Effectively Stopping Cyberbullying

Submitted in partial fulfillment of the requirements for the Award of the Degree

of

Master of Computer Applications

of

APJ Abdul Kalam Technological University

Submitted by

SACHIN K

Reg No: TVE17MCA039

Department of Computer Applications

COLLEGE OF ENGINEERING, TRIVANDRUM

MAY 2020

Declaration

We hereby declare that the project report titled **"Effectively Stopping Cyberbullying"** submitted by SACHIN K of **College Of Engineering,Trivandrum** for the partial fulfillment of the requirement for the award of the degree of Masters of Computer Applications of **APJ Abdul Kalam University** is a bonafide record done on by the guidance of **Prof.Pooja J.P**. We also declare that any part of this project either partial or full is submitted by any organisation or institute for awarding the degree from any other university.

Place : Trivandrum **Sachin K**

Date :

DEPARTMENT OF COMPUTER APPLICATIONS

COLLEGE OF ENGINEERING
TRIVANDRUM

CERTIFICATE

This is to certify that the project work entitled, ***"Effectively Stopping Cyberbullying"*** *is a bonafide record of the work done by "Sachin K" (Reg No: TVE17MCA039) student of College of Engineering, Trivandrum in partial fulfillment of the requirements for the award of the degree of Master of Computer Applications from APJ Abdul Kalam Technological University during the year 2020.*

Pooja J P
Project Guide

Dr. Sabitha S
Head of the Department

Acknowledgement

First and for most we thank **GOD** almighty and to my parents for the success of this project. We owe a sincere gratitude and heart full thanks to everyone who shared their precious time and knowledge for the successful completion of my project.

We am extremely thankful to **Dr Jiji C V**, Principal, College of Engineering Trivandrum for providing me with the best facilities and atmosphere which was necessary for the successful completion of this project.

We am extremely grateful to **Dr. Sabitha S**, HOD, Dept of Computer Applications, for providing me with best facilities and atmosphere for the creative work guidance and encouragement.

We express our sincere thanks to **Prof. Pooja J P**, Asst. Professor, Department of Computer Applications, College of Engineering Trivandrum for his valuable guidance, support and advice that aided in the successful completion of my project.

We profusely thank other Asst. Professors in the department and all other staffs of CET, for their guidance and inspirations throughout my course of study.

We owe my thanks to my friends and all others who have directly or indirectly helped me in the successful completion of this project. No words can express my humble gratitude to my beloved parents and relatives who have been guiding me in all walks of my journey.

Sachin K

Abstract

Online bullying and aggression against social media users have grown abruptly, causing serious consequences to victims of all demographics in recent years. It affects social media users worldwide, suffering from prolonged digital harassment.To prevent cyberbullying, a method is proposed to detect abusive contents in user input like post or comments in social media.Here Deep Neural network model is used to detect abusive content as it become increasingly popular for text processing in recent years. In particular, the emergence of word embeddings within deep learning architectures has recently attracted a high level of attention. Deep neural network model used to analyze user input and force them to change their input if found abusive before posting it on social media.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

Cyberbullying has been defined by the National Crime Prevention Council as the use of the Internet, cell phones or other devices to send or post text or images intended to hurt or embarrass another person. Various studies have estimated that between to 10% to 40% of internet users are victims of cyberbullying.Effects of cyberbullying can range from temporary anxiety to suicide. Many high profile incidents have emphasized the prevalence of cyberbullying on social media. On networking sites like Myspace, mostly teenager have reveal the exasperation as it creates too much emotional stress.

Individuals are given compete free hand as to what they can post when online. They are also given the power to post offensive remarks or pictures regardless of what disastrous implications it may create. These Interactive media also fails in filtering these adverse comments and status which can be uploaded in public platforms, but they are equipped with reporting systems, which enables the user to report the contents as abuse, and such contents can then be removed from social platform, but it is not fully effective to stop cyberbullying. So using Long Short Term Memory(LSTM), a deep neural network architecture with word embedding technique used to create a model to check whether the user input is abusive or not. If the user input contain abusive content, then the user is forced to change their content before posting it online.

Chapter 2

Requirement Analysis

2.1 Purpose

The purpose of this technique is to provide a better and safe usage of social media. Instead of reporting against abusive contents and waiting for the officials to review and take actions, the abusive detection system automatically analyze the user input and detect if it contain abusive content and force the user to change it before posting it online. This system helps in avoiding abusive contents in online.

2.2 Overall Description

The proposed system is implemented taking into account of deep learning technique and word embedding concept. This technique can be used as a feature in social media platform, the input from user can be analyzed using this feature. An alert will be given to the user if their input contain abusive content. The abusive words or contents are marked in red and force the user to change those. Deep learning technique used to analyze and detect user input. The user is only allowed to post their content in online if it is non-abusive.

2.2.1 Product Functions

- Accept text input from user.
- Analyze the input and check for abusive content.

- If the input found abusive, an alert will be given to the user.
- The user is only allowed to post their content if it is safe to the public.

2.2.2 Hardware Requirements

- Processor : Intel Core i5
- Storage : 1 TB Hard Disk space
- Memory : 8 GB RAM

2.2.3 Software Requirements

- Operating System : Linux
- Programming Language : Python
- Web framework : Flask
- Libraries used : Nlkt, Pandas, Sklearn, Keras

2.3 Functional Requirements

The functional requirements includes all the activities or processes that should be achieved by the proposed system. It includes

2.3.1 abusive text detection

The abusive text detection system accept the user text input. Analyze the user input using trained model, an alert will be given to the user if input contain abusive content. The abusive contents are highlighted and force the user to change the input. The system will only allow the user to post their content if it is non abusive.

2.4 Non Functional Requirements

2.4.1 Performance Requirements

- Accuracy : Accuracy in functioning and the nature of user-friendly should be maintained by the system.
- Speed : The system must be capable of offering speed.
- Low cost: This system is very cheap to implement and is also user-friendly.
- Time consuming: It is highly time consuming so that the user doesn't need to wait for their turn in the toll booth.

2.4.2 Quality Requirements

- Scalability : The software will meet all of the functional requirements.
- Maintainability : The system should be maintainable. It should keep backups to atone for system failures, and should log its activities periodically.
- Reliability : The acceptable threshold for down-time should be large as possible. i.e. mean time between failures should be large as possible. And if the system is broken, time required to get the system backup again should be minimum.
- Availability: This system is easily available as the core equipment in building the software is easily obtained.

2.4.3 Assumptions

- GPU availability for training.
- Text quality is up to mark.
- No limitation over Memory and Latency.
- Only English language is considered.

Chapter 3

Design And Implementation

The proposed system reduces the abusive contents in online platform. Real time analysis of the user input is an important feature of the proposed system. The proposed system will eliminate the need of human beings for monitoring and reviewing of the user contents. The user input are analyzed by the deep learning model for detection abusive content in user input.

3.1 Overall Design

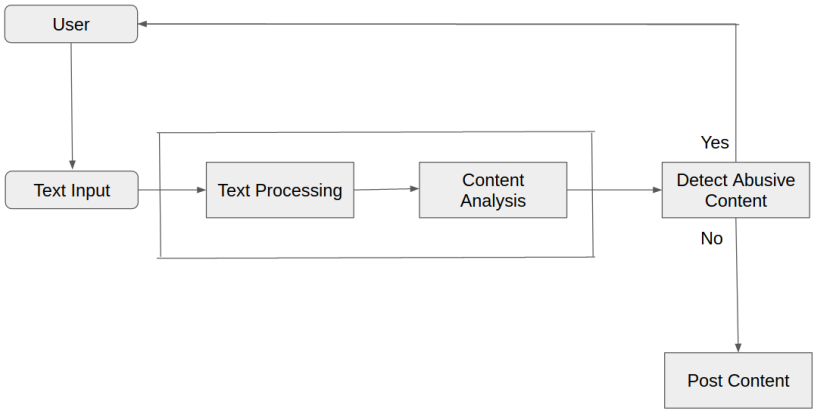

Figure 3.1: Block diagram of the proposed system

Abusive text detection analyze the user input, the words and their sequence order. The user input is converted into vectorized form and the deep neural network model analyse the vector and the sequence to predict whether the input is abusive or not.

3.1.1 Module Description

The main module and the functions of each of these module of the proposed system are given below:

- **Django:** Django is a high-level Python Web framework that encourages rapid development and clean pragmatic design. A Web framework is a set of components that provide a standard way to develop websites fast and easily. Django's primary goal is to ease the creation of complex database-driven websites. Some well known sites that use Django include PBS, Instagram, Disqus, Washington Times, Bitbucket and Mozilla.
- **Abusive content Detection:** The user input is analyzed to detect abusive content, Since it detects the abusive contents, the user cannot post their input without changing it.
- **Alert Message** The message will be displayed if user input contain abusive content.

3.2 Understanding the Problem

- **Abusive content Detection:** This problem can be tackled using Deep neural network approach where we need to train our model using the abusive and non-abusive text contents. The training data are converted to vector to represent them and to understand their sequences. The sequence learned by the model is used to detect abusive content in text input.

3.3 Data Flow Diagram

A data flow diagram (DFD) is a graphical representation of the "flow" of data through an information system, modelling its process aspects. A DFD is often used as a preliminary step to create an overview of the system without going into great detail, which can later be elaborated. DFDs can also be used for the visualization of data processing (structured design).

A DFD shows what kind of information will be input to and output from the system, how the data will advance through the system, and where the data will be stored. It does not show information about process timing or whether processes will operate in sequence or in parallel,

unlike a traditional structured flowchart which focuses on control flow, or a UML activity workflow diagram, which presents both control and data flows as a unified model.

Level 0

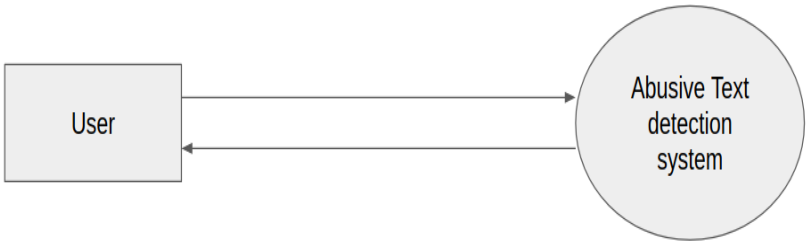

Figure 3.2: Level 0 DFD

Level 1

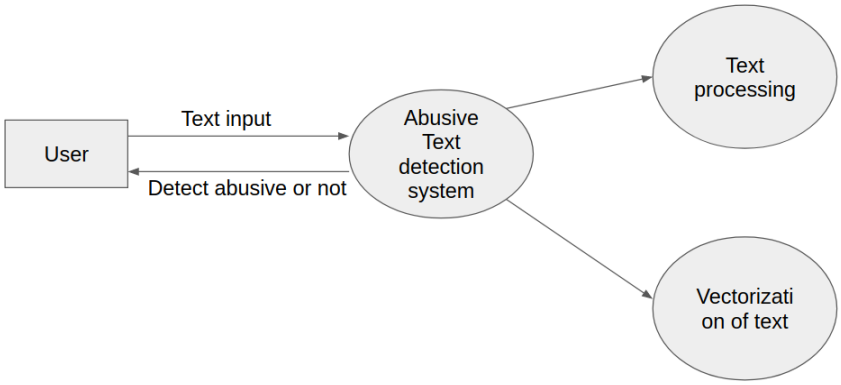

Figure 3.3: Level 1 DFD

Figure 3.4: Level 1.2 DFD

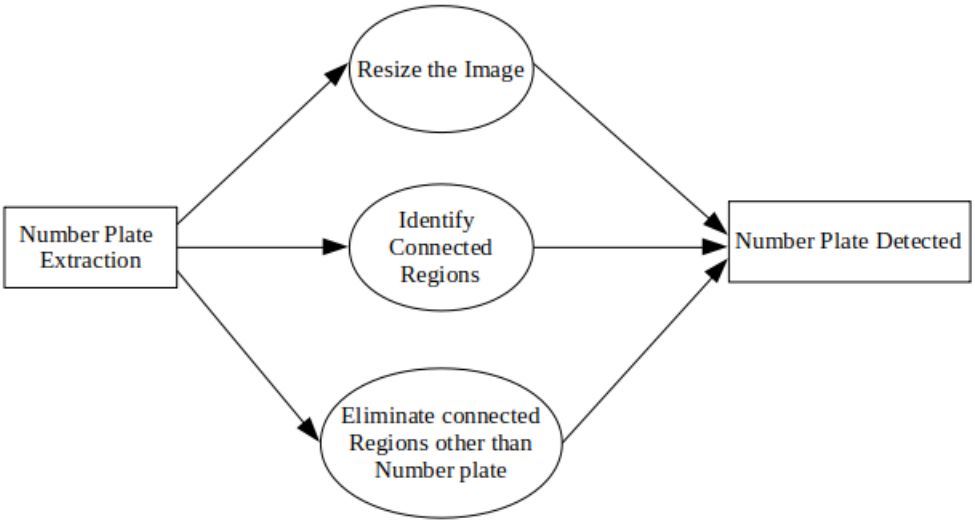

Figure 3.5: Level 2.1 DFD

3.4 Screen Shots

Home Page

User Registration

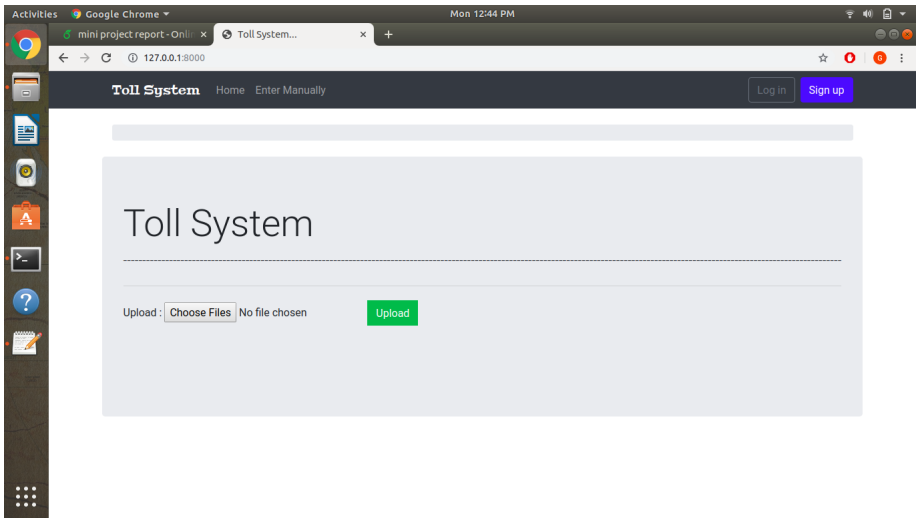

Figure 3.6: Home Page

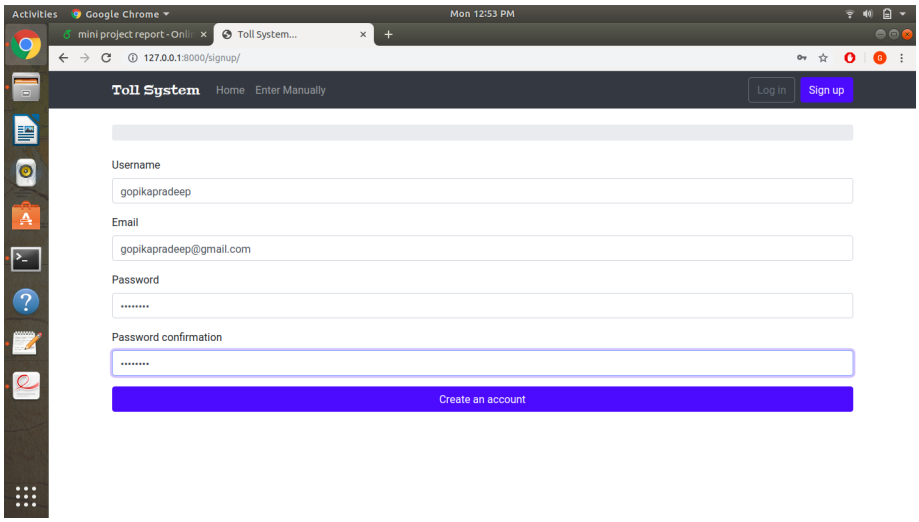

Figure 3.7: User Registration Page

Add Vehicle

Add Bank

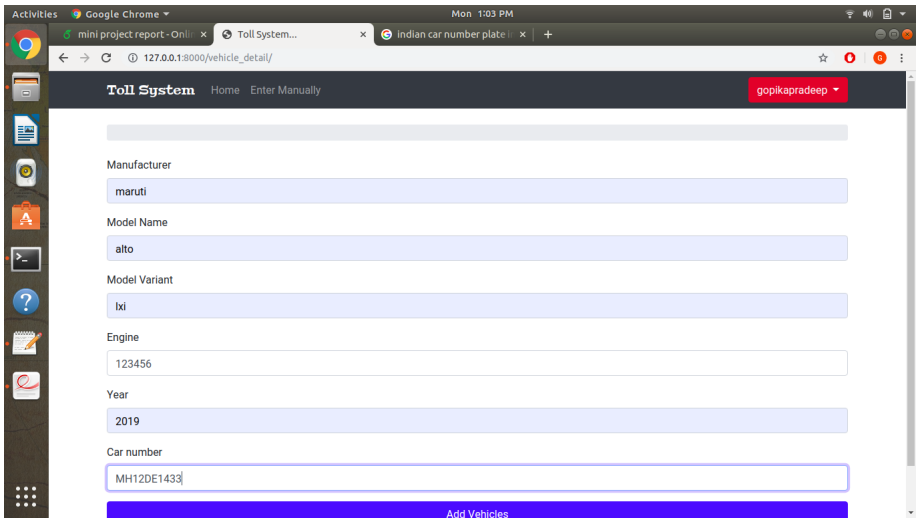

Figure 3.8: Add Vehicle

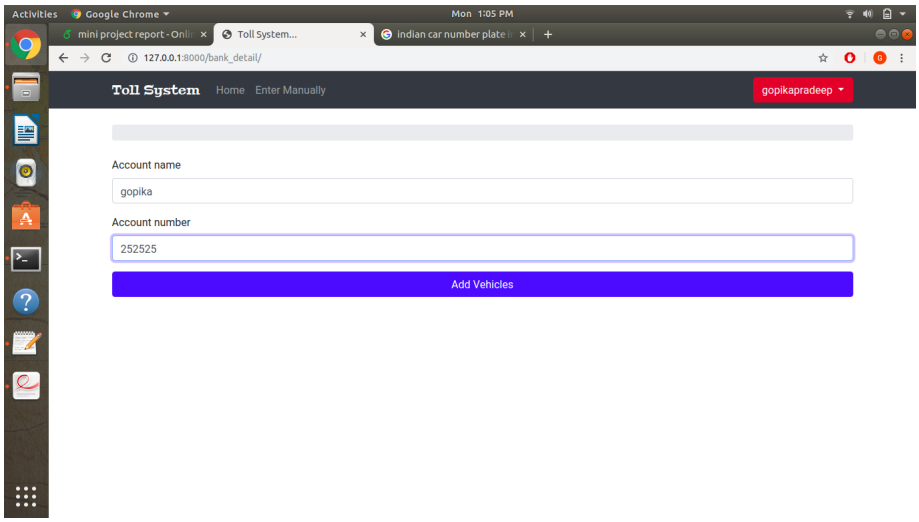

Figure 3.9: Add Bank

Update Balance Details

Vehicle Detection

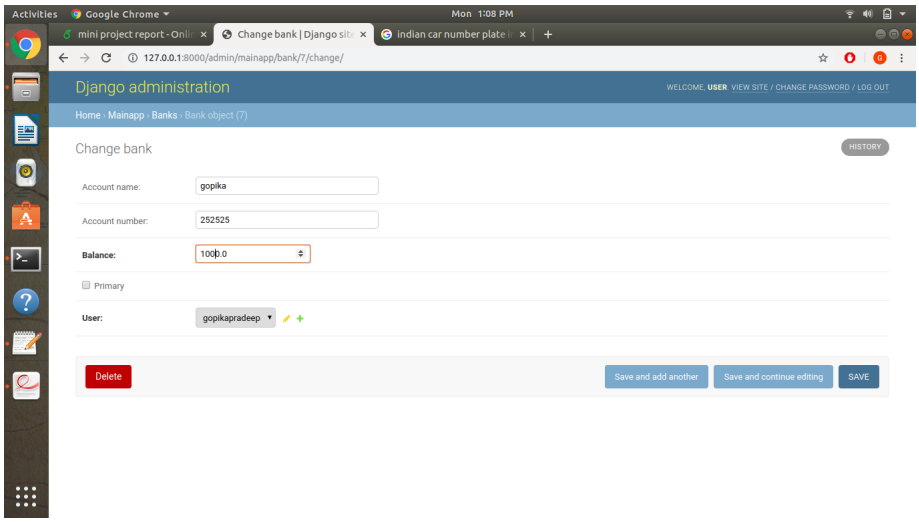

Figure 3.10: Update Bank Balance

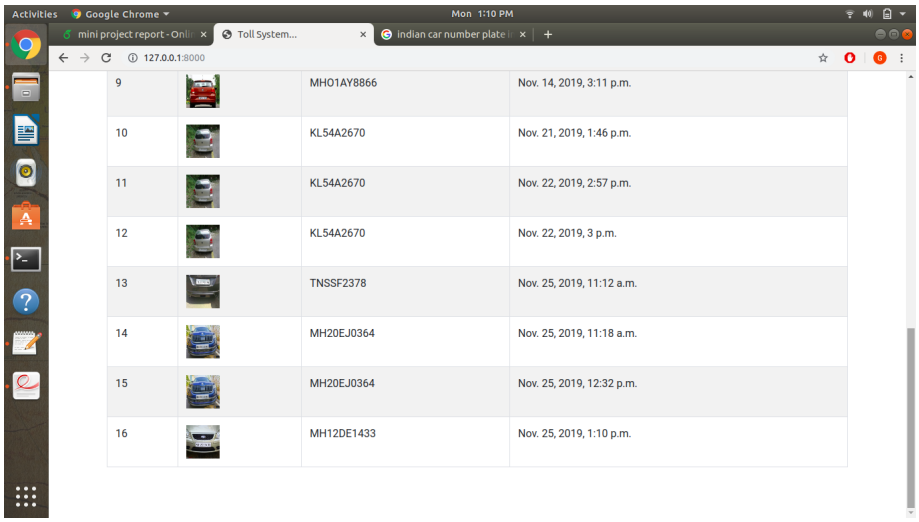

Figure 3.11: Registered vehicle identified

Updated Balance

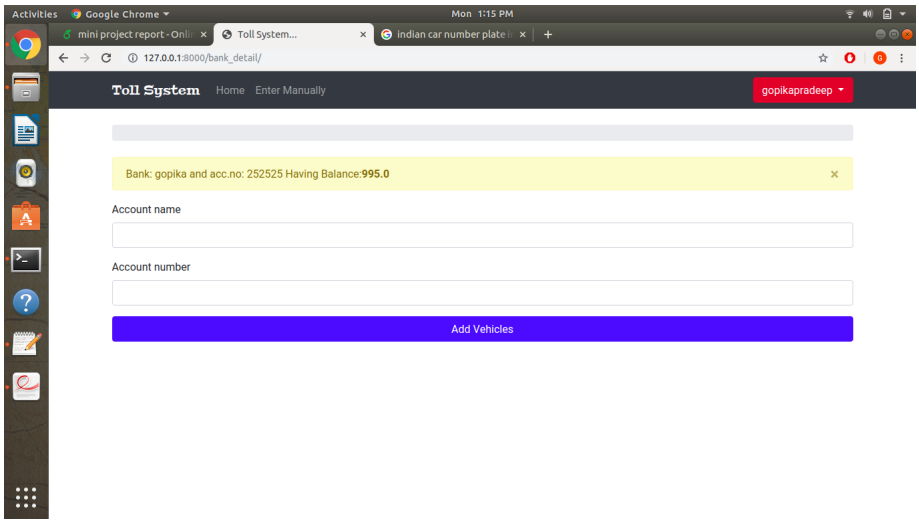

Figure 3.12: User balance after crossing toll gate

Chapter 4

Coding

Algorithm 1

1: Vehicle image is obtained from the camera at the toll gate.

2: By using edge detection process we have to find the number plate region from an image.

3: Character from the number plate is extracted and categorised using the KNN algorithm and it is stored in the database.

4: Number plate extraction is done by template matching. For this generate the template of digit from 0-9 and alphabets A-Z.

5: RFID tag is scanned and uniquely categorise the vehicle.

6: The tag reads the data and analyze whether the vehicle is registered in the database or not.

7: Fetch the details from the database and compare the input value with it.

8: If the vehicle is not a registered user the further process cannot be initiated.

9: After successful completion of the process, the toll amount is deduced from the user account.

Chapter 5

Testing and Implementation

5.1 Testing methods done for the project

System testing is the stage of implementation which is aimed at ensuring that the system works accurately and efficiently before live operation commences. Testing is the process of executing the program with the intent of finding errors and missing operations and also complete verification to determine whether the objective are met and the user requirements are satisfied.

The ultimate aim is quality assurance. Tests are carried and the results are compared with the expected document. In that case of erroneous results,debugging is done. Using detailed testing strategies a test plan is carried out on each module. The test plan defines the unit,integration and system testing approach. The test scope includes the following: A primary objective of testing application systems is to assure that the system meets the full functional requirements, including quality requirements (Non functional requirements).

At the end of the project development cycle, the user should find that the project has met or exceeded all of their expectations as detailed in requirements. Any changes, additions or deletions to the requirements document, functional specification or design specification will be documented and tested at the highest level of quality allowed within the remaining time of the project and within the ability of the test team.

The secondary objective of testing application systems will be to identify and expose all issues and associated risks, communicate all known issues are addressed in an appropriate matter

before release. This test approach document describes the appropriate strategies, process, work flows and methodologies used to plan, organize, execute and manage testing of software project.

5.2 Testing Methodology

5.2.1 Number Plate Detection

- **License Plate Recognition:** The vehicle license plate recognition system commonly combines 2 sub-systems: license plate detection, which aims to locate the vehicle and its license plate; and license plate recognition, which aims to recognize the characters on the plate. It is at this stage that the position of the license plate is determined. The input at this stage is an image of the vehicle and the output is the license plate.

- **Character Segmentation:** This stage convert the characters on the license plate that are mapped out and segmented into individual images. Segmentation is one of the most important processes for the automatic identification of license plates, because any other step is based on it. If the segmentation fails, recognition phase will not be correct.To ensure proper segmentation, preliminary processing will have to be performed.

- **Character Recognition:** The characters earlier segmented are identified here. We'll be using machine learning for this the classifier CNN. We have 30–40 images of 28X28 pixel dimesion PNG formats; numbers from 0 to 9.

5.2.2 Deep Learning

Deep learning is a class of machine learning algorithms that use several layers of nonlinear processing units for feature extraction and transformation. Each successive layer uses the output from the previous layer as input. Deep structured learning or hierarchical learning or deep learning in short is part of the family of machine learning methods which are themselves a subset of the broader field of Artificial Intelligence. Deep Learning Algorithms and Networks are based on the unsupervised learning of multiple levels of features or representations of the data.

5.2.3 PyTesseract

Python-tesseract is an optical character recognition (OCR) tool for python. That is, it will recognize and "read" the text embedded in images. Python-tesseract is a wrapper for Google's Tesseract-OCR Engine. It is a stand-alone invocation script to tesseract, as it can read all image types supported by the Pillow and Leptonica imaging libraries, including jpeg, png, gif, bmp, tiff,and others. Additionally, if used as a script, Python-tesseract will print the recognized text instead of writing it to a file. Optical Character Recognition involves the detection of text content on images and translation of the images to encoded text that the computer can easily understand. An image containing text is scanned and analyzed in order to identify the characters in it. Upon identification, the character is converted to machine-encoded text.

5.3 Testing Algorithm

5.3.1 Transfer Training

This system is actually trained in the method of Transfer Training algorithm. Since this technique helps us to easily compare the test data with another pre trained set. Transfer learning is the most popular approach in deep learning. In this, we use pre-trained models as the starting point on computer vision. Also, natural language processing tasks given the vast compute and time resource. Although, we need to develop neural network models. As transfer learning is related to many problems. Such as multi-task learning and concept drift.

5.3.2 Solution Approach

High Approach

As mentioned above, The problem has been divided into 2 parts:

- Number Plate Detection.
- OCR over extracted Number Plates extracted.

The solution has been approached in the following ways:

- Dataset preparation for training (Training/Validation) of Object Detection model.
- Setting up Darknet for YOLOv3 on Google Colaboratory(due to the availability of free GPU).
- Once satisfactory results are obtained after changing certain configurations according to the problem, validated the model.
- Now, the test dataset was fed and the objects were detected and coordinates stored.
- Extracted the coordinates and cropped the objects accordingly
- Fed the cropped Number Plates to Pytesseract which in turn performs OCR over the image.

5.3.3 Finalized Model/Algorithm Used

- The solution was finally implemented using YOLOv3 with the neural network framework Darknet.
- For text extraction, Pytesseract is considered.
- For certain image processing steps, a combination of PIL and OpenCV has been used.
- Dataset Preparation The dataset used has been a mix of two datasets.
- No other vehicle considered other than cars in training.

5.4 Unit Testing

Sl No	Procedures	Expected result	Actual result	Pass or Fail
1	User Login	User login and manage account	Same as expected	Pass
2	Admin page	Control all activities	Same as expected	Pass
3	Number plate recognition	Detects the number plate	Same as expected	Pass
4	Vehicle detection	Identifies the vehicle	Same as expected	Pass
5	Database fetching	Stores the data	Same as expected	pass
6	Notification receiving	Account balance is shown	Same as expected	pass

Table 5.1: Unit test cases and results

Chapter 6

Results and Future Scope

It is observed that the system performs all the functionalities as expected. The main aim behind this project was to solve the issues related to the manual toll paying system and this project has achieved most of its requirements successfully.

6.1 Advantages and Limitations

The proposed system features a lot of advantages over the existing system. It solves the problems of the existing system. The proposed system is way more secure than the existing system in many ways. Like any other system, this system also has its own advantages and limitations. The point is that its advantages override its limitations.

6.1.1 Advantages

- The system proposes an automated toll payment system taking into account of machine learning and RFID technology.
- The system is capable of detecting the number plate and recognising the characters in the plate successfully with an accuracy of 90 percent.
- In electronic toll collection system deceleration, acceleration and idling is completely eliminated. This helps in saving fuel for future purposes. Not only this, it also plays a vital role in reduction of operating cost of the vehicles.

- Electronic toll system users do not stop for paying toll, this helps a lot for saving the travel time

6.1.2 Limitations

- No other vehicle is considered other than cars in training due to unavailability of data set.
- Bad quality of images do not work with this system.
- Also video input do not work for this system but can be achieved by training with proper data set.

6.2 Future Extensions

The system can be generalized so that in future every type of vehicle could register for electronic toll system and use the facility. Apart from using vehicle images, the system can be designed so that it could use video of cctv footage so that real time image processing can be easily implemented and it could enhance the toll system more adequetly. The proposed system can even implement in traffic system to reduce the speed of the vehicle and to detect the other traffic rule violation.

Chapter 7

Conclusion

The most obvious advantage of this technology is the opportunity to eliminate congestion in tollbooths, especially during festive seasons when traffic tends to be heavier than normal. It is also a method by which to curb complaints from motorists regarding the inconveniences involved in manually making payments at the tollbooths. Other than this obvious advantage, applying ETC could also benefit the toll operators

By Performing RFID based Electronics Toll tax Project we can Save Time, Money, Fuel. Reduced accident rates because of it's a totally automatic. This allows us to easily collect the toll amount without much congestion in the traffic. The number plate is uniquely identified with total accuracy being proposed to the system. Since the toll amount is easily deduced from the user account and the notification will be sent to their respective virtual machines. RFID tag cannot be cloned, so cannot be cheated. This system provide greater automated tracking capability than existing technologies, and thus create the opportunity to reduce abhor, improve inventory management and generate better market intelligence, leading to lower operational costs and increased revenue generation

Bibliography

[1] *https://ieeexplore.ieee.org/document/8675073*, A Unique Identity based Automated Toll Collection System using RFID and Image Processing

[2] https://towardsdatascience.com/automatic-license-plate-detection-recognition-using-deep-learning-624def07eaaf

[3] *https://www.rfidjournal.com/blogs/experts/entry?10743*, How Do RFID-based Toll-Collection-Systems-Work

[4] https://blog.devcenter.co/developing-a-license-plate-recognition-system-with-machine-learning-in-python-787833569ccd