Journal of Data Science 0 (0), 1–20
??? 2025
DOI: 10.6339/25-JDS1195
Data Science in Action

Comparative Analysis of VADER and TextBlob on Financial News Headlines

Keshab Raj Dahal[1,*], Ankrit Gupta[2], and Nirajan Budhathoki[3,4]

[1]*Department of Mathematics, State University of New York Cortland, Cortland, NY, USA*
[2]*Department of Computer Science, Central Michigan University, Mount Pleasant, MI, USA*
[3]*Henry Ford Health + Michigan State University Health Sciences, Detroit, MI, USA*
[4]*Dept of Public Health Sciences, Henry Ford Health, Detroit, MI, USA*

Abstract

Financial news headlines serve as a rich source of information on financial activities, offering a wealth of text that can provide insights into human behavior. One key analysis that can be conducted on this text is sentiment analysis. Despite extensive research over the years, sentiment analysis still faces challenges, particularly in handling internet slang, abbreviations, and emoticons commonly found on many websites that cover financial news headlines, including Bloomberg, Yahoo Finance, and Financial Times. This paper compares the performance of two sentiment analyzers—VADER and TextBlob—on financial news headlines from two countries: the USA (a well-developed economic nation) and Nepal (an underdeveloped economic nation). The collected headlines were manually classified into three categories (positive, negative, and neutral) from a financial perspective. The headlines were then cleaned and processed through the sentiment analyzers to compare their performance. The models' performance is evaluated based on accuracy, sensitivity, specificity, and neutral specificity. Experimental results reveal that VADER performs better than TextBlob on both datasets. Additionally, both models perform better on financial news headlines from the USA than Nepal. These findings are further validated through statistical tests.

Keywords *finance; news; sentiment analysis; text mining*

1 Introduction

The proliferation of financial news headline platforms has significantly transformed how internet data is generated and consumed (Kolbitsch and Maurer, 2006). By providing concise and easily accessible information, these platforms cater to a wide audience seeking quick updates on various subjects. Users no longer need to sift through lengthy articles or complex reports; instead, they can rely on succinct headlines that deliver key information at a glance. This shift not only saves time but also enhances the efficiency of information dissemination, ensuring that individuals and businesses stay informed and make timely decisions. The revolution in data consumption driven by these platforms underscores the growing demand for streamlined and rapid access to relevant news and insights (van Ooijen et al., 2019).

The wealth of conversations on financial news headline platforms offers a rich resource for understanding human behavior through sentiment analysis. By examining the myriad opin-

*Corresponding author. Email: keshabraj.dahal@cortland.edu.

Received September 14, 2024; Accepted June 27, 2025

ions and discussions, researchers can gain insights into public sentiment and trends. Sentiment analysis, a key tool in this endeavor, can be approached through machine learning-based methods, which involve training algorithms on large datasets to recognize patterns, or lexical-based methods, which rely on predefined dictionaries of words associated with positive or negative emotions (Al-Qablan et al., 2023; Agbehadji and Ijabadeniyi, 2021). Both approaches enable a deeper understanding of the nuances in public opinion, aiding in more accurate predictions and decision-making in various fields (Shayaa et al., 2018).

The effectiveness of machine learning methods is often hampered by the scarcity of labeled data, which significantly limits their applicability to new data sets. This challenge arises because labeling data, even in small quantities, is an expensive and labor-intensive process. For simple tasks, the cost of annotating data can be disproportionately high, making it difficult to gather the necessary volume of labeled examples required for training robust machine learning models. Consequently, the limited availability of labeled data constrains the ability of these models to generalize and perform accurately on new, unseen data, thereby reducing their overall utility and effectiveness in practical applications (Pennebaker et al., 2001).

On the flip side, lexical-based sentiment analysis is often considered better than machine learning-based methods for sentiment analysis (Talpada et al., 2019). Unlike machine learning approaches, lexical-based methods do not require large amounts of labeled data for training, making them easier to implement and more accessible, especially for smaller projects or those with limited resources. These methods rely on predefined dictionaries of words associated with specific sentiments, allowing for straightforward identification of positive or negative language without the need for complex algorithms (Padmaja et al., 2014). Additionally, lexical-based approaches are more transparent and interpretable, providing clear reasoning for sentiment classifications, which can be crucial for understanding and refining the analysis process (Bharadwaj, 2023). This makes lexical-based sentiment analysis a practical and efficient choice for many applications.

Several studies have compared the effectiveness of various lexicons in conducting sentiment analysis on text (Gonçalves et al., 2013; Musto et al., 2014; Khoo and Johnkhan, 2018; Al-Shabi, 2020; Bonta et al., 2019). However, these studies primarily focus on the use of general lexicons such as the General Inquirer and SentiWordNet, which do not account for internet slang and emoticons. In contrast, this study aims to compare the performance of Valence Aware Dictionary for sEntiment Reasoning (VADER) and TextBlob in analyzing sentiment from financial news headlines. The concise nature of news headlines often necessitates the use of internet slang, acronyms, and emoticons for expression. Therefore, because both VADER and TextBlob accommodate these linguistic styles, the research seeks to determine which lexicon yields better results. Moreover, recent concerns have been raised regarding the obsolescence of traditional rule-based and lexicon-based Natural Language Processing (NLP) methods as large language models (LLMs) have gained dominance across nearly all NLP tasks, including sentiment analysis (Zhang et al., 2023). These developments necessitate a re-evaluation of lightweight tools such as VADER and TextBlob, particularly in contexts where resource constraints or explainability remain paramount. The findings from this study will provide insights into selecting the appropriate lexicon for analyzing short social media texts effectively.

Our comprehensive vision to achieve the stated goal is illustrated in the schematic diagram in Figure 1. The proposed study constructs the model using financial news headlines. Initially, we manually determine the actual sentiments (positive, negative, and neutral) from these headlines. During the data cleaning process, we utilize various built-in functions such as split(), tokenize(), and lemmatize(). Sentiment prediction is then computed using VADER and TextBlob. Finally,

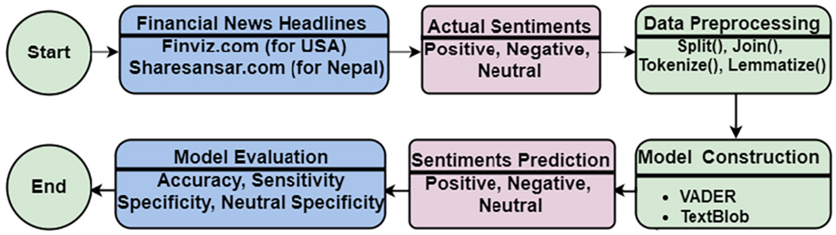

Figure 1: Schematic diagram of the proposed research framework.

we evaluate the accuracy and robustness of the proposed model using four distinct performance metrics: accuracy, sensitivity, specificity, and neutral specificity.

The primary contributions of this research are as follows: (a) Addressing the core inquiry: Given uniform circumstances, which model—VADER or TextBlob—is the better choice for predicting sentiment when using financial news headlines as input? (b) Investigating whether news headlines from well-developed and underdeveloped economic nations produce different outcomes in sentiment analyzers, specifically VADER and TextBlob. (c) Conducting a series of statistical hypothesis tests to confirm the experiment's reliability.

The remainder of this paper is structured as follows: Section 2 delves into related work within this domain. Section 3 provides a concise overview of the sentiment analyzers used in this study. Standard assessment metrics are explained in Section 4. Section 5 discusses the experimental design, including data description, data preparation, model comparison, followed by statistical analysis. Finally, Section 6 presents the discussion and future work, followed by acknowledgments and the list of references.

2 Related Work

In recent years, sentiment analysis in financial news has become a pivotal tool for investors, analysts, and policymakers seeking to gauge market trends and make informed decisions. Sentiment analysis methods such as VADER and TextBlob have been widely adopted due to their ability to process and interpret large volumes of text data efficiently (Sanyal and Barai, 2021; Asderis, 2022; Illia et al., 2021; Min et al., 2020). This section focuses on reviewing prior studies that have utilized various sentiment analyzers.

VADER has consistently demonstrated superior performance over TextBlob in sentiment analysis across a wide range of datasets and applications. Min et al. (2020) examines the performance of VADER and TextBlob, two lexicon-based sentiment analyzers, on a dataset of 7,997 tweets, with 300 tweets manually classified by experts into three polarity categories. The study found VADER to have an accuracy rate of 79%, compared to TextBlob's 73%, making VADER the superior lexicon. Illia et al. (2021) analyzed public sentiment towards an application using Twitter data from August 31 to September 7, 2021, and found VADER more effective for social media sentiment analysis. Asderis (2022) compared TextBlob and VADER using social media data from the COVID-19 pandemic, identifying discrepancies in sentiment classification and highlighting the need for improved sentiment analysis tools. Sanyal and Barai (2021) compared VADER and TextBlob using Amazon product review data, finding VADER more accurate in classifying negative opinions. Abiola et al. (2023) analyzed 1,048,575 tweets with the hashtag 'COVID-19' using TextBlob and VADER, finding varying sentiment distributions: VADER detected 39.8% positive, 31.3% neutral, and 28.9% negative sentiments, while TextBlob found 46.0% neutral, 36.7% positive, and 17.3% negative sentiments. Elbagir and Yang (2020) used VADER to classify sentiments in Twitter data related to the 2016 US election, demonstrating VADER's effectiveness in detecting ternary and multiple sentiment classes.

Comparative studies highlight VADER's strong performance in sentiment classification, while advanced models like BERT and RNNs excel in complex domains such as financial analysis. Bonta et al. (2019) employs Natural Language Toolkit (NLTK), TextBlob, and VADER sentiment analysis tools to classify movie reviews from Rotten Tomatoes, provided by Cornell University, and determines that VADER outperforms both NLTK and TextBlob in sentiment classification. Singh and Verma (2021) aims to simplify aspect-based sentiment analysis by training an Natural Language Processing (NLP) model on text and reviews, using VADER, TextBlob, and spaCy libraries to extract sentiment and aspects, asserting that this approach will expedite aspect-based sentiment analysis and alleviate the challenges of aspect extraction. Nemes and Kiss (2021) performed sentiment analysis on company news headlines using Bidirectional Encoder Representations from Transformers (BERT), VADER, TextBlob, and a recurrent neural network (RNN), finding that BERT and the RNN were more accurate in identifying stock value change timings by comparing sentiment results with stock changes, noting significant differences in the impact of emotional values across models.

Research consistently shows that VADER is a reliable sentiment analysis tool, with its effectiveness further amplified when integrated into combined-method approaches across diverse datasets. Gonçalves et al. (2013) introduced a novel combined-method that integrates seven pre-existing techniques: PANAS-t, Emoticons, SentiStrength, SentiWordNet, SenticNet, SASA, and the Happiness Index. By evaluating the harmonic mean of precision and recall, they demonstrated that their combined method outperformed eight established methods, including LIWC. Das et al. (2021) focused on sentiment analysis across four datasets obtained from Twitter, Facebook, Economic Times headlines, and stock market-related news articles, comparing seven sentiment analysis tools: Stanford, support vector classifier (SVC), TextBlob, Henry, Loughran-McDonald, logistic regression (LR), and VADER. The data scraped from Facebook generated the highest overall positive sentiment score of 38.17%, with VADER performing the best among the tools, calculating an overall positive sentiment score of 56.63%. Al-Shabi (2020) employed a lexicon-based methodology for sentiment analysis, focusing on five prominent sentiment analyzers: VADER, SentiWordNet, SentiStrength, Liu and Hu opinion lexicon, and AFINN-111, evaluating their performance in classifying Twitter sentiment by comparing their overall classification accuracy and F1-measure, finding that VADER achieved higher accuracy in classifying positive and negative sentiments.

Among sentiment analysis models, machine learning techniques like SVM and BERT consistently outperform lexicon-based methods, with VADER leading within the lexicon-based approaches. Srivastava et al. (2022) explored machine learning and lexicon-based algorithms for sentiment calculation and model performance using support vector machine (SVM) and LR for the machine learning models, and AFINN-111 and VADER for the lexicon-based models. The dataset comprised 'Trip Advisor Hotel Reviews', and classifier accuracy was assessed using various evaluation metrics. Among the machine learning models, SVM demonstrated superior accuracy, achieving a 95.2% classification rate with Bag of Words and 96.3% with TF-IDF. In the lexicon-based approach, VADER outperformed AFINN-111 with an accuracy of 88.7% compared to 86.0%. Mujahid et al. (2023) performed sentiment analysis and topic modeling on ChatGPT-based tweets using Latent Dirichlet Allocation (LDA) to identify the most frequently discussed topics. For sentiment analysis, a deep transformer-based BERT model with three dense neural network layers was proposed. Additionally, machines and deep learning models with fine-tuned parameters were utilized for comparative analysis. Experimental results demonstrated the superior performance of the proposed BERT model, achieving an accuracy of 96.49%.

Recent research highlights the superior performance of BERT and advanced preprocessing techniques in sentiment analysis of social media data, particularly during the COVID-19 pandemic, while also noting the varying strengths of tools like TextBlob and VADER. Saha et al. (2022), the author crawled tweets from Twitter about COVID-19 and Omicron, creating two datasets. They applied VADER and BERT to determine the sentiment polarity of the tweets and used five supervised ML algorithms—Naive Bayes, Random Forest (RF), Gradient Boosting (GBC), Extreme Gradient Boosting (XGBoost), and SVM—to analyze classification performance. They found that most tweets expressed negative sentiments, with SVM outperforming other algorithms with a 92% accuracy using BERT on the Omicron dataset. BERT improved classification performance for all algorithms except Naive Bayes. Pano and Kashef (2020) developed various text preprocessing strategies to correlate Twitter sentiment scores with Bitcoin prices during the COVID-19 pandemic, examining 13 preprocessing functions, features, and data time lengths. They found that splitting sentences, removing Twitter-specific tags, or combining both generally enhanced the correlation between sentiment scores, volume polarity scores, and Bitcoin prices, with better correlations over shorter timespans. Ccoya and Pinto (2023) utilized five Python and R-based sentiment analyzers—NLTK, TextBlob, VADER, Transformers (GPT and BERT), and Tidytext—to compute sentiment from social media text data, then compared four machine learning models—decision tree, SVM, Naive Bayes, and k-nearest neighbor (KNN)—using precision, recall, and F1 score metrics, concluding that the BERT transformer method was superior with an accuracy of 0.973. Dahal et al. (2024) compared LR, SVM, RF, XGBoost, and artificial neural network (ANN) to predict the next day's movement direction of the Nepal Stock Exchange (NEPSE) index closing price using sentiment scores from VADER and TextBlob. The results showed all models performed equally well with TextBlob sentiment scores and nearly identical performance with VADER sentiment scores, with overall better performance using TextBlob.

In conclusion, these studies collectively demonstrate the effectiveness of various sentiment analysis tools and machine learning algorithms across different contexts and datasets. Some authors used pre-trained sentiment analyzers such as VADER and TextBlob, while others used machine learning models such as LR and SVMs for sentiment analysis. Together, these findings emphasize the critical role of advanced sentiment analysis techniques and preprocessing strategies in enhancing the accuracy and reliability of sentiment-based predictions.

3 Modeling Approach

3.1 VADER

VADER is a simple rule-based model for general sentiment analysis (Hutto and Gilbert, 2014). It first uses a combination of qualitative and quantitative methods to create and validate a sentiment lexicon for microblog-like context. These lexical features are then integrated with rules that reflect how humans express sentiment intensity. Hutto and Gilbert (2014) claim that the VADER lexicon performs exceptionally well in the social media domain. With an F1 score of 0.96, it had a superior classification ability than individual human raters (F1 = 0.84) at classifying the sentiments of tweets into positive, neutral, or negative classes.

Table 1: Example to demonstrate VADER's scoring method.

Sentiment	Individual Words								Full Sentence
	Stocks	open	in	the	green	after	cool	CPI	
Negative	0	0	0	0	0	0	0	0	0
Neutral	1	1	1	1	1	1	0	1	0.7530
Positive	0	0	0	0	0	0	1	0	0.2470
Compound	0	0	0	0	0	0	0.3182	0	0.3182

VADER computes a vector of sentiment scores with negative, neutral, positive, and compound polarities from each body of text. The negative, neutral, and positive scores are normalized between 0 and 1 since they represent the proportion of text that fall under each category. The compound score is obtained as a sum of the valence scores of each word in the lexicon and is normalized to be between -1 (most extreme negative) and $+1$ (most extreme positive) (Hutto, 2020). VADER contains a predefined list of words (lexicon) that are assigned a sentiment valence. Each word in the lexicon is associated with a sentiment score between -4 (most negative) and $+4$ (most positive). To obtain a sentiment score, any given text is broken down into individual words (tokens), and each word is looked up in the lexicon. The compound score is the normalized sum of all the lexicon ratings in the sentence. It is computed as

$$\text{Compound} = \frac{\sum \text{valence scores}}{\sqrt{\sum(\text{valence scores})^2 + \alpha}}, \tag{1}$$

where α is a normalization constant ($\alpha = 15$ by default). This normalization constant approximates the maximum expected value of the valence scores. The default value of 15 increases the denominator of the normalization function and makes it less sensitive to small changes in the maximum expected value of the valence scores.

As an example to see how VADER works, let us consider a news headline that reads, "Stocks open in the green after cool CPI". Table 1 shows the break down of how VADER computes sentiment scores for individual words and the entire sentence.

Based on VADER's documentation, the word 'cool' has a positive sentiment score of 1.3. Therefore, from Equation (1), the compound score is found to be $\frac{1.3}{\sqrt{1.3^2+15}} = 0.3182$. For the full sentence, the scores are the proportion of individual words that carry the given sentiment. For example, seven of the eight words carry a neutral sentiment which translates to a raw proportion of 0.8750. However, VADER also considers the intensity and context of words, adjusting the raw proportion to produce a more nuanced score of 0.7530. The compound score of 0.3182 for the full sentence implies a moderately positive sentiment and is directly derived from the valence score of the word 'cool'. The neutral and positive scores provide additional context about the distribution of sentiment within the text but do not directly affect the compound score calculation.

VADER offers several advantages for real-world use compared to sophisticated machine learning techniques. To list a few, it is computationally economical, nicely interpretable, and is easily extended to several domains without requiring an extensive set of training data. VADER's suitability for use in this study followed by several successful applications in the past (Ekaputri and Akbar, 2022; Pokhrel et al., 2024; Maqbool et al., 2023; Dahal et al., 2023) where the researchers performed sentiment analysis from financial news.

3.2 TextBlob

TextBlob is a robust and versatile Python library designed for NLP tasks, making it an essential tool for sentiment analysis in various domains, including the analysis of news sentiment related to stock market indices. Its simplicity and efficiency stem from leveraging the capabilities of the NLTK and Pattern (De Smedt and Daelemans, 2012) libraries, providing a user-friendly interface for both novice and experienced researchers. TextBlob's suite of functionalities includes part-of-speech tagging, noun phrase extraction, sentiment analysis, and tokenization, which are pivotal for dissecting the linguistic nuances present in news articles. For instance, the sentiment analysis feature assigns polarity and subjectivity scores to text, allowing researchers to quantitatively assess the positivity or negativity of news content. This quantitative assessment is crucial for correlating news sentiment with stock market movements, as it enables the identification of sentiment trends that may influence investor behavior and market volatility.

In the context of stock market analysis, TextBlob's sentiment analysis can be employed to gauge the impact of news sentiment on stock indices (Koukaras et al., 2022; Nemes and Kiss, 2021). By analyzing historical news data and corresponding stock market responses, researchers can develop predictive models to forecast market trends based on current news sentiment. TextBlob's ability to handle various text preprocessing tasks, such as tokenization and part-of-speech tagging, further enhances its utility in preparing news data for analysis. The library's simplicity in syntax and integration with other Python tools also facilitates the creation of comprehensive NLP pipelines, enabling efficient and accurate sentiment analysis. Consequently, TextBlob serves as a valuable asset for financial analysts and researchers aiming to understand and predict the intricate relationship between news sentiment and stock market dynamics.

TextBlob uses rule-based sentiment analysis method to calculate text sentiments. For any given sentence, TextBlob returns two properties: Polarity and Subjectivity. The polarity determines polarity of the input statement. It is a float within the range $[-1.0, 1.0]$, where -1 indicates negative sentiment and 1 indicates good sentiment. The subjectivity is a float within the range $[0.0, 1.0]$ where 0.0 is very objective and 1.0 is very subjective (Loria, 2024). Higher subjectivity means that the text contains personal opinion rather than facts.

As an example, we again consider the sentence "Stocks open in the green after cool CPI" to calculate the polarity and subjectivity score. For this example, polarity is found to be 0.05 indicating a mildly positive sentiment, and subjectivity is 0.48 indicating a moderately subjective statement. These scores are driven by only three words: 'open', 'green', 'cool'. Based on TextBlob's documentation, the individual polarity and subjectivity of these words are presented in Table 2. The aggregate score for the full sentence is the mean score (sum of scores/3).

Table 2: Example to demonstrate Textblob's scoring method.

Properties	Individual Words								Full Sentence
	Stocks	open	in	the	green	after	cool	CPI	
Polarity	0	0	0	0	-0.2	0	0.35	0	0.05
Subjectivity	0	0.5	0	0	0.3	0	0.65	0	0.48

Table 3: Confusion matrix.

	Predicted positive	Predicted negative	Predicted neutral
Actual positive	True positive (TP)	False negative1 (FN_1)	False neutral1 ($FNEU_1$)
Actual negative	False positive1 (FP_1)	True negative (TN)	False neutral2 ($FNEU_2$)
Actual neutral	False positive2 (FP_2)	False negative2 (FN_2)	True neutral (TNEU)

4 Performance Measures

Sentiment prediction from the constructed model is assessed through the performance metrics: accuracy, sensitivity, specificity, and neutral specificity. These metrics help determine the best model in terms of accuracy and reliability (Hameed et al., 2022; Joshi et al., 2022; Shamrat et al., 2022). The elements of the confusion matrix are utilized to find four important metrics: accuracy, sensitivity, specificity, and neutral specificity. The analytical form of a confusion matrix is defined in Table 3.

Sensitivity indicates the proportion of true positive sentiments correctly identified as positive. Specificity reflects the proportion of true negative sentiments accurately identified as negative. Neutral specificity denotes the proportion of true neutral sentiments correctly classified as neutral. Accuracy measures the overall proportion of sentiments accurately predicted. The analytical definitions of these metrics are as follows:

$$\text{Sensitivity} = \frac{\text{True positive (TP)}}{\text{True positive (TP)} + \text{False negative1 } (\text{FN}_1) + \text{ False neutral1 } (\text{FNEU}_1)}, \tag{2}$$

$$\text{Specificity} = \frac{\text{True negative (TN)}}{\text{True negative (TN)} + \text{False positive1 } (\text{FP}_1) + \text{False neutral2 } (\text{FNEU}_2)}, \tag{3}$$

$$\text{Neutral Specificity} = \frac{\text{True neutral (TNEU)}}{\text{True neutral (TNEU)} + \text{False positive2 } (\text{FP}_2) + \text{False negative2 } (\text{FN}_2)}, \tag{4}$$

$$\text{Accuracy} = \frac{\text{TP} + \text{TN} + \text{TNEU}}{\text{TP} + \text{FN}_1 + \text{FNEU}_1 + \text{FP}_1 + \text{TN} + \text{FNEU}_2 + \text{FP}_2 + \text{FN}_2 + \text{FNEU}}. \tag{5}$$

Sensitivity and specificity are inversely related. The preferred model is chosen based on prioritizing the following metrics: accuracy (highest priority), followed by sensitivity, specificity, and neutral specificity (which are equally important but have a lower priority compared to accuracy).

5 Experimental Design

This study compared the performance of the VADER and TextBlob sentiment analyzers on news data from Nepal Stock Exchange (NEPSE) and National Association of Securities Dealers Automated Quotations (NASDAQ). The computing environment was set up using Google Colab, with Python 3.6.0, TensorFlow, and Keras APIs. The architecture employed for the analysis included VADER and TextBlob for text processing tasks. The experiment was conducted in four phases: data description, data preparation, model comparison, followed by statistical analysis.

5.1 Data Description

This study aims to compare the performance of sentiment analyzers, specifically VADER and TextBlob, on financial news headlines. To ensure a fair comparison, we analyze news headlines from Nepal, defined as an underdeveloped economic country (United Nations, 2025) and the United States of America, a well-developed economic country.

Nepal is a culturally diverse and landlocked country that occupies about 0.03% of the total area of the world. It is located between two major economies, China and India. Due to its geographic proximity and open borders, Nepal's financial news headlines are significantly influenced by economic activities in these neighboring countries. The economy of Nepal is highly dependent on agriculture, remittances from its diaspora, and tourism, making these sectors pivotal in shaping economic narratives. Political stability and government policies, particularly those that affect trade and foreign investment, have a substantial impact on financial news. Nepal's dependency on imports, especially from India and China, alongside infrastructural developments and international aid, are also crucial factors. In addition, natural disasters and their economic repercussions, as well as efforts towards economic diversification and sustainability, play a key role in shaping Nepal's financial news landscape (Shrestha and Lamichhane, 2022; Pokhrel et al., 2022).

NEPSE is the sole stock market in Nepal, facilitating the trading of securities and promoting investment opportunities within the country. Established in 1993 under the Securities Exchange Act, NEPSE plays a crucial role in fostering economic growth and development by providing a platform for companies to raise capital and for investors to participate in the financial markets. As the sole stock exchange in Nepal, NEPSE oversees the listing and trading of various financial instruments, including stocks, bonds, and mutual funds, which contribute significantly to the economic infrastructure and market stability of the country (Shrestha and Lamichhane, 2021, 2022).

To analyze financial news for Nepal, we extract headlines from the Share Sansar portal (sharesansar.com) using a web crawler. This process collects 27,773 headlines spanning from Thursday, January 1, 2015, to Wednesday, June 19, 2024. Share Sansar is a prominent online platform in Nepal dedicated to stock market news and analysis, offering a wealth of data including stock prices, market trends, company financials, and news articles. Of the total of 27,773 headlines, we filter using the keyword "NEPSE", resulting in 1,738 relevant headlines.

The United States of America boasts one of the world's largest and most dynamic economies, driven by a diverse range of industries including technology, finance, healthcare, and manufacturing (Gordon, 2002). The country's financial system is highly developed, with Wall Street serving as a global hub for capital markets, investment banking, and asset management (Gordon, 2002). The U.S. dollar is the world's primary reserve currency, reflecting the nation's economic stability and influence. Robust financial regulations, a strong banking sector, and a culture of innovation contribute to the prominent position of the United States in global finance, attracting investors from around the world (Berger, 2013).

The NASDAQ, or National Association of Securities Dealers Automated Quotations, is one of the world's largest and most influential stock exchanges, known for its high concentration of technology and internet-based companies. Founded in 1971, it was the first electronic exchange, allowing for automated trading and providing real-time price quotations. The NASDAQ is home to many of the most significant and innovative companies, including tech giants like Apple, Microsoft, Amazon, and Google. It operates under a dealer market structure, with market makers playing a crucial role in facilitating trades. The NASDAQ's emphasis on technology and innovation has made it a symbol of the modern digital economy and a vital barometer of tech industry performance (nasdaq.com).

To analyze financial news for the USA, we use a web crawler to extract headlines from the Finviz portal (finviz.com). This process gathers 56,198 headlines from Wednesday, January 1, 2020, to Friday, November 5, 2021. Finviz is a leading online platform in the USA that focuses on news and analysis of the stock market, providing extensive data such as stock prices, market trends, company financials, and news articles (Lemieux et al., 2014). From the total of 56,198 headlines, we filter for those containing the keyword "NASDAQ", resulting in 1,882 relevant headlines.

5.2 Data Preparation

For each news headline obtained from the previous Subsection 5.1, we conducted a thorough review from a financial perspective. We then manually categorized the news headlines into three distinct categories: Positive, Negative, and Neutral. For clarity, we refer to the news headlines from sharesansar.com for Nepal as NEPSE news data and those from finviz.com for the USA as NASDAQ news data. A snapshot of the NEPSE news data and the NASDAQ news data with their respective sentiments is provided in Table 4.

The sentiment analysis of financial news headlines for NEPSE and NASDAQ news data reveals distinct distributions. For NEPSE, out of 1,738 headlines, 770 were classified as Positive, 838 as Negative, and 130 as Neutral. In contrast, NASDAQ news data contains 1,882 headlines, with 1,166 categorized as Positive, 378 as Negative, and 388 as Neutral. These results highlight the differences in sentiment distributions, reflecting the contrasting financial environments and news contexts of Nepal and the United States.

Table 4: Glimpse of raw financial news headlines with sentiments collected for NEPAL and USA.

Raw news headlines	Sentiment
NEPSE above 900 level throughout this week; (Review period: Dec 28, 2014 to Jan 1, 2015)	Positive
NEPSE closes at 2 months high of 917.5 level; surges 13.82 points	Positive
NEPSE Trips Down 0.80% While Market Cap Declined to Rs. 33.22 Kharba: Overall Turnover Stood at Rs. 5.47 Arba	Negative
NEPSE Index Slides 0.89%, Barun Hydropower Tops Turnover Charts	Negative
NEPSE to remain closed today on the public holiday of Sonam Losar	Neutral
No early public holiday for Dashain; NEPSE to open till Kartik 02	Neutral
If You Had Bought InterGroup (NASDAQ:INTG) Stock Five Years Ago, You Could Pocket A 109% Gain Today	Positive
ACM Research's (NASDAQ:ACMR) three-year earnings growth trails the fantastic shareholder returns	Positive
China SXT Pharmaceuticals, Inc. Received Nasdaq Notification Letter Regarding Bid Price Deficiency	Negative
Strong week for Aerie Pharmaceuticals (NASDAQ:AERI) shareholders does not alleviate pain of three-year loss	Negative
Calculating The Intrinsic Value Of Autoscope Technologies Corporation (NASDAQ:AATC)	Neutral
What Kind Of Investors Own Most Of Allegiance Bancshares, Inc. (NASDAQ:ABTX)?	Neutral

Before computing sentiment using the sentiment analyzers discussed in Section 3, raw text data should undergo preprocessing steps such as removing stop words, handling punctuation, and performing stemming or lemmatization. Preprocessing raw text data is essential to ensure accurate and meaningful sentiment analysis. Removing stop words eliminates irrelevant terms that do not contribute to the sentiment, while handling punctuation reduces noise that can distort analyzer results. Stemming or lemmatization ensures consistency by converting words to their base or root form, thereby enabling the sentiment analyzers to focus on the core meaning of the text rather than its variations. These steps collectively improve the quality of input data and enhance the performance of sentiment analyzers like VADER and TextBlob. We clean the raw text data (news headlines) using a series of functions from the *nltk* library (Bird, 2006): $split()$, $fix()$, $join()$, $word\_tokenize()$, $lower()$, $string.punctuation()$, $stopwords.words('english')$, $pos\_tag()$, and $lemmatize()$.

Now, the text data is ready (clean) for sentiment analysis. VADER (Hutto and Gilbert, 2014) and TextBlob (Loria, 2024) are the most common and reliable tools for sentiment analysis (Bonta et al., 2019; Aljedaani et al., 2022; Al-Natour and Turetken, 2020; Araci, 2019). These tools have their strengths and weaknesses, and the choice between them depends on factors such as the specific use case, the type of text data being analyzed, and the desired level of accuracy and flexibility. After preprocessing, the data are fed into VADER, and TextBlob to determine their corresponding sentiment.

5.3 Model Comparison

VADER and TextBlob models were tested on NEPSE and NASDAQ news data. This section will focus on the performance of these models for these news data, discussing their strengths and weaknesses.

Table 5 presents the confusion matrix for the VADER and TextBlob sentiment analyzer using NEPSE news data. The VADER model correctly identified 554 out of 770 actual positive instances, 411 out of 838 actual negative instances, and 69 out of 130 actual neutral instances. However, the model misclassified 143 negative instances as positive and 284 negative instances as neutral. Additionally, 195 positive instances were misclassified as neutral. The total counts of predicted sentiments show a higher number of predicted positives (749) and neutrals (548) compared to negatives (441), indicating a potential bias towards positive and neutral classifications. Overall, while the VADER model shows a reasonable performance in identifying positive and negative sentiments, its accuracy for neutral sentiments is relatively lower, suggesting room for improvement in handling neutral classifications. Similarly, the TextBlob model accurately classified 354 out of 770 actual positive instances, 188 out of 838 actual negative instances, and 101 out of 130 actual neutral instances. However, it misclassified a significant number of

Table 5: Confusion matrix for NEPSE news data.

	VADER Predicted				TextBlob Predicted		
Actual	Positive	Negative	Neutral	Total	Positive	Negative	Neutral
Positive	554	21	195	770	354	143	273
Negative	143	411	284	838	202	188	448
Neutral	52	9	69	130	17	12	101
Total	749	441	548	1738	573	343	822

Table 6: Confusion matrix for NASDAQ news data.

Actual	VADER Predicted				TextBlob Predicted		
	Positive	Negative	Neutral	Total	Positive	Negative	Neutral
Positive	881	58	227	1166	493	48	625
Negative	97	209	72	378	27	138	213
Neutral	111	4	223	338	11	5	322
Total	1089	271	522	1882	531	191	1160

instances: 143 positives as negative, 273 positives as neutral, 202 negatives as positive, and 448 negatives as neutral. Additionally, there were 17 neutral instances misclassified as positive and 12 as negative. The total counts of predicted sentiments show a higher number of predicted neutral instances (822) compared to positives (573) and negatives (343), indicating a strong bias towards neutral classifications. Compared to the VADER model, TextBlob shows weaker performance in distinguishing positive and negative sentiments, but it demonstrates a better ability to correctly classify neutral instances. The overall performance suggests that while TextBlob can handle neutral classifications better, it struggles significantly with differentiating between positive and negative sentiments.

Table 6 presents the confusion matrix for the VADER and TextBlob sentiment analyzer using NASDAQ news data. The VADER model accurately classified 881 out of 1,166 actual positive instances, 209 out of 378 actual negative instances, and 223 out of 338 actual neutral instances. However, it misclassified 97 negatives as positive, 72 negatives as neutral, and 227 positives as neutral. Additionally, 111 neutral instances were misclassified as positive and 4 as negative. The total predicted sentiments show a higher number of predicted positives (1,089) compared to negatives (271) and neutrals (522), indicating a bias towards positive classifications. Compared to its performance on NEPSE news data, the VADER model demonstrates stronger performance on NASDAQ data, with higher accuracy in identifying positive and neutral sentiments and relatively fewer misclassifications. However, the misclassification of a substantial number of positive instances as neutral and vice versa highlights the model's difficulty in accurately distinguishing between these two sentiments. Similarly, the TextBlob model accurately classified 493 out of 1,166 actual positive instances, 138 out of 378 actual negative instances, and 322 out of 338 actual neutral instances. However, it misclassified 48 positive instances as negative and 625 positive instances as neutral. Among negative instances, 27 were misclassified as positive and 213 as neutral. Only a small number of neutral instances were misclassified as positive (11) or negative (5). The total predicted sentiments show a pronounced bias towards neutral classifications (1,160), with fewer predicted positives (531) and negatives (191). Compared to its performance on NEPSE news data, TextBlob continues to exhibit a strong tendency to classify sentiments as neutral, reflecting a significant challenge in accurately distinguishing between positive and neutral sentiments. Despite this, TextBlob shows a high accuracy in correctly identifying neutral instances, indicating its strength in handling neutral sentiment but a notable weakness in differentiating positive sentiments from neutral ones.

Table 7 presents the performance metrics of two sentiment analysis models, VADER and TextBlob, applied to both NEPSE and NASDAQ news data. For NEPSE news data, the VADER model outperforms TextBlob across most metrics, with an accuracy of 0.5949 compared to TextBlob's 0.3700. VADER also shows higher sensitivity (0.7195) and specificity (0.4905) rel-

Table 7: Model performance matrices.

Data	Models	Accuracy	Sensitivity	Specificity	Neutral specificity
NEPSE newsdata	VADER	0.5949	0.7195	0.4905	0.5308
NEPSE newsdata	TextBlob	0.3700	0.4597	0.2243	0.7769
NASDAQ newsdata	VADER	**0.6977**	**0.7556**	**0.5529**	0.6598
NASDAQ newsdata	TextBlob	0.5064	0.4228	0.3651	**0.9527**

Table 8: Examples of misclassification by TextBlob: neutral sentiment errors.

Cleaned news headlines	Actual	TextBlob
china sxt pharmaceutical inc receive nasdaq notification letter regard bid price deficiency	Negative	Neutral
abeona therapeutic nasdaq abeo investor sit loss 90 invest three year ago	Negative	Neutral
9.4 return week take aclaris therapeutic nasdaq acrs shareholder one-year gain 510	Positive	Neutral
political influence market turnover touch rs 863 million nepse gain 30 point	Positive	Neutral
political confrontation weaken investor rsquo confidence nepse drop 12.65 point	Negative	Neutral
nepse dwindles double figure market close 1,206.43 point lose 17.96 point	Negative	Neutral
nepse green long time turnover 19 crore	Positive	Neutral

ative to TextBlob's sensitivity of 0.4597 and specificity of 0.2243, indicating that VADER is better at correctly identifying positive sentiments and true negative cases. However, TextBlob surpasses VADER in neutral specificity, scoring 0.7769 against VADER's 0.5308, suggesting that TextBlob is more accurate in recognizing neutral sentiments. Similarly, for NASDAQ news data, VADER demonstrates superior overall performance, achieving an accuracy of 0.6977 compared to TextBlob's 0.5064. VADER excels in sensitivity (0.7556) and specificity (0.5529), indicating a stronger ability to correctly identify both positive sentiments and true negative cases than TextBlob, which has a sensitivity of 0.4228 and specificity of 0.3651. However, TextBlob significantly outperforms VADER in neutral specificity, with a score of 0.9527 versus VADER's 0.6598, suggesting TextBlob's greater accuracy in detecting neutral sentiments. This comparative analysis highlights VADER's overall superior performance in sentiment classification for both news data, while acknowledging TextBlob's strength in recognizing neutrality.

At first glance, Table 7 indicates that VADER outperforms TextBlob in accuracy, sensitivity, and specificity, but not in neutral specificity. However, a closer examination reveals that VADER also performs better than TextBlob in terms of neutral specificity. This is because TextBlob has a higher misclassification rate for neutral sentiments, often misclassifying many positive and negative sentiments as neutral. Table 8 below provides some counterexamples that illustrate this point.

TextBlob's sentiment analyzer struggles to classify the sentiments of the headlines in Table 8 due to its reliance on a simple lexicon-based approach that lacks context awareness. Headlines

Table 9: Examples of misclassification by both VADER and TextBlob for NEPSE cleaned news headlines.

NEPSE cleaned news headlines	Actual	Predicted
know company decrease give best return fy 2071/72 complete analysis nepse last fy	Neutral	Positive
nepse slip 960 level plunge 7.53 point	Negative	Neutral
nepse fall 2 day row	Negative	Neutral
nepse plunge point close 951.29 level	Negative	Neutral
nepse face double digit plunge close 971.47 level	Negative	Neutral
early public holiday dashain nepse open till kartik 02	Neutral	Positive
nepse decline 0.10 point last day week settle 1,386.98 point	Negative	Neutral

often involve financial terminology, subtle market indicators, or emotionally neutral language that requires domain-specific understanding to infer sentiment accurately. For example, phrases like "nasdaq notification letter regard bid price deficiency" or "nepse gain 30 point" carry significant sentiment implications within a financial context, but TextBlob may interpret these phrases literally without understanding their connotations. Additionally, the use of technical terms such as green, dwindles, and slip in the text can confuse the model, causing it to default to 'Neutral' rather than discerning the underlying positive or negative sentiment. This highlights the limitations of lexicon-based sentiment analysis when applied to complex or domain-specific content.

Table 5 highlights that both VADER and TextBlob struggle to accurately predict negative and neutral sentiment in NEPSE news headlines, largely due to the subtle and nuanced language of financial news. Such headlines often blend neutral tones with sentiment-laden keywords, making it difficult for lexicon-based models to accurately interpret the underlying intent or sentiment. Their reliance on prebuilt lexicons leads to misalignment with financial terminology, where words like "plunge" and "decline" strongly signal negativity but are frequently misclassified as neutral. Additionally, these models fail to effectively handle numerical data, such as "plunge 7.53 points" or "settle 1,386.98 points", missing the critical sentiment signals embedded in numerical trends. Domain-specific language, including terms like "slip", "row", and "level", further confuses the models, which struggle to distinguish their financial implications from neutral usage. Moreover, the models overgeneralize positive words like "holiday" or "open", leading to misclassifications of neutral headlines as positive. They also have difficulty detecting subtle negativity, as in "nepse fall 2 day row", where negative sentiment is implied without overtly emotional language. Table 9 illustrates these challenges, emphasizing the need for domain-specific sentiment models trained on financial news tailored to the NESPE.

5.4 Statistical Analysis

To assess the reliability of the models' outcomes, a statistical analysis was performed to identify significant performance differences among them. This analysis followed a hypothesis testing approach for two populations, as recommended in the literature (Wald and Wolfowitz, 1940; Dahal et al., 2020; Agresti and Caffo, 2000). The widely used Wald test, known for its simplicity, was applied to compare population proportions between models. The purpose of this test is to detect significant differences in proportions between groups, with statistical significance determined by a p-value derived from the standard normal distribution. In this study, the effectiveness of the

Table 10: Test statistic and p-values of the Wald test for pairwise comparisons of model accuracy.

Dataset	Hypothesis test	Test statistic	P-value
NEPSE	VADER NEPSE vs TextBlob NEPSE	13.2719	< 0.0001
NASDAQ	VADER NASDAQ vs TextBlob NASDAQ	11.9879	< 0.0001
NEPSE & NASDAQ	VADER NEPSE vs VADER NASDAQ	-6.4669	< 0.0001
NEPSE & NASDAQ	TextBlob NEPSE vs TextBlob NASDAQ	-8.2588	< 0.0001

Wald method depended on having a sufficiently large sample size—1,738 for NEPSE news data and 1,882 for NASDAQ news data—ensuring that the product of the sample size, accuracy percentage, and its complement is at least 10. The null hypothesis (H_0) states that the prediction percentage accuracies of the models are the same, while the alternative hypothesis (H_1) suggests that the prediction percentage accuracies of the models are significantly different. For instance, the first and third hypotheses outlined in Table 10 are as follows: (a) For NEPSE news data, the null hypothesis (H_0) asserts that the prediction percentage accuracies of VADER and TextBlob are the same, whereas the alternative hypothesis (H_1) indicates that their prediction percentage accuracies are significantly different, and (c) For VADER model predictions across NEPSE and NASDAQ news data, the null hypothesis (H_0) posits that the prediction percentage accuracies of VADER on NEPSE and NASDAQ datasets are the same, whereas the alternative hypothesis (H_1) asserts that there is a significant difference in their prediction percentage accuracies. The second (b) and fourth (d) hypothesis are similar to the first (a) and the third (c), respectively, except for the dataset used. The results of multiple pairwise comparisons using the Wald test, along with corresponding test statistics, and the p-values, are presented in Table 10.

The results of the hypothesis tests presented in Table 10, which displays the test statistics and p-values of the Wald test for pairwise comparisons of model percentage accuracy, indicate significant differences across all tested pairs. For NEPSE data, the comparison between the VADER and TextBlob models yielded a p-value of less than 0.0001, suggesting a statistically significant difference in accuracy. Similarly, for NASDAQ data, the VADER vs. TextBlob comparison also produced a p-value of less than 0.0001, indicating a significant difference. When comparing VADER's performance on NEPSE data to its performance on NASDAQ data, the p-value remained less than 0.0001, signifying a significant discrepancy. Likewise, the comparison between TextBlob's performance on NEPSE and NASDAQ data resulted in a p-value of less than 0.0001, underscoring a significant difference. These results collectively reject the null hypothesis, affirming that there are significant differences in model accuracy across the various datasets and model comparisons.

6 Discussion

This study analyzed the comparative performance of VADER and TextBlob sentiment analyzers on financial news headlines from the USA and Nepal. Sentiment analysis plays a critical role in shaping market trends and guiding investor decisions. Identifying the more effective sentiment analyzer is essential for improving accuracy in financial sentiment analysis and supporting informed decision-making in the financial sector.

Our findings demonstrated that VADER would outperform TextBlob in sentiment analysis of financial news headlines, and that both analyzers would perform better on headlines from a well-developed economic country (USA) compared to an underdeveloped economic country (Nepal). VADER demonstrated superior accuracy, sensitivity, and specificity across both datasets, while TextBlob showed strength in classifying neutral sentiments. These results highlight the influence of economic and contextual differences in financial news from well-developed and underdeveloped economic countries and emphasize the importance of selecting the appropriate tool for sentiment analysis based on the context. While machine learning-based methods may provide better performance given sufficient labeled data, this research underscores the practicality and effectiveness of lexical-based methods like VADER and TextBlob, particularly in resource-constrained environments. Moreover, recent developments have shown that smaller Large Language Models (sLLMs) such as BERT and bigger Large Language Models (bLLMs) such as GPT variants dramatically outperform traditional tools on benchmark sentiment analysis tasks (Zhang et al., 2025). However, these gains come with higher computational costs and less interpretability. Tools like VADER and TextBlob remain attractive for real-time applications and in low-resource settings. This trade-off underscores the continued relevance of lexicon-based methods, particularly when explainability and deployment efficiency are key priorities.

Based on our findings, we suggest financial analysts and researchers may adopt VADER for sentiment analysis of financial news headlines, particularly in applications where quick, accurate sentiment assessments are needed. However, future research could explore hybrid models that integrate the interpretability of lexical-based methods with the robustness of machine learning approaches to further enhance sentiment analysis accuracy. Domain-specific adaptations, real-time sentiment analysis systems, and the inclusion of financial news from diverse economic and linguistic contexts would provide more comprehensive insights into the performance of sentiment analysis models.

In conclusion, this research contributes to the field by offering a detailed comparative analysis of VADER and TextBlob and providing actionable recommendations for their application in financial sentiment analysis. By addressing the challenges identified in this study, future research can advance the accuracy, scalability, and applicability of sentiment analysis techniques, thereby enhancing their utility in financial decision-making.

Supplementary Material

Python codes as well as datasets used in the study are available in a supplementary file.

Acknowledgement

The authors would like to express their gratitude to ShareSansar for providing online access to financial news headlines for the Nepal Stock Exchange (NEPSE) via their website. Additionally, the authors would like to thank Finviz for making financial news headlines for the National Association of Securities Dealers Automated Quotations (NASDAQ) available online through their platform.

Funding

This research received no external funding.

References

Abiola O, Abayomi-Alli A, Tale OA, Misra S, Abayomi-Alli O (2023). Sentiment analysis of COVID-19 tweets from selected hashtags in Nigeria using VADER and TextBlob analyser. *Journal of Electrical Systems and Information Technology*, 10(1): 5. https://doi.org/10.1186/s43067-023-00070-9

Agbehadji IE, Ijabadeniyi A (2021). Approach to sentiment analysis and business communication on social media. In: Simon James Fong, and Richard C. Millham, editors, *Bio-inspired Algorithms for Data Streaming and Visualization, Big Data Management, and Fog Computing* 169–193.

Agresti A, Caffo B (2000). Simple and effective confidence intervals for proportions and differences of proportions result from adding two successes and two failures. *American Statistician*, 54(4): 280–288. https://doi.org/10.1080/00031305.2000.10474560

Al-Natour S, Turetken O (2020). A comparative assessment of sentiment analysis and star ratings for consumer reviews. *International Journal of Information Management*, 54: 102132. https://doi.org/10.1016/j.ijinfomgt.2020.102132

Al-Qablan TA, Mohd Noor MH, Al-Betar MA, Khader AT (2023). A survey on sentiment analysis and its applications. *Neural Computing & Applications*, 35(29): 21567–21601. https://doi.org/10.1007/s00521-023-08941-y

Al-Shabi M (2020). Evaluating the performance of the most important lexicons used to sentiment analysis and opinions mining. *International Journal of Computer Science and Network Security*, 20(1): 1.

Aljedaani W, Rustam F, Mkaouer MW, Ghallab A, Rupapara V, Washington PB, et al. (2022). Sentiment analysis on twitter data integrating TextBlob and deep learning models: The case of US airline industry. *Knowledge-Based Systems*, 255: 109780. https://doi.org/10.1016/j.knosys.2022.109780

Araci D (2019). Finbert: Financial sentiment analysis with pre-trained language models. arXiv preprint: https://arxiv.org/abs/1908.10063.

Asderis GA (2022). Sentiment analysis on twitter data, a detailed comparison of TextBlob and VADER.

Berger S (2013). *Making in America: From Innovation to Market*. MIT Press.

Bharadwaj L (2023). Sentiment analysis in online product reviews: Mining customer opinions for sentiment classification. *International Journal for Multidisciplinary Research*, 5(5): 1–34.

Bird S (2006). NLTK: The natural language toolkit. In: *Proceedings of the COLING/ACL 2006 Interactive Presentation Sessions*, 69–72.

Bonta V, Kumaresh N, Janardhan N (2019). A comprehensive study on lexicon based approaches for sentiment analysis. *Asian Journal of Computer Science and Technology*, 8(S2): 1–6. https://doi.org/10.51983/ajcst-2019.8.S2.2037

Ccoya W, Pinto E (2023). Comparative analysis of libraries for the sentimental analysis. arXiv preprint https://arxiv.org/abs/2307.14311.

Dahal KR, Amezziane M, et al. (2020). Exact distribution of difference of two sample proportions and its inferences. *Open Journal of Statistics*, 10(03): 363. https://doi.org/10.4236/ojs.2020.103024

Dahal KR, Gupta A, Pokhrel NR (2024). Predicting the direction of NEPSE index movement with news headlines using machine learning. *Econometrics*, 12(2): 16. https://doi.org/10.3390/econometrics12020016

Dahal KR, Pokhrel NR, Gaire S, Mahatara S, Joshi RP, Gupta A, et al. (2023). A comparative study on effect of news sentiment on stock price prediction with deep learning architecture. *PLoS ONE*, 18(4): e0284695. https://doi.org/10.1371/journal.pone.0284695

Das N, Gupta S, Das S, Yadav S, Subramanian T, Sarkar N (2021). A comparative study of sentiment analysis tools. In: *2021 International Conference on Innovative Computing, Intelligent Communication and Smart Electrical Systems (ICSES)*, 1–7. IEEE.

De Smedt T, Daelemans W (2012). Pattern for Python. *Journal of Machine Learning Research*, 13(1): 2063–2067.

Ekaputri AP, Akbar S (2022). Financial news sentiment analysis using modified VADER for stock price prediction. In: *2022 9th International Conference on Advanced Informatics: Concepts, Theory and Applications (ICAICTA)*, 1–6. IEEE.

Elbagir S, Yang J (2020). Sentiment analysis on Twitter with Python's natural language toolkit and VADER sentiment analyzer. In: *IAENG Transactions on Engineering Sciences: Special Issue for the International Association of Engineers Conferences 2019*, 63–80. World Scientific.

Gonçalves P, Araújo M, Benevenuto F, Cha M (2013). Comparing and combining sentiment analysis methods. In: *Proceedings of the First ACM Conference on Online Social Networks*, 27–38.

Gordon RJ (2002). Technology and economic performance in the American economy. *Working Paper 8771*, National Bureau of Economic Research.

Hameed Z, Garcia-Zapirain B, Aguirre JJ, Isaza-Ruget MA (2022). Multiclass classification of breast cancer histopathology images using multilevel features of deep convolutional neural network. *Scientific Reports*, 12(1): 15600. https://doi.org/10.1038/s41598-022-19278-2

Hutto C (2020). *VADER-Sentiment-Analysis*, GitHub. Available at: https://github.com/cjhutto/vaderSentiment [Accessed: 2-Jul-2024].

Hutto C, Gilbert E (2014). VADER: A parsimonious rule-based model for sentiment analysis of social media text. In: *Proceedings of the International AAAI Conference on Web and Social Media*, volume 8, 216–225.

Illia F, Eugenia MP, Rutba SA (2021). Sentiment analysis on pedulilindungi application using TextBlob and VADER library. In: *Proceedings of The International Conference on Data Science and Official Statistics*, volume 2021, 278–288.

Joshi RC, Singh D, Tiwari V, Dutta MK (2022). An efficient deep neural network based abnormality detection and multi-class breast tumor classification. *Multimedia Tools and Applications*, 81(10): 13691–13711. https://doi.org/10.1007/s11042-021-11240-0

Khoo CS, Johnkhan SB (2018). Lexicon-based sentiment analysis: Comparative evaluation of six sentiment lexicons. *Journal of Information Science*, 44(4): 491–511. https://doi.org/10.1177/0165551517703514

Kolbitsch J, Maurer HA (2006). The transformation of the web: How emerging communities shape the information we consume. *Journal of Universal Computer Science*, 12(2): 187–213.

Koukaras P, Nousi C, Tjortjis C (2022). Stock market prediction using microblogging sentiment analysis and machine learning. In: *Telecom*, volume 3, 358–378. MDPI.

Lemieux VL, Fisher B, Dang T (2014). The visual analysis of financial data. In: Margarita S. Brose, Mark D. Flood, Dilip Krishna, and Bill Nichols, editors, *The Handbook of Financial Data and Risk Information II* 279–326.

Loria S (2024). *TextBlob: Simplified Text Processing*. Available at: https://textblob.readthedocs.io/en/dev/quickstart.html#sentiment-analysis [Accessed: 28-Jun-2024].

Maqbool J, Aggarwal P, Kaur R, Mittal A, Ganaie IA (2023). Stock prediction by integrating sentiment scores of financial news and MLP-regressor: A machine learning approach. *Procedia*

Computer Science, 218: 1067–1078. https://doi.org/10.1016/j.procs.2023.01.086

Min WNSW, Zulkarnain NZ, et al. (2020). Comparative evaluation of lexicons in performing sentiment analysis. *Journal of Advanced Computing Technology and Application*, 2(1): 1–8.

Mujahid M, Rustam F, Shafique R, Chunduri V, Villar MG, Ballester JB, et al. (2023). Analyzing sentiments regarding ChatGPT using novel BERT: A machine learning approach. *Information*, 14(9): 474. https://doi.org/10.3390/info14090474

Musto C, Semeraro G, Polignano M, et al. (2014). A comparison of lexicon-based approaches for sentiment analysis of microblog posts. In: *DART@AI*IA*, 59–68. Citeseer.

Nemes L, Kiss A (2021). Prediction of stock values changes using sentiment analysis of stock news headlines. *Journal of Information and Telecommunication*, 5(3): 375–394. https://doi.org/10.1080/24751839.2021.1874252

Padmaja S, Fatima SS, Bandu S (2014). Evaluating sentiment analysis methods and identifying scope of negation in newspaper articles. *International Journal of Advanced Research in Artificial Intelligence*, 3(11): 1–6. https://doi.org/10.14569/IJARAI.2014.031101

Pano T, Kashef R (2020). A complete VADER-based sentiment analysis of bitcoin (BTC) tweets during the era of COVID-19. *Big Data and Cognitive Computing*, 4(4): 33. https://doi.org/10.3390/bdcc4040033

Pennebaker JW, Francis ME, Booth RJ (2001). Linguistic inquiry and word count: LIWC 2001. *Mahway: Lawrence Erlbaum Associates*, 71(2001): 2001.

Pokhrel NR, Dahal KR, Rimal R, Bhandari HN, Khatri RK, Rimal B, et al. (2022). Predicting NEPSE index price using deep learning models. *Machine Learning with Applications*, 9: 100385. https://doi.org/10.1016/j.mlwa.2022.100385

Pokhrel NR, Dahal KR, Rimal R, Bhandari HN, Rimal B (2024). Deep-SDM: A unified computational framework for sequential data modeling using deep learning models. *Software*, 3(1): 47–61. https://doi.org/10.3390/software3010003

Saha S, Showrov MIH, Rahman MM, Majumder MZH (2022). VADER vs. BERT: A comparative performance analysis for sentiment on coronavirus outbreak. In: *International Conference on Machine Intelligence and Emerging Technologies*, 371–385. Springer.

Sanyal S, Barai MK (2021). Comparative study on lexicon-based sentiment analysers over negative sentiment. *International Journal of Electrical, Electronics and Computers*, 6(6): 1–13. https://doi.org/10.22161/ijeec.66.1

Shamrat FJM, Azam S, Karim A, Islam R, Tasnim Z, Ghosh P, et al. (2022). Lungnet22: A fine-tuned model for multiclass classification and prediction of lung disease using X-ray images. *Journal of Personalized Medicine*, 12(5): 680. https://doi.org/10.3390/jpm12050680

Shayaa S, Jaafar NI, Bahri S, Sulaiman A, Wai PS, Chung YW, et al. (2018). Sentiment analysis of big data: Methods, applications, and open challenges. *IEEE Access*, 6: 37807–37827. https://doi.org/10.1109/ACCESS.2018.2851311

Shrestha PM, Lamichhane P (2021). Macroeconomic factors and stock market performance in Nepal. *PYC Nepal Journal of Management*, 14(1): 79–92. https://doi.org/10.3126/pycnjm.v14i1.41061

Shrestha PM, Lamichhane P (2022). Effect of firm-specific variables on stock returns: Evidence from Nepal. *Quest Journal of Management and Social Sciences*, 4(2): 249–259. https://doi.org/10.3126/qjmss.v4i2.50320

Singh AK, Verma A (2021). An efficient method for aspect based sentiment analysis using SpaCy and VADER. In: *2021 10th IEEE International Conference on Communication Systems and Network Technologies (CSNT)*, 130–135. IEEE.

Srivastava R, Bharti P, Verma P (2022). Comparative analysis of lexicon and machine learning approach for sentiment analysis. *International Journal of Advanced Computer Science and Applications*, 13(3): 71–77.

Talpada H, Halgamuge MN, Vinh NTQ (2019). An analysis on use of deep learning and lexical-semantic based sentiment analysis method on twitter data to understand the demographic trend of telemedicine. In: *2019 11th International Conference on Knowledge and Systems Engineering (KSE)*, 1–9. IEEE.

United Nations (2025). Least developed country category: Nepal. Available at: https://www.un.org/development/desa/dpad/least-developed-country-category-nepal.html [Accessed: 12-Jan-2025].

van Ooijen C, Ubaldi B, Welby B (2019). A data-driven public sector: Enabling the strategic use of data for productive, inclusive and trustworthy governance. *Technical Report 33*, OECD Publishing, Paris.

Wald A, Wolfowitz J (1940). On a test whether two samples are from the same population. *The Annals of Mathematical Statistics*, 11(2): 147–162. https://doi.org/10.1214/aoms/1177731909

Zhang T, Irsan IC, Thung F, Lo D (2025). Revisiting sentiment analysis for software engineering in the era of large language models. *ACM Transactions on Software Engineering and Methodology*, 34(3): 1–30.

Zhang W, Deng Y, Liu B, Pan SJ, Bing L (2023). Sentiment analysis in the era of large language models: A reality check. arXiv preprint https://arxiv.org/abs/2305.15005.