**Tribhuvan University**

**Institute of Science and Technology**

**Central Department of Computer Science and Information Technology**

**Kirtipur, Kathmandu**

**2024**

**Seminar Report on**

**“Sentiment Analysis”**

**In partial fulfillment of the requirement for a Master’s degree in Computer Science and Information Technology (M.Sc. CSIT), 1st Semester**

**Submitted to:**

Central Department of Computer Science and Information Technology, Tribhuvan University, Kirtipur, Kathmandu, Nepal

**Submitted By:**

Sagar Timalsena (Roll No)

# **ACKNOWLEDGEMENT**

I would like to express my sincere gratitude to everyone who has helped with technical and informational support and guidance.

I begin by extending my gratitude to my mentor Mr. Bikash Balami, Assistant Professor, Central Department of Computer Science and Information Technology (CD.CSIT) Tribhuvan University for his continuous guidance and support during the research and learning period. Without his guidance, this research would not have been successful all staff members for their cooperation and guidance during the entire research period. The success and outcome of this research required a lot of guidance and assistance from many people and I am extremely fortunate to have them all along the completion of this research period. Such guidance was indispensable for the accomplishment of this endeavor.

I am thankful to our project supervisor for providing me with his continuous guidance, advice, motivation, and knowledge he imparted to me regarding every perspective of the report. I also express my gratitude to the Central Department of Computer Science and Information Technology MSc.CSIT department and library for providing me with the necessary resources during the time of research and report.

I would also like to acknowledge, my parents, for providing financial and psychological support throughout my study and all those who contributed directly and indirectly.

Thanking You,

Sagar Timalsena

(TU Exam Roll No.: )

# **ABSTRACT**

This report delves into the intriguing field of sentiment analysis, a crucial component of comprehending language, to unravel the intricacies of emotional expression within a given paragraph. Utilizing sophisticated computational methodologies, this study endeavors to discern, classify, and interpret emotions such as happiness, sadness, or neutrality embedded within textual content. By harnessing the power of advanced technology, this research aims to shed light on the nuances of emotional communication in various contexts, including social media interactions, product reviews, and customer feedback. The insights gleaned from this investigation hold the potential to enrich decision-making processes and enhance user experiences across digital platforms.

**Keywords:** Natural Language Processing (NLP), Language Semantics, Text Analysis, Textual Meaning Extraction

# **Table of Contents**

[**ACKNOWLEDGEMENT** ii](#_Toc168331088)

[**ABSTRACT** iii](#_Toc168331089)

[**Table of Contents** iv](#_Toc168331090)

[**List of Figure** vi](#_Toc168331091)

[**List of Table** vii](#_Toc168331092)

[**List of Abbreviations** viii](#_Toc168331093)

[**Chapter 1: Introduction** 1](#_Toc168331094)

[**1.1** **Introduction** 1](#_Toc168331095)

[**1.2** **Problem Statement** 1](#_Toc168331096)

[**1.3** **Objective** 1](#_Toc168331097)

[**Chapter 2: Background Study and Literature Review** 2](#_Toc168331098)

[**2.1 Background Study** 2](#_Toc168331099)

[**2.1.1 Recurrent Neural Network** 2](#_Toc168331100)

[**2.1.2Gated Recurrent Unit** 2](#_Toc168331101)

[**2.2 Literature Review:** 2](#_Toc168331102)

[**Chapter 3: Methodology** 3](#_Toc168331103)

[**3.1 Data Loading** 3](#_Toc168331104)

[**3.2Text Tokenization and Sequencing** 3](#_Toc168331105)

[**3.3 Padding:** 3](#_Toc168331106)

[**3.4 Train-test split:** 3](#_Toc168331107)

[**3.5 GRU model:** 3](#_Toc168331108)

[**3.6 Performance Evaluation:** 3](#_Toc168331109)

[**Chapter 4: Implementation** 4](#_Toc168331110)

[**4.1 Implementation** 4](#_Toc168331111)

[**4.2 Implementation details** 4](#_Toc168331112)

[**4.2.1Train–Test Split** 4](#_Toc168331113)

[**4.2.2 Text Sequencing and Padding** 4](#_Toc168331114)

[**4.2.3 GRU Architecture** 4](#_Toc168331115)

[**4.2.4 Model Training and Adam Optimizer** 4](#_Toc168331116)

[**4.2.5 Model Evaluation** 4](#_Toc168331117)

[**4.2.6 Sentiment Analysis Function** 4](#_Toc168331118)

[**Chapter 5 Result and Findings** 5](#_Toc168331119)

[**5.1 Overfitting and Generalization** 5](#_Toc168331120)

[**5.2 Confusion Matrix** 5](#_Toc168331121)

[**5.3 Model Performance Measure** 5](#_Toc168331122)

[**References** 7](#_Toc168331123)

# **List of Figure**

# **List of Table**

# **List of Abbreviations**

**MSC.CSIT** Master of Science in Computer Science and Information Technology **NLP** Natural Language Processing

# **Chapter 1: Introduction**

## **Introduction**

Sentiment analysis, also referred to as opinion mining, is a computational process that involves analyzing and interpreting the emotions and opinions expressed in textual data. In today's digital age, where the amount of textual content generated online is growing exponentially, sentiment analysis has become increasingly important across various domains.

Understanding public opinion, consumer feedback, and social media trends is critical for businesses, governments, and organizations to make informed decisions and stay relevant in a highly competitive landscape. Sentiment analysis provides valuable insights into the sentiments and attitudes of individuals towards specific topics, products, services, or events.

Traditional methods of sentiment analysis often rely on simplistic approaches such as rule-based systems or bag-of-words models. While these methods can provide basic sentiment classification, they often fail to capture the nuances and context-dependent nature of language. For example, a simple approach might classify the word "good" as positive, but it might overlook subtle variations in meaning, such as "good" in the context of being sarcastic or as part of a negation ("not good"). In recent years, deep learning techniques, particularly recurrent neural networks (RNNs), have emerged as powerful tools for sentiment analysis. RNNs are a class of artificial neural networks designed to effectively process sequential data, making them well-suited for tasks involving natural language processing.

Unlike traditional methods, which require handcrafted features or predefined rules, RNNs can learn to capture complex patterns and dependencies within textual data automatically. This ability to learn from data makes RNNs highly adaptable and capable of capturing the subtle nuances of language, including sentiment.

Gated recurrent units (GRUs) are a type of RNN architecture that has gained popularity for sentiment analysis tasks due to their simplicity and effectiveness. GRUs address some of the limitations of traditional RNNs, such as the vanishing gradient problem, by introducing gating mechanisms that control the flow of information through the network.

By leveraging deep learning techniques like RNNs and GRUs, sentiment analysis models can achieve higher levels of accuracy and robustness, enabling more nuanced and context-aware sentiment analysis. This, in turn, empowers businesses, researchers, and policymakers to gain deeper insights into public opinion, consumer sentiment, and social media dynamics, ultimately facilitating more informed decision-making and strategic planning.

## **Problem Statement**

As the amount and intricacy of textual data continue to grow, accurately deciphering sentiments presents a significant challenge. Current sentiment analysis models often falter in grasping contextual cues and subtle emotional nuances, resulting in less-than-ideal outcomes. Therefore, there's a pressing demand for resilient and effective sentiment analysis frameworks adept at managing vast quantities of text data and precisely categorizing sentiments with heightened accuracy.

## **Objective**

The primary objective of this study is to develop a sentiment analysis framework using gated recurrent units (GRUs), a variant of recurrent neural networks, to effectively classify sentiments in text data. Specifically, we aim to:

* Implement a GRU-based sentiment analysis model
* Evaluate the performance of the model using appropriate metrics
* Analyze the effectiveness and limitations of the proposed framework

# **Chapter 2: Background Study and Literature Review**

## **2.1 Background Study**

### **2.1.1 Recurrent Neural Network**

Recurrent Neural Networks (RNNs) are a class of artificial neural networks designed to handle sequential data by maintaining a memory state. Unlike feedforward neural networks, which process data instantaneously, RNNs can retain information about previous inputs, making them well-suited for tasks such as language modelling, time series prediction, and speech recognition. However, traditional RNNs suffer from the vanishing gradient problem, which hampers their ability to capture long-term dependencies in sequential data.

### **2.1.2Gated Recurrent Unit**

Gated Recurrent Units (GRUs) are a type of RNN variant that address some of the limitations of traditional RNNs. Introduced by Cho et al. in 2014, GRUs incorporate gating mechanisms to regulate the flow of information within the network. This allows GRUs to selectively update their memory states, making them more effective at capturing long-range dependencies and mitigating the vanishing gradient problem. GRUs have gained popularity in various applications, including natural language processing, time series analysis, and image captioning.

## **2.2 Literature Review:**

Sentiment analysis, also known as opinion mining, is the task of automatically determining the sentiment expressed in a piece of text. This area of research has garnered significant attention due to its wide range of applications, including social media monitoring, customer feedback analysis, and market research. Researchers have employed various techniques to perform sentiment analysis, ranging from traditional machine learning algorithms to deep learning models.

Early approaches to sentiment analysis often relied on lexicon-based methods, where sentiment polarity was determined based on the presence of positive or negative words in the text. While these methods are simple and interpretable, they often struggle with nuanced expressions and context-dependent sentiments.

With the advent of deep learning, researchers began exploring neural network-based approaches for sentiment analysis. Convolutional Neural Networks (CNNs) and Recurrent Neural Networks (RNNs), including variants like Long Short-Term Memory (LSTM) and Gated Recurrent Units (GRUs), have been applied to sentiment analysis tasks with promising results. These models can automatically learn relevant features from raw text data, capturing both syntactic and semantic information.

Recent advancements in sentiment analysis have also seen the integration of techniques such as attention mechanisms, which enable models to focus on important parts of the input sequence, and transfer learning, where pre-trained models are fine-tuned on sentiment-specific tasks to improve performance on limited datasets.

In summary, sentiment analysis has evolved from lexicon-based approaches to sophisticated deep learning models, enabling more accurate and nuanced sentiment classification. However, challenges such as domain adaptation, handling sarcasm and irony, and understanding context remain areas of active research in the field.

# **Chapter 3: Methodology**

## **3.1 Data Loading**

## **3.2Text tokenization and Sequencing**

## **3.3 Padding:**

## **3.4 Train-test split:**

## **3.5 GRU model:**

## **3.6 Performance Evaluation:**

# **Chapter 4: Implementation**

## **4.1 Implementation**

## **4.2 Implementation details**

### **4.2.1Train–Test Split**

### **4.2.2 Text Sequencing and Padding**

### **4.2.3 GRU Architecture**

### **4.2.4 Model Training and Adam Optimizer**

### **4.2.5 Model Evaluation**

### **4.2.6 Sentiment Analysis Function**

# **Chapter 5 Result and Findings**

## **5.1 Overfitting and Generalization**

## **5.2 Confusion Matrix**

## **5.3 Model Performance Measure**

# **Chapter 6: Conclusion and Future Recommendations**

## **6.1 Conclusion**

## **6.2 Future Recommendation**

# **References**

[1] V. a. S. V. a. P. A. Sahayak, "Sentiment analysis on twitter data," International Journal of Innovative Research in Advanced Engineering (IJIRAE), vol. 2, pp. 178-183, 2015.

[2] S. a. K. C. a. K. H. a. M. K. a. K. P. Seo, "Comparative study of deep learning-based sentiment classification," IEEE Access, vol. 8, pp. 6861--6875, 2020.

[3] J. D. a. V. N. a. S. S. N. Bodapati, "Sentiment Analysis from Movie Reviews Using LSTMs," Ing{\'e}nierie des Syst{\`e}mes d Inf., vol. 24, pp. 125--129, 2019.

[4] B. C. a. M. M. a. F. J. T. a. A. R. a. C. A. M. Mateus, "Comparing LSTM and GRU models to predict the condition of a pulp paper press," Energies, vol. 14, p. 6958, 2021.

[5] T. B. S. M Yasser H, "IMDB Movie Ratings Sentiment Analysis," vol. 49, 2022.