

Tribhuvan University
Institute of Science and Technology



**A WEB BASED SENTIMENT-CENTRIC NEWS PORTAL
USING SVM AND TF-IDF WITH TIME SERIES ANALYSIS
THROUGH LDA**

A FINAL PROJECT REPORT

Submitted to

Department of Computer Science and Information Technology

*In partial fulfillment of the requirements for the Bachelor's Degree in Computer
Science and Information Technology*

Submitted by

Prayusha Acharya

5-2-1175-23-2020

15/11/2024

Under the supervision of

Shyam Sunder Khatiwada

DEERWALK INSTITUTE OF TECHNOLOGY



Tribhuvan University
Institute of Science and Technology

SUPERVISOR'S RECOMMENDATION

I hereby recommend that this project prepared under my supervision by PRAYUSHA ACHARYA entitled “**A WEB BASED SENTIMENT-CENTRIC NEWS PORTAL USING SVM AND TF-IDF WITH TIME SERIES ANALYSIS THROUGH LDA**” in partial fulfillment of the requirements for the degree of B.Sc. in Computer Science and Information Technology be processed for the evaluation.

.....

Shyam Sunder Khatiwada
Project Coordinator/Supervisor
DWIT College
Deerwalk Institute of Technology

DWIT College
DEERWALK INSTITUTE OF TECHNOLOGY

STUDENT'S DECLARATION

I hereby declare that I am the only author of this work and that no sources other than those listed here have been used in this work.

.....

Prayusha Acharya

5-2-1175-23-2020

15/11/2024

DWIT College
DEERWALK INSTITUTE OF TECHNOLOGY

LETTER OF APPROVAL

This is to certify that this project prepared by PRAYUSHA ACHARYA entitled “**A WEB BASED SENTIMENT-CENTRIC NEWS PORTAL USING SVM AND TF-IDF WITH TIME SERIES ANALYSIS THROUGH LDA**” in partial fulfillment of the requirements for the degree of B.Sc. in Computer Science and Information Technology has been well studied. In our opinion it is satisfactory in the scope and quality as a project for the required degree.

<p>.....</p> <p>Shyam Sunder Khatiwada</p> <p>Project Supervisor</p> <p>DWIT College</p>	<p>.....</p> <p>Mr. Hitesh Karki</p> <p>Chief Academic Officer</p> <p>DWIT College</p>
<p>.....</p> <p>[External Examiner]</p> <p>[Academic designation]</p> <p>IOST, Tribhuvan University</p>	<p>.....</p> <p>Shyam Sunder Khatiwada</p> <p>Project Coordinator</p> <p>DWIT College</p>

ACKNOWLEDGEMENT

I would like to extend my thanks to Mr. Shyam Sunder Khatiwada for his guidance and assistance. His valuable feedback and encouragement improved the quality of my work.

I am also grateful to Deerwalk Institute of Technology for their dedication to advancing technical knowledge by pushing our limits.

I also want to acknowledge the support and understanding of my family and friends who helped me with their valuable expertise and insights to make the project better.

I wish to express my appreciation to all those who contributed to the completion of this project, whether directly or indirectly. Their support, insights, and encouragement have played a significant role in shaping the outcome of this scientific endeavor. I am sincerely grateful for their involvement and recognize their role in the successful completion of this project.

Prayusha Acharya

5-2-1175-23-2020

15/11/2024

ABSTRACT

The “Sentiment-centric News Portal” project addresses the challenges of navigating Nepali news content in the digital age, particularly for non-Nepali speakers and those seeking deeper insights into the emotional context of news. This platform utilizes Natural Language Processing (NLP) and machine learning techniques to provide sentiment analysis of Nepali news article, for a more insightful and accessible news consumption platform. The system employs TF-IDF for feature extraction and combines it with Support Vector Machine (SVM) for achieving high accuracy in determining whether news articles convey positive, negative, or neutral sentiments. A key feature of the portal is its ability to automatically translate Nepali articles into English, expanding accessibility to a global audience. The platform focuses on generating visual representations of entities (such as people, organizations, or topics) mentioned in the news, tracking how their emotional tone evolves over time. This time-series sentiment analysis is powered by Latent Dirichlet Allocation (LDA) to reveal trends and shifts in public sentiment on specific entities. By combining sentiment analysis with dynamic visualizations, the Sentiment-centric News Portal provides valuable insights enabling a more informed, context-driven approach to news consumption. The portal not only bridges linguistic barriers but also fosters greater awareness of how news narratives and public opinions evolve in Nepal.

Keywords: *Emotion Detection, Entity Visualization; News Portal; Natural Language Processing; Time-Series Analysis;*

TABLE OF CONTENTS

SUPERVISOR’S RECOMMENDATION	ii
STUDENT’S DECLARATION	iii
LETTER OF APPROVAL	iv
ACKNOWLEDGEMENT	v
ABSTRACT.....	vi
TABLE OF CONTENTS.....	vii
LIST OF FIGURES	ix
LIST OF TABLES	x
LIST OF ABBREVIATIONS.....	xi
CHAPTER 1: INTRODUCTION	1
1.1. Introduction.....	1
1.2. Problem Statement	1
1.3. Objectives	2
1.4. Scope and Limitation	2
1.5. Development Methodology	3
1.6. Report Organization.....	4
CHAPTER 2: BACKGROUND STUDY AND LITERATURE REVIEW	5
2.1. Background Study.....	5
2.2. Literature Review.....	6
2.4. The problem with Current System.....	9
CHAPTER 3: SYSTEM ANALYSIS.....	10
3.1. Requirement Analysis.....	10
3.1.1. Functional Requirement.....	10
3.1.2. Non-Functional Requirement.....	11
3.2. Feasibility Analysis.....	11

3.2.1. Technical Feasibility	11
3.2.2. Operational Feasibility	11
3.2.3. Economic Feasibility	11
3.2.4. Schedule Feasibility	11
3.3. Analysis.....	13
3.3.1. Object Modeling with Class and Object Diagrams.....	13
3.3.2. Dynamic Modeling with State and Sequence Diagrams.....	13
CHAPTER 4: SYSTEM DESIGN.....	15
4.1. Design	15
4.1.1. Refinement of Class, Object, State, Sequence and Activity diagrams	15
4.1.2 Component diagram.....	18
4.1.3 Deployment Diagram.....	18
4.2. Algorithm Details.....	18
REFERENCES	21

LIST OF FIGURES

Figure 1: Waterfall Model	3
Figure 2: Home Page of The Rising Nepal	8
Figure 3: Home Page of Online Khabar	8
Figure 4: Use Case Diagram for Sentiment-centric News Portal	10
Figure 5: Gantt Chart for the Sentiment-centric News Portal.....	12
Figure 6: Class diagram of Sentiment-centric News Portal.....	13
Figure 7: Sequence diagram of Sentiment-centric News Portal	13
Figure 8: State diagram of Sentiment-centric News Portal	14
Figure 9: Refined Class diagram of Sentiment-centric News Portal	15
Figure 15: Working of SVM.....	19
Figure 16: Working of TF-IDF	19
Figure 17: Working of LDA	20
Figure 18: Blueprint of LDA Model.....	20

LIST OF TABLES

Table 1: Outline of document	4
Table 2: Gantt Chart Table for the Sentiment-centric News Portal	12

LIST OF ABBREVIATIONS

IDF	Inverse Document Frequency
LDA	Latent Dirichlet Algorithm
NLP	Natural Language Processing
SVM	Support Vector Machine
TF-IDF	Term Frequency – Inverse Document Frequency

CHAPTER 1: INTRODUCTION

1.1. Introduction

The "Sentiment-centric News Portal" project aims to revolutionize how we consume and analyze news by leveraging advanced natural language processing and machine learning techniques. This innovative system will crawl multiple Nepali news portals, automatically translating the content into English to ensure accessibility for a broader audience. The core of the sentiment analysis will be powered by algorithms combining Support Vector Machines (SVM) with TF-IDF (Term Frequency-Inverse Document Frequency), offering high accuracy in discerning the emotional tone of news articles. By performing this sentiment analysis on the aggregated news, the platform will provide users with a deep understanding of the emotional context surrounding various events and topics in Nepal and beyond.

A key feature of this project is its time-series analysis capability, which employs Latent Dirichlet Allocation (LDA) to track and visualize sentiment trends over time. This approach will offer valuable insights into the evolution of public opinion on specific topics, entities, or general news trends within the Nepali-speaking world.

Additionally, the platform allows users to save and bookmark news articles for future reference, ensuring they can revisit content that interests them. Users will have the ability to create blogs and express their own opinions on news topics, facilitating a platform for interactive and personal discourse. This feature aims to encourage public engagement and offer a space for thoughtful commentary on current events.

By combining these diverse functionalities and focusing on Nepali news sources, the Sentiment-centric News Portal will empower users to navigate the modern information landscape of Nepal with greater clarity. This tool will be valuable for anyone interested in gaining insights into Nepali public opinion and news trends, ultimately fostering more informed decision-making and public discourse both within Nepal and internationally.

1.2. Problem Statement

In today's digital age, the overwhelming volume of Nepali news content poses significant challenges for readers, researchers, and decision-makers. Language barriers for non-Nepali speakers, a lack of effective sentiment analysis, and inefficient news categorization make it difficult to extract meaningful insights. Additionally, the constant consumption of news can negatively impact individual well-being, causing stress and anxiety. Current news

systems fail to address these issues, often presenting a one-sided view of events and neglecting sentiment trends, especially in the context of Nepali language news.

The Sentiment-centric News Portal project aims to tackle these challenges by utilizing advanced machine learning techniques, including TF-IDF for feature extraction and SVM for sentiment analysis. The system will automatically translate Nepali content into English, making it accessible to a wider audience while preserving the original meaning. Rather than classifying news into categories, the portal will track and visualize sentiment trends over time, offering users valuable insights into the emotional context of news. Additionally, the platform will include features like article bookmarking, saving, and blogging, allowing users to engage more mindfully with the news. By providing a more balanced perspective and enhancing the understanding of sentiment shifts, the portal will bridge the gap between Nepali news sources and a global audience, ultimately promoting informed decision-making and healthier news consumption.

1.3. Objectives

- To create an advanced sentiment analysis model using CNN with SVM that accurately interprets the emotional tone of news articles.
- To develop a time-series analysis feature using LDA to track and visualize sentiment trends over time.

1.4. Scope and Limitation

The scope of this project is:

- Developing a web crawler to collect news articles from various Nepali news portals.
- Implementing a translation module to convert Nepali articles into English, ensuring accessibility for a global audience.
- Integrating sentiment analysis algorithms combining SVM and TF-IDF to accurately assess the emotional tone of news articles.
- Implementing LDA for tracking and visualizing sentiment trends over time.
- Allow users to express their opinion by creating custom blog posts, like articles, bookmark, and comment.
- Designing a user-friendly interface to display aggregated news articles, blog posts, sentiment analysis results, and sentiment trends.

The limitations of this project include:

- The sentiment analysis may struggle to fully capture the context of the Nepali language.
- While the system uses advanced TF-IDF and SVM models, the accuracy of sentiment analysis may vary depending on the quality and complexity of the news content.

1.5. Development Methodology

The waterfall methodology was selected for the development of the Sentiment-centric News Portal due to its systematic progression through development phases. This approach ensures a clear sequence of steps, from requirements analysis to deployment, aligning with the project's structured goals. The emphasis on thorough planning, well-defined stages, and minimal changes post-development suits the project's characteristics, where a stable set of requirements and a systematic approach are prioritized.

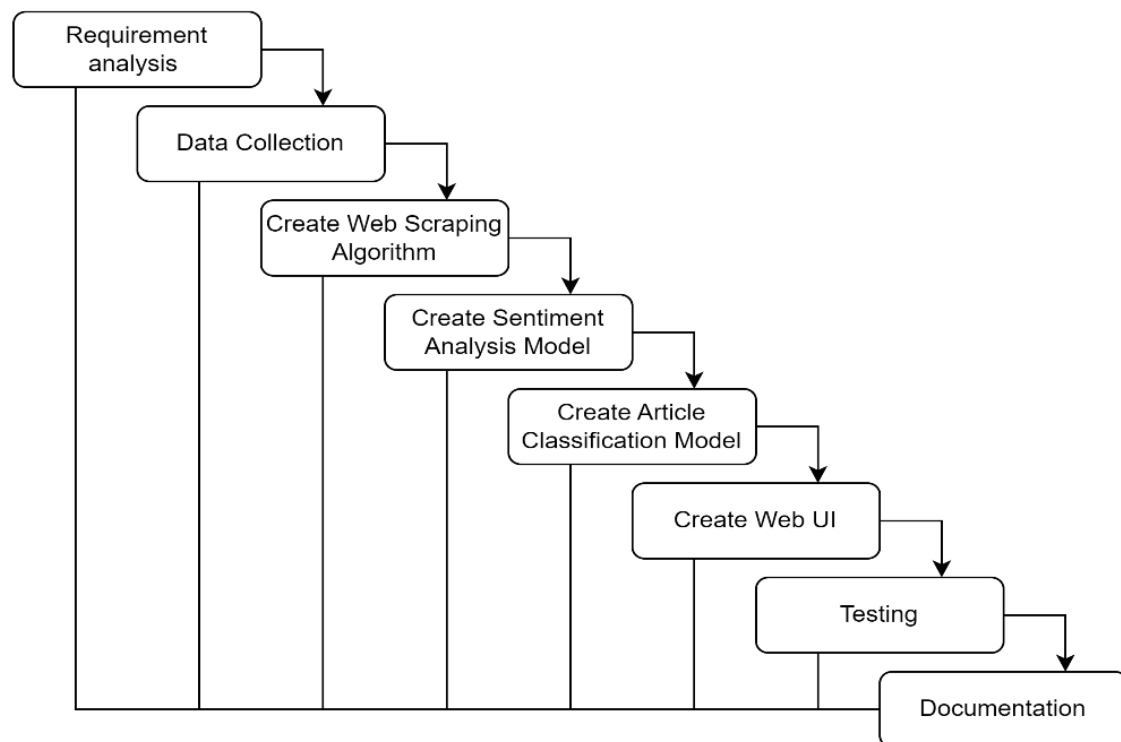


Figure 1: Waterfall Model

1.6. Report Organization

This report is organized as follows:

Table 1: Outline of document

Chapter 1	Overview of the "Sentiment-centric News Portal" project, its objectives, and the need for this solution in addressing the information overload in Nepali news.
Chapter 2	Fundamental theories and reviews of similar projects, and research in the areas of sentiment analysis, machine learning for news aggregation, and sentiment trend tracking.
Chapter 3	Detailed description of system requirements, feasibility analysis, and system modelling. Contains implementations and test results.
Chapter 4	Outlines the design approach covering database, and interface. Describes the algorithms used for sentiment analysis, including the implementation of TF-IDF and SVM.
Chapter 5	Process of implementation and testing and is described along with all the tools used for the development.
Chapter 6	Summary of the project's outcomes, how it addresses the identified problems, and its overall success.

CHAPTER 2: BACKGROUND STUDY AND LITERATURE REVIEW

2.1. Background Study

The Sentiment-centric News Portal project operates at the intersection of several key fields: natural language processing (NLP), machine learning, and data visualization. These fields provide the theoretical foundation for analyzing and interpreting large volumes of text data, such as news articles, and for understanding the sentiment and emotions behind them. Below are the fundamental theories and concepts that underpin this project:

- Sentiment Analysis

Sentiment analysis is a key area within NLP that involves determining the emotional tone or sentiment expressed in a piece of text. In this project, sentiment analysis focuses on identifying whether the sentiment of news articles is positive, negative, or neutral. The analysis of sentiment provides deeper insights into public opinion and emotional context around specific news events.

- TF-IDF (Term Frequency-Inverse Document Frequency)

TF-IDF is a statistical technique used in NLP to evaluate the relevance of words in a document relative to a corpus of documents. It is a common approach for feature extraction in text mining. TF-IDF assigns higher weight to words that appear frequently in a particular document but are rare across the entire corpus, making it useful for understanding the key themes in news articles.

- Support Vector Machines (SVM)

SVM is a supervised machine learning algorithm used for classification tasks, such as classifying news articles based on sentiment. In this project, SVM will be used to distinguish between different sentiment categories (positive, negative, and neutral) by learning patterns in the text data from a training set. SVM is known for its ability to handle high-dimensional spaces, making it effective for text classification tasks.

- Entity Tracking and Time-Series Analysis

Entity tracking involves monitoring how specific entities (people, organizations, locations) appear and are discussed across different news articles over time. Time-series analysis is the technique used to visualize changes in sentiment over a specific period. This is crucial for understanding how public sentiment evolves, particularly in relation to specific topics or events.

2.2. Literature Review

The following literature review examines key studies in areas relevant to the Sentiment-centric News Portal project. These reviews synthesize the research to highlight the relevance to our objectives and how these findings contribute to our project.

a. Emotional Impact of News on Social Judgments (Baum, 2021)

Baum's study in Social Cognitive and Affective Neuroscience investigates how emotional content in news influences social judgments and decision-making. This research is particularly relevant to our project as it underscores the importance of considering emotional content in news analysis. The finding aligns with the core objective of the Sentiment-centric News Portal project analyzing emotional tones within news articles to provide users with insights into various events [1].

b. Social Media News Headlines and Well-Being (Mousoulidou et al., 2024)

Mousoulidou et al.'s research explores the psychological impact of social media news headlines on emotional states and well-being. This study, published in the European Journal of Investigation in Health, Psychology, and Education, discusses the potential psychological harm caused by continuous exposure to emotional news. Their findings are crucial for our project, as they emphasize the need to provide a more mindful and balanced approach to news consumption [2].

c. Ensemble Machine Learning for Twitter Sentiment Analysis (Radiuk et al, 2022)

Radiuk et al. present an ensemble machine learning model for sentiment analysis on Twitter, demonstrating that combining two different algorithms results in more accuracy than using these algorithms alone. Our system uses TF-IDF in conjunction with SVM for sentiment classification, inspired by the success of ensemble techniques. The findings suggest that combining different models enhances the accuracy of sentiment analysis, which is a key component of our news analysis system [3].

d. N-Gram, TF-IDF, and Ensemble Methods in Sentiment Classification (Rahman et al., 2020)

Rahman et al.'s work explores various methods for sentiment classification, comparing N-Gram, TF-IDF, and ensemble techniques. Their study supports our choice of using TF-IDF for feature extraction, which is a highly effective method for identifying the most important terms in a corpus. The comparison of different methods for sentiment analysis helps to justify the choice of TF-IDF combined with SVM for accurate sentiment classification in Nepali news [4].

e. Multilingual Sentiment Analysis (Das et al., 2023)

Das et al.'s research on multilingual sentiment analysis highlights the significance of effective text preprocessing, especially for non-English languages. Their study demonstrates how techniques like the Porter Stemming Algorithm can improve the accuracy of sentiment models. Given the linguistic nuances of Nepali, preprocessing steps are critical for ensuring that sentiment analysis is both accurate and contextually relevant, a challenge that our project aims to overcome [5].

f. LDA Topic-Based Sentiment Analysis in Tourism (Ali et al., 2022)

Ali et al. apply Latent Dirichlet Allocation (LDA) for topic-based sentiment analysis in the tourism sector, demonstrating how LDA can be combined with sentiment analysis to track sentiment trends over time. By incorporating LDA into our sentiment analysis pipeline, the Sentiment-centric News Portal will enable users to visualize how sentiment around specific topics or entities evolves, helping to identify trends in public opinion [6].

g. Comparative Analysis of Sentiment Analysis Techniques: SVM, Logistic Regression, and TF-IDF Feature Extraction (Jadia, 2023)

Jadia's study compares different sentiment analysis techniques, including SVM, Logistic Regression, and TF-IDF feature extraction. The paper evaluates the performance of these methods in sentiment classification tasks. The research finds that SVM combined with TF-IDF outperforms Logistic Regression in terms of accuracy and efficiency in handling complex datasets. This study reinforces our choice of using SVM with TF-IDF for sentiment analysis, supporting the idea that this combination is particularly effective for processing large texts and achieving high accuracy [7].

2.3. Current Systems

In analyzing the current landscape of Nepali news portals, two prominent systems—The Rising Nepal and Online Khabar were examined.

a. The Rising Nepal

The Rising Nepal provides a major contribution to Nepal's digital media landscape, but its focus on English-language content restricts its reach to the broader Nepali-speaking population. Additionally, technical issues, such as broken links and faulty navigation buttons, degrade the user experience. The portal also lacks the integration of sentiment analysis, which could otherwise offer valuable insights into public opinion and emotional responses to news events.

NATION'S 1ST ENGLISH BROADSHEET

THE RISING NEPAL

ALL BE HAPPY, ALL BE WELL

HOME NATION WORLD PROVINCES REPUBLIC DAY SPORTS BUSINESS HEALTH SOCIETY SCIENCE & TECH MORE

President Paudel observes Bhoto Jatra

TRN Online Sun, 4 August 2024

Shah appointed Sudurpashchim CM for second time



Figure 2: Home Page of The Rising Nepal [8]

b. Online Khabar

Online Khabar, while being a dominant player in the Nepali-language news space, faces its own set of challenges. Its exclusive focus on Nepali content limits its accessibility for non-Nepali speakers, hindering its potential to reach a more global audience. Furthermore, like The Rising Nepal, Online Khabar does not incorporate sentiment analysis, missing the opportunity to understand how the public feels about specific news topics, which could help shape more responsive and informed journalism.

onlinekhabar
२० साउन २०८१, आइतबार

होमपेज समाचार बिजनेस जीवनशैली मनोरंजन विचार खेलकुद अन्य health markets. EN

किमोथेरापी इजरायल क्रिकेट आगलागी गौजा तस्करि कर्मचारी Search Keywords

Trending..	Nepal Finance	Nepal Finance	Samridhi Finan...	Nepal Hydro Dev...	Shree Investmen...	Naraya...	View Stock Live
CL 675.00	NFS 2,154.00	SFCL 670.00	NHDL 704.00	SIFC 686.00	NABF		Market Closed
+3% -24.00	+9.96% +195.10	-4.49% -31.50	+10.00% +64.00	+8.27% +52.40	+0.15		

प्रधानमन्त्रीको निर्देशन : दुई लेनभन्दा साना सडक नबनाउनुस्

अनलाइनखबर ४५ मिनेट अगाडि

तीपाको उक्त विषयमा आफैं गम्भीर भएको

Figure 3: Home Page of Online Khabar [9]

These observations highlight significant gaps in the existing news delivery systems, particularly in terms of language inclusivity and the lack of sentiment-driven insights. This study underscores the need for a more inclusive, multilingual, and sentiment-aware news portal capable of addressing these gaps in the current landscape.

2.4. The problem with Current System

The analysis of The Rising Nepal and Online Khabar highlights several key issues that hinder the effectiveness and inclusivity of these news portals.

a. Language Limitations

Language is one of the most significant factors in determining the accessibility and reach of digital news platforms [10]. In the case of The Rising Nepal, offering content exclusively in English limits its accessibility to non-English speaking or less proficient audiences.

In Nepal approximately 75% of the population in Nepal prefers to consume content in Nepali, while only about 25% can understand or prefer English [11]. This data underscores the importance of offering multilingual support, which both The Rising Nepal and Online Khabar fail to adequately address.

b. User Experience and Technical Issues

Poor user-experience due to navigation issues, broken links, and unresponsive pages significantly impacts user retention and engagement. Up to 30% of visitors will leave a site if they encounter technical problems that hinder easy navigation [12].

For The Rising Nepal, where technical issues like broken links and poor navigation have been noted, these problems create barriers to news access. Improving navigability and addressing technical glitches are crucial for increasing user satisfaction and engagement on news portals.

c. Absence of Sentiment Analysis

The absence of sentiment analysis in both The Rising Nepal and Online Khabar is a notable gap, as sentiment analysis has become a critical tool in understanding public opinion and emotional reactions to news events. Sentiment analysis is increasingly being integrated into newsrooms to gauge public reaction and tailor content. This feature allows news platforms to provide content that resonates more deeply with their audience by adjusting tone, coverage, and even editorial focus based on the emotions and sentiments expressed by readers [13].

CHAPTER 3: SYSTEM ANALYSIS

3.1. Requirement Analysis

3.1.1. Functional Requirement

- The system shall automatically translate the extracted Nepali content into English.
- The system shall classify the sentiment of news articles as positive, negative, or neutral.
- The system shall visualize sentiment trends over time for specific topics and entities.
- The user shall be able to filter news articles by sentiment, genre, and date.
- The user shall be able to create and view blog posts based on specific categories.
- The user shall be able to like, bookmark and comment on articles.
- The user shall be able to login to the website to view their saved news articles.

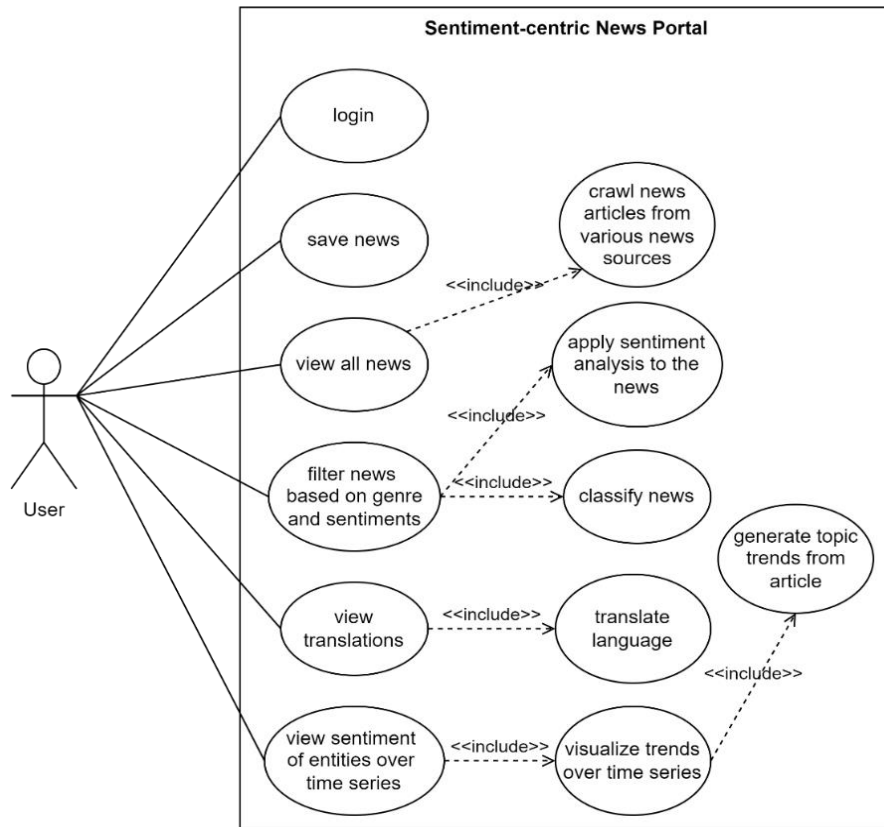


Figure 4: Use Case Diagram for Sentiment-centric News Portal

In this diagram, we have an actor who interacts with the news portal system. The actor could be any individual engaging with the platform. This diagram outlines the functionalities available to users, with an emphasis on sentiment analysis for filtering and viewing news content.

3.1.2. Non-Functional Requirement

- The system must handle large volumes of data efficiently.
- The sentiment analysis and translation processes must achieve high accuracy rates.
- The categorization and time-series analysis must be precise and reliable.
- The translation process must ensure the translated text retains the original meaning and context.

3.2. Feasibility Analysis

3.2.1. Technical Feasibility

The proposed application Sentiment-centric News Portal is a web-based application that uses python-based frameworks. JavaScript, CSS and HTML pages for development of front end while Python is used in backend. Internet connection is required to function properly. For the operation, the application supports any Operating System. The application can be declared technically feasible as all the technical resources are easily available and accessible.

3.2.2. Operational Feasibility

The proposed system will be designed in a way such that a layman can operate it without having to care about technical knowledge. It will enable a person to read news from a single Sentiment Analysis of News 11 website and gives users a clear understanding of the nature of the news beforehand. The simple design makes it an easy-to-use application. This application can be accessed from any place that has an internet connection. Hence, it is understood that Sentiment-centric News Portal is an operationally feasible application.

3.2.3. Economic Feasibility

For the sentiment analysis of Nepali news headlines, there was no monetary expenses and was built by using the existing resources. Therefore, this project can be declared too economically feasible.

3.2.4. Schedule Feasibility

The schedule feasibility of the project was analyzed using Gantt-chart representation for different tasks involved in the development of the system. The project activities are planned to be completed within following time frames: -

Table 2: Gantt Chart Table for the Sentiment-centric News Portal

S.N.	Tasks	Duration
1	Research and Requirement Analysis	7 days
2	Dataset Collection	6 days
3	Create Web Scraping Algorithm	6 days
4	Create Sentiment Analysis Model	15 days
5	Create Article Classification Model	10 days
6	Create Web UI for Output	9 days
7	Testing	4 days
8	Documentation	5 days

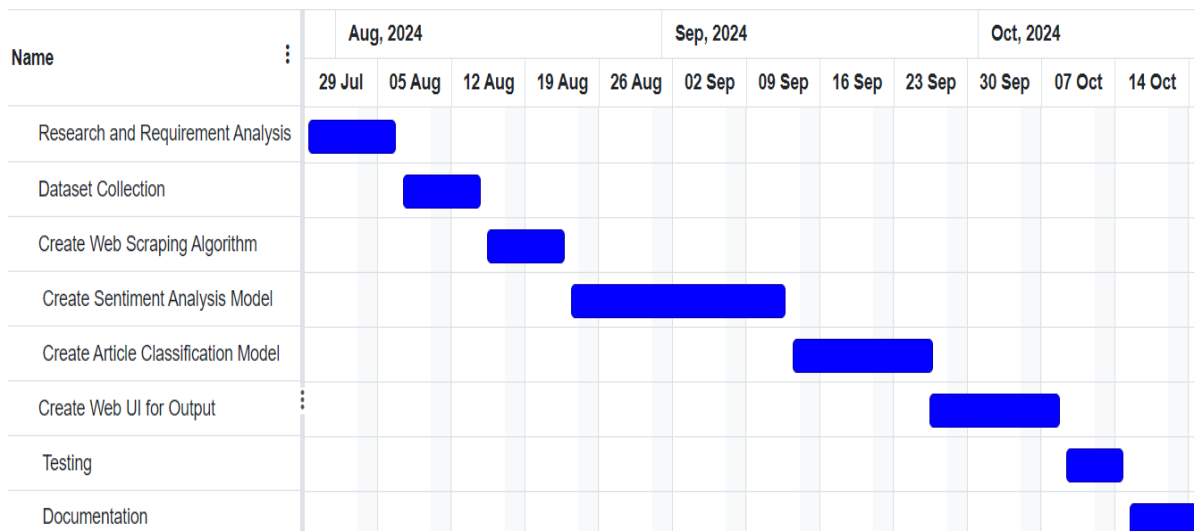


Figure 5: Gantt Chart for the Sentiment-centric News Portal

3.3. Analysis

The system of Sentiment-centric News Portal is based on object-oriented approaches.

3.3.1. Object Modeling with Class and Object Diagrams

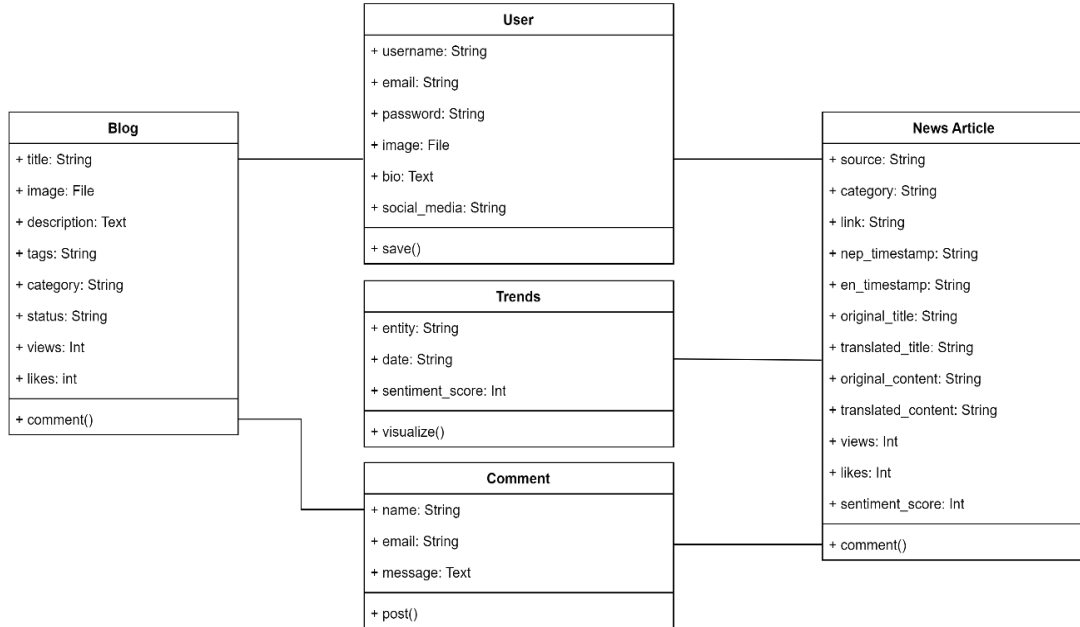


Figure 6: Class diagram of Sentiment-centric News Portal

3.3.2. Dynamic Modeling with State and Sequence Diagrams

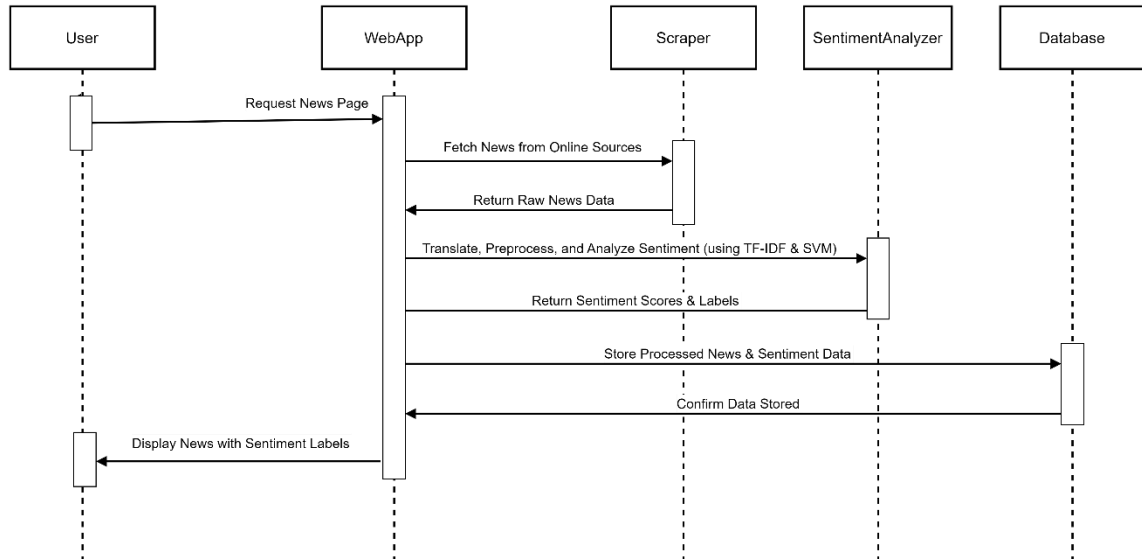


Figure 7: Sequence diagram of Sentiment-centric News Portal

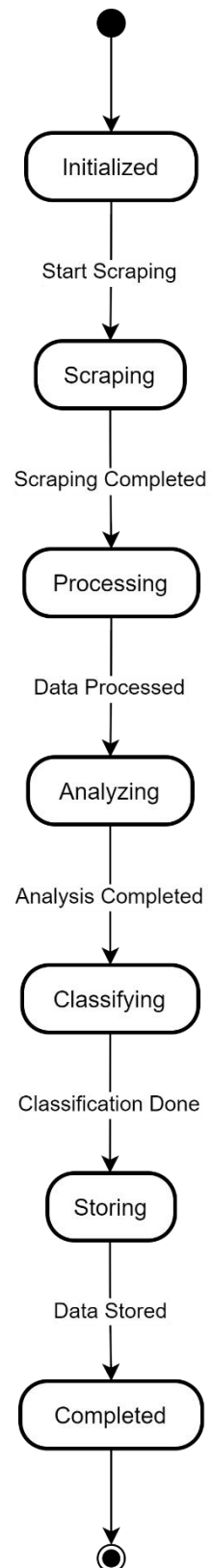


Figure 8: State diagram of Sentiment-centric News Portal

CHAPTER 4: SYSTEM DESIGN

4.1. Design

4.1.1. Refinement of Class, Object, State, Sequence and Activity diagrams

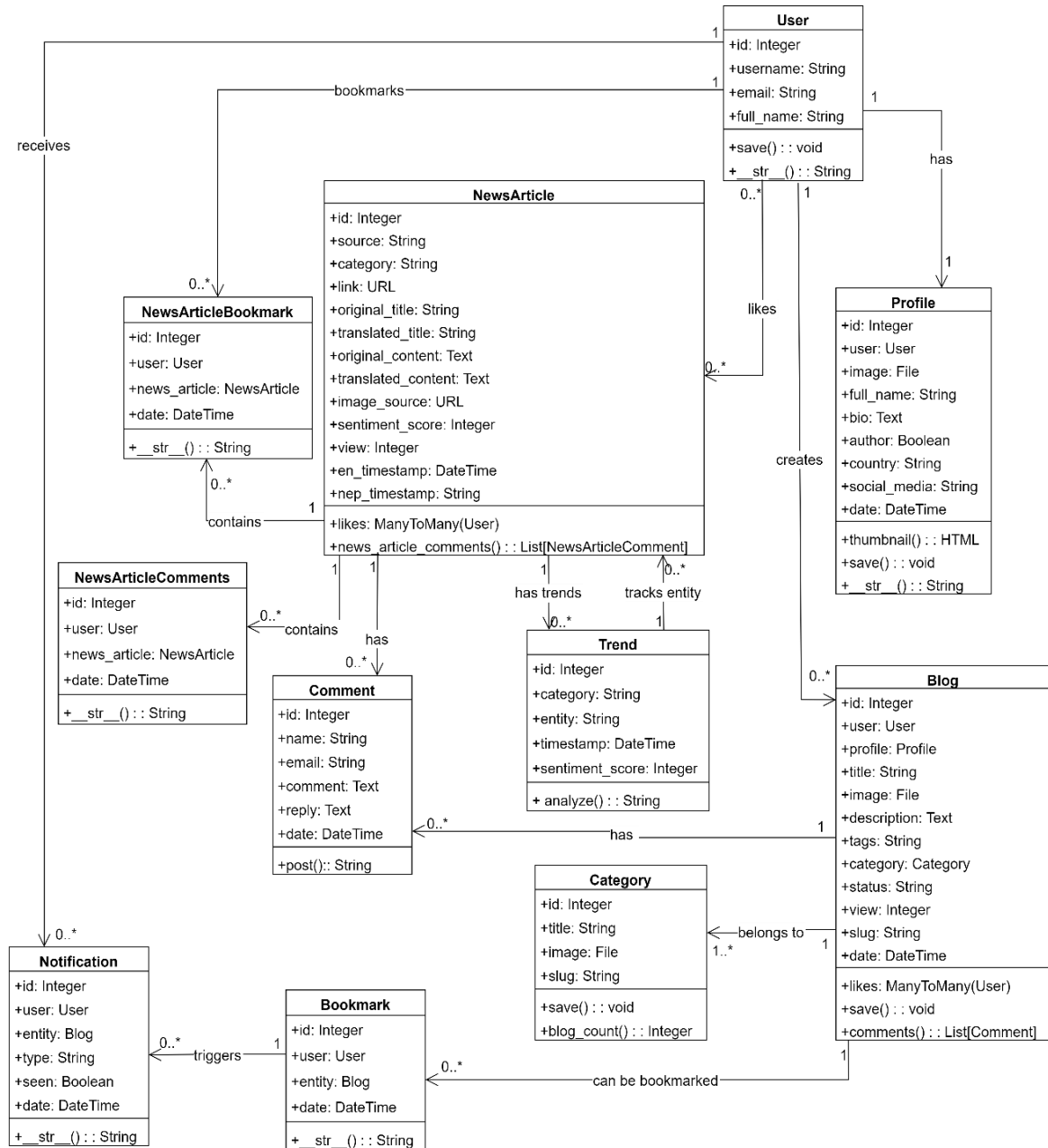


Figure 9: Refined Class diagram of Sentiment-centric News Portal

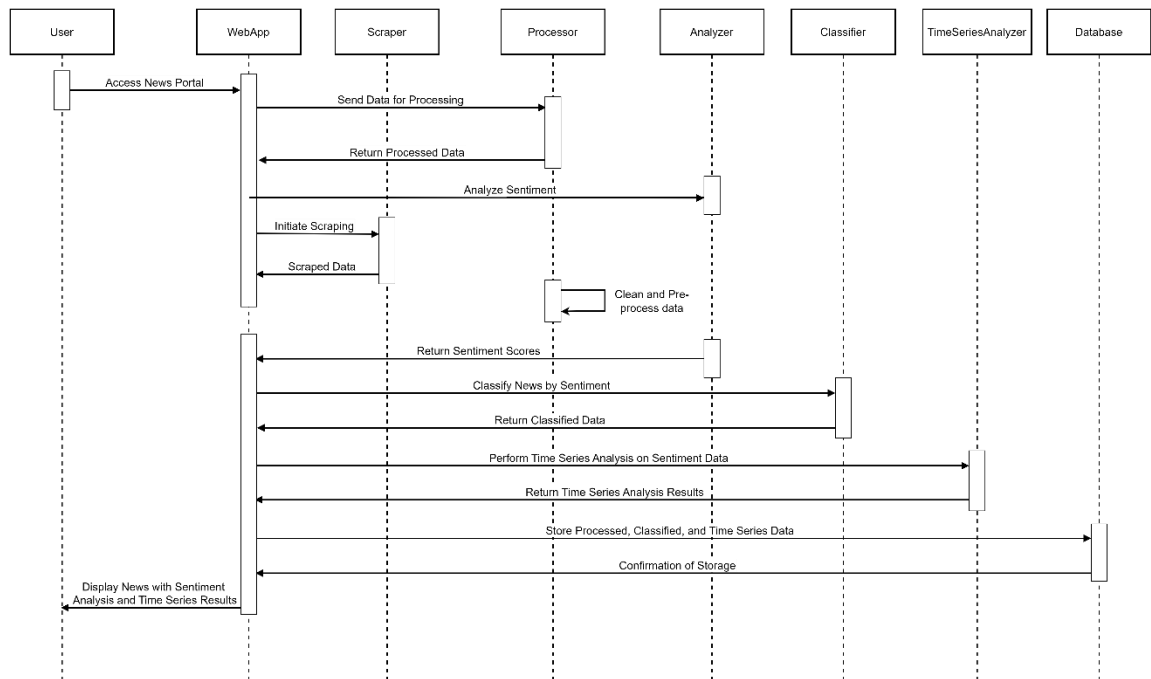


Figure 10: Refined Sequence diagram of Sentiment-centric News Portal

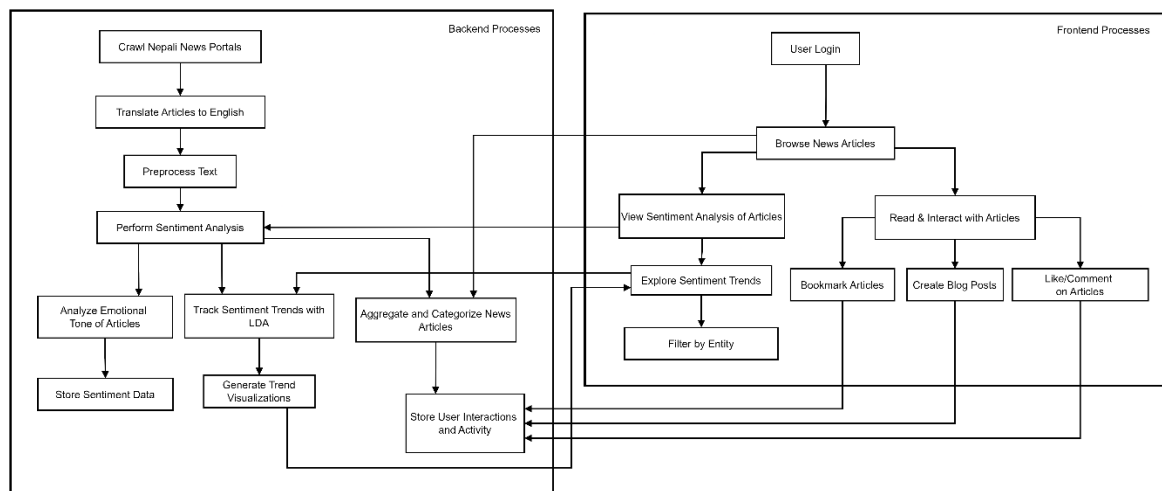


Figure 11: Activity diagram of Sentiment-centric News Portal



Figure 12: Refined State diagram of Sentiment-centric News Portal

4.1.2 Component diagram

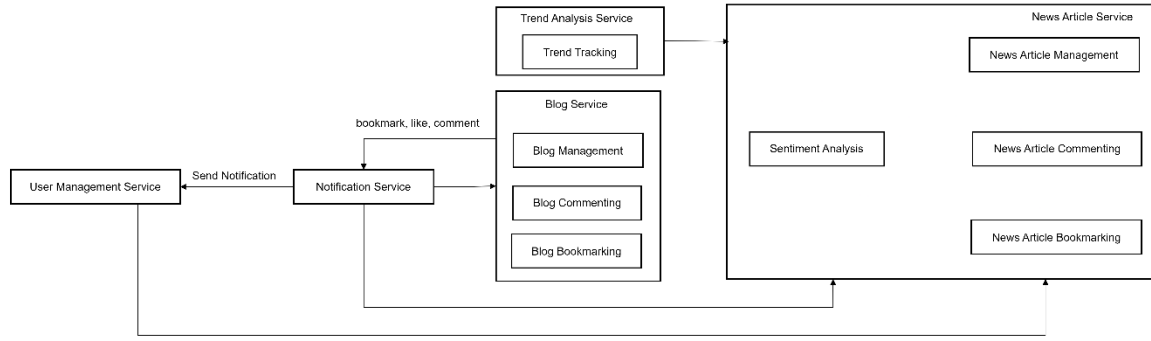


Figure 13: Component diagram of Sentiment-centric News Portal

4.1.3 Deployment Diagram

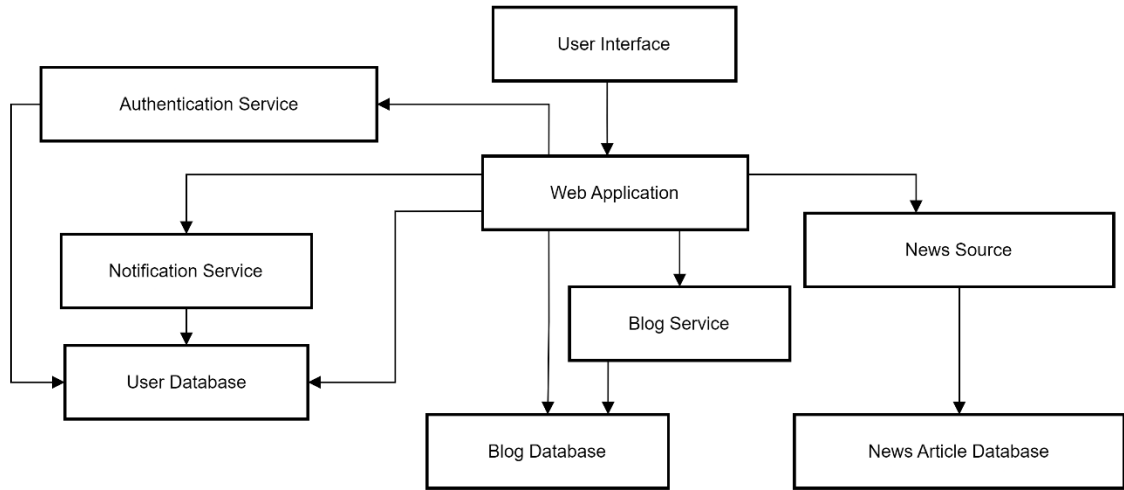


Figure 14: Deployment diagram of Sentiment-centric News Portal

4.2. Algorithm Details

a. SVM

SVM helps in classifying news articles based on their sentiment, especially when dealing with complex, non-linearly separable data. The decision function of SVM can be represented as:

$$f(x) = \omega^T x + b$$

Where:

- ω is the weight vector
- x is the input vector
- b is the bias term

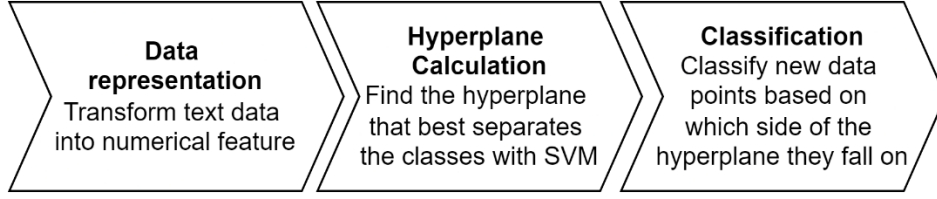


Figure 10: Working of SVM

b. TF-IDF

TF-IDF is used to extract meaningful features from the text of news articles, capturing important terms that contribute significantly to the sentiment analysis. The TF-IDF score is calculated as:

$$TF-IDF(t, d) = TF(t, d) * IDF(t)$$

Where:

- $TF(t, d)$ is the term frequency of term t in document d .
- $IDF(t)$ is the inverse document frequency of term t , calculated as

$$IDF(t) = \log \frac{N}{df(t)}$$

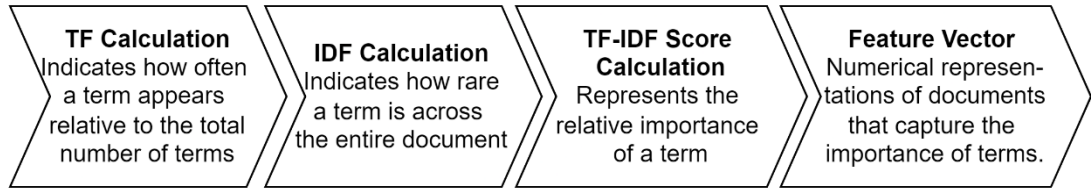


Figure 11: Working of TF-IDF

c. LDA

LDA will be used to track and visualize how sentiment evolves for specific topics or general news trends. The joint probability distribution of LDA can be represented as:

$$P(\omega, z, \theta, \phi, \beta) = P(\theta) \prod_{k=1}^K P(\phi_k) \prod_{n=1}^N P(z_n | \theta) P(\omega_n | \phi_{z_n})$$

Where:

- ω_n are the words in the documents.
- z are the topic assignments.
- θ is the topic distribution for a document.
- ϕ is the word distribution for a topic.
- K is the number of topics.
- N is the number of words in a document.

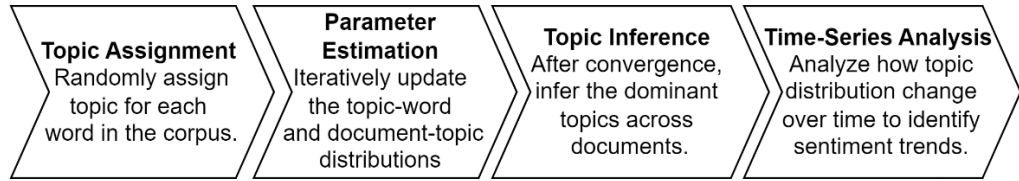


Figure 12: Working of LDA

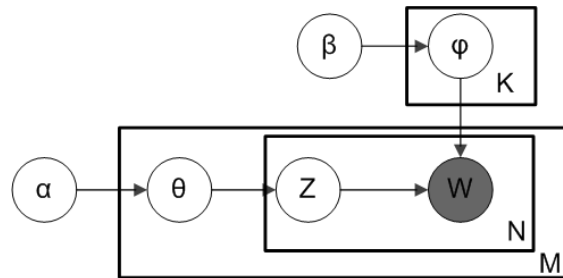


Figure 13: Blueprint of LDA Model

This algorithmic pipeline will enable the Sentiment-centric Intelligent News Portal to conduct thorough sentiment analysis, and trend analysis on Nepali news content, providing users with comprehensive insights into news sentiment and themes.

REFERENCES

- [1] R. A. R. Julia Baum, "Emotional news affects social judgments independent of perceived media credibility," *Social Cognitive and Affective Neuroscience*, vol. 16, no. 3, 2021.
- [2] L. T. A. C. Marilena Mousoulidou, "Social Media News Headlines and Their Influence on Well-Being: Emotional States, Emotion Regulation, and Resilience," *European Journal of Investigation in Health, Psychology and Education*, vol. 14, no. 6, 2024.
- [3] O. P. N. H. Pavlo Radiuk, "An Ensemble Machine Learning Approach for Twitter Sentiment Analysis," in *6th International Conference on Computational Linguistics and Intelligent Systems (CoLInS-2022)*, Gliwice, Poland, 2022.
- [4] K. B. M. B. B. M. H. R. K. S. & T. I. Sheikh Shah Mohammad Motiur Rahman, "An Investigation and Evaluation of N-Gram, TF-IDF and Ensemble Methods in Sentiment Classification," in *International Conference on Cyber Security and Computer Science*, Springer, Cham, 2020.
- [5] M. I. M. M. H. S. R. M. H. S. A. K. Rajesh Kumar Das, "Sentiment analysis in multilingual context: Comparative analysis of machine learning and hybrid deep learning models," *Heliyon*, vol. 9, no. 9, 2023.
- [6] B. O. K. S. Twil Ali, "Analyzing tourism reviews using an LDA topic-based sentiment," *MethodsX*, vol. 9, 2022.
- [7] H. Jadia, "Comparative Analysis of Sentiment Analysis Techniques: SVM, Logistic Regression, and TF-IDF Feature Extraction," *International Research Journal of Modernization in Engineering Technology and Science*, vol. 05, no. 10, 2023.

- [8] "The Rising Nepal | Nepal's First English Broadsheet Daily.," [Online]. Available: <https://risingnepaldaily.com/>. [Accessed 4 August 2024].
- [9] "Online Khabar," [Online]. Available: <https://www.onlinekhabar.com/>. [Accessed 4 August 2024].
- [10] L. F. M. & P. Z. Jenkins, "Language barriers in digital news accessibility," *Journal of Communication*, 2020.
- [11] "Survey on Digital News," Nepal Telecommunications Authority, 2022. [Online]. Available: nta.gov.np. [Accessed 2024].
- [12] "The effects of UX design on website abandonment rates," Statista, 2021. [Online]. Available: [statista.com](https://www.statista.com). [Accessed 2024].
- [13] A. J. M. & B. T. Hughes, "Sentiment analysis in digital journalism.," *Journal of Digital Media & Policy*, vol. 10, no. 1, 2019.