Sentiment Analysis Using RNN, LSTM, and GRU on Twitter Data

1. Introduction

This report presents a comparative analysis of sentiment classification models using recurrent neural networks (RNN, LSTM, and GRU) on a Twitter sentiment analysis dataset. Additionally, we evaluate a lexicon-based approach (VADER) for sentiment analysis and discuss its advantages and limitations compared to deep learning models.

2. Methodology

2.1 Dataset and Preprocessing

- The dataset consists of labeled tweets categorized into positive, negative, and neutral sentiments.
- Preprocessing steps include:
    - **Text Cleaning:** Removing URLs, mentions, hashtags, emojis, and special characters.
    - **Tokenization:** Converting tweets into sequences of words.
    - **Sequence Padding:** Ensuring uniform input length.

2.2 Model Implementation

Three neural architectures were implemented:

1. **RNN:** A simple recurrent model with an embedding layer and a vanilla RNN cell.
2. **LSTM:** A long short-term memory network designed to handle long-range dependencies.
3. **GRU:** A gated recurrent unit model, offering a balance between RNN and LSTM in terms of complexity and efficiency.

Training Details:

- Optimizer: Adam
- Loss Function: Categorical Cross-Entropy
- Evaluation Metrics: Accuracy, Precision, Recall, F1-score
- Early Stopping and Learning Rate Scheduling were used to optimize training.

3. Results and Evaluation

3.1 Performance Metrics

Model	Test Accuracy	Precision	Recall	F1-score
RNN	**0.67**	**0.68**	**0.67**	**0.67**
LSTM	**0.70**	**0.71**	**0.70**	**0.70**
GRU	**0.70**	**0.73**	**0.70**	**0.70**

- **LSTM and GRU achieved the highest accuracy**, indicating strong performance on long sequences.
- **GRU performed comparably to LSTM** but trained faster.
- **RNN struggled** due to vanishing gradient issues and weaker memory retention.

4. Critical Analysis: VADER vs. RNN/LSTM/GRU

Feature	VADER (Lexicon-Based)	RNN	LSTM	GRU
Interpretability	High (clear word scores)	Low	Low	Low
Computational Cost	Low	High	Very High	High
Handling Context & Sarcasm	Poor	Moderate	Strong	Strong
Need for Labeled Data	No	Yes	Yes	Yes

5. Discussion & Reflection

- **LSTM struggled with short texts (e.g., tweets)** due to its need for sequential dependencies. Short texts lack enough context for it to leverage its memory advantage.
- **VADER is preferable when labeled data is unavailable** and when speed and interpretability are prioritized.
- **Neural networks (LSTM/GRU) are superior for complex language tasks** where deep contextual understanding is necessary.

6. Conclusion

- **LSTM is best suited for longer-sequence sentiment classification** but requires substantial training time.
- **GRU provides a good tradeoff between efficiency and performance.**

- **VADER remains a strong choice for lexicon-based, lightweight sentiment analysis.**

This study highlights the trade-offs between deep learning and rule-based models, guiding the choice of approach based on dataset characteristics and computational constraints.