AlertSphere: Proactive Disaster Tweet Detector

PROJECT SYNOPSIS
OF MAJOR PROJECT

BACHELOR OF TECHNOLOGY
CSE B

SUBMITTED BY

AL SAIM SHAKEEL (1800100635)
MOHD FARAZ KHAN (1800102166)
September 2023

INTEGRAL UNIVERSITY LUCKNOW

Table of Contents

Introduction

Welcome to AlertSphere: Proactive Disaster Tweet Detector! In an era dominated by social media, timely identification of disaster-related information is crucial for effective response and mitigation. This project aims to harness the power of machine learning to predict whether a given tweet is associated with a disaster or not.

AlertSphere is designed to be your go-to tool for real-time analysis of tweets, offering a proactive approach to disaster detection. By leveraging advanced natural language processing techniques, this system sifts through the vast ocean of tweets, providing valuable insights to aid emergency responders, researchers, and the general public.

Stay tuned as we delve into the realm of sentiment analysis, unraveling the narrative within tweets to enhance our understanding of disaster-related content. AlertSphere is not just a tool; it's a proactive sentinel, keeping you informed in the face of potential crises.

Rationale

In an age where social media serves as a primary source of information dissemination, the need for an advanced disaster tweet detection system becomes paramount. AlertSphere addresses this imperative by combining the power of machine learning and natural language processing.

The rationale behind this project lies in the inherent challenges posed by the sheer volume of tweets generated during disasters. Traditional methods often struggle to sift through this data efficiently. AlertSphere seeks to fill this gap, providing a proactive solution for early disaster detection. By analyzing the sentiment embedded in tweets, the system not only identifies potential disasters but also aids in gauging the urgency and severity of the situation.

This project is driven by the vision of creating a tool that empowers emergency responders with timely and accurate information, enabling swift and effective response strategies. AlertSphere is not just a technological innovation; it is a strategic asset, poised to make a tangible impact on disaster management in the digital age.

Objectives

1. Accurate Classification: Develop a robust machine learning model capable of accurately classifying tweets into "disaster" or "non-disaster" categories, ensuring a high level of precision.

2. Real-time Analysis: Implement a system capable of real-time analysis, allowing for prompt detection and response to disaster-related tweets as they emerge on social media platforms.

3. Adaptability to Emerging Trends: Incorporate mechanisms to adapt to evolving language patterns and emerging disaster-related keywords, enhancing the system's ability to detect novel crisis situations.

4. User-Friendly Interface: Develop an intuitive and user-friendly interface for easy accessibility, allowing a wide range of users, including emergency responders and researchers, to leverage the system effectively.

5. Performance Metrics: Establish comprehensive performance metrics, including accuracy, precision, recall, and F1 score, to continually evaluate and improve the system's predictive capabilities.

6. Community Impact: Strive to make a positive impact on disaster management by creating a tool that enhances the efficiency of response efforts and contributes to community safety.

Feasibility Study

1. Technical Feasibility:
   - Assess availability of relevant ML algorithms and frameworks.
   - Evaluate real-time processing feasibility for prompt tweet analysis.

2. Operational Feasibility:
   - Gauge user acceptance among emergency responders and researchers.
   - Explore integration feasibility with existing disaster management systems.

3. Economic Feasibility:
   - Conduct cost-benefit analysis for development and maintenance.
   - Assess potential economic benefits against operational expenses.

4. Scheduling Feasibility:
   - Evaluate project timeline for timely completion.
   - Consider phases, testing, and deployment in the schedule.

5. Resource Feasibility:
   - Assess team's skill set and expertise.
   - Confirm availability of required hardware and software resources.

6. Risk Analysis:
   - Identify potential risks such as data security threats.
   - Develop contingency plans for unforeseen challenges.

Methodology / Planning of Work

Steps to consider:

1. Read the dataset

2. Remove handle null values (if any).

3. Preprocess the disaster tweets data based on the following parameter:
    a) Tokenizing words
    b) Convert words to lower case
    c) Removing Punctuations
    d) Removing Stop words
    e) Stemming or lemmatizing the words

4. Transform the words into vectors using
    a) Count Vectorizer
    or
    b) TF-IDF Vectorizer

5. Select x(independent feature) as tweets after preprocessing and target as y(dependent feature).

6. Split data into training and test data.

7. Apply the following models on the training dataset and generate the predicted value for the test dataset
    a) Multinomial Naïve Bayes Classification
    b) Logistic Regression
    c) KNN Classification

8. Predict the target for test data

9. Compute Confusion matrix and classification report for each of these models

10. Report the model with the best accuracy.

Facilities required for proposed work

The successful execution of "AlertSphere: Proactive Disaster Tweet Detector" necessitates access to key facilities, including high-performance computing and ample storage for model training, datasets, and checkpoints. Essential software tools encompass machine learning frameworks, NLP libraries, and integrated development environments for efficient coding and debugging. A reliable internet connection is crucial for real-time data collection, while collaboration platforms ensure effective communication among team members. User interface development tools, privacy measures, and documentation software contribute to the creation of an intuitive and secure system. Testing tools are vital for evaluating functionality and performance, and a suitable deployment platform, whether on the cloud or on-premises, is required for the finalized system. Continuous improvement relies on mechanisms for collecting user feedback, and opportunities for team training enhance proficiency in cutting-edge ML and NLP techniques.

Expected Outcomes

The anticipated outcome of "AlertSphere: Proactive Disaster Tweet Detector" is a highly accurate and efficient system for classifying tweets as either disaster-related or not. The machine learning model, trained on diverse datasets, is expected to demonstrate robust performance in real-time tweet analysis, providing timely and precise information for emergency responders and researchers. The user-friendly interface will facilitate seamless interaction, making the tool accessible to a wide audience. Privacy and ethical considerations will be diligently upheld, ensuring responsible data handling. The integration of AlertSphere into existing disaster management systems is expected to enhance overall response capabilities. Continuous improvement mechanisms, informed by user feedback, will contribute to the long-term effectiveness and adaptability of the system, ultimately making a positive impact on proactive disaster management.

Supervisor's Remarks:

Approved **Not Approved**

Remarks or Suggestions by Supervisor:

……
……
……
……
……

Signature:

Name:

Date: