INSURANCE FRAUD DETECTION, SENIOR DESIGN ENHANCED PROJECT

by

NISCHAL RANA

Presented to the Faculty of the Honors College of

The University of Texas at Arlington in Partial Fulfillment

of the Requirements

for the Degree of

HONORS BACHELOR OF SCIENCE IN COMPUTER SCIENCE

THE UNIVERSITY OF TEXAS AT ARLINGTON

May 2022

ACKNOWLEDGMENTS

First and foremost, I would like to show great gratitude and gratefulness to the beloved almighty God. Actually, to all the trillions of Gods that my religion Hinduism celebrates. Their blessings and guidance are why I am in the current position to write this and conclude my year-long project and the enhanced version of it for graduation from the Honors College.

Then I would like to send my utmost regards and thanks to The University of Texas at Arlington for this wonderful opportunity to be part of a Senior Design Team working on a very promising project. This project has been the most challenging and engaging project that I have ever been part of in my life. Certainly, Senior Design projects are the most amazing ones for all the Engineers, and mine was no different. I want to send my utmost gratitude to State Farm, who sponsored our project and gave us the opportunity to work on a State Farm sponsored project. Special thanks to the mentors from State Farm, Amy Simone, and Richins Dawsen, and our mentor of the CSE Department, Prof. Dr. Christopher McMurrough, for their constant mentorship and supervised guidance throughout the two semesters that we worked together. I highly doubt that the completed project would be this good without their utmost supervision and guidance. Big applause and appreciation to my Senior Design Team composed of Rabindra Yadav, Ai Vui Nguyen, Robert Farro, and Alexander Rollision for their constant great effort every week of school to ensure we get the major milestones the project delivered. The major hurdle with group projects is workflow and division. Certainly, some members will have to do more work

than others, and dynamics will affect the process. While we had a lot of hurdles while working, the team was determined to make this work, and every member was highly aware of their role and contribution to the project. Nobody really left anybody out, and the determination and vision of each and every member made this project a possibility. I would like to applaud each and every one of them.

Lastly, I would like to send a big thanks, a sense of gratitude, and a big chunk of gratefulness to the Honors College here at UTA for giving me such a big opportunity to be part of a group of highly intellectual students working collectively to get that Honors stamped on the degree. I feel nothing but sheer honor to be part of the Honors College. While being a part of the esteemed group and working along the way, I certainly learned a lot about myself and the various extra efforts and work we had to put in along the way. I am very glad that I went through all this in life.

May 1, 2022

ABSTRACT

INSURANCE FRAUD DETECTION, SENIOR DESIGN ENHANCED PROJECT

Nischal Rana, B.S Computer Science

The University of Texas at Arlington, 2023

Faculty Mentor: Christopher McMurrough

Insurance is a major service provided to consumers that provides some relief concerning their valuable assets in the modern world. With so many valuable objects on the market, insurance claims are at an all-time peak, and with this comes attention, both wanted and unwanted. Insurance Fraud Detection is a sophisticated web application integrated with various modules that help insurance agents determine whether a claim by a client is fraudulent or not. Insurance agents can utilize this application to web scrape the internet for clues about the item in question to determine if it is still in the original owner's possession. The application then evaluates the evidence of the case. It compares the features with the clues found via the internet to solidify the search and help the agent decide if the claim is fraudulent.

TABLE OF CONTENTS

LIST OF ILLUSTRATIONS

LIST OF TABLES

CHAPTER 1

INTRODUCTION

1.1 Origin

With the growing increase in investments in insurance among the general population and the development of a healthy habit of insuring even the smallest of valuable objects, fraudulent attempts at claiming insurance are now common and a major hurdle for Insurance companies. So, this Project is a sponsored project by State Farm to utilize technology to help mitigate this problem and reduce it to a low extent. Being one of the top-tier businesses in the insurance market, State Farm had an eye out for this problem for quite a while. With the realization that mixing up technological aspects in the modern generation is key to a modern solution, State Farm, in affiliation with UTA, made a plan and formulated a milestone to present to one of the UTA Senior Design Team of achieving a modern-day solution incorporating various modern-day technological modules. The application is built with the slogan of State Farm in its core concept and pretty much does act according to it: "Like a Good Neighbor, State Farm is there." [1]

1.2 Background

As the statistics show, a huge amount of loss is present due to fraudulent cases. Over the course of 7 years, a survey from January 2011 to December 2017, over 2000 potential fraudulent cases were registered. A total of 90 million dollars was lost due to potential fraud, which is a great deal of money being lost because of criminal actions. Adding to that, the map of the nation shows how fraudulent claims are prevalent throughout

the country, and almost half of the insurance claims made in these specific states are potential frauds. This represents a serious problem in the insurance market and a huge obstacle in the working of insurance companies. When insurance companies with their insurance market are supposed to work for the welfare of our society, providing a safe pillow to their customers regarding their valuables, they are faced with the major dilemma of processing a claim without knowing if the customer actually faced an issue with their valuable object or not. They are made to pay up or cover for the wrong demographic of customers, making the life of regular honest customers miserable with added laws and regulations.

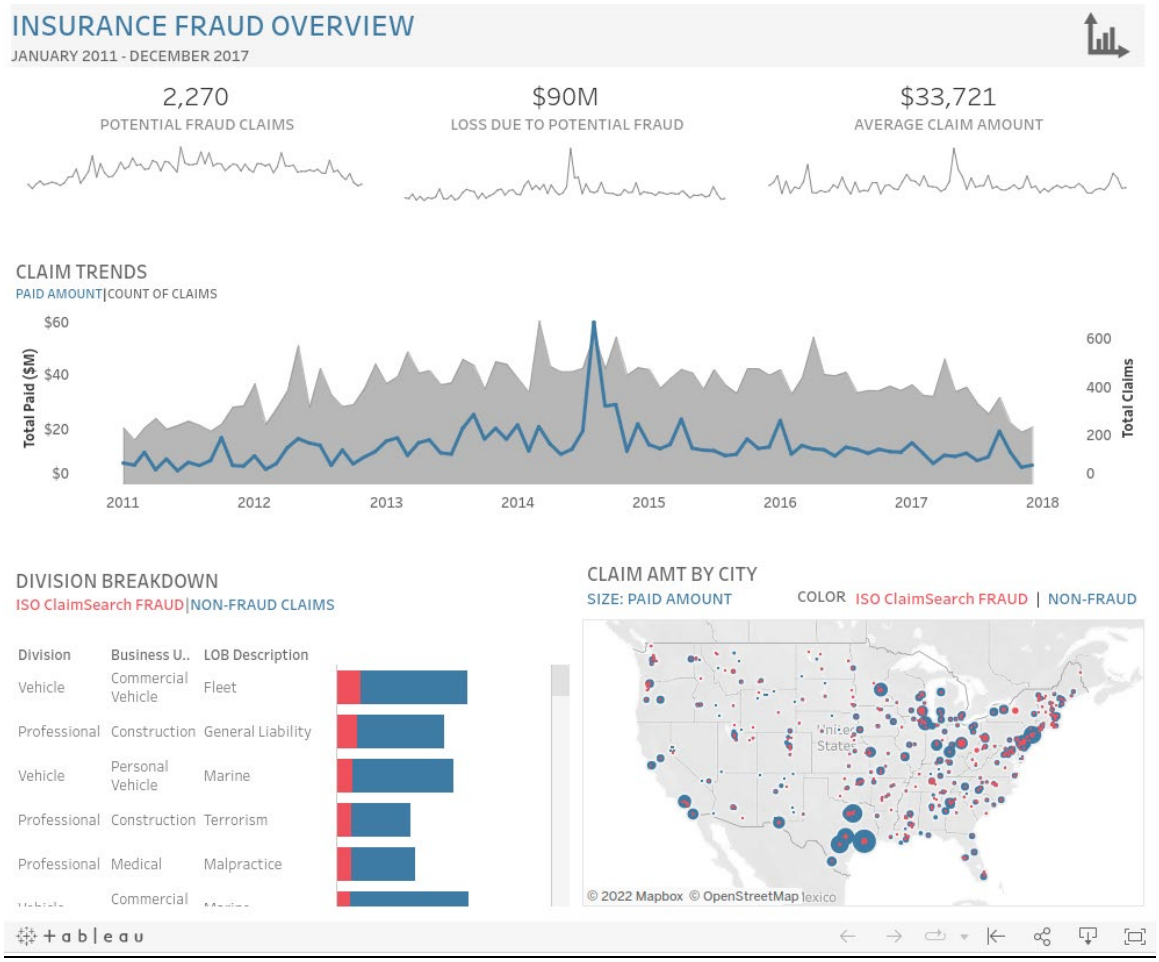

Figure 1.1: Insurance fraud overview through graphs and charts

1.3 Introduction

Insurance Fraud Detection is a sophisticated web application designed with the major problem in the insurance market in mind, integrated with various modules that help Insurance agents determine whether a claim by a client about their belongings is fraudulent or not. The application is a sophisticated web application integrated with object detection and feature extraction modules, database modules, web scraping modules, and machine learning and deep learning modules in binary classifiers. Originally, State Farm presented the team with a simple plan and formulation that they are aiming to have a web application that has object detection and Feature extraction abilities to sort of taking the input of photographic evidence from the agent about the object in question and scrape the internet for valid locations such as any social media website and e-commercial markets where the client might have posted any sort of post to boast about their valuable item after they filed the claim or they might have posted it on markets to list them for sell after they filed the claim. Subsequently, the end goal was to gather enough evidence from the internet regarding any hints of the current claim being fraudulent. The hints included primarily photographic features and the other attributes of a post, such as a date/time, behavior of the post, the nature of the post, and many other attributes.

Along with the main application, the enhanced edition or extra feature that I added to the entire project as part of the Enhanced Capstone Project was the ability of the application to work with videographic contents. Initially, our setting was just plainly for photographic images of the original item that the application would work on to make findings of any sort. But this might not always be the case in the real-world setting. For instance, certain types of evidence may not always be available. Sometimes clients might

show up with only videographic content. Therefore, my goal was to integrate the application with additional modules that would enable it to work with videographic evidence as well as photographic evidence that the client would show up with.

CHAPTER 2

LITERATURE REVIEW

While State Farm was debriefing the team initially at the start regarding what the project was going to be and what measures were to be taken, they were explicitly keen on not planning too much as to get on it right away. The major idea was to formulate different modules and take each of them one step at a time, along with planning at the same time. So, the concept was to keep solving and plan along with solving to get to the milestone. Therefore, the team did not really spend a lot of time on Literature Review to look for similar works that were done, but this was also because of the reason that not exactly similar applications have been attempted in the past.

Here are some of the works that were done in the past that we were able to find that were similar to our work or modules that we were trying to work with.

SDD - similar drawing detection: Identify similarities in 2D drawings. The algorithms find similar drawings with "SDD - similar drawing detection." The evaluation is performed on neutral formats such as .pdf and is highly precise. This enables customers to analyze even older data - where no 3D models exist. This is an algorithm that would work on 2D drawings, but the requirements are specifically for real objects in photos. [5]

• CamFind - CamFind is an image search tool. Once the picture is uploaded, the app matches it against others on the internet and identifies the object. You can look for related images, shop for the item, watch related videos, search on the web, or go through related posts. [2]

• Google Lens - Google Lens initially arrived as a Pixel exclusive; the company later integrated the feature into Google Photos. Now both iOS and Android users can use Google Lens to reverse search images. On Android, Google Lens is available as a standalone app. iOS users can access Lens within the Google Photos app. Open an image you have taken and tap the Lens icon (second from the right, next to the trash icon). Google's visual search engine will analyze the image; the results include a link to the Google image search page. [3]

• Veracity - Veracity is an intuitive visual search engine app. It lets you choose images from your Camera Roll or Photo Library, plus it can link to your Dropbox account. Veracity offers a basic image editor, but you must pay to unlock it. [6]

• Reverse Image - Reverse Image Search App provides another minimalist reverse search engine experience. Take images from your Camera Roll or Photo Library to reverse image search via Google Image Search, Yandex Image Search, and Bing Image Search. You can also crop, rotate images, and save them to your device for free. All the above applications do not take geographic locations into the calculation, which is one important aspect of narrowing the search. [4]

CHAPTER 3

METHODOLOGY

The major objective of this project was to assemble a web application that would help insurance companies search the internet for any clues regarding the object in claim on whether the claim was valid. This was determined because the internet is so prevalent today; relevant information about an individual's belongings would be present, with social media being the primary destination where clients would post about those belongings. Then it had to produce a report with appropriate findings and conclusions based on its operations.

3.1 System Overview

The user can upload the description and/or images of the item along with the information of the claimer. Then, the claim scanner scans social media and returns images with items similar to the lost one. Here, the Object detection algorithm will help detect the object present in social media. The feature extraction algorithm will then extract the physical properties, such as size, color, brand, etc., of the object from the images. If the features match the claimed item, the program will analyze the context and check the readability to the owner, like the owner of the post, date of the post, etc. If any connection exists, the claim scanner flags the claim as potentially fraudulent.

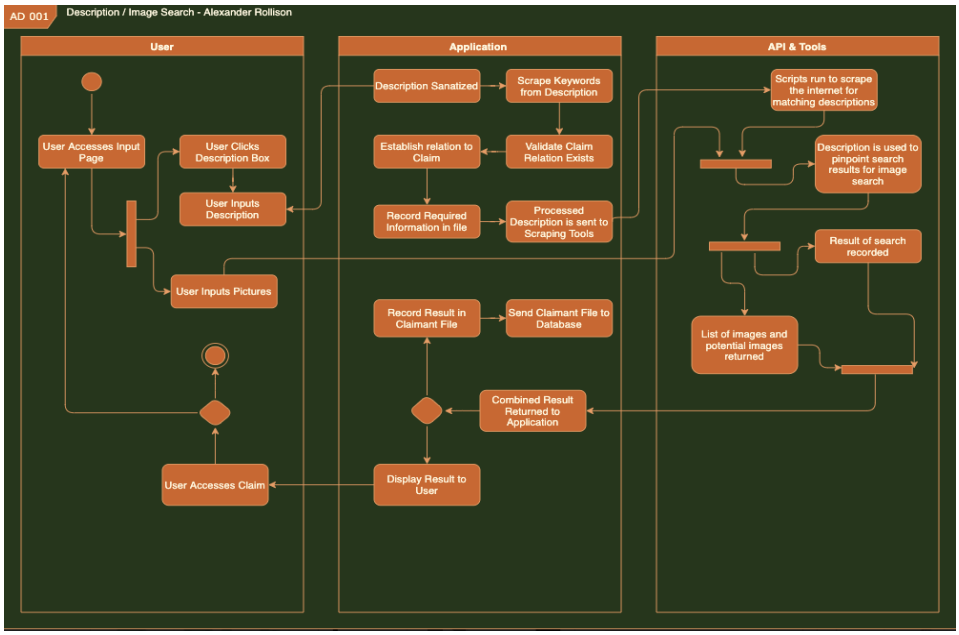

Figure 3.1: Activity diagram of the entire system

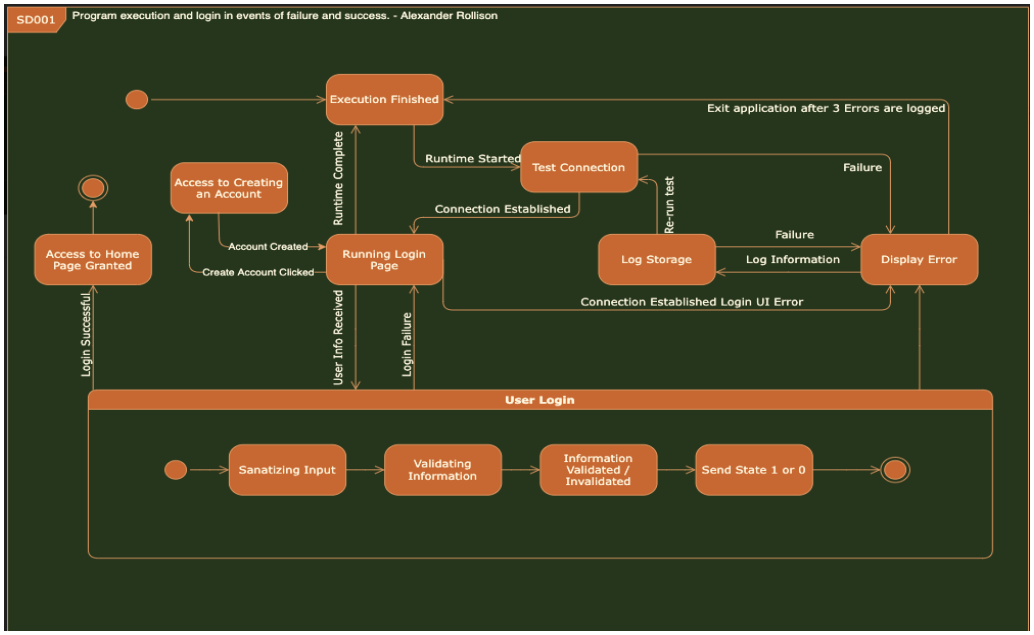

Figure 3.2: State transition diagram of the login/registration feature

3.2 Main Methodology

The application has a Scanner page that comes with a form for the insurance agent to go ahead and put-up necessary information regarding the claim in question, including certain social media accounts and e-commerce platform profiles, along with picture evidence of the original object in question. Then the application takes all the information and stores them in its database modules. It then web scrapes the internet using a web scraping application at provided locations, including potential locations. The application determines the best places to gather related image evidence that could provide a basis for comparison. Then, when certain evidence is found, the object detection and feature extraction module would compare such images with the original image to determine whether the original one is present on the internet.

Suppose there are solid hints at the original item being on the internet. In that case, further analysis is done regarding the different elements of the one found over the internet, such as the posted date and time, the nature of the post, and what regards it was posted. Analyzing all these elements would help the application make a report regarding all the evaluations it includes along with the match it was able to find along with the search and present it to the insurance agents with plenty of information to make a judgment upon.

3.3 Extra Component Methodology

Regardless of the source of evidence, the logical flow from the jump start was simple: The application would take the evidence into its account, store it, and then utilize it to figure out matching evidence over the internet at all the possible locations. Regardless of the form of evidence, in the end, the application would make conclusions by comparing features that it could extract from the original content and the content it was able to find on

the internet. Therefore, the only tweak it required was simply for the application to accept videographic evidence. For this, the major component to work with was the object detector module, as it is the one that is responsible for making adequate comparisons between the original content and the ones it could gather.

Our Object Detector was OpenCV and its library YOLO, which initially simply took the picture and processed its features. Therefore, the only work required was to present whatever evidence we get as pictures and frames of the object visibly present in the picture. So, I simply added some components such that the OpenCV would be able to take the given video input, process it, and capture frames of pictures with the object in proper visibility. It wasn't a lot of work as OpenCV already came with added functionalities of working with videos. So, it basically knew how to work on it. I simply had to redirect the video evidence to the OpenCV specifically labeled as videos, and then only the module would know how exactly to work on it. After it is able to extract some picture frames from the video evidence, then the rest of the work is as it was before and therefore would continue its work regularly.

CHAPTER 4

DISCUSSION

In the team's regard and our mentor's regard, the work we tried to achieve was actually very useful in the sense of practicality in the outside world, given the circumstances in the Insurance Business world. With the fast-paced outside world where everybody is super busy and constantly moving, to have a technological aspect to the issue that could provide relevant information and evidence to help determine potentially fraudulent claims in quick response time is a mind-blowing asset to have by your side, especially if you are an Insurance company. The quick response time is key as many Insurance companies cannot afford to conduct deep analyzed research with time constraints and a lot of resources to be invested for a long period of time to determine a case simply. They cannot afford, mainly to invest heavy resources and time heavily on such investigations, which could be numerous in numbers. They cannot have their client wait oddly for a decision to be made when the client simply wants the investment they made in return. Also, the degree of accuracy matters a lot, as one small false claim would mean the client will terminate the relationship furiously, and the company's reputation will be at risk because of it. Therefore, having a modern answer to the problem is a huge boon with all these constraints in mind. The modern answer comes with all the constraints in mind as the modern generation is actually fond of technology because of its characteristics to address all the problems discussed earlier.

4.1 Roles and Responsibilities

The main stakeholders of the project would be the professor who would grade our project, the sponsors who supervise our workflow, the team working on the fraud detection web application, and any who may use or test the application.

The point of contact (PoC) from the sponsor or customer side is Alexander Rollison, but a Teams group chat will be created for the team and the sponsors. As for the roles of the rest of the team, everyone is to work with a near-equal workload. Robert Fabbro is the team lead and scrum master working heavily in machine learning. Alexander Rollison will handle most UI work, database operations, and result processing. Rabindra Yadav and Nischal Rana will be working on UI and parts of the application to connect different components such as the scraping programs to the Team Anti-Fraud - Fall 2021 classification programs. Ai Nguyen is the team's data-set collector who will also provide documentation for the project. When it comes to the process of building the application, the whole team comes together every Thursday and Friday in meetings to discuss progress as well as to help create/add to sprints before Monday class meetings. All tasks are mostly divided evenly, and if the team lead is unsatisfied with the work of one of the individuals, they will address these problems with the professor. The team will maintain the product owner and scrum master as the same person for the whole semester and change in the second semester based on need.

4.2 Product Description

This is an overview of the features and functions of the small claim fraud tool. From the perspective of the end users, maintainers and administrators, the primary operational

aspects of the product are defined here. The key features and functions found in the product and critical user interactions and user interfaces are described in detail.

4.2.1 Features & Functions

The product will determine whether a person is possibly committing small claims fraud on specific inputted items by searching social media accounts. The product will only look at images and items within those images on said accounts and compare them to a ground truth image or images of the product. The product will also scan for mentions of the item in social media posts. The product will not scan auctions or selling websites such as Craigslist or Facebook Marketplace. Cases that are tagged as possible fraud will appear in the "Cases" section as per figure 1. The scanning page will look roughly like figure 1, with it being subject to some change given changes in development priorities or design changes. The cases section will have folders or files that will be accessible by the user to view in more depth the social media posts that were flagged as suspicious. The application will require internet access and an external web server to host the scanning algorithms.

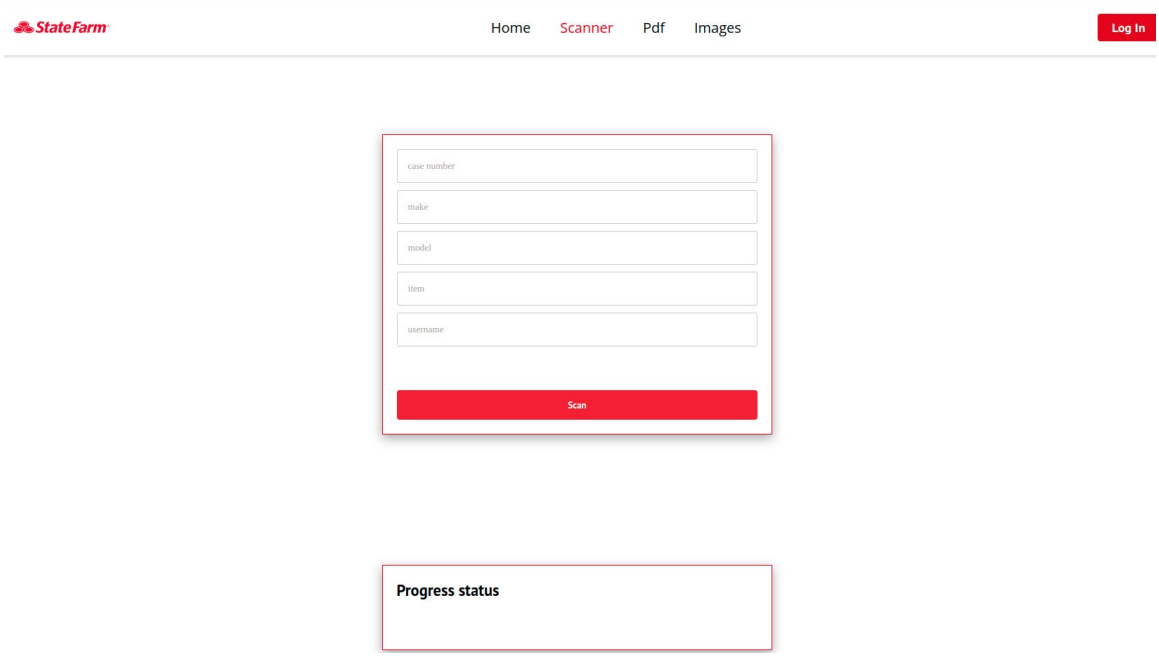

Figure 4.1: Scanner tab with the appropriate form

4.2.2 External Inputs & Outputs

Describe critical external data flows. What does your product require/expect to receive from end-users or external systems (inputs), and what is expected to be created by your product for consumption by end-users or external systems (outputs)? In other words, specify all data/information to flow into and out of your systems. A table works best here, with rows for each critical data element and columns for name, description, and use.

Inputs		
Name	Description	Use
Case Number	The case number that is associated with the claim.	Will be needed to associate flagged cases for further investigation.
Item	The item name.	Will be used to find images of the item server side.
Item Description	Description of the item.	Will be used to search text sub-missions on social media.
Date	The date when the claim was submitted.	Will be used to only search posts after this date for the item.

Table 4.1: Various input elements for the system

Output		
Name	Description	Use
Claim Files	File section that contains a report on whether a claim is suspicious or not.	Will be used by the user to determine possible fraud.
Claim Images	Images that the machine found to be suspicious.	Will be used by the user to easily determine if the item is actually there.
Claim Posts	Links to the suspicious posts.	Will be used to easily access any posts that contain images or text.

Table 4.2: Various output elements for the system

CHAPTER 5

CONCLUSION

In conclusion, the year-long project our team was working on for the whole last year over two semesters are finally over. I feel grateful and lucky to be a part of such a fun and interesting project. Certainly, I learned a lot along the whole process while working on the main project and implementing extra components to enhance the project. Overall, as a team and our mentors, we are very happy with the milestone we achieved with the project. The application integrated with such heavy modules was actually running smoothly, and the logical flow was excellent and never got dismissed. At the end of the whole project, we had a meeting with different State Farm personnel. We basically described to them how this application is useful for a big-time Insurance company like State Farm and why they should consider investing in the modern-day solution. We believe that the application can be a great asset to the companies. Given the right investment, research, and work, this application can be taken a step further to an even greater extent to solve the problem at hand.

REFERENCES

[1] State Farm Official Slogan. https://www.statefarm.com/

[2] Image Searcher Incorporated (2015). “CamFind - The First Visual Search Engine Goes Social.” Retrieved May 2, 2022, from https://www.prnewswire.com/news-releases/camfind---the-first-visual-search-engine-goes-social-300067412.html.

[3] Chris Hall (2022). “What is Google Lens and what can it do?” Retrieved May 2, 2022, from https://www.pocket-lint.com/apps/news/google/141075-what-is-google-lens-and-how-does-it-work-and-which-devices-have-it.

[4] Matt Golowczynski (2020). “Google reverse image search: Everything you need to know.” Retrieved May 2, 2022, from https://smartframe.io/blog/google-reverse-image-search-everything-you-need-to-know/

[5] “SID - similar item detection | Reduce the variety of parts”. Retrieved May 2, 2022, from https://shouldcosting.com/en/services/sid-similar-item-detection-en

[6] Nicole Martins Ferreira (2020). “8 Best Image Search Engine: How to Reverse Search Images on Google Easily.” Retrieved May 2, 2022, from https://www.oberlo.com/blog/image-search-engine.

BIOGRAPHICAL INFORMATION

Nischal Rana is currently a junior at UTA presently majoring in Computer Science. Originally from Nepal, a small landlocked country located in the Southern part of East Asia, he comes from a heavy STEM background. The country's education system pushes its students heavily on the STEM aspect more than anything. Upon graduating from high school back in Nepal, the degree and career in Engineering were set by him for his future from an early age.

As of right now, he is rigorously aiming to get that bachelor's degree in Computer Science with Honors. Over the course of six semesters, he definitely has learned a lot in the field of Computer Science. While being heavily invested and interested in coding, he also has a wide variety of interests in the theoretical aspect of modern Computer technology. Therefore, he elected to change from Computer Engineering to Computer Science. His interests mainly range from Software Development, Web Development to numerous Computer Science topics of Artificial Intelligence, Machine Learning, Theoretical Concepts including Compilers, and many more. He has been part of many project groups, among which the recently completed Senior Design Project is his biggest feat. It was his first-time being part of such a huge project group. It was also sponsored by a big company such as State Farm. His future plans involve having a career as a Web developer/Software developer with expertise in other fields as well.