Project Name:

Sentiment Analysis for Sexual Assault Call Center

Statement of Work

Student Names:

Gal Druker – 318636701

Liron Golan - 207253428

Supervised by Doctor Sharon Yalov Handzel

Partner names: Dr. Adi Fux, Shiran Carmeli, Moriya Greenhut

Date of Submission:

Table of contents

Introduction

In an era defined by digital interactions, this project analyzes calls from sexual harassment support centers. Through sentiment analysis, we explore the vital task of classifying these calls as urgent or non-urgent.

- **Addressing Urgency in Support Systems:**

  Survivors increasingly turn to digital platforms, emphasizing the critical need to understand sentiments expressed in these conversations.
  This project aims to enhance support systems for survivors of sexual assault.

- **Methodologies, Challenges, and Ethical Considerations:**

  Beyond technical advancements, our exploration encompasses research methodologies, encountered challenges, and ethical dimensions.
  We contribute insights to the dialogue on utilizing technology for societal betterment.

- **Sentiment Analysis Unveiled:**

  At the core of our project is sentiment analysis, a tool in natural language processing.
  By dissecting sentiments in text conversations, we aim to revolutionize how sexual harassment support centers prioritize and respond to calls.

Project Goals

Our project aims to harness Natural Language Processing (NLP) to refine support services for survivors of violence against women and girls. By training an NLP model on pre-classified data—spanning legal, counselling, and emergency needs—we aim to categorize new text conversations swiftly and accurately. A vital feature of this initiative is the development of a user-friendly interface to streamline the analysis process. This effort is directed towards ensuring survivors receive timely and appropriate assistance, laying the groundwork for an in-depth exploration of each project phase.

1. Develop a Natural Language Processing (NLP) Model: Utilize the already classified text conversation data to train an NLP model. The training process will focus on enhancing the model's ability to accurately understand and interpret the nuanced communications from survivors of violence against women and girls.

2. User Interface (UI) Creation: Design and build a user-friendly UI allowing easy input of new text conversations. This interface will be integral in facilitating the immediate analysis of conversations.

3. Automated Support Categorization: Implement the trained NLP model within the UI to automatically categorize new conversations into predefined support needs: legal assistance, counselling, or emergency intervention. This feature aims to streamline the decision-making process regarding the type of support required.

4. Accuracy and Efficiency Improvement: The NLP model aims to enhance the precision and speed of identifying the appropriate support services for survivors, ensuring they receive the necessary assistance promptly.

Project Metrics

To effectively evaluate our project, we have outlined specific, quantifiable objectives as follows:

1. Model Performance Metrics:
    - Achieve at least an 80% accuracy rate in correctly classifying calls as urgent or non-urgent, ensuring the model can reliably identify the sentiment and urgency of conversations.
    - Attain precision and recall rates of at least 75% in distinguishing between different types of support needs (legal assistance, counseling, emergency intervention), to minimize misclassification and ensure survivors receive the appropriate support.

2. User Interface (UI) Usability and Engagement:
    - Secure a user satisfaction score of at least 4 out of 5 in usability tests for the UI, confirming its ease of use, intuitiveness, and effectiveness in facilitating the analysis of text conversations.

Project Gantt

Initial requirements from the system (High Level)

System goals

As outlined in previous sections, the primary objective of this project is to develop a Natural Language Processing (NLP) model capable of analyzing calls from sexual harassment support centers. This model will classify calls as urgent or non-urgent based on sentiment analysis.

Stakeholders

Stakeholders in this project are:

- Survivors of sexual assault.
- Support center staff.
- System administrator.
- Legal and counseling professionals.
- Developers.

Following our consultation with The Aid Center for Victims of Sexual Assault in Haifa and the North, it became evident that their primary challenge is the timely identification of urgent messages. This capability is crucial for enhancing their responsiveness and providing more effective and immediate support to those in need.

Functional requirements

Requirements from the NLP Model

- The NLP model should accurately classify the urgency of calls by analyzing sentiment, ensuring high reliability in distinguishing urgent from non-urgent conversations by these classifications - legal assistance, counselling, or emergency intervention.

- The model will continuously learn and improve its accuracy by training on new data collected from text conversations.

System Requirements

- Ensure seamless communication between the server hosting the NLP model and the user interface, facilitating real-time analysis and classification of text conversations.

- Perform all NLP model training, text analysis, and classification processes on the server to minimize the computational load on the end user's device.

- The system must efficiently store and retrieve classified conversation data in a database, including urgency status and sentiment analysis results.

Requirements from the User Interface (UI)

- The UI will present a clear and intuitive dashboard for users to input new text conversations and view analysis results.

- Display detailed sentiment analysis results, including the urgency classification for each conversation.

Non-functional Requirements

Design Requirements

- The interface should be user-friendly, allowing users with no technical background to navigate and utilize the system effectively from their first interaction.

- Adopt a minimalist design approach to ensure clarity and prevent user overwhelm, facilitating a straightforward analysis process.

Performance Requirements

- The UI should deliver analysis results promptly, with specific benchmarks for response times to be determined based on testing.

Maintenance Requirements

- Design the system for easy updates and compatibility with future versions, ensuring long-term sustainability.

- Implement a scalable database architecture capable of handling growing data volumes without performance impact.

Reliability Requirements

- Ensure system stability and reliability, maintaining high availability for users to access and use the application as needed.

Privacy Requirements

- Wherever possible, data will be anonymized to protect the identity of survivors.
- The data collected will only be data that is strictly necessary for the classification and analysis of calls.

Use cases

Use Case 1: User Authentication

- Actors: Support center staff
- Preconditions: The user has been registered in the system with appropriate credentials.
- Basic Flow:
  - The user navigates to the login page.
  - The user enters their username and password.
  - The system authenticates the user and grants access based on their role.
  - Postconditions: The user can access the system functionalities relevant to their role.
- Alternate Flows:
  - Invalid credentials: The system displays an error message and prompts the user to try again.

Use Case 2: Conversation Input and Analysis

- Actors: Support center staff

- Preconditions: The user is logged in and can input new conversations.
- Basic Flow:
    - The user inputs a new text conversation.
    - The system processes and analyzes the conversation to determine its sentiment and urgency.
    - The system categorizes the conversation as urgent or non-urgent and displays the results to the user.
- Postconditions: The conversation is categorized.
- Alternate Flows:
    - Analysis error: If the system cannot analyze the conversation, it prompts the user to try again.

Use Case 3: Continuous Learning and Model Improvement

- Actors: developers
- Preconditions: New data from analyzed conversations is available for training.
- Basic Flow:
    - Developers collect and prepare new training data from recent conversations.
    - The NLP model is retrained with the updated dataset to improve accuracy.
    - The updated model is deployed, enhancing future sentiment analysis and urgency detection.
- Postconditions: Based on the latest conversational data, the system's ability to accurately identify urgent messages is improved.

Architectural Requirements

System Architecture

The system will adopt a simple client-server architecture, with the server handling sentiment analysis and data storage and the client providing a user interface for input and results display.

Performance

Ensure a response time that will be determined based on the model for analyzing and classifying text inputs to keep user wait times minimal.

Security

Essential security measures to protect privacy include secure login mechanisms.

Data Architecture

Use a simple relational database for storing user data and conversation analysis results, ensuring data integrity and straightforward backup procedures.

User Experience

A user-friendly web interface that is responsive and intuitive, allowing easy navigation and operation without extensive training.

Technology Stack

A straightforward stack using Python and Flask for the server side to handle NLP tasks and sentiment analysis, with HTML, CSS, and JavaScript for the client side.

Scalability

While initially designed for lower traffic, the system architecture should allow for future expansion, such as adding more server resources or optimizing the database for faster queries as user demand increases.

Technological Requirements

Programming Languages and Frameworks

Python for backend development and model development, leveraging libraries for ML and NLP, and libraries like Flask or Django for creating web applications.

React for the front end to create a dynamic and responsive user interface, considering their extensive community support and component-based architecture that facilitates scalability.

Database Technology

PostgreSQL, as the primary DBMS, was selected for its reliability, feature richness, and support for complex queries and full-text search, which is essential for the processing and storing conversation data.

Frontend Technologies

Use HTML, CSS, and JavaScript to ensure a modern and accessible web application.

Backend Technologies

Server-side logic will be implemented in Python with Flask or Django as the web framework for its simplicity and flexibility in building RESTful APIs.

Development Tools

Git for version control.

PyCharm or Jupiter notebook for model development and backend.

Visual Studio Code for UI development.

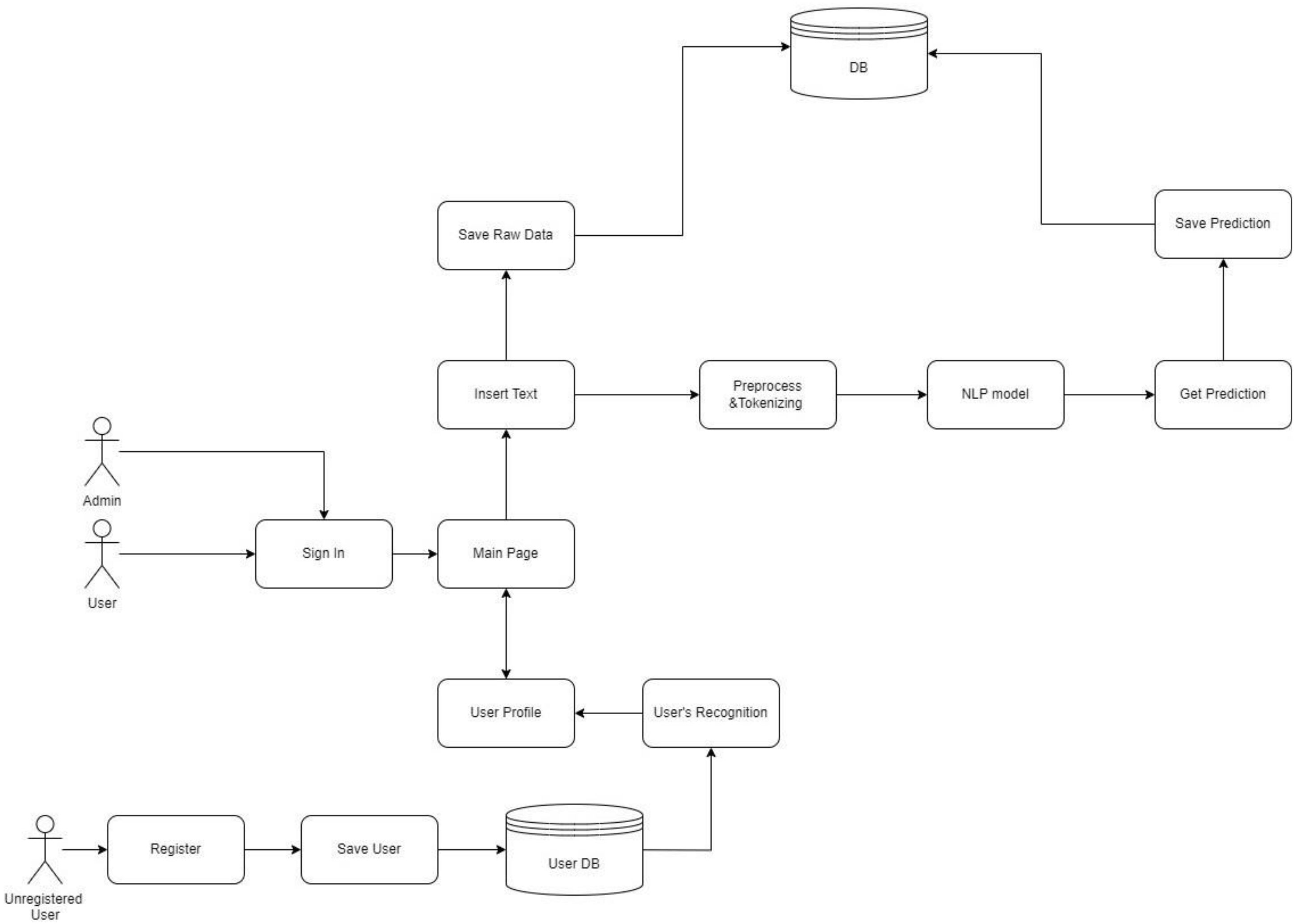
Compliance and Standards

Adhere to RESTful API standards and use JSON for data interchange between the front and back end.

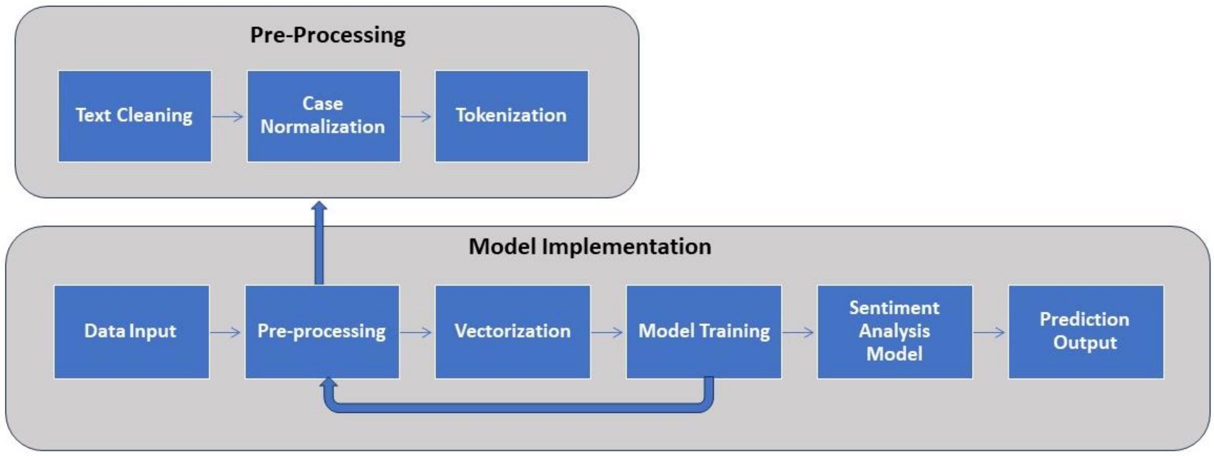
Documentation Standards

Maintain clear code documentation and in-line comments adhering to best practices.

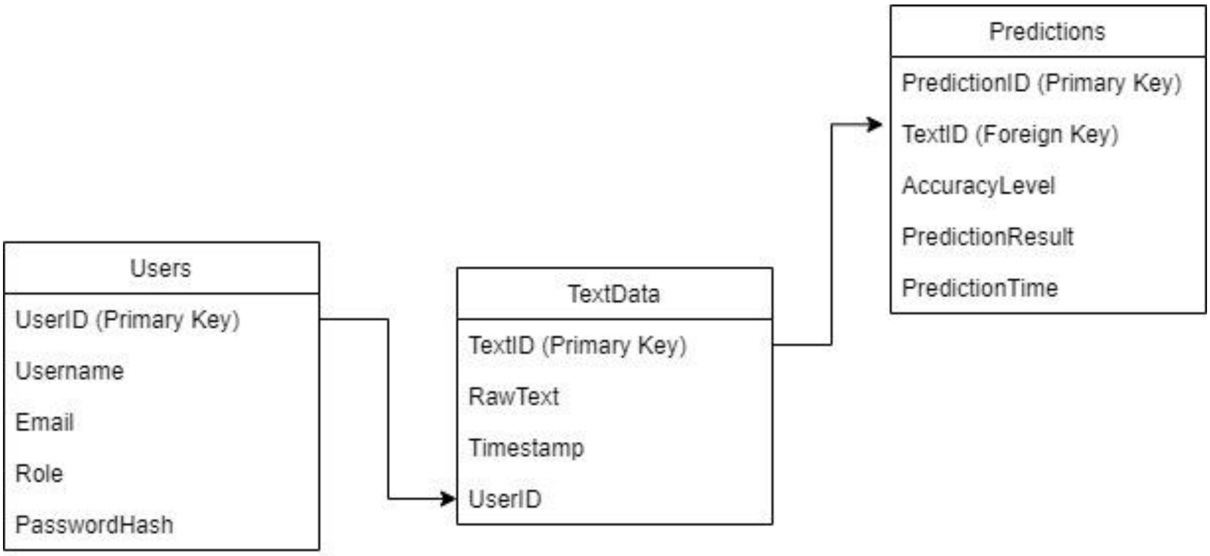
System Flow



NLP Model Flow

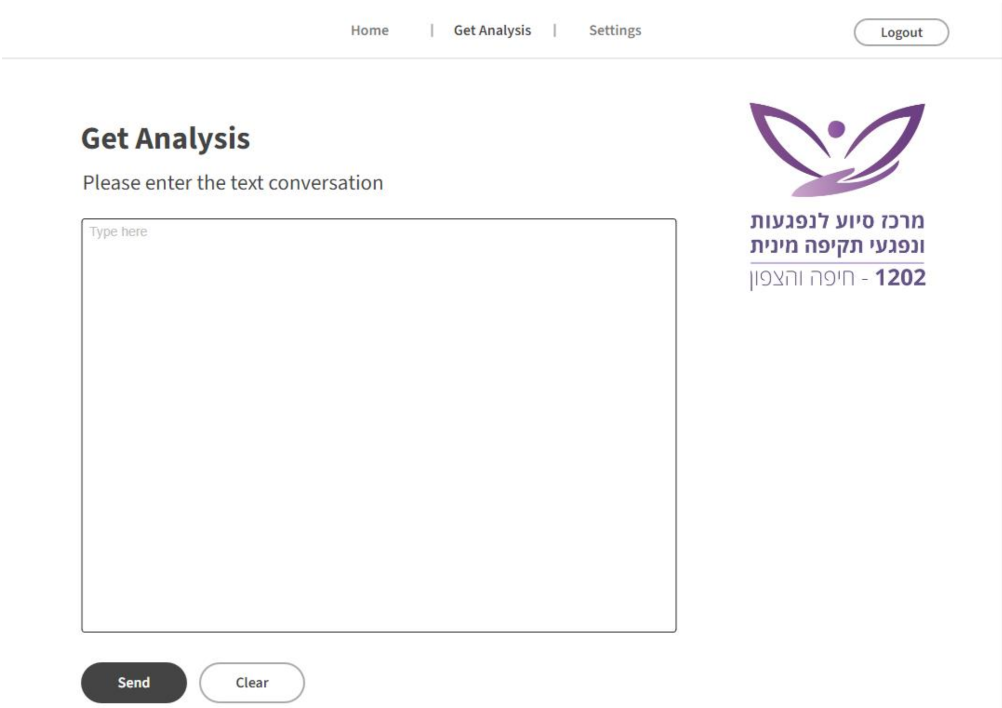


DB Scheme

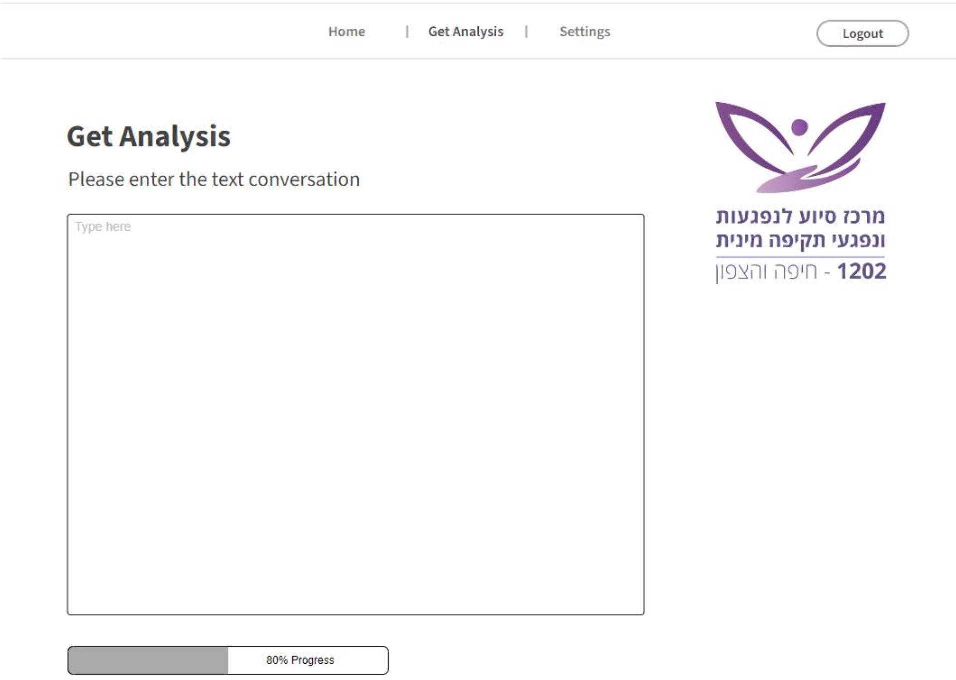


UI Mockups

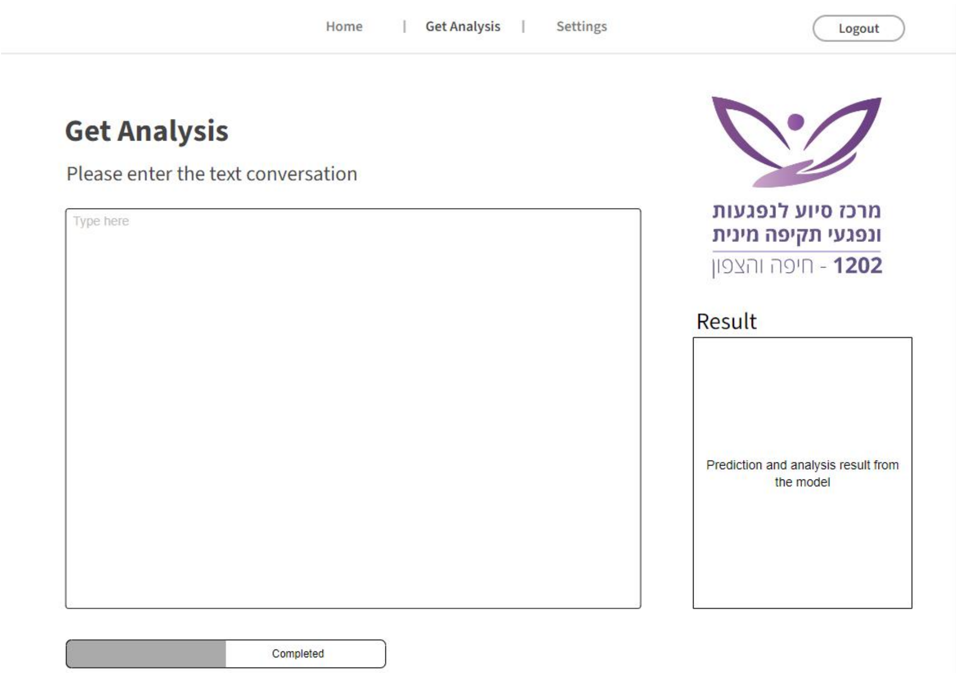
The user enters a text conversation.



The user waits for the model.



The user gets prediction results from the model.



Literature review

Background

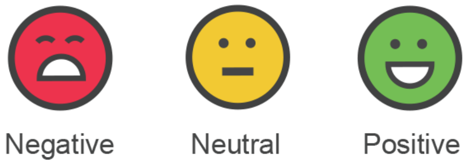
In natural language processing and machine learning, sentiment analysis has become a powerful tool for understanding human emotions through textual data.
This literature review focuses on applying sentiment analysis to text conversations originating from sexual assault message centers.
Analyzing the sentiments expressed in these conversations becomes crucial in a society where survivors often turn to digital platforms to seek help and share their experiences.
This review aims to explore existing research methodologies, challenges, and ethical considerations within this context, shedding light on the potential of sentiment analysis in enhancing support systems for survivors of sexual assault.

Sentiment analysis

Sentiment analysis, also known as opinion mining, is a branch of natural language processing (NLP) that focuses on the computational analysis of text data to determine the sentiment or emotional tone expressed within the text. The primary objective of sentiment analysis is to automatically classify a given text, such as a review, tweet, or news article, into predefined categories representing different sentiments or emotions, typically positive, negative, or neutral.

The sentiment analysis process is usually composed of this:

1. Text Input: Sentiment analysis begins with inputting a piece of text, ranging from short sentences to longer documents.
2. Text Preprocessing: The text often undergoes preprocessing steps to clean and prepare the data before analyzing sentiment. This may include tasks like removing special characters, tokenization (splitting text into words or phrases), and stemming or lemmatization (reducing words to their base form).
3. Sentiment Classification: The core of sentiment analysis involves classifying the sentiment of the text into one or more predefined categories. The most common categories are:
    a. Positive: Indicates a favorable sentiment.
    b. Negative: Signifies an unfavorable sentiment.
    c. Neutral: Represents a lack of solid sentiment, often used when the text does not express a clear emotional tone.



First Paper: #MeTooMaastricht: Building a chatbot to assist survivors of sexual harassment (Baue, et al., 2019)

Introduction

The paper discusses the development of a chatbot designed to assist survivors of sexual harassment by identifying the type of harassment, extracting spatio-temporal information, and facilitating dialogue. The initiative, inspired by the #MeToo movement, aims to support survivors and enhance the documentation of incidents.

Methodology

The paper's methodology involves leveraging machine learning techniques to build a chatbot capable of processing reports of sexual harassment. This involves two main tasks: text classification and named entity recognition. For text classification, the chatbot is trained to categorize the input into different types of harassment using data from SafeCity. Named entity recognition is employed to extract specific details from the reports, such as the location and time of the incident. The chatbot is developed on the Telegram platform, allowing interactive communication with users to gather detailed and structured reports of harassment incidents.

Results & Discussions

The chatbot achieved over 98% accuracy in identifying harassment cases and specified types of harassment with around 80% accuracy. Spatio-temporal details were extracted with high precision, demonstrating the chatbot's potential as a supportive tool for survivors.

Research Gap or Area of Improvement

Further improvement is suggested in chatbot interaction to enhance user experience and data collection accuracy. Collaboration with social scientists could refine the dialogue flow based on focus group feedback.

Strengths and Weaknesses of the paper

The paper successfully demonstrates the application of machine learning in supporting sexual harassment survivors. However, the chatbot's dialogue flow and user interaction could be enhanced for better engagement and effectiveness.

Second Paper: Detecting the Presence of Mental Illness Using NLP Sentiment Analysis (Yadav, 2022)

Introduction

This study explores the potential of Natural Language Processing (NLP) and sentiment analysis to detect mental health issues from social media data. The premise is that digital texts, such as tweets, can reveal insights into a user's mental health status by analyzing the emotional tone of their posts.

Methodology

The methodology in the paper involves using Natural Language Processing (NLP) techniques and sentiment analysis to analyze social media data for signs of mental illness. It employs a hybrid approach that combines rule-based methods with machine learning algorithms to classify the sentiment of texts as positive, negative, or neutral. This classification helps in identifying potential mental health issues based on the emotional tone of online posts. The study emphasizes the use of tweets as a primary data source for detecting early signs of depression or other mental health conditions.

Results & Discussions

The research highlights the feasibility of using sentiment analysis for early detection of mental illness through social media posts. By analyzing tweets, the system can identify users showing signs of depression, potentially aiding in early intervention and support.

Research Gap or Area of Improvement

The paper suggests further refinement in the accuracy of sentiment analysis techniques and emphasizes the need for privacy-preserving methods in analyzing sensitive personal data.

Strengths and Weaknesses

A significant strength is the innovative use of NLP for mental health monitoring. However, relying on social media data may introduce biases and privacy concerns that need addressing.

Third Paper: Supervised Machine Learning Models for Depression Sentiment Analysis (Obagbuwa, Danster, & Chibaya, 2023)

Introduction

The paper highlights the global increase in mental health issues, particularly depression, and the potential of machine learning and sentiment analysis techniques to predict depression levels in social media users early.

Methodology

The methodology section describes the datasets obtained from Twitter posts. It explores various supervised machine learning models such as Support Vector Machine (SVM), Naive Bayes, and Random Forest, comparing their performance based on accuracy, precision, recall, and F1-score. Additionally, it incorporates feature selection methods to improve model efficiency and discusses the training and testing split to validate the models' predictive capabilities on unseen data.

Results & Discussions

Results indicate that SVM and Logistic Regression models were most accurate, with Logistic Regression showing a slight edge in accuracy and significantly lower execution time. The discussion emphasizes the effectiveness of these models for early detection of depression, suggesting their practical applicability in real-world scenarios.

Research Gap or Area of Improvement

The study suggests future research to reduce computational time while improving model accuracy and to test the model on new datasets for detecting depression, indicating a gap in optimizing performance and generalizability.

Strengths and Weaknesses of the paper

The study's strengths lie in its comprehensive analysis and comparison of different machine learning models, demonstrating high accuracy and efficiency in detecting depression sentiments on social media.
However, the study might be limited by its focus on Twitter data only, which may not fully represent the broader spectrum of social media behaviors and sentiments.

Fourth Paper: LAW-U: Legal Guidance Through Artificial Intelligence Chatbot for Sexual Violence Victims and Survivors (Socatiyanurak, 2021)

Introduction

The research introduces LAW-U, an AI chatbot designed to provide legal guidance to sexual violence survivors by recommending relevant Supreme Court decisions based on the users' situations. Developed using NLP pipelines and trained with mock-up dialogs from Thai Supreme Court cases, LAW-U aims to empower victims by improving their understanding of legal rights and the judicial process.

Methodology

The methodology involved developing NLP pipelines for LAW-U, using 182 Thai Supreme Court cases related to sexual violence. These cases were processed to create a database for training the chatbot, focusing on matching user inputs to relevant legal precedents, achieving an 88.89% accuracy rate in matching Supreme Court cases to user scenarios.

Results & Discussions

The chatbot showed high accuracy in recommending relevant legal cases to survivors' situations. Its design and functionality demonstrate potential as a precedent for similar tools globally, aiming to raise awareness of sexual violence and support victims in affirming their rights.

Research Gap or Area of Improvement

While LAW-U shows promise, the study suggests further research to expand its capabilities, including emergency services for immediate victim support and continuous updates to reflect legal changes.

Strengths and Weaknesses of the paper

Strengths include the innovative use of AI in legal guidance for sexual violence survivors and high accuracy in matching legal cases. Weaknesses might involve broader testing and expansion to include more diverse scenarios and legal changes.

Fifth Paper: Combining Psychological Theory with Language Models for Suicide Risk Detection (Izmaylov, Segal, Gal, Grimland, & Levi-Belz, 2023)

Introduction

This research introduces a novel language model, SR-BERT, aimed at automatically detecting suicide risk in online chat sessions. It combines hierarchical BERT language modelling with psychological theories to enhance detection, outperforming previous non-hierarchical models in a Hebrew language setting.

Methodology

SR-BERT incorporates a base layer for conversation text encoding and an additional layer to capture conversation structure. It also integrates a domain-specific Suicide Risk Factors (SRF) lexicon developed by psychologists, aiming for improved prediction accuracy.

Results & Discussions

The SR-BERT model performed better than others, achieving a 0.76 F2 score and 0.92 ROC-AUC. Its ability to analyse conversations early on, even in a low-resource language, demonstrates its potential for real-time detection and support in suicide prevention efforts.

Research Gap or Area of Improvement

Future work may focus on incorporating more conversation aspects, like prosody and mental state dynamics, and providing explanations for predictions to support counsellors’ decision-making processes.

Strengths and Weaknesses of the paper

Strengths include the innovative combination of psychological theory with advanced language modelling and significant performance improvements. Limitations involve its evaluation solely in Hebrew and reliance on psychological lexicons requiring extensive human effort.

Market Search

Introduction

In the evolving landscape of technology and human interaction, applying data-driven insights is pivotal in shaping our understanding of various facets of society and human behaviour.
This market research embarks on a journey to explore the practical implications of applying sentiment analysis to text conversations and how it is used in similar cases to ours.

WoeBot: (Woebot)

Woebot is a mental health technology company that provides accessible and convenient support to individuals experiencing challenges with anxiety and depression. The company's approach is based on decades of Cognitive Behavioral Therapy (CBT) research and is powered by natural language processing (NLP) technology. Woebot aims to bridge the gap in mental health care by offering an engaging AI-powered platform that delivers evidence-based therapeutic approaches to users.

Key Features:

- Demand for Convenient and On-Demand Mental Health Support:
  - In recent years, there has been a growing awareness of the importance of mental health and well-being.
  - Many adults experience challenges related to anxiety and depression, and there is a strong demand for accessible and on-demand support.
- Shortcomings of Traditional Mental Health Care:
  - Traditional mental health care often faces challenges such as limited availability, long waitlists, and geographic barriers.
  - On-demand teletherapy programs are not always available 24/7, leading to patient frustration and increased provider costs.
- WoeBot is currently only available in the United States.

Wysa: (Wysa)

Wysa AI Coach is an artificial intelligence-based emotional support service that utilizes evidence-based cognitive-behavioral techniques (CBT), Dialectical Behavioral Therapy (DBT), meditation, breathing exercises, yoga, motivational interviewing, and micro-actions to help individuals build mental resilience skills and improve their emotional well-being. It offers a unique combination of AI-driven support and access to human emotional well-being professionals.

Key Features:

- Growing Awareness of Mental Health:
  - In recent years, there has been a significant increase in awareness regarding mental health and emotional well-being.

    - Individuals are increasingly recognizing the importance of seeking support and tools to manage stress, anxiety, and depression.
- Complementary Support to Traditional Therapy:
    - Wysa is a complementary resource to traditional therapy and mental health care.
    - It recognizes that individuals may benefit from non-judgmental conversations, self-help tools, and formal therapy.
- Global User Base:
    - Wysa's global reach with users from more than 30 countries highlights the international demand for accessible mental health resources.
    - The platform's broad appeal indicates a diverse user base seeking emotional support and resilience-building techniques.

rAInbow: (rAInbow)

rAInbow is a social enterprise that has built an intelligent, ethical, and scalable solution to tackle the lack of support and the loneliness faced by domestic violence survivors.
rAInbow launched in January 2018 with a chatbot that can deliver tailored conversations to women facing domestic abuse in South Africa. Survivors and those at risk can access the rAInbow chatbot through Facebook Messenger to converse with Bo. Bo is a friendly and empathetic bot that can provide training on the signs of abuse and resources available to those in need via a conversational exchange online.

Key Features:

- Supportive Platform:
    - Rainbow Chatbot offers a confidential and safe environment for individuals to talk about domestic abuse and seek guidance.
- Accessibility:
    - Easily accessible through Facebook Messenger, it requires no additional downloads, allowing for straightforward access to support.
- Information and Education:
    - The chatbot addresses common concerns related to unhealthy relationships, including signs of control and financial abuse. It educates victims and their friends, family, and colleagues on how to offer support.
- Potential for Governance Reform:
    - By collecting anonymized data, Rainbow Chatbot can provide insights that may inform policymaking and improve support mechanisms for victims.

Bibliography

Baue, T., E. D., Glazunov, M., Jaramillo, W. L., Mohan, B., & Spanakis, G. (2019). #MeTooMaastricht: Building a chatbot to assist survivors of sexual harassment. *Machine Learning and Knowledge Discovery in Databases: International Workshops of ECML PKDD 2019*.

Izmaylov, D., Segal, A., Gal, K., Grimland, M., & Levi-Belz, Y. (2023). Combining Psychological Theory with Language Models for Suicide Risk Detection. *Findings of the Association for Computational Linguistics: EACL 2023*.

Obagbuwa, I. C., Danster, S., & Chibaya, O. C. (2023). Supervised machine learning models for depression sentiment analysis. *Frontiers in Artificial Intelligence*.

rAInbow. (n.d.). *Home*. From https://hirainbow.org/.

Socatiyanurak, V. K. (2021, 12 2). Law-u: Legal guidance through artificial intelligence chatbot for sexual violence victims and survivors. *IEEE Access*. From Scorebuddy:

Woebot. (n.d.). *Home*. From Woebot: https://woebothealth.com/

Wysa. (n.d.). *Home.* From wysa: https://www.wysa.com/

Yadav, S. (2022). Detecting Presence of Mental Illness Using NLP Sentiment Analysis. *International Research Journal of Modernization in Engineering Technology and Science*.