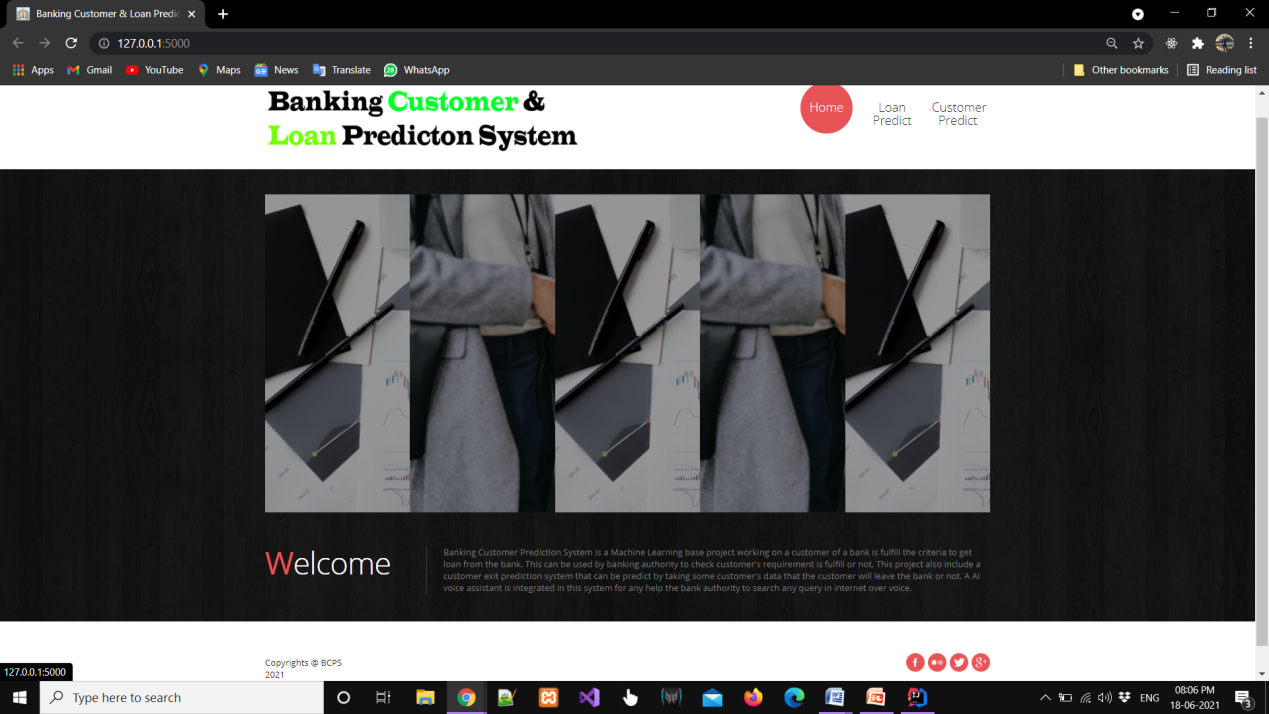


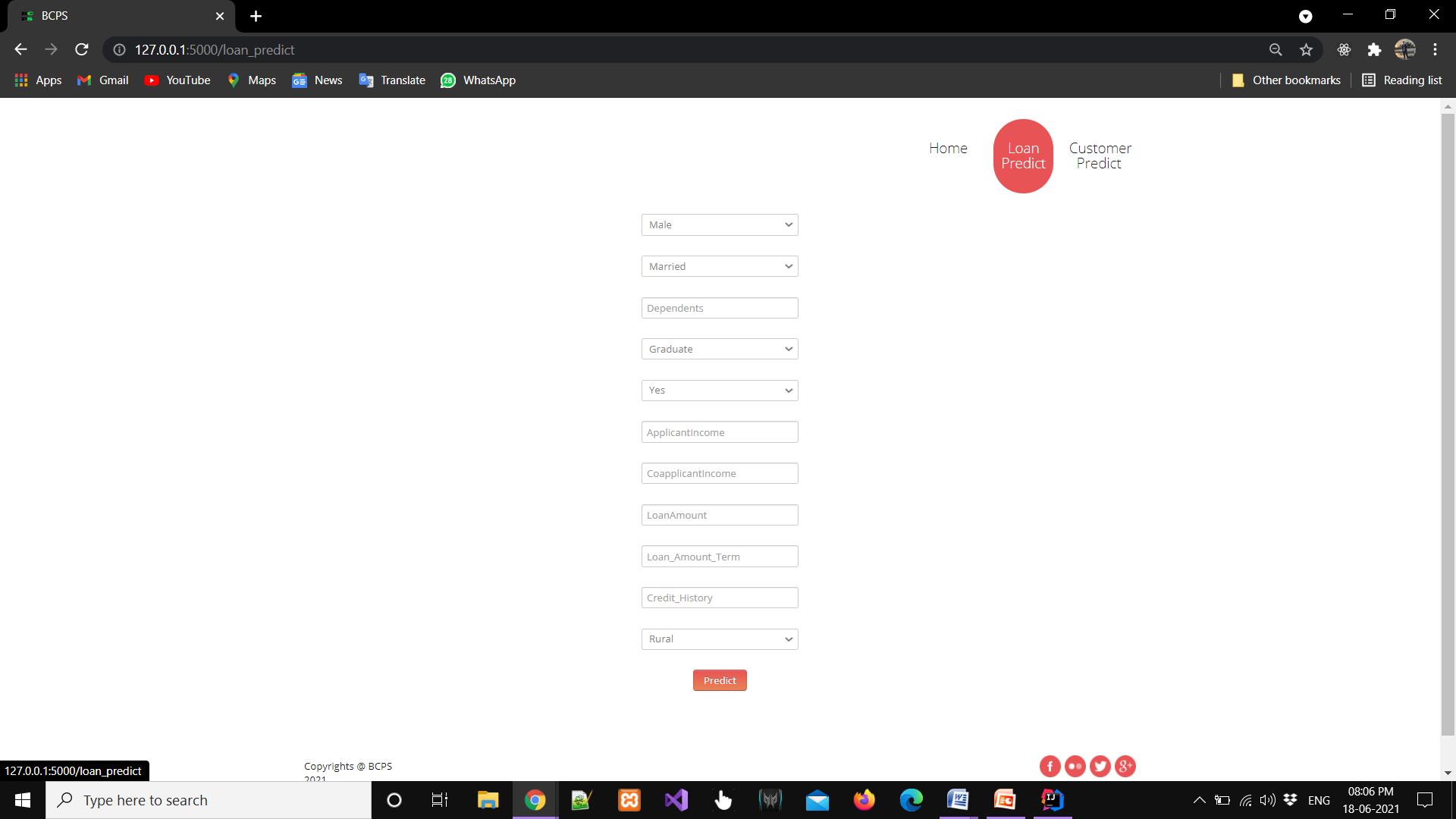
### *Step 4: Test our model with test dataset*

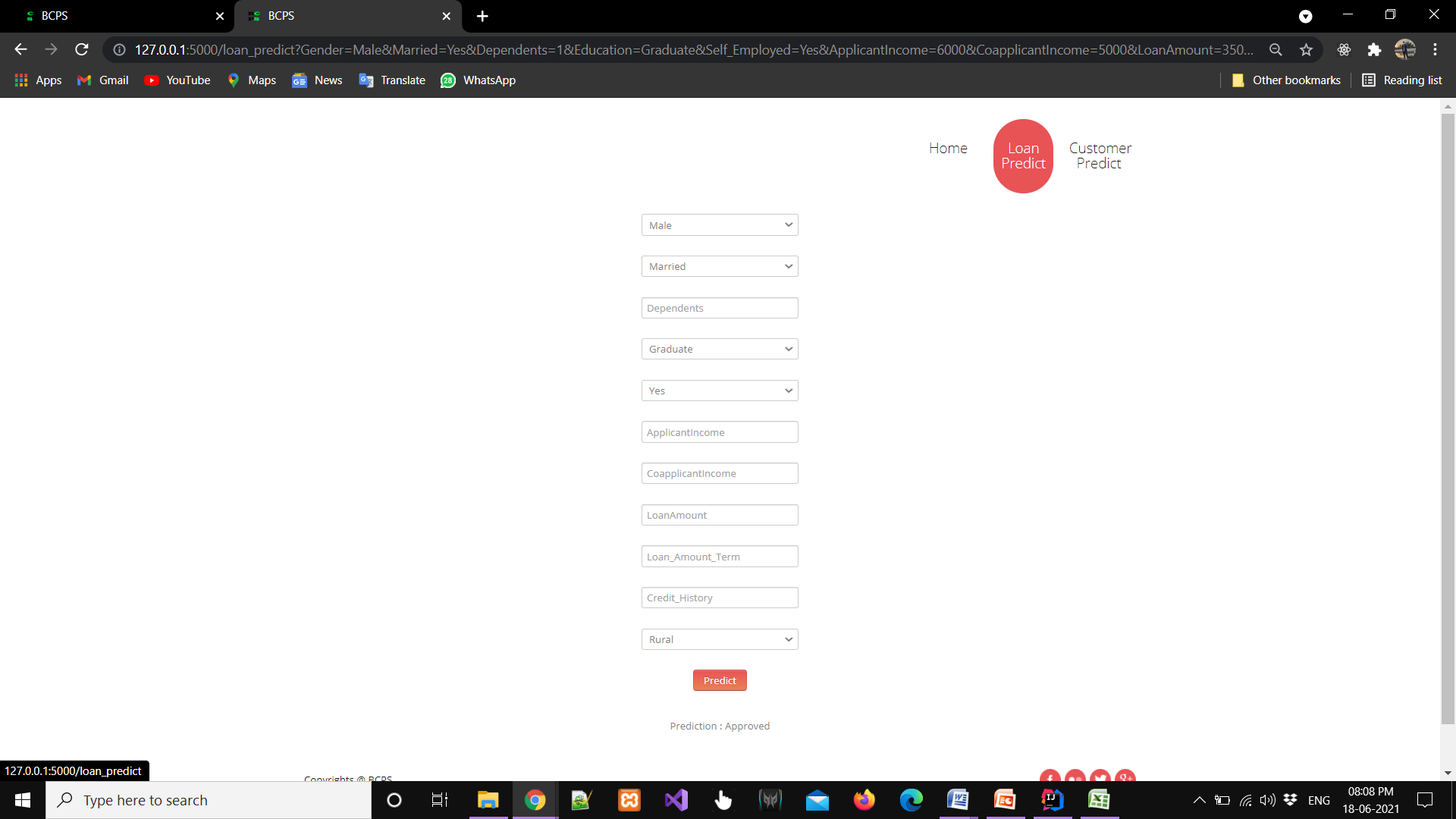
Our dataset contains a test folder and in a test.csv file, we have the details related to the image path and their respective class labels. We extract the image path and labels using pandas. Then to predict the model, we have to resize our images to 30×30 pixels and make a numpy array containing all image data. From the sklearn.metrics, we imported the accuracy\_score and observed how our model predicted the actual labels. We achieved a 91% accuracy in this model.

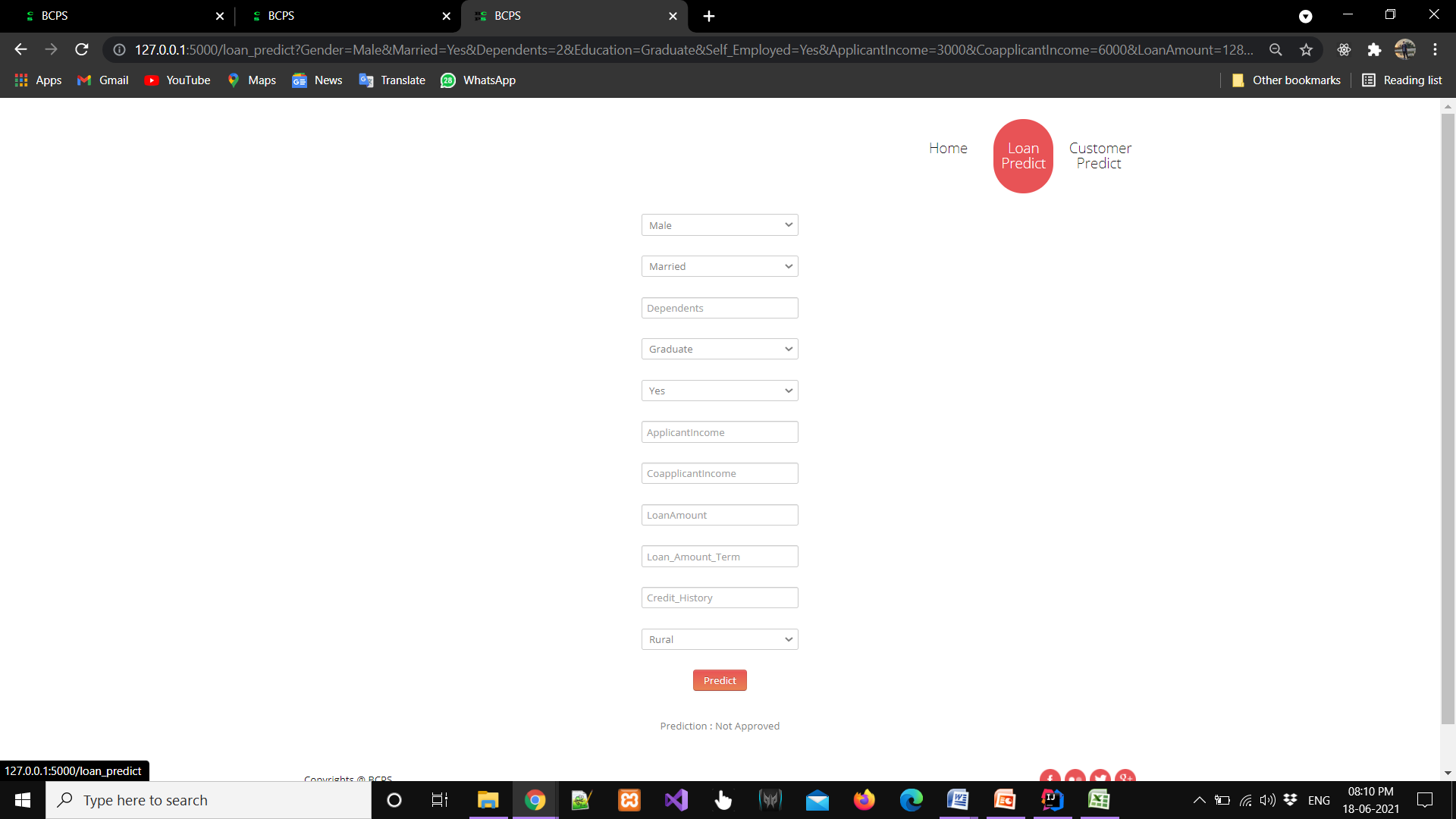
In the end, we are going to save the model that we have trained using the Keras model.save() function.

# Chapter 6: SNAPSHOTS









# Chapter 7: CONCLUSION

## *7.1 Conclusion*

From a proper analysis of positive points and constraints on the component, it can be safely concluded that the product is a highly efficient component. There have been numbers cases of computer glitches, errors in content and most important weight of features is fixed in automated prediction system, So in the near future the so –called software could be made more secure, reliable and dynamic weight adjustment .In near future this module of prediction can be integrate with the module of automated processing system. the system is trained on old training dataset in future software can be made such that new testing date should also take part in training data after some fix time.

## *7.2 Benefits*

Applicants with Credit history not passing fails to get approved, Probably because that they have a probability of a not paying back. Most of the Time, Applicants with high income sanctioning low amount is to more likely get approved which make sense, more likely to pay back their loans. Some basic characteristic gender and marital status seems not to be taken into consideration by thecompany.

## *7.3 Future scope*

In future I wish to develop a Data mining application using wrapper-fisher feature selection algorithm and surely it helps the bank officers to take proper decision when a new customer approaches the bank for taking the loan. The proposed hybrid feature selection algorithm for classifying the loan credibility behavior of a customer in a banking industry can also be used for several other applications in the future especially binary classification problems such as prediction of various diseases, prediction of various examination results etc.

# BIBLIOGRAPHY

[1]Cowell,R.G.,A.P.,Lauritez,S.L.,and Spiegelhalter,D.J.(1999). Graphical models and

Expert Systems. Berlin: Springer. This is a good introduction to probabilistic graphical models.

[2] Kumar Arun, Garg Ishan, Kaur Sanmeet, MayJun. 2016. Loan Approval Prediction based on Machine Learning Approach, IOSR Journal of Computer Engineering (IOSR-JCE)

[3] Wei Li, Shuai Ding, Yi Chen, and Shanlin Yang, Heterogeneous Ensemble for Default Prediction of Peer-to-Peer Lending in China, Key Laboratory of Process Optimization and Intelligent DecisionMaking, Ministry of Education, Hefei University of Technology, Hefei 23009, China

[4] Clustering Loan Applicants based on Risk Percentage using K-Means Clustering Techniques, Dr. K. Kavitha, International Journal of Advanced Research in Computer Science and Software Engineering.

[5] Research on bank credit default prediction based on data mining algorithm, The International Journal of Social Sciences and Humanities Invention 5(06): 4820-4823, 2018.

[6] Short-term prediction of Mortgage default using ensembled machine learning models, Jesse C. Sealand on july 20, 2018.