



Tezpur University

Minor Project Progress Report on - “Malware Detection Using a Chatbot”

Abhilash Chutia [CSM23010]

Bishal Sharma [CSM23024]

Submitted to-
Dr. Dhruva K Bhattacharyya,
Sr Professor, Department of CSE,
Tezpur University

Acknowledgment

We are very thankful to everyone who worked with us in bringing our project, "**Malware Detection Using a Chatbot**," to our next milestone.

We first, all of us, would like to thank **Dr. Dhruba K Bhattacharyya, Sr. Professor, Department of Computer Science and Engineering, Tezpur University**, for his overall guidance, encouragement, and very much insightful feedback that this project had through him; his expertise and advice led to shaping our work's direction.

We also gratefully appreciate the resources and environment provided by the **Department of Computer Science and Engineering, Tezpur University**, that were helpful in research and development.

We would specifically like to thank our peers and friends at various stages of this project work. Their constructive discussions and feedback added richness to the outcomes of this work.

Furthermore we would like to extend our deepest gratitude to our families for their constant encouragement and understanding, which motivated us to achieve our goals.

This project has been a great learning experience, and we are indeed indebted to everyone who made this journey both memorable and rewarding.

Table of Contents

1. Project Overview

- Goal and Objective
- Technologies Used
- Problem Scope

2. Progress Summary

- Chatbot Basics and Applications
- Chatbot Bias and Mitigation
- Indian General Knowledge Bot
- Malware Analysis

3. Detailed Summary

- Presentations
 - Chatbot Basics
 - Chatbot Bias
 - Malware Concepts
- Development of Bots
 - Basic Chatbot
 - Knowledge-Based Bot
 - Malware Detection Bot

4. Sherlock – Malware Detective

- Design and Libraries
- Key Features
- Workflow

5. Features to be Implemented

- Assamese Language Support
- Language Selection Module

6. Summary

- Achievements
- Future Directions

Project Overview

The goal of this project is to achieve an intelligent working Chatbot with the capabilities to consult an user through a series of questions at the end of which the Bot will be able to correctly identify if the user machine is infected by a malware and the type of malware.

By leveraging Natural Language processing, the bot is able to assist as a cybersecurity helper in identifying suspicious behaviour or indicator of compromise depending on the user inputs.

This project is a combination of chatbots, machine learning and cybersecurity.

Progress Summary

Understanding and Presenting Chatbot Basics: Analysing chatbot technology, its applications, and potential in cybersecurity.

Addressing Chatbot Bias:

Exploring how biases can affect chatbot interactions and ensuring an unbiased user experience.

Developing Knowledge-Based and Detection Chatbots:

Implementing a General knowledge bot using the Gemini API to answer questions about India.

Malware Analysis and Presentation:

Researching and presenting key concepts of malware to inform the chatbot's detection capabilities.

Developing Malware Detection Chatbots:

Using the knowledge from the previous trial with the Gemini API, develop another Bot that can handle user interaction and identify malware threats.

Detailed Summary

1. Presentation on Chatbot

This presentation covered the following topics:

1. What is a chatbot?

2. Types

- Rule Based Chatbot

- Open Ended Chatbots

3. Hardware/ Software Requirements

4. Python Libraries

- Natural Language Toolkit [nltk]

- Transformers [Transformer Models]

- Scikit-learn [Machine Learning Algorithms]

- TensorFlow [Deep Learning Models]

- Django [Web Interface]

5. Advantages

6. Use Cases

7. Challenges faced by Chatbots

2. Basic Chatbot

A basic chat bot has the following requirements to function:

1. A intents data file containing the interaction expected of the bot. This is in the form of a json file
2. Word Lemmatizer is used to reduce words to their base forms
3. Punctions are ignored
4. Words are extracted from the patterns and stored in a list called “Words”. Unique intent tags are stored in a list called “Classes”
5. The pickle library is used to save the vocabulary and classes for future use.

Model Training

1. TensorFlow's Keras library is used to create a sequential model.
2. This model has two hidden layers with 128 and 64 neurons respectively. These layers extract features from the given data and learn about the complex relationships between the words and intents.
3. Dropout layers prevent overfitting of the training data.
4. The Stochastic Gradient Descent [SGD] is used to train the model.
5. The model is saved as "chatbot_model.h5" to be used again.

Libraries Used

1. Random
2. Json
3. Pickle
4. Numpy
5. TensorFlow
6. Nltk


```
GO! Bot is running!
Hello
1/1 ————— 0s 123ms/step
Good to see you again
What can you do?
1/1 ————— 0s 50ms/step
I am a general purpose chatbot. My capabilities are :
1. I can chat with you. Try asking me for jokes or riddles!
2. Ask me the date and time
3. I can google search for you. Use format google: your query
4. I can get the present weather for any city. Use format weather: city name
5. I can get you the top 10 trending news in India. Use keywords 'Latest News'
6. I can get you the top 10 trending songs globally. Type 'songs'
7. I can set a timer for you. Enter 'set a timer: minutes to timer'
8. I can get the present Covid stats for any country. Use 'covid 19: world' or 'covid 19: country name'
For suggestions to help me improve, send an email to ted.thedlbot.suggestions@gmail.com . Thank you!!
|
```

```
Tell me a joke
1/1 ————— 0s 41ms/step
o what if I don't know what 'Armageddon' means? It's not the end of the world.
shut up
1/1 ————— 0s 27ms/step
Well that hurts :(
who is Abhilash?
1/1 ————— 0s 31ms/step
Head over to his any of his social profiles to find out! Linkedin: www.linkedin.com/in/abhilash-chutia
you are awesome
1/1 ————— 0s 23ms/step
Thank you!
|
```

```
you are awesome
1/1 ————— 0s 23ms/step
Thank you!
Who inspires you?
1/1 ————— 0s 35ms/step
Personally, I find Chat GPT very inspiring. I might not be very fair though..
ask me a riddle
1/1 ————— 0s 41ms/step
What two things can you never eat for breakfast?.....Lunch and Dinner!
how old are you?
1/1 ————— 0s 33ms/step
I was made in 2024, if that's what you are asking!
bye
1/1 ————— 0s 39ms/step
```

3. Presentation on Bias

This presentation included the following topics

1. What are Biases?

2. Types of Biases

Data Bias

Data Collection

Data Representation

Historical Bias

Selection Bias

User Interaction Bias

Algorithmic Bias

Presentation and Ranking Biases

Gender Bias

3. Impact of Biases on Chatbot

4. Strategies for mitigating Biases in Chatbots.

4. Indian General Knowledge Bot

The libraries used here are:

1. Flask
2. OS
3. Google Generative AI

The chatbot is designed as a web application.

The users can interact with the bot by typing into a text field.

The Flask framework is used to handle user input and communication with the GenAI model.

GenAI Integration:

1. API key is configured to access the large language model capabilities.
2. Gemini 1.5 flash model is used for response generation, fluency and factual accuracy.
3. A chat session is initialized to provide a history of previous questions with the bot.

Flask Application:

A Flask application is created to handle web requests and responses.

Routing:

The root path is mapped to the index function which renders the interface Html template for the chatbot.

The ask path is mapped to the ask function which handles user interactions.

Overview:

The chatbot allows user to ask questions related to India. The Gemini Ai model utilizes its knowledge and understanding to generate the most informative response.

If the user asks questions unrelated to the topic of India, the model kindly lets the user know of its limitations and redirect the user to questions related to India.

INDIA-BOT

Ask a question about India

Ask

Your Question:

Who is the president of India?

India-Bot's Answer:

The current President of India is "Droupadi Murmu". She assumed office on 25th July 2022, becoming the 15th President of India. She is the first tribal woman to hold this prestigious position and is also the second woman to be elected as the President after Pratibha Patil.

INDIA-BOT

Ask a question about India

Ask

Your Question:

What is the largest river in India?

India-Bot's Answer:

The largest river in India, by volume of water, is the "Ganges River". It's a sacred river for Hindus, and plays a vital role in the culture and economy of northern India. It's also one of the longest rivers in Asia, flowing for over 2,500 kilometers (1,550 miles). Do you have any other questions about the Ganges River or other rivers in India?

INDIA-BOT

Ask a question about India

Ask

Your Question:

When was the Great Wall of China built?

India-Bot's Answer:

I'm designed to answer questions about India. I'm sorry I cannot answer your question. The Great Wall of China is an incredible structure with a long history, but it's not related to India. Would you like to learn about any fascinating historical structures in India? We have a rich and diverse history, with many ancient forts, temples, and palaces to explore!

5. Presentation on Malware

The Presentation on Malware consisted of the following topics:

1. What is Malware?

2. Types of Malware

Viruses

Ransomware

Worms

Trojans

Spyware

Botnets

3. Signs of Malware infection

4. Steps to take on Malware Detection

6. Sherlock – Malware Detective

The libraries used here are:

1. Flash
2. OS
3. Google generative AI

The chatbot is designed as a web application.

The users can interact with the bot by typing into a text field.

The user input is stored as history so that the the AI model can remember the context and generate appropriate responses.

The Flask framework is used to handle user input and communication with the GenAI model.

An API key is configured to access the large language model capabilities.

Gemini 1.5 flash model is used for response generation, fluency and factual accuracy.

A chat session is initialized to provide a history of previous questions with the bot.

Overview

Introduction:

The chatbot introduces itself as "Sherlock" and starts by asking questions about the user's operating system and system setup.

Interactive Q&A:

Sherlock poses a series of questions to understand user symptoms, system characteristics, and any recent abnormal behaviour.

Symptom Analysis:

Based on user responses, Sherlock provides insights on whether the system might be infected.

Session Continuity:

Sherlock retains the chat history for a seamless conversational experience, allowing it to remember previous responses and build on them.

The "Malware Detection Using Chatbot" project provides an innovative, user-friendly way to assist individuals in identifying potential malware issues on their systems. This project demonstrates the potential of AI-driven tools to assist in cybersecurity awareness and initial malware assessment.

Chat with Sherlock: The Malware Detective

You: Hello

Sherlock: Greetings! I am Sherlock, the malware detective. I've been created by Abhilash Chutia and Bishal Sharma from Tezpur University to help you understand if your computer might be infected by malware.

To start our investigation, could you tell me what operating system you are using? (e.g., Windows, macOS, Linux)

Send

Chat with Sherlock: The Malware Detective

You: Windows 10

Sherlock: Thank you. Windows 10, noted. Now, could you tell me what version of Windows 10 you're running? (e.g., 21H2, 22H2)

Send

Chat with Sherlock: The Malware Detective

You: 22H2

Sherlock: Excellent. Windows 10, 22H2. Do you have any antivirus software installed on your computer? And if so, which one?

Send

Chat with Sherlock: The Malware Detective

You: None

Sherlock: Interesting. No antivirus software. Have you noticed any unusual behavior on your computer recently? This could be anything from slow performance to unexpected pop-ups or changes in your files.

Send

Chat with Sherlock: The Malware Detective

You: My files are locked with the .coos extension

Sherlock: Ah, that is very concerning. Files locked with the .coos extension are a strong indicator of a ransomware infection. Ransomware is a type of malware that encrypts your files, making them inaccessible, and then demands a ransom payment to unlock them.

Let's gather some more information to be sure. Have you received any messages or pop-ups demanding payment?

Send

Chat with Sherlock: The Malware Detective

You: There is a text file demanding cryptocurrency in exchange for the decryption key

Sherlock: That confirms it. This is a classic ransomware attack. You are correct to be concerned.

Without an antivirus solution, the likelihood of ransomware infection is higher. And the presence of the .coos extension and ransom demand leaves little doubt.

I strongly advise against paying the ransom. It is not guaranteed to work, and it will only encourage further attacks.

Instead, I recommend the following:

1. ****Disconnect from the internet immediately.**** This will prevent further spreading of the ransomware and potential data theft.
2. ****Do not attempt to decrypt the files yourself.**** This can further damage your files.
3. ****Contact a reputable cybersecurity firm or IT professional.**** They can help you recover your files and secure your computer from future attacks.

I know this is a stressful situation, but rest assured, there are ways to recover. Remember to take immediate action to mitigate further damage.

Send

Chat with Sherlock: The Malware Detective

You: Thank you

Sherlock: You're welcome. Remember, I'm here if you have any further questions. Good luck with your recovery.

Send

Future Direction:

[To be implemented]

1. **Assamese** being our mother tongue meant that we could try to focus this piece of software to help people who can read and speak Assamese.

To achieve this goal,

- ❖ A translation library is required.
- ❖ A pathway to send the data from the Gemini API to the translation app and back to the local project.
- ❖ A translate function to handle the message translation.

2. A selection page to select between Assamese and English.

Using flask to route the selection to their respective chatbot.

Summary

The "Malware Detection Using a Chatbot" project, undertaken by us, Abhilash Chutia and Bishal Sharma under Dr. Dhruba Kr Bhattacharya, aims to develop a chatbot that identifies malware on a user's system. This chatbot uses natural language processing to interact with users, helping detect suspicious behaviour that could indicate malware. Key stages of the project include:

Chatbot Basics and Bias: We studied chatbot functionalities, types, and applications in cybersecurity, addressing potential biases that could impact user interactions.

Chatbot Development: Initial versions included a general knowledge chatbot using the Gemini API to answer questions about India. This served as a foundation for building a more advanced malware-detecting bot.

Presentations on Malware: Presentations on different malware types (viruses, ransomware, etc.) and indicators of infection helped enhance the chatbot's diagnostic capabilities.

Sherlock – The Malware Detective: Named "Sherlock," this chatbot functions as an interactive web application using Flask and the Gemini AI model. "Sherlock" asks users about system details and symptoms, analysing responses to determine potential malware presence. It remembers chat history for coherent, context-aware responses.

This project exemplifies how AI-powered tools can aid in cybersecurity by identifying malware threats, creating an accessible and user-friendly method for initial malware detection.

