Dataset Name : Suicidal Tweet Detection Dataset

Dataset Description:

This dataset contains tweets labeled as either expressing suicidal sentiments or not. It is designed for training machine learning models to classify tweets into two categories: "Not Suicide post" and "Potential Suicide post."

Usage: Ideal for training models to detect potentially harmful content on social media, enabling early intervention and mental health support.

Potential Applications:

- Suicidal ideation detection
- Mental health support tools
- Sentiment analysis research
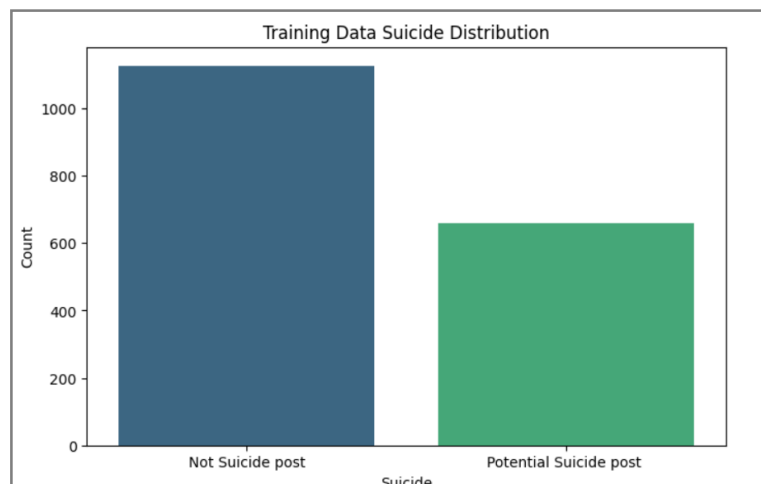- Public health awareness initiatives

Dataset structure:

The dataset consists of a CSV files:

1. **Suicide_Dataset.csv**– Used for model training

The dataset contain two columns:

1. **Tweet**: The text content of the tweet.
2. **Suicide**: The label indicating if the tweet expresses suicidal thoughts ("Potential Suicide post") or not ("Not Suicide post").

Statistics:

Model Selection:

1.Recurrent Neural Network (RNN):

RNNs are designed for sequential data, making them suitable for tasks like sentiment analysis. They feature a memory mechanism that allows information to persist across time steps. While RNNs are effective for short-term dependencies in text, they face challenges such as the vanishing gradient problem, which limits their ability to learn long-range dependencies in sequences.

2.Long Short-Term Memory (LSTM):

LSTMs are an advanced variant of RNNs that address the vanishing gradient problem by using gates (input, forget, and output gates) to manage information flow over long sequences. They are highly effective in NLP tasks, including sentiment analysis, as they capture long-range dependencies and contextual information, making them well-suited for detecting nuanced emotions in text, such as suicidal ideation.
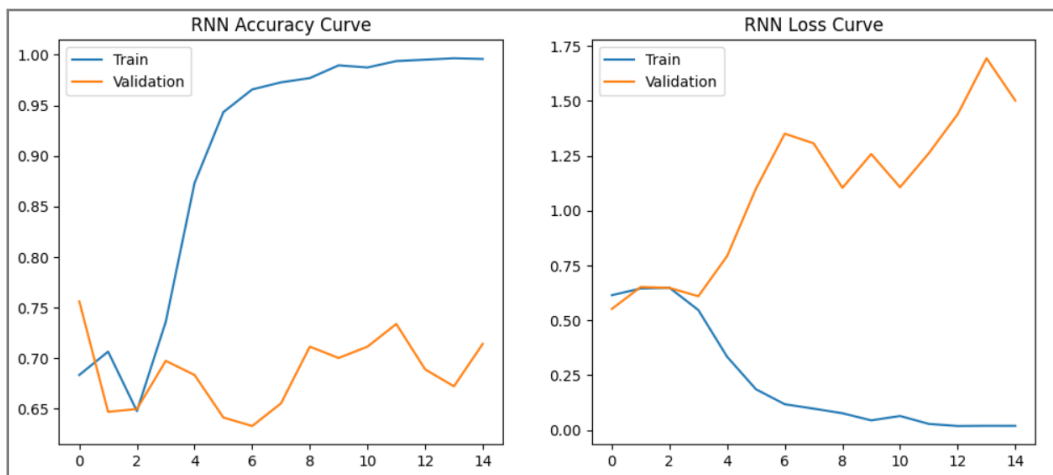
Result :

1. RNN:

First Experiment :

Test Accuracy :
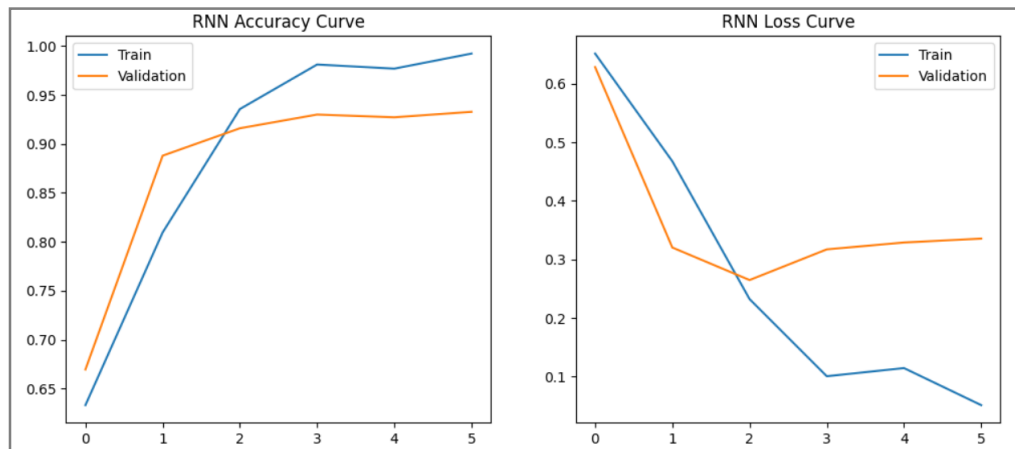
RNN Model Accuracy: 71.43%

Accuracy & Loss Curve:

Second Experiment :

Test Accuracy :

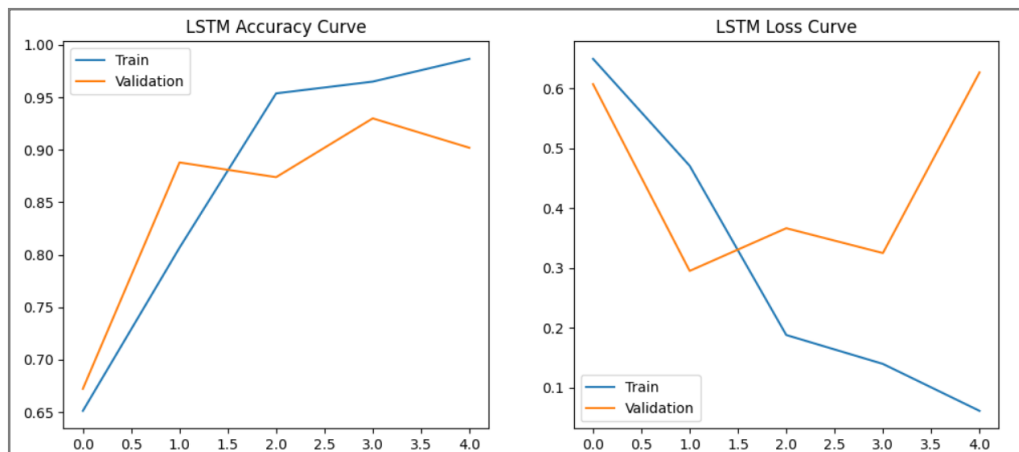RNN Model Accuracy: 91.60%

Accuracy & Loss Curve:

2.LSTM:

First Experiment :

Test Accuracy :

LSTM Model Accuracy: 88.80%
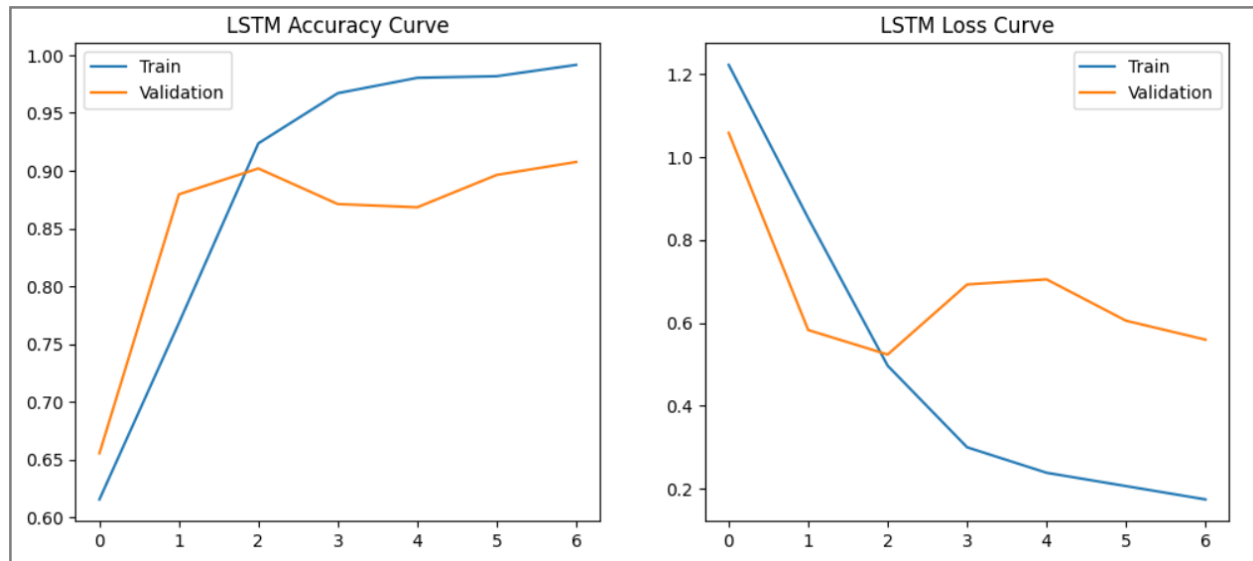
Accuracy & Loss Curve:

Second Experiment :

Test Accuracy :

LSTM Model Accuracy: 90.20%

Accuracy & Loss Curve:

Discussion:

In suicidal tweet detection, the choice between Recurrent Neural Networks (RNNs) and Long Short-Term Memory networks (LSTMs) is essential. RNNs are effective for sequential data but struggle with the vanishing gradient problem, limiting their ability to capture long-term dependencies in text. This can hinder the detection of subtle signs of distress spread across longer sequences.

LSTMs, on the other hand, overcome this limitation by using gates to retain and manage long-range dependencies. This makes them more capable of understanding the emotional nuances in tweets, which is crucial for identifying suicidal ideation. While LSTMs are more computationally demanding, their ability to capture long-term context makes them a better choice for this task.

In summary, LSTMs are more suited for suicidal tweet detection due to their superior handling of long-term dependencies, making them more effective at identifying emotional cues in text.