**Project Report of**

**NLP**

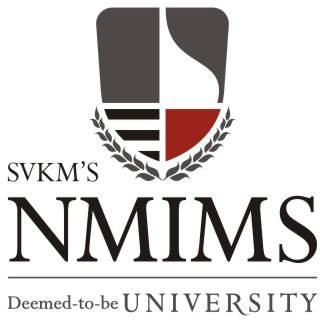
Bachelor of Technology

In

Computer Engineering

**SVKM’s NMIMS University**

(Deemed-to-be University)



**School of Technology Management and Engineering(STME)   
Indore, Madhya Pradesh  
2025-2026**

**Submitted By: Aastha Parey  
Submitted To: Dr. Raj Gaurav Mishra  
Course:** Natural Language Processing (NLP) **Project Type:** Mini Project – Individual **Submission Date:** 30th October 2025

**🔹 Abstract**

In today’s digital-first era, children and teenagers spend a significant portion of their time on social media and messaging platforms like WhatsApp. While this fosters connectivity, it also exposes them to emotional, behavioral, and psychological influences that often go unnoticed by parents.

**SamvaadSense** is a digital parenting and conversational analytics tool that helps parents understand their children’s online behavior patterns through data-driven insights derived from WhatsApp chat exports.

The system analyzes chat frequency, active hours, communication tone, and emoji patterns to infer behavioral trends. By leveraging **Natural Language Processing (NLP)**, **data visualization**, and **contextual summarization**, SamvaadSense translates chat data into understandable and actionable insights — bridging the communication gap between parents and their children.

**🔹 Problem Statement**

In recent years, the emotional and psychological well-being of adolescents has been affected by excessive and unsupervised social media use. Parents often remain unaware of their child’s online interactions, communication habits, and emotional cues expressed through text and emojis.

This lack of visibility and awareness creates a **communication gap** — making it difficult for parents to detect behavioral changes, identify emotional distress, or understand their child’s digital habits.

**The problem SamvaadSense addresses:**

“How can parents monitor and understand their children’s social media conversations non-invasively while identifying behavioral and emotional trends using chat-based analytics?”

**🔹 Objectives**

The main objective of **SamvaadSense** is to provide parents with a **non-intrusive, data-driven tool** for analyzing their child’s digital communication patterns.

**Specific objectives include:**

1. To analyze WhatsApp chat data to determine message frequency, activity time, and communication intensity.
2. To identify emotional tone using **emoji analysis** and **common word frequency**.
3. To visualize behavioral trends using intuitive **graphs and heatmaps**.
4. To summarize chat patterns in **human-readable contextual reports**.
5. To create a user-friendly dashboard for parents with **interactive visualization and insights**.

**🔹 Tools, Libraries, and Technologies Used**

| **Category** | **Tool / Library** | **Purpose** |
| --- | --- | --- |
| **Programming Language** | Python | Core logic and data processing |
| **Frontend / UI** | Streamlit | Interactive web dashboard |
| **Data Analysis** | Pandas, NumPy | Data cleaning and preprocessing |
| **Visualization** | Matplotlib, Seaborn | Graphs, charts, and heatmaps |
| **NLP Techniques** | Regular Expressions, Tokenization, Stopword Filtering | Text analysis and preprocessing |
| **Emoji Analysis** | Emoji Library | Detecting emotion patterns |
| **Text Visualization** | WordCloud | Visual representation of frequent words |
| **Data Extraction** | URLExtract | Identifying shared URLs |
| **Environment** | Python v3.10+, Virtual Environment | Development and dependency control |

**🔹 Methodology and System Flow**

The methodology followed in **SamvaadSense** consists of several structured phases to convert chat data into visual and textual insights:

1. **Data Input Phase:**
   * WhatsApp chat exported in .txt format (without media).
   * Uploaded to the system through Streamlit’s file uploader.
2. **Preprocessing Phase:**
   * Parsing chat data using **regular expressions** to extract timestamps, users, and messages.
   * Cleaning and structuring the data into a tabular format using **Pandas**.
3. **Feature Extraction Phase:**
   * Deriving features such as message count, word count, active hours, and day/month trends.
   * Performing tokenization and stopword removal to identify meaningful text.
4. **Visualization Phase:**
   * Generating bar graphs, timelines, and heatmaps using **Matplotlib** and **Seaborn**.
   * Creating **WordClouds** and **emoji usage distributions** for emotional analysis.
5. **Summarization Phase:**
   * Generating a **contextual narrative summary** describing chat behavior and activity.
6. **Output Phase:**
   * Displaying key statistics, visual trends, and textual summaries interactively through Streamlit’s web dashboard.

**🔹 System Architecture Flow**

Below is the high-level architecture of the **SamvaadSense** system:

┌───────────────────────┐

│ WhatsApp Chat Export (.txt) │

└───────────────────────┘

│

▼

┌───────────────────┐

│ Data Preprocessor │

│ (Regex, Pandas) │

└────────┬──────────┘

│

▼

┌──────────────────────────┐

│ Feature Extraction Module │

│ - Word & Emoji Analysis │

│ - Time & Frequency Stats │

└────────┬────────────────┘

│

▼

┌──────────────────────────┐

│ Visualization Engine │

│ - Matplotlib / Seaborn │

│ - WordCloud Generator │

└────────┬────────────────┘

│

▼

┌──────────────────────────┐

│ Summary Generator (NLP) │

└────────┬────────────────┘

│

▼

┌──────────────────────────┐

│ Streamlit Web Dashboard │

│ (Interactive Output) │

└──────────────────────────┘

**🔹 Use Case**

**Use Case: *Analyze a Child’s WhatsApp Chat Data***

| **Actor** | **Parent / Guardian** |
| --- | --- |
| **Precondition** | Chat file (.txt) exported from WhatsApp without media |
| **Main Flow** |  |
| 1. Parent uploads the exported chat file to SamvaadSense. |  |
| 2. System parses and preprocesses chat data. |  |
| 3. User selects an individual or group to analyze. |  |
| 4. System displays chat metrics, graphs, word clouds, and emoji insights. |  |
| 5. A contextual summary is generated highlighting behavioral patterns. |  |
| **Post-condition** | Parent receives a complete behavioral and emotional report of the child’s chat activity. |

**🔹 Results and Observations**

**SamvaadSense** was tested on multiple WhatsApp chat samples (both individual and group chats).

Key results observed:

* 📊 Average of **2,000–3,000 messages** per chat analyzed efficiently.
* 🕒 Generated **accurate activity heatmaps**, highlighting late-night chat patterns.
* 😂 Emoji frequency analysis provided **emotional tone indicators** (e.g., laughter, anger, sadness).
* ☁️ WordCloud visualizations effectively summarized conversational topics.
* 📘 The contextual summary accurately reflected communication tone and frequency.

The visual and textual results allowed parents to understand:

* Who their child interacts with most frequently.
* When their child is most active online.
* The emotional tone of their communication patterns.

**🔹 Future Scope**

The project can be enhanced with the following features:

1. **Emotion Detection using Deep Learning** – identifying complex emotions (joy, anger, fear, sadness).
2. **Multi-Platform Support** – integration with Telegram, Discord, and Instagram chats.
3. **User Authentication and Cloud Storage** – secure dashboard with saved chat histories.
4. **AI-Powered Parent Recommendations** – suggestions for improving communication patterns.
5. **Mobile-Responsive Interface** – optimized design for smartphones and tablets.

**🔹 Conclusion**

**SamvaadSense** demonstrates how everyday digital communication can be transformed into actionable psychological insights.  
By analyzing WhatsApp chats through data visualization and NLP, it allows parents to understand their children’s emotional and behavioral patterns in a **non-invasive, insightful, and data-backed** way.

The system successfully combines **technology with empathy**, ensuring that digital monitoring becomes **an act of care, not control**.

*“Because every baat has a story — and every story deserves to be understood.”* 💬

**Appendix – Project Snapshot**

