**<CHATYLITICS>**

**Submitted for**

**Statistical Machine Learning CSET211**

Submitted by:

**(<E23CSEU1438>) Harshita Nagpal**

**(<E23CSEU1419>) Navya Singh**

Submitted to

**Dr. Susmita Das**

**July-Dec 2024**

**SCHOOL OF COMPUTER SCIENCE AND ENGINEERING**

A close-up of a logo

Description automatically generated

1. **ABSTRACT**

**A chat analytics project that utilizes statistical machine learning and natural language processing techniques to analyze chat-based data, CHATYLITICS is designed to transform raw chat logs into actionable insights through user activity visualization, communication trend identification, and sentiment analysis. Main functionalities include word cloud generation, emoji analysis, activity heatmaps, and identification of most active participants in a conversation. CHATYLITICS supports trend analysis via timelines, allowing users to browse chat behavior over different periods. The project has thus availed a simple yet powerful dashboard for users to upload their chat files and result in visualizing interactively, making this an invaluable tool for personal, organizational, or research use.**

1. **Introduction**

**Today's digital world communicates significantly through messaging, thereby producing enormous volumes of unstructured data. This kind of data is rich in value but tends to lie fallow as it is too complex and does not have enough tools for meaningful analysis. CHATYLITICS addresses this challenge with its platform for effectively analyzing and interpreting chat data.**

**This project aims to extract insights from chat logs by way of text processing, finding the patterns among users, and visualizing trends. The approach is far beyond basic statistical measures as advanced NLP techniques are used for features such as sentiment analysis and word filtering.**

**CHATYLITICS seeks to guide the masses to embrace the power of conversational data, particularly in customer support, social media analytics, and team productivity monitoring, into better decision-making. Through an intuitive dashboard, results are made accessible to both technical and non-technical users.**

1. **Methodology**

**The project follows this workflow:**

* **Data Preprocessing: The raw chat data is parsed and preprocessed using Python libraries. This includes extracting all the information like users, messages, timestamps, and metadata.**
* **Statistical Analysis:**
* **Total messages, word count, and media shared are computed.**
* **User activity patterns are extracted.**
* **Visualizations:**
* **Word Cloud: The graphical word representation of frequently used words, excluding stop words.**
* **Emoji Analysis: The most frequently used emojis along with their frequency.**
* **Heatmaps: Trends in weekly and daily activities visualized to better understand them.**
* **Machine Learning Models Sentiment analysis and NLP techniques were applied for classification and interpretation of chat content**
* **Dashboard Creation: Interactive user interface based on Streamlit is displayed where users upload their chat files. Dynamically generate visualized results.**

1. **Hardware/Software Required**

* **Hardware:**
* **Computer with at least 8 GB RAM and 4-core processor.**
* **GPU-enabled system for faster model training (optional).**
* **Software:**
* **Python 3.x**
* **Libraries: Pandas, NumPy, Matplotlib, Seaborn, Word Cloud, Streamlit, re, and emoji.**

1. **Experimental Results**

* **Statistical Summary:**
* **Total messages, word count, media shared, and links are accurately calculated.**
* **Word Cloud: Displays the most commonly used words after filtering stop words.**
* **Emoji Analysis: Shows the frequency and proportion of emojis used in conversations.**
* **Activity Heatmaps: Visualizes user activity by day and time for detailed trends.**
* **Timeline Analysis: Graphs displaying activity trends over time (monthly and daily).**

1. **Conclusions**

**CHATYLITICS successfully demonstrates the potential of statistical machine learning in analyzing and visualizing conversational data. The platform effectively identifies trends, user activity, and sentiment, providing actionable insights. It has practical applications in customer service, e-commerce, and team productivity analysis.**

1. **Future Scope**

* **Advanced NLP Techniques: Sentiment Polarity and Intent Classification to Deep Analysis.**
* **Real-Time Analysis: Enabling real-time chat monitoring and visualization.**
* **Integration with Messaging Platforms: Support extending for direct integration with WhatsApp, Slack, and all other types of chat platforms.**
* **Scalability: Developed to scale with large-sized data as well as multi-user support.**

1. **GitHub Link of Your Complete Project**

https://github.com/navyaaaasingh/chatylitics