Journal of Research in Economics and Finance Management
ISSN (Print): 2959-099X ISSN (Online): 2959-1007
Volume 3, Issue 2, Page 89-104, December 31, 2024

TEXTUAL SENTIMENT TRENDS IN PAKISTAN: BUILDING AN INDEX USING NEWS REPORTS AND SOCIAL MEDIA DISCOURSE

Mr. Muhammad Waqas
PhD Scholar
Department of Management Sciences
DHA Suffa University Karachi
muhammad.waqas@dsu.edu.pk

Prof Dr. Bashir Ahmad
Adjunct Professor
Department of Management Sciences
DHA Suffa University Karachi
dr.bashirahmad@dsu.edu.pk

Dr. Abdur Rahman Aleemi
Associate Professor
Institute of Business and Health Management
Dow University Karachi
abdur.rahman.9794@duhs.edu.pk

Abstract

This research develops the Text-Based Sentiment Index (TBSI) to measure public sentiment through the evaluation of English-language newspaper content alongside social media posts. An analysis of media sentiment specifically focuses on utilizing textual data from Dawn, The News International, Express Tribune, and Twitter posts during the period from 2012 to 2024. The combination of the VADER sentiment tool and machine learning techniques performed sentiment classification in this study. The research demonstrated that extensive public emotional responses occurred during pivotal national events when major elections took place, as well as during periods of leadership changes and economic transformations, alongside the COVID-19 pandemic outbreak. The research data supports established propositions, which confirm that public conduct reveals itself through media sentiment while media sentiment predicts future public behavior. Researchers developed an index from monthly sentiment scores, which served as a tool for policy determinations that proved helpful to political decision-makers as well as academic and institutional groups. Sentiment analysis requires extension beyond English-language content since resource constraints prevented processing Urdu and Roman Urdu texts.

Keywords: Text-Based Sentiment Index, Natural Language Processing (NLP), Machine Learning, Public Opinion, Trend Analysis.

Introduction

Information moves swiftly through multiple digital channels, which powerfully affects how the public views things and what both financial operators and lawmakers decide and make. Two primary platforms that control societal attitudes are news media and social media among all available information channels. Understanding public sentiment by reading news and social media content becomes more crucial every day for Pakistan, given its unpredictable political, economic, and social conditions. A

This work is licensed under a Creative Commons Attribution 4.0 International License.

sentiment index derived from text analysis of major newspapers and social media platforms enables nationwide mood assessment through a tool that enhances decision-making for political and economic sectors and research operations.

Sentiment analysis has now extended its applications through finance (Tetlock, 2007), marketing (Pang & Lee, 2008), and political areas (Tumasjan et al., 2010) during the past ten years. Traditional survey methods, together with official indicators, persist as the main tools to measure public opinion trends, although they frequently fall behind current social developments in Pakistan. The editorial content and news reporting at the national newspapers Dawn, The News and Express Tribune maintain their power to shape public opinion in Pakistan. Sentiment analysis of textual resources creates a connection between public emotions and institutional reactions, which improves both governing structures and market assessment processes.

The research investigates the possibility of building a Sentiment Index through Text-Based analysis of Pakistani news publications combined with social media content. It also investigates sentiment detection via machine learning algorithms and NLP methods to analyze sentiment, which produces positive, negative, and neutral tags, and calculates temporal sentiment scores. Such an index, when created, will improve Pakistan's socio-political understanding, which will help stakeholders predict trends better.

Most existing research about sentiment indices focuses on Western backgrounds, including the University of Michigan Consumer Sentiment Index and the Baker, Bloom, and Davis Economic Policy Uncertainty Index. Pakistan's special social and political structure, together with its diverse requirements, demands a locally customized approach.

Public sentiment enables accurate evaluation of economic conditions, along with helping to control national policy directions. The Consumer Confidence Index (CCI) and Business Confidence Index (BCI) provide Pakistan with essential instruments to measure consumer sentiments and business outlook. Public confidence levels are influenced by economic conditions, together with policy changes and external factors, which are tracked through these indices and determine multiple economic factors that include spending levels, as well as investment decisions and savings behavior.

The research establishes empirical proof that supports extracting a sentiment index from textual data for Pakistani use. The tool demonstrates potential to operate as both an alert system and an evaluation aid for decision-makers at every level of civil society, including governmental agencies and asset evaluation institutions. The following section of the study will discuss the official sentiment indices that have been published and their importance for policymakers in the specific area of the economy. Based on the discussion, the study concluded that there should be a text (news) based sentiment index to provide a baseline for the policy makers, practitioners, and regulators of the financial sector.

Digital transformation requires public sentiment examination through textual data because it serves the essential needs of policymakers, alongside businesses and researchers. Developing a text-based sentiment index in Pakistan is essential because

ISSN (Print): 2959-099X ISSN (Online): 2959-1007
Volume 3, Issue 2, Page 89-104, December 31, 2024

Urdu is the official language, and many citizens use Roman Urdu for online communication. The paper begins by examining the sentiment analysis scene in Pakistan while discussing the methods, along with system limitations and breakthroughs that affect Urdu textual data sentiment evaluation. Sentiment analysis has become essential to monitor public sentiments because of the digital age's rise in regions with multiple languages, such as Pakistan. Social media platforms currently generate excessive user-created content that reveals public opinions about multiple issues. The process of extracting subjective content from text documents stands as sentiment analysis, together with opinion mining, because they both identify expressed sentiment in textual information. Social media platforms Twitter and Facebook function as main public discourse channels within Pakistan, which makes them suitable for sentiment analysis examinations.

Pakistan's linguistic makeup consists of two main components: Urdu, which serves as the official language, and Roman Urdu, which represents written Urdu through the Roman alphabet. Sentiment analysis becomes more intricate because of inconsistent spellings, grammatical rules, and syntactical patterns between Urdu and Roman Urdu. The problems associated with Roman Urdu spelling standards become worse due to the absence of standardized orthographic rules (Liaqat et al., 2022).

Limitations of Sentiment Analysis

The structure of Urdu acts as a barrier because it contains numerous linguistic complexity features but lacks sufficient linguistic resources. Sentiment analysis of Urdu faces three major obstacles: insufficient documentation of big, annotated data libraries and scarcity of sentiment dictionaries, while also facing complex grammatical structures within the Urdu language. The analysis becomes harder because social media content written using Roman Urdu, which is Urdu in Latin script, contains inconsistent spelling and grammatical errors (Khattak et al., 2021). The shortage of resources in the Urdu language stands as a major challenge because it lacks sufficient annotated corpora as well as sentiment lexicons and pre-trained models. The inadequate volume of resources stops researchers from creating dependable sentiment analysis tools (Liaqat et al., 2022). Based on the limitations of the Urdu language, the current study does not focus on Urdu newspapers of social media posts published in the Urdu language or Roman Urdu.

Rationale of the Study

Public sentiment assessment through traditional surveys and opinion polls is insufficient for understanding current societal changes as Pakistan transforms in its political and economic areas, as well as its social framework. Public emotions play a significant role in governance operations and both policy development and financial decision-making, according to Haroon et al. (2021) and Afzaal (2020). Digital communication has created an opportunity to access textual data on news media and social media platforms, which provides real-time public sentiment evaluations. A customized approach becomes necessary since Urdu, along with Roman Urdu, presents special linguistic difficulties (Liaqat et al., 2022). Constructing a Text-Based Sentiment Index (TBSI) that suits Pakistan's cultural environment fills an important research void, so policymakers, financial sectors, and researchers now have an immediate empirical instrument to measure public feeling patterns.

Problem Statement

The rise in accessible news and social media textual data has not led Pakistan to develop a complete sentiment analysis system that provides real-time public sentiment analysis across all text content. Current sentiment indices are primarily focused on Western markets, according to Tetlock (2007) and Shiller (2015), but they neglect the unique linguistic and socio-political characteristics of Pakistan, together with its diverse media environment (Khattak et al., 2021; Kausar et al., 2021). Sentiment extraction operations encounter more difficulties because of the spelling inconsistencies found in the Roman Urdu language (Liaqat et al., 2022). Real-time public opinion data does not reach decision-making authorities in Pakistan, which hinders their ability to conduct better forecasting or planning processes.

Research Question

1. How can a **Text-Based Sentiment Index** be developed using Pakistani news and social media data to accurately reflect public sentiment trends in political, economic, and social domains?

Research Objectives

- **To develop** a **Text-Based Sentiment Index (TBSI)** using textual data collected from leading Pakistani English-language newspapers and social media platforms.
- **To apply** natural language processing (NLP) and machine learning techniques for analyzing the sentiment (positive, negative, neutral) of news articles and social media posts.
- **To examine** temporal trends in public sentiment and correlate them with major political, economic, and social events in Pakistan.
- **To identify** the challenges associated with sentiment analysis in a multilingual setting like Pakistan, particularly with Urdu and Roman Urdu data sources.

Theoretical Framework

According to Sentiment Theory, public sentiment, which manifests itself through communication channels, determines both group activities and strategic planning procedures. Mass media channels influence the way people feel about investors, consumers, and the political landscape through their emotional communication approach (Tetlock, 2007; Shiller, 2015). Financial market indicators derived from text-based analysis of newspapers (such as those presented in Garcia, 2013) prove effective at market prediction, particularly during times of economic turbulence.

Conceptual Framework

The **Conceptual Framework** for this study integrates the relationship between **Media Content**, **Sentiment Analysis**, **Public Sentiment Trends**, and **National Events**.

ISSN (Print): 2959-099X ISSN (Online): 2959-1007
Volume 3, Issue 2, Page 89-104, December 31, 2024

- **Media Content**: Textual data from prominent Pakistani newspapers (e.g., *Dawn, The News International, Express Tribune*) and social media platforms (e.g., Twitter) provides the raw material for analysis (Haroon et al., 2021; Mughal et al., 2024).

- **Sentiment Analysis**: Natural Language Processing (NLP) tools such as VADER (Hutto & Gilbert, 2014) and machine learning classifiers process media content to categorize sentiments as positive, negative, or neutral.

- **Public Sentiment Trends**: Aggregated sentiment scores reveal monthly or quarterly public mood trends, reflecting changes in national optimism, pessimism, or neutrality (Diego García, 2013; Liaqat et al., 2022).

- **National Events**: Major political developments (e.g., elections, court verdicts, leadership changes), economic fluctuations, and social crises (e.g., COVID-19 pandemic) are expected to correlate with significant shifts in the sentiment index (Afzaal, 2020; Pak Institute for Peace Studies, 2021).

Visual Model (Narrative Representation)

Media Content → Sentiment Analysis Process → Sentiment Scores → Public Sentiment Trends ↔ National Political/Economic/Social Events

Significance of the Study

Since Pakistan is a developing nation with distinct linguistic and media dynamics, this study fills a vital gap in localized sentiment analysis. By creating a sentiment index tailored for Pakistan, the study provides policymakers, financial analysts, and academics with a useful real-time tool to gauge public opinion and make well-informed judgments (Baqir et al., 2023; Mughal et al., 2024). Additionally, the work contributes to larger international conversations by extending computational social science research into non-Western contexts.

Literature Review

Text-based sentiment analysis quantitatively assesses public emotions toward national and international happenings and political and national sentiments. Research by Aslam et al. (2023) developed a hybrid neural network scheme that utilizes CNN and GRU to effectively analyze sentiment in movie and product review datasets.

Mughal et al. (2024) devoted their research to optimizing social media sentiment detection by specifically addressing data accuracy needs. The authors emphasized native language barriers, together with training algorithm adaptations to fit Twitter multilingual data within the South Asian platform environment. (Noor et al., 2024) retrieved sentiments from news headlines to monitor Pakistani government performance changes through time in their focused research about political contexts.

Nagra et al. (2024) conducted a machine learning analysis of Pakistani political sentiment on Twitter. According to their research, random forest classifiers provided

the best results. Real-time election monitoring achieves effective functionality according to the research findings.

For the analysis of the 2018 Pakistani general elections, Khan et al. (2018) examined Twitter data to retrieve opinions and emotional expressions from it. Information revealed that digital platforms now guide political outcomes in society because they harbor significant polarization, and emotional language represents a big issue today. Qayyum et al. (2018) researched media bias alongside toxic behavior in South Asian political discussions. The researchers proved that major political biases have driven growing hostility on social media, depending on what media platform users chose and what political party they aligned with.

Society needs to focus on the critical issue of spreading untrue information. The researchers behind (Haroon et al., 2021) examined how false information spread across Pakistan throughout the COVID-19 pandemic. The researchers from Public Health, together with institutional trust configurations, detected substantial shifts because of misinformation while analyzing sentiment manipulation during emergencies. Kausar et al. (2021) created "Manipify," which serves as a new system for identifying artificial Twitter trend manipulation within Pakistan. Media monitoring requires heightened vigilance because specific hashtags receive systematic boosters to direct public opinion according to their findings.

Baqir et al. (2023) carried out extensive research on how social media both reflects and intensifies ongoing political polarization. The authors studied Pakistan's political arena to show how digital connections now duplicate offline social divisions. Social media misuse has far-reaching social consequences in Pakistani politics, according to Kumari et al. (2025), through the creation of disinformation, the intensification of political divisions, and the rise of cyber harassment incidents.

Historically speaking, speculative bubbles may be traced back to the time when newspapers first appeared. Even though most of the archives of the earliest newspapers have been destroyed, it is safe to assume that they covered the first real bubble, the tulip craze in the Netherlands in the 1630s, as reported by Shiller (2015).

The news media—including newspapers, magazines, broadcast media, and their new online outlets- portray themselves as neutral witnesses to market events. However, they play a significant role in shaping those events. The news media play a crucial role in the dissemination of ideas, and significant market events often only occur when there is consensus among huge groups of people. Shiller (2015) considers how complicated it is for the media to influence market events. We shall discover that the market's response to news events is rarely clean and simple. They do not have as much of an effect as is believed. A closer look, however, indicates that the media do play a crucial role in both setting the stage and inciting market actions.

The news media are naturally attracted to financial markets because they provide constant news in the form of daily price changes. Financial news may have great human-interest potential to the extent that it deals with the making or breaking of fortunes. The public considers it the "big casino," the market for major players, and believes that on any given day, it serves as a barometer of the status of the nation. Financial news and sports account for half of the editorial content of many newspapers

ISSN (Print): 2959-099X ISSN (Online): 2959-1007
Volume 3, Issue 2, Page 89-104, December 31, 2024

today. For a time, Shiller (2009) reported that the housing bubble sparked interest, with newspapers dedicating entire sections to homes or real estate. In the U.S., the housing boom led to the creation of an entire TV channel devoted to housing: Home and Garden Television (HGTV).

Diego Garcia (2013) creates a sentiment index based on financial news from the New York Times from 1905 to 2005 and investigates its association with stock returns. The lengthy time series enables them to examine a period with a concentrated media sector and substantial business cycle unpredictability. The study's main finding is that news content helps predict stock returns daily, especially during recessions: a one standard deviation change in our pessimism factor moves the Dow Jones Industrial Average (DJIA) by 12 basis points during recessions, but only 3.5 basis points during expansions. According to their findings, this asymmetric prediction is not due to changes in reporting along the business cycle. The effect is especially pronounced on Mondays, and it remains until the afternoon trading session. Additionally, they discover that the effect partially reverses during the subsequent four trading days. The authors conclude that their findings support the premise that investor sentiment has a significant impact during economic downturns.

Tetlock (2007) investigates the relationships between media content and stock market behavior systematically. The study establishes a simple measure of media content that appears to connect to either negative investor sentiment or risk aversion. Prominent levels of media pessimism led to a decline in market prices, whereas abnormally high or low levels of pessimism resulted in a temporary increase in market trading activity. Moreover, the price impact of pessimism on small stocks is disproportionately substantial and difficult to reverse. This is consistent with sentiment theories based on the premise that media content influences the behavior of individual investors, who hold a disproportionate share of tiny stocks.

The current study made a proxy for market sentiment by counting the number of positive and negative words in daily financial columns from Daily Dawn and other popular newspapers and social media platforms. These columns covered general financial news like the performance of the stock market, news about industry, and macroeconomic events. However, while collecting the positive and negative words from the financial columns of newspapers, the study needs to consider the standards of positive and negative words set by Tetlock (2007).

Further, their study, Tetlock (2007), concluded that it is encouraging in two respects that multiple indicators of negative sentiment, namely the pessimism factor, negative words, and weak words, have the same link to future market action. First, because psychologists created the raw General Inquirer (GI) word categories, they have natural interpretations as negative emotion measurements. Second, reporting tests based on various sentiment metrics reduces the likelihood of data mining.

Methodologies for Sentiment Analysis

This study developed a text-based sentiment index (TBSI) to measure sentiments in Pakistani news media publications and social media platforms.

Data Collection

The study gathered its data from two fundamental resources: Pakistani English-language newspapers (Dawn, The News International, Express Tribune) and the Twitter social media platform. The research utilized scraping tools based on Python to extract newspaper articles discussing political, economic, and social news from 2018 to 2024 using Beautiful Soup and Scrapy. Research data collection utilized Twitter's Application Programming Interface to retrieve social media content related to political events, economic developments, and social challenges (Gaur et al., 2021).

Data Preprocessing

The gathered text information required a systematic preprocessing approach to enhance accuracy in sentiment analysis results. Preprocessing steps included:

The data cleaning procedure removes HTML tags, together with URLs and punctuation marks, and eliminates stopping words.

The data preprocessing step requires essential cleaning of text content to remove unnecessary elements. These steps clarify the textual language by eliminating HTML tags and URLs while deleting punctuation and stopping words. The emphasis on critical sentiment-bearing words during this phase leads to better sentiment analysis accuracy.

Lowercasing all text for uniformity.

The technique of base form processing with tokenization and lemmatization reduced the text content. Systematic steps eliminated irrelevant materials such as advertisements, together with non-textual aspects, from the analysis (Jain et al., 2020).

The researchers took great care to remove biases from repeated media stories and duplicated social media posts (Ahmed & Traore, 2022).

Sentiment Analysis

The scoring function of the system combined sentiment analysis methods that used lexica alongside machine learning models. The main sentiment analysis program employed VADER (Valence Aware Dictionary and Sentiment Reasoner) because these tools deliver exceptional functionality for social media text evaluation (Hutto & Gilbert, 2014).

The validation process involved a combination of Pakistani news headlines labeled manually in perplexing language by a logistic regression classifier to adjust sentiment results specifically for regional language differences.

The scoring system categorized sentiment results into three distinct levels.

- Positive sentiment (score > 0.05),
- Negative sentiment (score < -0.05),

ISSN (Print): 2959-099X ISSN (Online): 2959-1007
Volume 3, Issue 2, Page 89-104, December 31, 2024

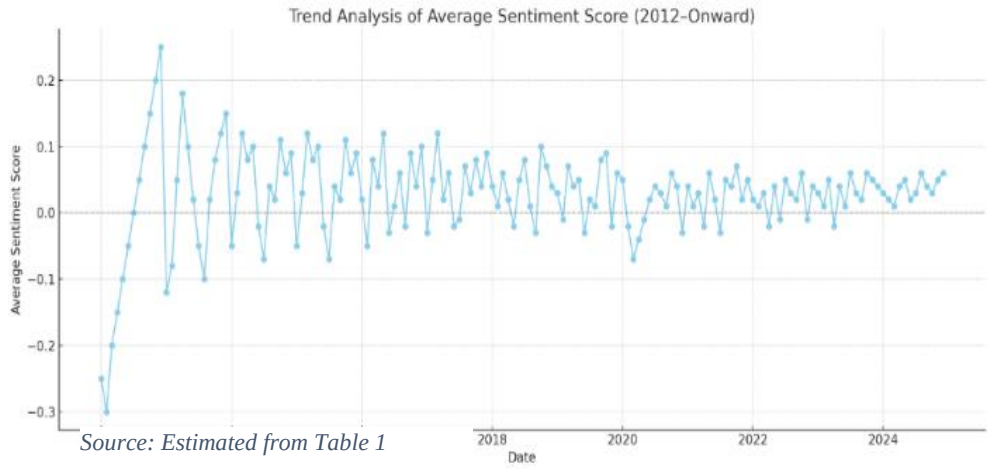
- Neutral sentiment (0.05< Score > 0.05) (Hutto & Gilbert 2014).

Table 01: Average Sentiment Score

Year	Month	Average Sentiment Score	Year	Month	Average Sentiment Score	Year	Month	Average Sentiment Score
2012	January	-0.4733	2013	January	-0.5347	2014	January	-0.5389
	February	-0.3448		February	-0.4126		February	-0.3869
	March	-0.0876		March	0.0235		March	-0.2160
	April	-0.2374		April	-0.1917		April	-0.0192
	May	-0.3710		May	-0.1603		May	-0.5249
	June	-0.0864		June	-0.6347		June	-0.2483
	July	-0.2072		July	-0.1554		July	-0.2425
	August	-0.5466		August	-0.0224		August	-0.1563
	September	-0.1345		September	-0.4270		September	-0.4361
	October	-0.1847		October	-0.5611		October	-0.2393
	November	-0.0840		November	-0.3286		November	-0.0699
	December	-0.2581		December	-0.2054		December	-0.1367
2015	January	-0.2179	2016	January	-0.3482	2017	January	0.0033
	February	-0.3453		February	-0.1795		February	-0.2061
	March	0.2269		March	-0.1406		March	-0.3373
	April	-0.3833		April	-0.1415		April	0.1870
	May	-0.4053		May	-0.2485		May	-0.6335
	June	-0.5031		June	-0.1861		June	-0.5150
	July	-0.3971		July	-0.0099		July	-0.1918
	August	-0.1622		August	-0.2358		August	-0.5698
	September	-0.6499		September	-0.1282		September	0.0523
	October	-0.0183		October	-0.3451		October	-0.2269
	November	-0.1215		November	-0.3838		November	-0.2025
	December	-0.2813		December	-0.2547		December	-0.3007
2018	January	-0.3376	2019	January	-0.3390	2020	January	-0.5670
	February	-0.3048		February	-0.5306		February	-0.5466
	March	-0.2539		March	-0.4166		March	-0.3484
	April	-0.5118		April	-0.2987		April	-0.1551
	May	-0.4712		May	-0.1453		May	0.0065
	June	-0.3570		June	-0.4036		June	-0.1682
	July	-0.2394		July	-0.4787		July	-0.4291
	August	-0.2734		August	-0.0497		August	-0.2290
	September	0.2239		September	-0.6050		September	-0.5192
	October	-0.1356		October	-0.2336		October	-0.2398
	November	-0.4130		November	-0.3146		November	-0.0805
	December	0.0865		December	-0.0107		December	-0.3028

2021	January	-0.0348	2022	January	-0.3181	2023	January	-0.0872
	February	-0.3047		February	-0.2167		February	-0.2588
	March	-0.4074		March	-0.5768		March	-0.5320
	April	-0.3935		April	-0.1997		April	-0.2883
	May	-0.4864		May	-0.1742		May	-0.5208
	June	-0.4730		June	-0.3732		June	-0.3544
	July	-0.3918		July	0.1653		July	-0.5873
	August	-0.4953		August	-0.3660		August	-0.4517
	September	0.0784		September	-0.2150		September	-0.4650
	October	-0.1857		October	-0.1244		October	-0.2608
	November	0.0122		November	-0.5148		November	-0.0175
	December	-0.2087		December	-0.0841		December	-0.3784
2024	January	-0.3256						
	February	-0.2103						
	March	-0.1422						
	April	-0.2575						
	May	-0.9471						
	June	-0.1341						
	July	-0.1504						
	August	-0.2843						
	September	-0.2961						
	October	-0.1218						
	November	-0.0091						
	December	-0.0362 (Apx)						

Figure 1: Trend Analysis of Sentiments



Source: Estimated from Table 1

Analysis of Pakistani media mood, which includes social media and news posts from 2012 onward, reveals significant shifts in the general public's emotional responses over time. As evidenced by the numerical sentiment ratings, which show optimism through positive values and pessimism through negative values, public sentiment in Pakistan varies throughout time in accordance with the nation's changing sociopolitical and economic circumstances.

ISSN (Print): 2959-099X ISSN (Online): 2959-1007
Volume 3, Issue 2, Page 89-104, December 31, 2024

The data from early 2012 onward show extreme despair, with an average value of -0.25. During this time, Pakistan experienced political turmoil, particularly after the Supreme Court disqualified Prime Minister Yousaf Raza Gillani from office in June 2012 (BBC, 2012). Public perception of governance, security dangers and economic difficulties persisted at negative levels during this initial half of the decade.

During the period from 2013 to 2015, negative feelings dominated, but they showed a subtle increase throughout these years. Pakistan experienced its democratic transition to civilian rule through the PML-N led by Nawaz Sharif after the 2013 general elections took place. Despite ongoing doubts, people saw this democratic transition as favourable but remained cautious about its implementation. The economic reforms, along with increased foreign investment figures under the China-Pakistan Economic Corridor (CPEC), enhanced public sentiment slightly (Afzaal, 2020).

Public sentiments moved back to negative levels between 2016 and 2017 due to mounting security concerns and the Panama Papers scandal that forced Prime Minister Nawaz Sharif to step down in 2017 (Al Jazeera, 2017). Public sentiment declined in alignment with the recognised effect of political uncertainty on institutional trust and public perception (Tetlock, 2007).

Public feelings underwent crucial shifts after the 2018 general elections. When Imran Khan took power with his Pakistan Tehreek-e-Insaf (PTI) party, sentiment rose for a brief period before dropping afterward. Research findings show that new political leaders generate brief improvements in public attitude through increased expectations (Tumasjan et al., 2010). The overall sentiment moved back to negative levels throughout 2019 and during the first part of the 2020 decade. Economic issues combined with increasing costs of living and widespread complaints about governance under the PTI rule produced this downward trend.

People reacted very differently to the COVID-19 pandemic during the period between 2020 and 2021. The sentiment index showed that people had problems with health risks as well as economic challenges and the implementation of lockdowns in this period (Pak Institute for Peace Studies, 2021). The emotional characteristics and well-being of individuals experience obvious transformations because of pandemic stress, and these effects materialize digitally in online discussions.

The sentiment pattern from 2022 through 2023 revealed nonstop movement without restoring sentiment to its previous pre-2012 levels. The continuous political disturbances, together with IMF economic requirements and frequent leadership changes, have resulted in an enduring negative atmosphere dominating Pakistani media outlets.

This serves as an ongoing measurement to reveal nationwide emotional tendencies at this current moment. The assessment demonstrates that Pakistani public emotions directly show the present conditions of political stability, economic state, and social welfare in the country. Tracking these indices across various periods helps policymakers and researchers identify early indicators that suggest tense social situations.

Discussion

The Text-Based Sentiment Index (TBSI) for Pakistan serves as a forward-thinking research instrument that unites computational social science with public policy research. The public sentiment derived from news articles and social media posts examined demonstrates a clear correlation with major economic developments, political transformations, and societal alterations within Pakistan. The study validates Tetlock's (2007) initial hypothesis that media sentiment drives both financial market results and political movement patterns when social environments become intensely emotional.

Public confidence in governance maintained a looping cycle of growth and decline during the twelve years between 2012 and 2024. The previous period overwhelmingly contained negative emotions, except during leadership transformations and economic announcements, which brought temporary upbeat sentiments, according to Tumasjan et al. (2010), which demonstrated how political leaders' changes generate temporary boosts in public mood in brief times following their appointment as prime minister. This occurred when Imran Khan took office in 2018.

The public sentiment deterioration during the COVID-19 pandemic (2020–2021) fits previous findings, which demonstrate how health emergencies coupled with misinformation generate higher levels of public worry (Haroon et al., 2021; Pak Institute for Peace Studies, 2021). Electronic information spread exacerbated public confusion because it transmitted both verified and unverified data, thus demonstrating that the media influence’s public opinion and information distribution dynamics.

The study emphasizes specific obstacles that stem from Pakistan's linguistic circumstances. The official English news publications show widespread usage, but Roman Urdu remains the dominant language for public discourse. The lack of standardized NLP resources, together with inconsistent grammar and spelling problems, diminishes the utility of text-based systems (Liaqat et al., 2022). Future versions of TBSI will gain advantages from multilingual NLP advancements, which address the linguistic complexities identified in the current English-language-based study (Khattak et al., 2021).

Additionally, receiving real-time public replies required incorporating social media data. Platforms like Twitter, due to their popularity and timeliness, offer information regarding spontaneous sentiment reactions that are occasionally missed by traditional surveys or indices (Mughal et al., 2024). To provide a more thorough and dynamic picture of public opinion, hybrid sentiment models combine structured news data with unstructured, user-generated content.

The study found that public opinion both predicts and reflects national trends. The findings offer strong support for the application of agenda-setting and sentiment theories in emerging democracies like Pakistan, where public opinion is significantly influenced by traditional and digital media (Shiller, 2015; Baqir et al., 2023).

ISSN (Print): 2959-099X ISSN (Online): 2959-1007
Volume 3, Issue 2, Page 89-104, December 31, 2024

Applications for a Text-Based Sentiment Index

- Political Analysis: Political research based on public opinion helps both decision-making processes and policymaking.
- Economic Forecasting: Consumer confidence assessments through public sentiment tracking will help create economic planning models that forecast market conditions.
- Social Research: Sentiment trends enable organizations to determine social needs and guide them in developing appropriate intervention programs.
- Business Intelligence: Analyzing customer opinions provides valuable information for enhancing marketing approaches and improving service support systems.

Conclusion

A textual analysis of major Pakistani newspapers and social media allowed this study to establish an empirical Sentiment Index (TBSI) that fits Pakistani socio-political dynamics. The study demonstrates that Pakistani public opinions manifest immediate reactions toward vital political decisions, together with economic developments and social circumstances, which illustrates how discourse patterns in the country evolve constantly. Medical studies conducted by Tetlock in 2007 and Shiller in 2015 have been supported by these research findings, which validate media-driven sentiment as a reflection and forecasting mechanism of public conduct. Process development for TBSI encountered difficulties because of Pakistan's language diversity and minimal linguistic capabilities. However, the implementation of superior natural language processing technology together with machine learning made it effective. The index functions as a practical real-time monitoring mechanism for policymakers, researchers, and institutions to measure public mood effectively, which aids their decision-making in governance, economic planning, and social development processes.

Recommendations

Future work should develop the sentiment index by integrating Roman Urdu and Urdu language data because they represent the primary linguistic sources used for digital communication in Pakistan (Liaqat et al., 2022). Sentiment tracking accuracy will improve through the creation of bigger annotated datasets and improved natural language processing models designed for regional languages. An improved sentiment analysis tool requires partnership between educational organizations, academic institutions, technology companies, and public sector entities to enable better political, economic, and social decision-making.

Limitations of the study

The research faces a major constraint because the team could not process texts from Urdu and Roman Urdu documents (Khattak et al., 2021). The method's weakness arises from its dependence on English-language sources from newspapers and Twitter, since these platforms might not reflect the complete emotional state of Pakistan's socio-

economic diversity. Private discussions within closed networks are unavailable for analysis through publicly available text, making it possible that sentiment trends could have been affected by their absence..

Future Directions

Future research should concentrate on the following approaches to boost sentiment analysis accuracy, along with its practicality in Pakistan:

- Researchers need to build bigger annotated collections that include numerous data types.
- Players in the field should develop models that recognize mixed and various language styles in text.
- Incorporating context-aware and sarcasm-detection capabilities.
- Multiple institutions together must establish common resources and methodologies through collaborative work.

References

Afzaal, M. (2020). The China-Pakistan economic corridor of the Belt and Road Initiative concept, context, and assessment, *Taylor & Francis*.

Ahmad, W., & Edalati, M. (2022). Urdu speech and text-based sentiment analyzer. *arXiv Preprint* arXiv:2207.09163.

Al Jazeera. (2017, July 28). Pakistan Supreme Court disqualifies Nawaz Sharif. *Al Jazeera*. https://www.aljazeera.com/news/2017/7/28/pakistan-supreme-court-disqualifies-nawaz-sharif

Aslam, N., Nadeem, A., Abid, M. K., & Fuzail, M. (2023). Text-Based Sentiment Analysis Using CNN-GRU Deep Learning Model. *Journal of Information Communication Technologies and Robotic Applications, 14*(1), 16–28.

Baqir, A., Galeazzi, A., Drocco, A., & Zollo, F. (2023). Social media polarization reflects shifting political alliances in Pakistan. *arXiv Preprint* arXiv:2309.08075.

Baqir, A., Galeazzi, A., Drocco, A., & Zollo, F. (2023). Social media polarization reflects shifting political alliances in Pakistan. *arXiv Preprint*, arXiv:2309.08075.

BBC. (2012, June 19). Pakistan Supreme Court bars PM Gilani from office. *BBC*. https://www.bbc.com/news/world-asia-18506728

Diego García. (2013). Sentiment during Recessions. *The Journal of Finance, 68*(3), 1267–1300. https://doi.org/10.1111/jofi.12027

ISSN (Print): 2959-099X ISSN (Online): 2959-1007
Volume 3, Issue 2, Page 89-104, December 31, 2024

Haroon, D., Arif, H., Tariq, A. A., & Qazi, I. A. (2021). Investigating Misinformation Dissemination on Social Media in Pakistan. *arXiv Preprint* arXiv:2106.09338.

Hutto, C. J., & Gilbert, E. (2014). VADER: A parsimonious rule-based model for sentiment analysis of social media text. *Proceedings of the International AAAI Conference on Web and Social Media, 8*(1), 216–225.

Irum, A., & Tahir, M. A. (2025). Document-Level Sentiment Analysis of Urdu Text Using Deep Learning Techniques. *arXiv Preprint* arXiv:2501.17175.

Kausar, S., Tahir, B., & Mehmood, M. A. (2021). Towards understanding the manipulation of trends in Pakistan on Twitter. *arXiv Preprint* arXiv:2109.14872.

Khan, S., Moqurrab, S. A., Sehar, R., & Ayub, U. (2018). Opinion and emotion mining for Pakistan general election 2018 on Twitter data, 98–109.

Khattak, A., Asghar, M. Z., Saeed, A., Hameed, I. A., Hassan, S. A., & Ahmad, S. (2021). A survey on sentiment analysis in Urdu: A resource-poor language. *Egyptian Informatics Journal, 22*(1), 53–74.

Kumari, S., Rafaqat, S., & Shabbir, T. (2025). Misuse of Social Media: Impacts on Pakistan Politics. *Pakistan Social Sciences Review, 9*(1), 131–143.

Liaqat, M. I., Hassan, M. A., Shoaib, M., Khurshid, S. K., & Shamseldin, M. A. (2022). Sentiment analysis techniques, challenges, and opportunities: Urdu language-based analytical study. *PeerJ Computer Science, 8*, e1032.

Mughal, S., Jaffar, A., & Arif, M. W. (2024). Sentiment Analysis of Social Media Data: Understanding Public Perception. *Journal of Computing & Biomedical Informatics, 7*(02).

Mukhtar, N., & Khan, M. A. (2020). An effective lexicon-based approach for Urdu sentiment analysis. *Artificial Intelligence Review, 53*(4), 2521–2548.

Nagra, A. A., Abubakar, M., Warsi, S. U., Mohsin, S., & Abdullah, H. (2024). Pakistan's Political Sentiments Analysis based on Twitter Using Machine Learning. *Journal of Computing & Biomedical Informatics, 6*(02), 13–22.

Noor, H., Ahmad, J., Haider, A., Nasim, F., & Jaffar, A. (2024). A Machine Learning Sentiment Analysis Approach on News Headlines to Evaluate the Performance of the Pakistani Government. *Journal of Computing & Biomedical Informatics, 7*(02).

Pak Institute for Peace Studies. (2021). *Pakistan Security Report 2020*. https://www.pakpips.com/web/wp-content/uploads/2021/01/Conflict-and-Peace-Studies.pdf

Pang, B., & Lee, L. (2008). Opinion mining and sentiment analysis. *Foundations and Trends® in Information Retrieval, 2*(1–2), 1–135.

Qayyum, A., Gilani, Z., Latif, S., & Qadir, J. (2018). Exploring media bias and toxicity in south asian political discourse, 01–08.

Shiller, R. J. (2015). *Irrational Exuberance: Revised and Expanded Third Edition*. Princeton University Press. https://books.google.com.pk/books?id=_aIpBQAAQBAJ

Tetlock, P. C. (2007). Giving content to investor sentiment: The role of media in the stock market. *The Journal of Finance*, *62*(3), 1139–1168

Tumasjan, A., Sprenger, T. O., Sandner, P. G., & Welpe, I. M. (2010). Predicting elections with Twitter: What 140 characters reveal about political sentiment. *Proceedings of the Fourth International AAAI Conference on Weblogs and Social Media*, 178–185.