Compress to 3 objectives only.
(1.) To train...
(2) to evaluate...
(3) to develop....

Comparative Analysis of Traditional Machine Learning and ~~Neural Network~~ Deep Learning Models for Sentiment classification in Rinconada Bikol Language Native Text

	...luding a comparative analysis of traditional odels and neural network-based approaches. The study aims to assess the effectiveness and performance of various classification models, including Support Vector Machine (SVM), Naïve Bayes, Logistic Regression (traditional models), and Long Short-Term Memory (LSTM), Recurrent Neural Networks (RNN), and Convolutional Neural Networks (CNN) (deep learning models). The primary objectives of the study are to: • To preprocess Rinconada-Bikol text data by applying text normalization techniques such as tokenization, stopword removal (if necessary), and word embedding methods. • To implement and train traditional classification models (SVM, Naïve Bayes, Logistic Regression) for sentiment classification of Rinconada-Bikol text. • To implement and train neural network models (LSTM, RNN, CNN) for sentiment classification and compare their effectiveness with traditional methods. • To evaluate and compare model performance using standard metrics such as accuracy, precision, recall, and F1-score. • To analyze the strengths and limitations of both traditional classifiers and deep learning approaches when applied to a low-resource language like Rinconada-Bikol. • To determine the most suitable model for sentiment classification in Rinconada-Bikol based on performance results and computational efficiency. The dataset consists of 12,000 labeled text samples written in Rinconada-Bikol, categorized into different sentiment classes (e.g., positive, negative, and neutral). The evaluation will focus on assessing the accuracy of various classification models in performing sentiment analysis on Rinconada-Bikol text. The study will compare traditional machine learning classifiers and neural network models to determine the most effective approach. Ethical considerations, including user privacy and data anonymization, will be strictly observed. This research will not extend to real-time sentiment analysis or the development of a fully

	deployable system but will instead serve as a proof of concept for future advancements in sentiment classification for low-resource languages.
Limitations of the Study:	This study focuses on the comparative analysis of traditional classification models (e.g., SVM, Naïve Bayes) and neural network-based models (e.g., LSTM, RNN, CNN) for sentiment analysis of the Rinconada-Bikol language. The primary objective is to evaluate and compare the effectiveness of these models in classifying sentiments from a labeled dataset. The study will not extend to advanced applications such as real-time sentiment analysis, predictive behavioral modeling, or the development of a fully integrated sentiment monitoring system. Data collection will be limited to an existing dataset of Rinconada-Bikol text, comprising 12,000 labeled rows. The study will not involve real-time data scraping or analysis of multimedia content such as images, videos, or audio. Additionally, texts with significant code-switching between Rinconada-Bikol and other languages will not be the primary focus. The evaluation will focus solely on sentiment classification accuracy, model efficiency, and comparative performance. The study will not delve into complex linguistic phenomena such as sarcasm, irony, or implicit sentiment expressions, which require deeper contextual understanding. Furthermore, ethical considerations, including anonymization and privacy safeguards, will be strictly observed. This research will serve as a proof of concept for sentiment classification in low-resource languages but will not include real-time monitoring or large-scale deployment. These limitations ensure that the project remains feasible within the given time frame and available computational resources.
Project Design/Development Plan(Software Methodology):	Module 1: Data Preprocessing and Preparation This module will focus on preparing the labeled dataset of 12,000 Rinconada-Bikol text samples for sentiment analysis. Data preprocessing steps will include text normalization, tokenization, stopword removal, and stemming/lemmatization to ensure consistency across all input samples. The dataset will be split into training, validation, and test sets to facilitate model evaluation.

Module 2: Implementation of Traditional Classification Models

This module will implement traditional machine learning models such as Support Vector Machines (SVM), Naïve Bayes, and Random Forest for sentiment classification. These models will be trained on the preprocessed dataset, and their classification performance will be evaluated using accuracy, precision, recall, and F1-score. Feature extraction techniques such as TF-IDF and word embeddings may be used to enhance model performance.

Module 3: Implementation of Neural Network Models

This module will involve training deep learning models, including Long Short-Term Memory (LSTM), Recurrent Neural Networks (RNN), and Convolutional Neural Networks (CNN), for sentiment classification. Word embeddings such as Word2Vec or FastText may be utilized to improve text representation. The models will be trained and tested using the same dataset as the traditional models to ensure a fair comparison.

Module 4: Model Evaluation and Comparative Analysis

This module will compare the performance of traditional classification models and neural network models based on various evaluation metrics, including accuracy, precision, recall, and F1-score. The analysis will highlight the strengths and weaknesses of each approach in classifying sentiments in Rinconada-Bikol text. Model performance will also be analyzed in terms of computational efficiency, training time, and resource consumption.

Module 5: Sentiment Classification System (Proof of Concept)

A prototype system will be developed to demonstrate the sentiment classification capability of the trained models. The system will allow users to input Rinconada-Bikol text and receive sentiment predictions. However, it will be limited to text-based inputs and will not support real-time analysis or large-scale deployment.

Software Specification

Operating System: Windows 10
Programming Language: Python
Machine Learning Libraries: scikit-learn, TensorFlow, Keras
NLP Libraries: Hugging Face Transformers, NLTK, spaCy
Data Processing Libraries: pandas, NumPy
Visualization Tools: Matplotlib, Seaborn

	Hardware Specification Personal Laptop/PC (For Development and Testing): • Processor: Intel Core i3 or higher • RAM: 4 GB DDR4 or higher • Storage: 500 GB SSD • Graphics: Integrated GPU • Input Devices: USB keyboard and mouse
[x] Recommended **ROSEL O. ONESA, MIT** Thesis Subject Adviser	[ ] Rejected **ROSEL O. ONESA, MIT** Thesis Subject Adviser