

# Text Mining for Public Sentiment Analytics in the National Bank of Rwanda

*Murera Gisa Jean de Dieu,  
Ntirushwamaboko Dominique  
and  
Muhongayire Rukundo Sheilla*

*Working Paper Series: CF009*

AFRICAN ECONOMIC RESEARCH CONSORTIUM  
CONSORTIUM POUR LA RECHERCHE ÉCONOMIQUE EN AFRIQUE



# **Text Mining for Public Sentiment Analytics in the National Bank of Rwanda**

By

*Murera Gisa Jean de Dieu,  
Ntirushwamaboko Dominique  
and  
Muhongayire Rukundo Sheilla*

AERC Working Paper Series: CF009  
African Economic Research Consortium, Nairobi  
November 2021

**THIS RESEARCH STUDY** was supported by a grant from the African Economic Research Consortium. The findings, opinions and recommendations are those of the author, however, and do not necessarily reflect the views of the Consortium, its individual members or the AERC Secretariat.

Published by: The African Economic Research Consortium  
P.O. Box 62882 - City Square  
Nairobi 00200, Kenya

© 2021, African Economic Research Consortium.

# Contents

List of tables

List of figures

List of abbreviations and acronyms

Abstract

1.	Introduction	1
2.	Literature review	3
3.	Methodology	10
4.	Results and discussion	19
5.	Conclusion, policy implications and limitations	25
	References	27

# List of tables

1.	Descriptive analysis of NBR Twitter account	19
2.	Performance measures of machine learning algorithms	22

## List of figures

1.	Social media usage in Rwanda	5
2.	The intermediate steps for pre-processing data	11
3.	Public emotions' categories and their scores	20
4.	Public sentiment classes	20
5.	Public sentiment evolution over time	21
6.	Word cloud for the period under review	22
7.	Receiver operating curve for sentiment prediction	23
8.	ANN predicted sentiment score	24

# List of abbreviations and acronyms

ANNs	Artificial Neural Networks
API	Application Programming Interface
AUC	Area Under Receiver Operating Characteristics
FOMC	Federal Open Market Committee
GDP	Gross Domestic Product
LDA	Linear Discriminant Analysis
MPC	Monetary Policy Committee
MPFSS	Monetary Policy and Financial Stability Statement
NBR	National Bank of Rwanda
PCA	Principal Component Analysis
PwC	Price Waterhouse Coopers
ROC	Receiver Operating Characteristics
URLs	Uniform Resource Locator

# Abstract

Today's internet has become an essential platform for accessing relevant information on various topics and policies; as well as on institution's daily activities. Countless individual users share their opinions and positions on social life, policies, and politics every day using several micro-blogging, including Twitter. The exponential increase of information sharing over Twitter makes it a rich source of data and an ultimate ocean of information for opinions mining and sentiment analysis. To gain insights on how such information benefits the central banks' monetary policy communication, this paper explores the Twitter posts of the National Bank of Rwanda for sentiment analysis by extracting subjective feelings, emotions and public audience perceptions and opinions as feedback to the National Bank of Rwanda's communication.

To understand the current trends, the National Bank of Rwanda's twitter popularity and influential rate, as well as the public perceptions and expectations towards the bank's tweets, we employed the natural language processing and text mining methodologies for text mining and analytics. From our analysis, the results showed that public perception and sentiment towards the central bank's tweets are widely positive, and predicted to remain predominantly positive in the near future by the artificial neural network classifier.

**Key words:** *Monetary policy; Text mining; Text analytics; Social media; Sentiment analysis.*

**JEL classification codes:** *C38, C45, C55, E52, E58*



# 1. Introduction

According to the remarks by the former Governor of the Bank of Italy in 1979, he highlighted that “*the actions of central banks are no longer cloaked in silence, and perhaps never will be again*” (Donato & Davide, 2016). This statement accentuated the change that has marked the evolution of central bank communication. Over the years, the central bank’s communication has evolved from an era of mystery to an era of transparency. Transparency in central banks increased rapidly since the early 1990s, and this trend began with the adoption and implementation of inflation targeting by the Bank of England, the Bank of Canada, the Reserve Bank of New Zealand, and Sweden’s Riksbank (Masciandaro et al., 2020). On the back of this evolution, central banks globally have adopted communication as a key instrument of bank’s transparency.

According to Babatunde and Mohammed (2019), central bank communication is about the quality and quantity of information availed to the public with regard to the central bank’s monetary policy strategy and objectives, its perception on economic future outlook, as well as signals about the next policy path. Today, there is a “general consensus among central bankers that transparency is, not only an obligation for a public entity, but also a real benefit to the institution and its policies”, because, not only does transparency make central banks more accountable, but also enhances the effectiveness of central bank’s monetary policy (Dmitriy, 2000).

Developments in the transparency of central banks corresponded with vivid technological advances (Dmitriy, 2020). The role of transparency and communications in central banking has risen significantly in recent decades. This trend accelerated with the advent of new information technologies, such as web sites and, more recently, social media, which allowed disseminating information faster and among a larger audience (Dmitriy, 2000)

Since the 2000s, following the internet revolution, web content became a norm rather than the pioneering approach in communication with the public (Dmitriy, 2020). However, even then, social media use by central banks was still low, though was much noticed. In 2010, according to central banking research, just a few central banks had a social media presence: five central banks had Facebook accounts, 37 had Twitter feeds, and 27 offered YouTube channels (Rachael & Karolina, 2020). A decade after, central banks are increasingly going digital and the use of social media has skyrocketed.

In the current era of data revolution, the blogs, social networks, and other media spread on the World Wide Web contribute a lot in the production of big data. This

enormous data contains critical opinions with regards to the information that can be used to benefit economic agents, as well as financial and economic stability. Economists and researchers consider Twitter as one of the biggest social media channels where the public expresses their opinion, sentiments and feelings towards organizations.

As central banks improve their online presence on the web and social media platforms, there are key questions to ask: “What is public sentiment or how does the public feel about central bank communication”? The paramount intention of this study is to describe, with a specific example, how data mining, text analytics, and natural language processing can be applied to social media to ultimately identify insights from such a data ocean.

By answering these research questions, we aim to meet the following objectives:

- Introduce novel machine learning methods to mine Twitter data and analyse public sentiments towards the central bank's social media communication.
- Describe social media usage and how the public responds to the central bank's tweets.
- Compare the performance of machine learning classifiers and artificial neuro network classifiers for sentiment prediction in central banks.

Precisely, this work will describe the quantitative analysis of Twitter posts related to the monetary policy communication of the National Bank of Rwanda (NBR). Hence, identifying and quantifying the public sentiments, perceptions, feelings and subjective emotions towards the NBR's tweets. The purpose of this work is, not only to undertake the sentiment analytics, but also to predict and illustrate future public sentiment scores behind the NBR's Twitter feeds on several policies and regulations through the usage of supervised machine learning techniques. Social media platforms have provided a strong way for central banks to connect with the public. Hence the importance for understanding and translating the emotions that come from the way the public reacts toward central bank communication.

In our study, we make four contributions. First, we extract information from a Twitter Application Programming Interface (API) and show how the public perceives the monetary policy communication of the NBR by understanding their sentiments, opinions and feelings. Second, we analyse the public sentiment over time. Third, we apply machine learning techniques to classify and predict the sentiment polarity behind the central bank's tweets. Last, we contribute to the existing stock of literature on use of novel machine learning techniques for text analytics, a method used for the first time at the National Bank of Rwanda to predict the public's sentiment towards the Bank's tweets.

The rest of the paper is outlined as follows: Section 2 covers the existing literature review; Section 3 presents the methodological framework and describes the data used in this study. Section 4 presents experimental results and findings; and further provides the interpretation and discussion of the findings and obtained results. Finally, Section 5 concludes and provides key recommendations.

## 2. Literature review

### Central banks adoption of social media

The origin of modern social media networks can be traced back to the early 2000s. Since the mid 2000s, a host of social networks have gained international prominence (Dmitriy, 2020). Today, we see the use of social media as a new norm to stay updated with current news. In December 2018, a study by the Pew Research Centre showed that social media sites for the first time surpassed newspapers as sources of news for people in the US—one-in-five adults consume news on social media, compared with 16% for newspapers (Vogels, Auxier, & Anderson, 2020). This revolution in social media changed, not only the source people use to stay updated with latest news, but also transformed the way people communicate with each other, and the central bank community has not stayed behind but rather adopted the new norm and grown their presence on social media platforms. Today, websites of central banks have definitely moved from being “mildly interesting new technology” to the core of communications policy (Dmitriy, 2000).

According to Dmitriy, looking at tweets uploaded by central banks, the Federal Reserve Bank of Dallas can be regarded as the first monetary institution to start using social media by uploading interview with Milton Friedman on its official YouTube channel on 11 November 2007 (Dmitriy, 2000). Half of world's central banks adopted social media by the year 2014—seven years after the first social media account was established (Dmitriy, 2000). In the last decade we see a growing trend of central bank's presence on social media, and with the way the whole world is interconnected today, this trend is yet to even grow enormously as social media provide a platform for interactions and connectivity. Looking at the content central banks post on social media reveal either their using of their presence on social networks to channel traffic to their main site, or using their presence to spread useful and informative content to different target groups in accessible way (Bholat, S. Hans, & Schonhardt-Bailey, 2015) (Bjelobaba et al., 2017). In addition, for central banks, transparency and communication with the public are of extreme importance. Hence, central banks' presence in social media networks can be considered as an additional indicator of central bank transparency (Dmitriy, 2000). Social media platforms provide different capabilities and content to its users. Thus, the decision of monetary authorities to adopt social media should depend on its preferences in communication process and

type of content (Dmitriy, 2000). Among different social media platforms used, Twitter can be considered as the most popular social media channel used by most central banks (Bjelobaba et al., 2017). Central banks are able to target a diverse audience that is demographically distinct from the general public. Specifically, they are younger and more interested in the current affairs than the population on average, and may be hard to reach with traditional media (Brechler, Hausenblas, Komárková, & Plašil, 2019) (Korhonen & Newby, 2019).

In addition, economists argue that twitter communication data is widely used to control economic and financial market fluctuation (Ausserhofer and Maireder, 2013). Since then, many organizations around the world use Twitter to monitor and promote their brands and campaigns (Varol, 2017) for profit optimization. Among other many social media platforms, Twitter has become one of the most important databanks, providing the main source of data for social media analysis to various audiences, including economists, journalists, and businesses.

The big challenge for central banks in Twitter is how to build or consolidate confidence, gain authority and be a reliable source for their followers, gain support for future initiatives and current issues of importance (Bholat, S. Hans, & Schonhardt-Bailey, 2015). However, to be able to translate and understand what is behind different tweets, likes, and retweets, twitter has enabled data miners via its three APIs to access its data and most researchers and analysts depend on it for their research and policy making (Pfeffer, 2018).

Hence an opportunity for data analysts to leverage on social media and web data to shade light on phenomenon previously unexplained, and analyse the relationship between data from a text known unstructured and structured data to explain and reveal hidden insights.

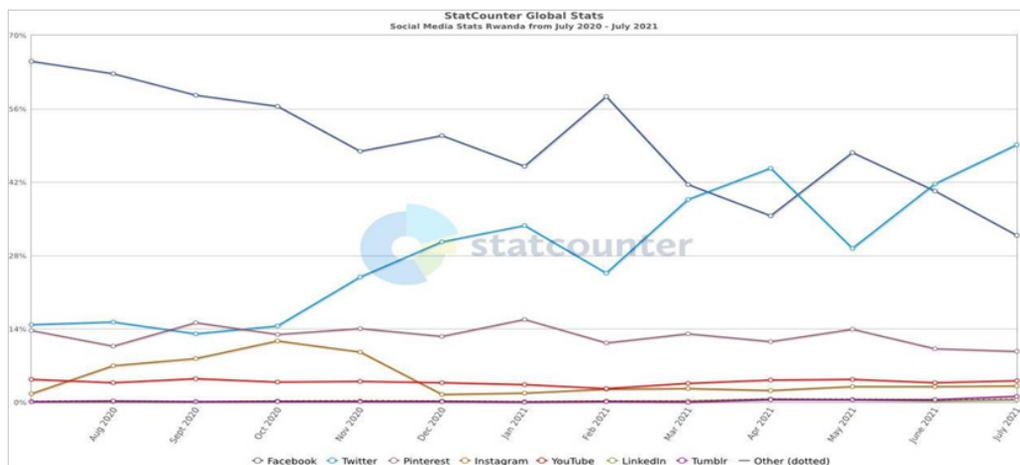
## **Overview of the central bank's communication in Rwanda**

Social media platforms are probably the biggest change in communication over the past two decades, following the advent of smartphones and the ubiquity of the internet. These exciting technologies brought along different social media platforms, and as a result, communication has been completely revolutionized making it much easier to reach and interact with the public directly.

The NBR communication is in line with the growth we are witnessing in the use of social media, especially in the young generation. According to stats counter as of July 2021, Twitter has 49.16% followed by Facebook with 31.3% of social media users, followed by Pinterest with 9.67% and YouTube with 4.07% users, the details are as shown in Figure 1 (StatsCounter, 2021). These statistics show that 11% of Rwanda's total population are active users of social media, and they spend on social media an average of 54 minutes daily (StatsCounter, 2021). Focusing on Twitter, we see that its Rwandan users tripled over the year from July 2020 to July 2021, which shows a

surge of interests in using this social media platform (StatsCounter, 2021). This surge in the use of Twitter in Rwanda unlocks big opportunities to gain insights on central banks communicate from Twitter users by leveraging its data.

**Figure 1: Social media usage in Rwanda**



Source: StatCounter, 2021.

Tracing back the journey of central bank communication, all indicators evaluating central banks communication show a sound increase since the mid 1990s, demonstrating that it has become a key issue, and the financial crisis strengthened this trend (Blot & Hubert, 2018). Central bank communication is known to be the channel through which information from the central bank is passed on to the public. This information can be transmitted using both qualitative communication instruments (e.g., public speeches, official publications, statements, and minutes), as well as quantitative communication instruments, such as inflation forecasts and other macro variables. Despite the channel used, according to Blinder, the important role of communication is to influence market expectations, which is key for financial markets and the entire economy.

Currently, when communicating to the public, the NBR targets the following: the financial system, the business community, households, think tanks, media, youths, academia, NBR Staff, parliament, and other stakeholders including public authorities and international organizations. Over the years, different communication channels have been adopted to reach different groups the NBR communicates to. Today the tools used include Monetary Policy Committee (MPC) press releases, press conferences by the Governor, the NBR website, central bank publications, the use of social media platforms such as Twitter, YouTube, Instagram, Facebook, Email, and WhatsApp Group with Journalists, Slido App, outreach programmes, and a semi-annual Monetary Policy and Financial Stability Statement held every February and August.

As stated earlier in this paper, we will be focusing on public sentiment towards communication by the NBR, specifically the Monetary Policy and Financial Stability

Statement (MPFSS). This is a statement made twice a year in February and August that highlights the development in the global economy, domestic economy (real and external sectors), monetary sector (monetary policy formation and implementation and outcome, financial market development), developments in inflation and outlook, financial sector stability and payment system development. The first MPFSS by the Governor was in March 2005 to an audience of cabinet members, representative of the business community, chief executive officers of financial institutions, representatives of international institutions in Rwanda, academicians and NBR staff. Since then, tremendous improvement has been made to reach a wider audience. As mentioned before, this statement is currently made twice a year. First in February to give policy orientation of the year after reviewing the achievements and challenges of the previous one, and later in August—a mid-term review of the performance of the first half of the year and set perspectives for the remaining period of the year.

In January 2019, the NBR adopted a price-based monetary policy from the use of a monetary targeting framework that was exercised since 1997. This new framework spearheaded the development of NBR's communication strategy, which was adopted in September 2019 to influence the market's expectations by creating awareness about the Bank's policies and how they relate to the public, enhance the bank's engagement with policy makers, sector players, and the general public while protecting the bank's image and institutional reputation, and ensuring the availability of a broad range of financial education initiatives that promote financial literacy, financial inclusion, and consumer protection.

Also, this communication strategy aligns with the bank's effort to increase its transparency, credibility and accountability, as well as making monetary policy decisions more predictable. Monetary policy predictability is the ability of financial markets to correctly anticipate the next monetary policy decision of a central bank (Nkuttner, 2001; Rasche, 2000). For the public to better understand the direction of monetary policy, the NBR has reinforced explaining their monetary policy framework, and are precise about what they want to achieve in terms of their policy goals. This is achieved by informing the public on how monetary policy decisions are taken, how monetary policy is conducted and what the bank's objectives are. This information is provided to the market in an effort of increasing transparency of the way the NBR operates, and in turn, this has been key in making the Bank's actions more predictable.

To improve the public understanding about central bank communication, the NBR is putting effort in engaging a wider audience. For example, in the bank's communication strategy, the NBR has adopted strategies to communicate to different audiences taking into consideration that their levels of education and knowledge about the use of financial markets are different depending on their academic background and level of education. Therefore, identifying the target audience is very crucial for the NBR for it has played a role in designing different channels of communication the bank uses today.

## Empirical literature

This section reviews widely the existing theoretical and empirical knowledge on the central bank's communication channels and text-based sentiment analysis. Sentiment analysis is an application of natural language processing to extract and process the text in order to identify subjective feelings and opinions from text data (Farhadloo & Rolland, 2016). The natural language processing enables the analysis of unstructured text using different techniques and groups sentiments into categories such as positive, negative, and neutral. Consequently, it defines the general sentiment tendency of the blogger regarding the topic in context.

To understand the central bank public sentiment, some recent studies utilized social media feedback on central banks' communication and found out that, as central banks increase their social media reach, their communication outreach increases and improves the likelihood to effect market expectations (Dmitriy, 2020; Eriksson, 2018; PwC, 2020). Historically, in the 1970s and 1980s, central bank's activities and monetary policies conduct were mystique and secrecy (Masciandaro et al., 2020). In that era, the silence was considered as an assurance of independence. Whereas, today we live in a time where actions of the central bank are no longer a mystery but rather defined by transparency, openness, and accountability. This is achieved by giving an explicit account of one's actions (Masciandaro et al., 2020). There is so much evidence and research that shows how the development in central bank communication has improved the transmission of monetary policies, and this is inevitable given that monetary policy strategy and its communication are inexorably linked (Orphanides, 2019).

Moreover, in the past, several monetary policy objectives, instruments, and policy actions were commonly not predictable (Masciandaro et al., 2020). With this regard, most central bankers believed that monetary policy makers should release as little information as possible to the public (Janikowski & Rzonca, 2018). In the last decade of the 20th century, this attitude shifted from secretiveness to transparency and accountability. Hence a tremendous change from a period where central banks were very hesitant to share any information to an era of timely updates. Research shows that this change has been attributed to the acknowledgement that monetary policy is essential in managing public expectations (Janikowski & Rzonca, 2018).

Today, transparency in central bank's communication has made it possible to analyse how communication by a central bank is perceived by the market. When central banks were cloaked in mystery measuring public sentiments towards the central bank communique would have been impossible given that, for a long time, many central banks never made their decisions public. Therefore, clarity in the way central banks communicate is key and clear communication of an independent central bank's strategy disciplines the use of discretion, which is essential for ensuring that policies remains systematic (Orphanides, 2019).

The provision of better information to market participants on the monetary policy ongoing activities and future intentions should enhance the point to which central bank policy decisions can affect public expectations and, thereby, improve the

effectiveness of monetary policy stabilization (Masciandaro et al., 2020). Therefore, it is of paramount importance for the public to comprehend current and future policies, and one way to identify market expectations is via measuring market sentiment. Central bank economists have argued that there are always inconsistencies between the expectations and the actual outcomes when it comes to central bank monetary policy communication and the public reactions (Blinder et al., 2008). To understand this phenomenon, some attempts were made to measure how much the market reacts towards monetary policy actions.

In many central banks, the interest in modelling the monetary policy communication is on the rise. Hubert and Labondance of the science politique of Université de Bourgogne Franche analysed the aggregate effects of optimism or pessimism on monetary policy by testing in the context of monetary policy the theoretical prediction that sentiment may have on aggregate effects as hypothesized by Angeletos et al. (Angeletos & Jennifer, 2013). They quantified the concept of central bank sentiment and tested its likely importance in economic decisions. They also used the tone conveyed by the Federal Open Market Committee (FOMC) policy makers in their statements using computational linguistics methods.

Besides, they identified sentiment as the unpredictable factor of tone, orthogonal to fundamentals, expectations, investors' sentiment, monetary shocks, and central bank communication about future policy. Lastly, they estimated the impact of sentiment on the term structure of private interest rate expectations using a high-frequency methodology and an ARCH model. Their results revealed that the optimistic sentiment increases the term structure of interest rates primarily at the 1-year maturity. Interestingly, they also found out that central bank sentiment affects inflation and industrial production beyond monetary shocks and in opposite directions (Labondance, 2018).

In Korea, Juyoel Lea and Wook Sohn used text mining to quantify the monetary policy surprises. They proposed a new method to measure monetary policy shocks using sentiment analysis. To quantify the tones of monetary communication, Lea and Sohn employed 24,079 news articles and roughly 152 publications of the Bank of Korea's Monetary Policy Board meetings announced to the public from March 2005 to November 2017. In their study, they measured monetary policy surprises using the changes of those tones following monetary policy announcements and estimated the impact of monetary policy surprises on asset prices. In their analysis, they also found out that the new indicator relates to changes in long-term rates, while changes in the Bank of Korea's base rate are more closely associated with changes in short-term rates. The results of their study showed that text analytics methods could help understand monetary policy surprises on a timely manner to shed light on information related to forward guidance and market expectations on future monetary policies (Young Joon et al., 2019).

In literature, Twitter sentiment extraction is linked to economic fundamentals and its impact on central bank communication by using computational linguistics to infer diverse measurements of central bank communication (Hansen, 2014). In addition,

there is growth in literature that studies the influence of Twitter sentiment on financial services and systems fluctuations (Bollen, 2011; Zhang et al., 2011).

In the same light, Gholampour and van Wincoop (2017) shows that Twitter is a vital source of information to understand and predict the euro-dollar exchange rate. Gholampour and van Wincoop went further and illustrated that the current generation of traders share their market expectation and information on Twitter. Regarding monetary policy, in his research Azar performed sentiment analysis of tweets talking about the Federal Reserve and the results revealed that Twitter sentiment has a great impact on asset prices (Azar, 2016). Meinus and Tillmann (2017) in their study inferred the beliefs of users about monetary policy from Twitter influencing central bank policy. Meinus and Tillman investigated the degree and level to which long-term bond yields and the exchange rate are sensitive to fluctuations in hopes of the Federal Reserve's exit from quantitative easing. To proxy these beliefs, these authors labelled and aggregated tweets from April to October 2013, hence distinguishing the users' opinions on whether the Federal Reserve will tamper soon or late. Using a VAR-X model, they identified a belief shock. Their results indicated that changes in beliefs have strong and persistent effects on bond yields and exchanges rates.

It is worth noting that, the area of Twitter sentiment analysis in central banks is relatively new, and few studies are available in this area. Most studies are done on general central bank communication analysis, its impact on the monetary policy implementation and its relation to different economic variables. However, a few studies have attempted to analyse social media, particularly Twitter, to understand public sentiment towards the central bank's Twitter communication. Many of the studies available in the area of sentiment analysis use traditional econometric approaches (Angeletos & Jennifer, 2013; Labondance, 2018; Gholampour and van Wincoop, 2017), while a few recent studies use novel methods of machine learning to classify the sentiments of individual tweets as positive or negative employing supervised methods and lexicon-based methods (Zhang et al., 2011; Young Joon et al., 2019; Hassan et al., 2015).

This study employed a combination of lexicon-based and machine learning methods to analyse Twitter public sentiments vis-a-vis the National Bank of Rwanda. This study is the first of its kind to be conducted in the context of a central bank in Africa. We expect it to forge away for subsequent studies in social media and introduce social media text analytics in economic analysis and prediction in central banks.

## 3. Methodology

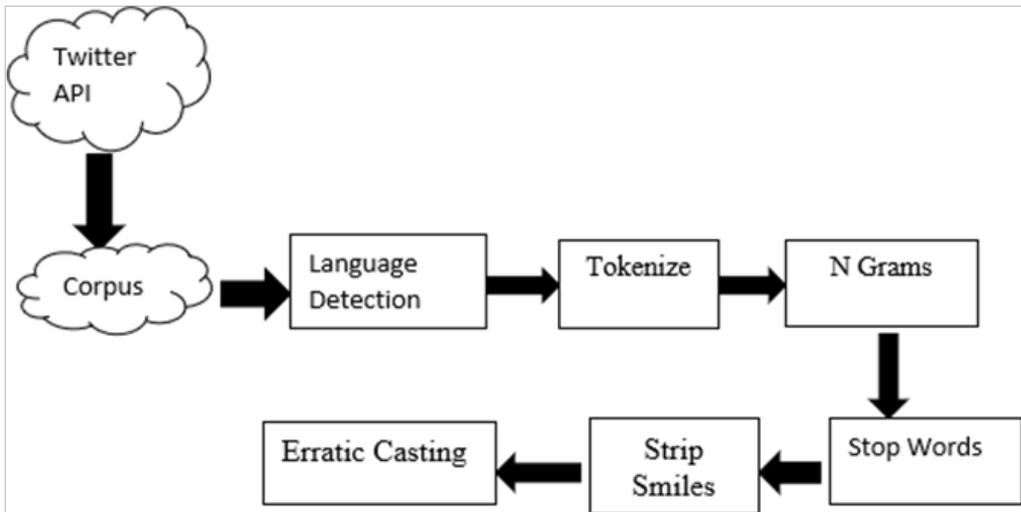
This section describes the methodology that was used for Twitter data analysis to explore public sentiment towards the NBR's monetary policy and financial statement issued every February and August. It also mentions the data processing, extraction of emotions and sentiments, modelling, and prediction utilizing machine learning methods.

### Data description

We extracted data from Twitter, a social networking and microblogging media that permits users to tweet real time messages. These tweets are short messages, restricted to 140 characters in length and most of the time contains abbreviations and emoticons. The data set employed consists of 35,223 tweets and 29,554 retweets that have been extracted using the Twitter Application Programming Interfaces (API) and were mined using the keyword (@CentralBankRw) for the NBR. The variables present in the data set are: The polarity of the tweet (positive or negative), which is the target variable; unique ID of the tweet; the date of the tweet; user referring to the username of the tweeted text. We restricted data collection to tweets posted only in English as other languages such Kinyarwanda, Swahili, and French used in Rwanda did not have their sentiment associated dictionaries for analysis. We employed the data extracted from 25 October 2016, a period when the NBR started to engage actively on social media to 23 August 2021, a time with the latest available data. The Python and R softwares were used in data mining, processing and analysis.

### Data processing process

The analysis of social media presents some special consideration as its data is unstructured and informal. The data presents various issues including misspelling, use of abbreviations, decrypted sort of messages, shortening long words, symbols and emoticons that do have a relevant meaning and implying additional attention in cleansing and wrangling the Twitter text data. Since the tweets extracted are not in a structured format, we had to perform certain text-specific steps. The list of such steps is discussed in Figure 2 to widely provide the sequential processes undertaken to carry out the entire pre-processing work.

**Figure 2: The intermediate steps for pre-processing data**

- (1) The Corpus shown above is termed as a collection of written texts. The tweets collected from the API are a mix of URLs, punctuations, symbols and other non-sentimental information like hashtags, annotations, and misspellings. To obtain n-grams, we have to tokenize the text input (corpus), and next normalize the text in the reviews by making all letters lower case, removing punctuations, stop-words and white spaces. In extracted URLs, we removed emojis/dodgy unicode, duplications, lemmatizing and tokenizing the text and stem the documents. Formally, Figure 2 details the segment explanations on the various intermediate processing steps.
- (2) In language detection we separated all tweets into English and non-English labels (Kinyarwanda and others) as we restricted our study on tweets in English. This has been possible by using NLTK's language detection by restricting other languages.
- (3) Tokenization takes a sample input text like "NBR reduces its central bank rate to 4.5percent", divides this statement into tokens which are substrings of a string. We tokenized a text to distil punctuations and symbols from words that can enhance sentiment polarity score of the text under consideration.
- (4) Construction of n-grams which are the number of words in the statement. We constructed n-grams out of consecutive NBR's tweet words. A negation (such as "no", don't, won't, and "not") is attached to a word which precedes it or follows it. For example, a sentence "I do not like this policy" will form two bigrams: "I do+not", "do+not", "like", "not+like this policy". Such a procedure allows for the improving of the accuracy of the emotion and sentiment polarities since the negation plays a special role in an opinion, expectations, feelings, and sentiment expression.

- (5) The stop word in information retrieval is a common tactic to ignore very common words such as “a”, “an”, “the”, etc. since their appearance in a post does not provide any useful information in classifying a text document, we removed them in the pre-processing phase.
- (6) Strip smileys: Many public microblogging posts make use of emoticons to convey emotions and feelings, making them very useful for sentiment analysis. A range of many emoticons, including “:);”, “:(;”, “:D”, “=]”, “:];”, “=)”, “=[”, “=(” was being replaced with either a SMILE or FROWN keyword. Also, we replaced emoticons with variations of laughter such as “haha” or “ahahaha” with a single laughter keyword.
- (7) Erratic casing: For addressing the problem of posts containing various casings for words (e.g., “HeLLo”, “Hey”, and “Hi”), we sanitized the input by lowercasing all words, which provides some consistency in the lexicon. Punctuation in microblogging posts, it is common to use excessive punctuation to avoid proper grammar and to convey emotion. By identifying a series of exclamation marks or a combination of exclamation and question marks before removing all punctuation, relevant features are retained while more consistency is maintained.

In addition to the above approaches to processing and preparation of the extracted tweets, we replaced a sequence of frequently used characters by three characters, for example, transformed coooooooooooooool into cool. We replaced the sequence by only two characters since we wanted to separate regularly utilized words from emphasized ones. After completing the above processing process, we moved to another step to employ the dictionaries accommodated by R libraries to attribute emotions and sentiments.

## **Sentiment scores and emotion classification**

Sentiment classification deals with the method to predict the direction of the message on blogs, webs or text documents (Brendan Tierney, 2016). Generally, it assumes that the text to be explored is found in sources such papers, policy reviews, and audience feedback templates. Thoughts and opinions might be classified as fitting to differing positive, negative or neutral sentiment polarity scores. Additionally, one might go in-depth to compute eight emotions and their corresponding frequency in a statement or entire document. Emotions and sentiments should both be classified and ordered according to a range of opinions and feedback about a product or policy communicated. They can also be presented using a numeric scale, to indicate the level of positive and negative intensity of the sentiment involved in a piece of a text document (Jockers et al., 2017). Intuitively, this study applied several linguistic resources, namely, Emoticon (NRC sentiment dictionary), Acronym and SentiWordNet dictionaries to accurately assign emotions and sentiments to the NBR's tweets. These linguistic resources are explained hereunder.

The WordNet is a lexical database of semantic relations between words in more than 200 languages. WordNet links words into semantic relations, including synonyms, hyponyms, and meronyms (Miller, 1995). It is widely used in text mining and natural language processing due to its important database that comprises of 155,327 words arranged in 175,979 synsets for a sum of 207,016 word-sense pairs. On the other hand, SentiWordNet is defined as an opinion lexicon derived from the WordNet lexical database synset where each term is associated with numerical scores indicating positive and negative sentiment information (Stefano Baccianella, 2006). SentiWordNet's structure makes it a useful tool for computational linguistics, text mining, and natural language processing.

The NRC Word-Emotion Association Lexicon (Emolex), which is a list of English words and their associations with eight basic emotions (anger, fear, anticipation, trust, surprise, sadness, joy, and disgust) and two sentiments (negative and positive), is also utilized in sentiment computation. Mohammad (2017) revealed in his study that this dictionary has made the following impacts which make it popular and most used in the area of sentiment analytics. It remains the largest lexicon dictionary that combines six emotions and two sentiments, though careful attention is paid to ensure appropriate annotations of words. The lexicon has been identified to be used for word, sentence, tweet, and document-level sentiment and emotion analysis. Beyond this, it has been used for abusive language detection, personality trait identification, stance and colloquial detection (Mohammad, 2017). The above linguistic resources and dictionaries have been widely used in this study to attribute and quantify the public emotions and sentiments behind the NBR tweets.

## **Machine learning algorithms for sentiment prediction**

We used the machine learning techniques to predict the public sentiments and expectations behind the NBR's tweets. The existing set of rules processes or algorithms have been developed to enable computers to learn such target mapping functions from data and make accurate predictions.

The selection of the algorithms to be used depends on various criteria such as the size, structure and data type, distribution of data entries, complexity, latency and speed of the model itself. The choice was also informed by the literature surveyed with selected authors using Latha (2019), Bannach-Brown (2018), and Baharudin (2010). The employed models are the following methods:

### ***i. Logistic regression***

The logistic model is a technique borrowed by machine learning from the field of statistics. It is also used for classification problems and model the occurrence probability of a certain class or existing event.

In our case, we have the data set called  $D = \{(x_1, y_1); \dots; (x_n, y_n)\}$  with  $x_i \in \mathbb{R}^{n \times p}$  and  $y_i \in \{Pos, Neg\}$ ; where  $n \gg p$ , the Logistic Regression will model the occurrence probability that an input  $x_i$  belongs to either positive (Pos) or negative (Neg) sentiment classes. Formally, we can write this as:

$$(x_i) = (Y = Pos \text{ or } Neg | x_i) \quad (1)$$

where,  $i = 1, 2, 3, \dots, n$ , with  $n$  representing the number of data observations.

## ii. Naïve Bayes

The Naïve Bayes classifier is in the family of probabilistic classifiers methods based on the application of Bayes' theorem, which suggests strong independence and equality assumptions between the variables. The Bayes' theorem finds the probability of an event to occur given the probability of another event that has already occurred. Mathematically, it is expressed as follows:

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)} \quad (2)$$

where, A and B are the events and P (A) is the occurrence probability of event A.

With the above-stated theorem, we want to find the probability of event A given that event B is true (has happened). Event B dubbed as evidence. Additionally, P (A) is the prior probability of the event before it happen known as the priori of A. The evidence is an attribute value of an unknown instance (here, it is event B).

Finally, P (A|B) is the probability of event after evidence happens (a posteriori probability of B). Then, in relation to our data set and sentiment classification task, we applied the Bayes' theorem as follows:

$$P(y|X) = \frac{P(X|y)P(y)}{P(X)} \quad (3)$$

where, y is class variables (positive and negative) and X is set of dependent feature vector of size n, where:

$X = (x_1, x_2, x_3, \dots, x_n)$  and P (X|y) means the probability of occurrence sentiment class given any BNR' tweets.

### ***iii. Simple artificial neural network***

Artificial neural networks (ANNs) get their origin from the human brain and are mimicked in computer programs to simulate how the human brain processes information (González-Bailón, Wang, Rivero, Borge-Holthoefer, & Moreno., 2014) . ANNs learn behaviours and association by discovering data patterns and connections and learn through experience, not from programming. ANNs consist of the input layer, hidden layers, and the output layer. The input layer is the first layer while the output layer is the last. In each of the layers in ANN, there are nodes called neurons. A node is the building block that processes the data in the network through a sum and transfer function. The ANN predicting accuracy depends on the number of neurons in the hidden layer. The ANN accuracy is directly proportional to the number of neurons in the hidden layer; however, this is not always applicable.

### ***iv. Linear discriminant analysis***

It is useful when the number of input features increases to the point where the predictive modelling function of a model becomes too difficult for the model to function. This is more commonly called the curse of dimensionality. It is widely used for dimensionality reduction and used as a pre- processing step in machine learning for pattern recognition and classification.

The reduction of dimensions refers to techniques that reduce the number of input variables in a data set which makes the model more robust. The model consists of statistical properties of the data that have been calculated for each class. Linear Discriminant Analysis (LDA) predicts by estimating the likelihood of a new set of inputs relating to each class. Also, the prediction is made simple by the use of the aforementioned Bayes' theorem which estimates the probability of the output class given the input features, and the output class is the class that gets the highest probability.

## **Modelling, features selection, and models performance evaluation**

After processing, the data are labelled as positive, negative and neutral. Before we proceed to model building, we split the data set into training, validation, and testing sets. The training and validation data sets were used for the model development process (learning process) while the testing data set has been used to test the model performance and its ability to achieve the accurate prediction of class prevalence. Hereafter, we trained the chosen four supervised machine learning classifiers on the training set and built a classification model.

Prior to training of models, we selected the principal predictors which contribute widely on the variance of the entire data set. We selected them by using the principal

component analysis (PCA) technique for dimension reduction. PCA is a powerful statistical tool for analysing large data sets, and is formulated in the language of linear algebra (Apoorv et al., 2010). The PCA results are computed utilizing an eigenvector analysis of a correlation matrix. Then we automatically selected the factors from a correlation matrix with eigenvalues greater or equal to one (Kaiser Criterion) and are considered as principal predictors to be used in our predictive models. The predictive performance of each model was estimated using algorithmic syntaxes supported by caret library in R. We trained each model and selected the optimal model across the hyper-parameter tuning through the utilization of 10-fold cross-validation repeatedly three times as re-sampling method. Furthermore, the pseudo-random number generator has been used in training every model to avoid results variability in obtaining trusted results.

Finally, the predictive power of built models was tested, evaluated and validated on the test data set to measure the models' quality and efficacy performance through different performance metrics. Model evaluation is extremely important as it quantifies each classifier's performance ability compared to others trained on the same data set. In this phase, we test the performance of our machine learning models through several evaluation metrics for classification problems. The performance capabilities of the models used were compared based on their performance. To measure the sentiment prediction accuracy of an individual model, we tested each classifier model on set of independent samples (test data set) in a way to incorporate all possible sources of statistical variability. We employed the k-fold cross-validation where the data set was divided into 10 k (k=10) times of model training, one of the subsets randomly set aside and used to test the next fold in effort to assess the classifier's performance. In this regard, all the possible causes of the entire data set undergo training and testing, leading to lower variance within the set estimator and less bias of the true rate estimator, which is the crucial advantage of using the k-fold re-sampling approach.

To effectively select the optimal model, the performance of each model was evaluated using the confusion matrix and receiver operating characteristics (ROC). In confusion matrix, for the case of two-classes "Positive" and "Negative"; there are four possible outputs for prediction, which are TR=True Positives, TN=True Negatives, FP=False Positives, and FN=False Negatives. The different performance metrics such as Accuracy, Sensitivity, Specificity, Cohen's Kappa, and F- measure (F1Score) that combine precision and Recall were computed using the four outcomes to evaluate the classifiers' performance. The computed performance metrics are defined below:

Accuracy: It measured the proportion of cases correctly classified and was computed by:

$$Accuracy = \frac{TP+TN}{n} \quad (4)$$

where,  $n$  is sample size and  $n = (TP + TN) + (FP + FN)$

**Sensitivity:** It measured the fraction of positive cases that were classified as positive and has been calculated as:

$$Sensitivity = \frac{TP}{TP+FN} \quad (5)$$

In our research case, maximizing sensitivity is more relevant since positive sentiment class matters more than a negative one and we do not need to miss out any positive sentiment because of being classified as a negative one.

**Specificity:** Measured the ratio of negative sentiments that were accurately classified as negative sentiment polarities, it was estimated by:

$$Specificity = \frac{TN}{TN+FP} \quad (6)$$

The combination of sensitivity and specificity makes the model's ability to accurately classify the text. The higher the detection rate, the good is the predictive model. The other great performance measure used is F-measure, which combines precision and recall. This approach is to find what percentage of the model's positive predictions (Precision) and negative predictions (Recall) is accurately predicted. It is computed as:

$$Fmeasure = \frac{2(Precision*Recall)}{Precision+Recall} \quad (7)$$

Finally, we computed Cohen's kappa to measure the classifier's performance compared to the performance of a classifier by chance. The Kappa score measures the difference between the accuracy measure and the null error rate, which is the general error made in prediction  $(TN+FP/n)$ . The higher the values of these statistical measure, the better the predictive performance of the machine learning classifiers.

The Receiver Operating Characteristics (ROC) curves were calculated also based on the predicted outcome and the true outcome. The area under the ROC was averaged for the test data sets to compare the discriminating powers of the ML algorithms (Zhang et al., 2011). Theoretically, the Area Under Receiver Operating Characteristic Curve (AUC) lies between 0 and 1, where a perfect classifier can take a maximum of value 1. However, (Zhang et al., 2011) agrees that the practical lower bound for random classification is 0.5, and classifier whose AUC is significantly greater than 0.5 has at least some ability to discriminate between cases and non-cases.

Cohen's kappa statistic is an extremely good measure to handle the quality of

binary classification and imbalanced class problems (Olteanu, Castillo, Diaz, & Kiciman., 2016). In its essence, it measures the correlation and agreement between the predicted and the actual classifications in a data set. It returns a coefficient between -1 and +1. (Bangdiwala, Munoz, & I., 2010) Landis and Koch (1977) provide a way to characterize the values of this statistic: a coefficient of +1 represents a perfect prediction, 0 not better than random prediction, and -1 indicates total disagreement between prediction and observed classes.

In this paper, five machine learning algorithms were trained using a sample size of 80% of the twitter data set (training data set,  $n=14,145$ ) and validated in the remaining 20% (test data set,  $n=732$ ). All models were trained based on 10-fold cross-validation repeated five times.

## 4. Results and discussion

This section presents descriptive statistics of the NBR tweets and empirical results from the model and prediction of future sentiments.

### Descriptive results

This study considers the period from 25 October 2016 to 23 August 2021. The choice of this period corresponds to the time when the central bank adopted its social media strategy to engage the public in the conduct of its monetary policy. Table 1 provides key statistics of the sample data. The sample comprises 35,223 tweets and 29,554 retweets. The data shows that, as of August 2021, the NBR had 74,924 followers, 82 friends, 150 subscribers, 7,859 statuses, and 235 tweet favourites. Even though Twitter data do not show the location of the twitterers that follow the NBR, we believe that the majority are Rwandans interested in the economic and monetary policies of the NBR.

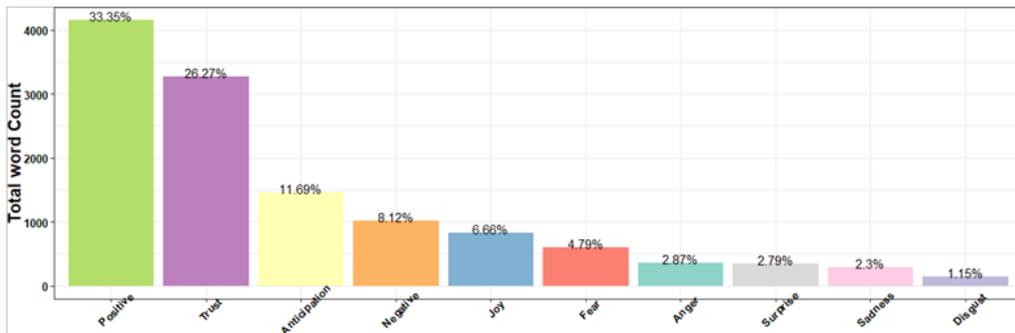
**Table 1: Descriptive analysis of NBR Twitter account**

Description	Count
Number of tweets	35,223
Number of retweets	29,554
Tweets without retweets	5,669
Number of Rwandan Twitter users (as of August.2020)	16.15% of population
NBR followers	74,924
NBR friends	82
Tweet favourites (tot. popular/ liked tweets)	235
Tweet listed (subscribers as interested in NBR posts)	150
NBR tweet statuses (authenticated users tweeted)	7,859

Source: Authors' calculations from NBR Twitter Account.

Using the natural language Emolex dictionary based on English language, eight basic emotions, namely, anger, fear, anticipation, trust, surprise, sadness, joy, and disgust were computed. As per the Figure 4, the largely dominant emotion is positive, representing 33.35% of all emotions, followed by trust, anticipation and negative expression with 26.27%, 11.69%, and 8.12%, respectively. The lowest emotions are disgust followed by surprise and anger with 1.15%, 2.3%, and 2.79%, respectively.

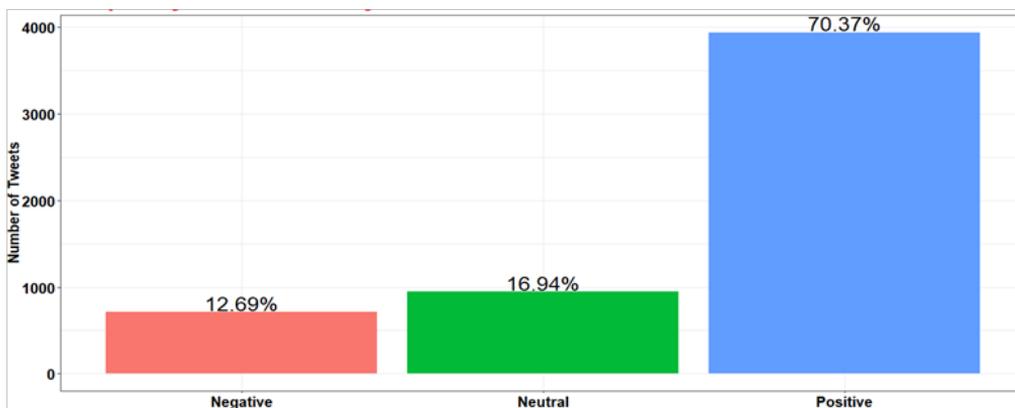
**Figure 3: Public emotions' categories and their scores**



Source: NBR, Author's calculation.

As discussed in the literature review, the common type of sentiment analysis is polarity detection that classifies a text as positive, negative, or neutral. In the period under study, the majority Twitter users expressed themselves positively with a 70.37% score, followed by neutral with 16.94%, while negative sentiment came last with 12.69%. These findings may indicate that the reactions of the public towards the NBR posts, events and policies are predominantly positive, which implies an implicit trust to the central bank’s activities and policies. Even though the majority sentiment is positive, there are other few users who are negative and neutral, their voice should also be heard so that the central bank can adjust or make inclusive policies.

**Figure 4: Public sentiment classes**



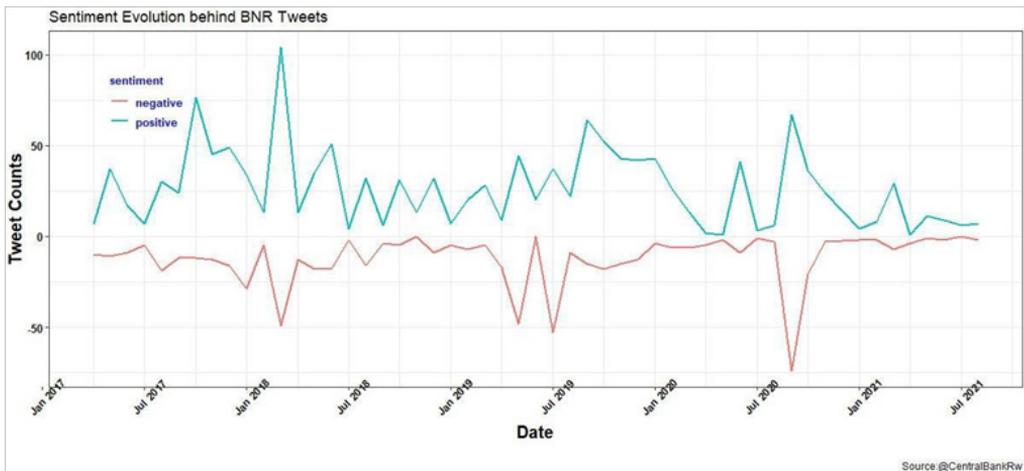
Source: NBR, Author's calculation.

We performed a time dependent sentiment analysis to understand the trends and patterns of how the public responded to NBR use of Twitter over time. As stated in the introduction, the objective (2) was to understand the public sentiment towards the

NBR tweets. To reach this goal, we created two visual displays for Twitter sentiment trends and the comparison word cloud. The sentiment trends explored the changes in public sentiments over time, which might correspond to specific events, new policies, announcements and publications. The comparison word cloud was found to be a powerful tool to understand the discussion or interests of the public in the period under review. Figure 5 presents Twitter sentiment trends, showing the percentages of positive and negative tweets per month, respectively, for the research period of 25 October 2016 to 23 August 2021. The spikes and vales in these trends reveal how the public sentiment changed in respect with the NBR announcements, events, and activities.

For example, we noticed a hike in positive sentiments in February of 2018 and that February corresponds to the time when the Governor presents the monetary and financial statement to the public. It also coincided with the promotion of the launch of the NBR engage event in which the NBR invites all the schools in the country to participate in the monetary policy challenge where the winners get awards and recognitions. On the other hand, we noticed a spike in negative sentiments in June of 2019, which might have resulted from public outcry for foreign exchange and the lending rate in the commercial banks. These observations show how the public sentiments could be determined by events, and that the NBR should take into consideration the public social media responses when making decisions and enacting policies. Similar findings were observed by (Hsuanwei et al., 2016) in Austin (US) study on government use of social media, where their results showed that the spikes and valleys in citizens' sentiments depended on government events and policies.

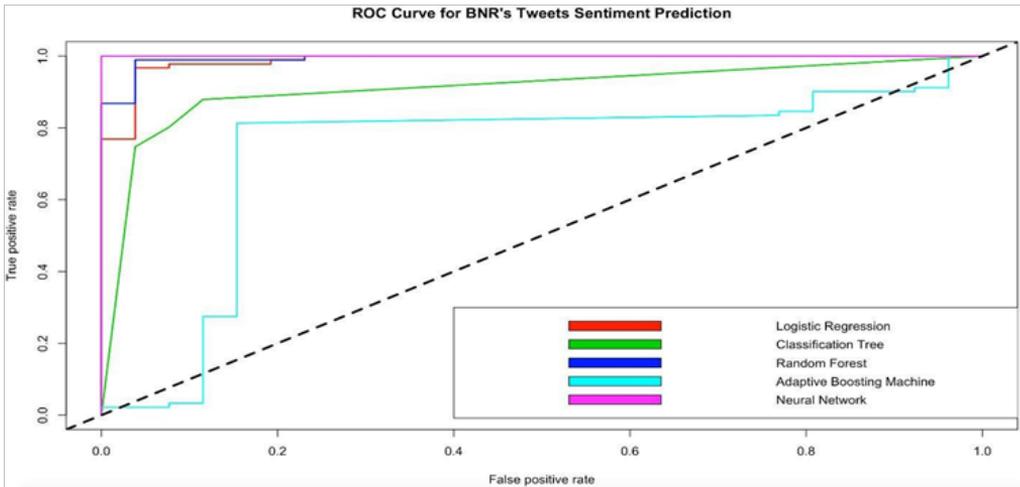
**Figure 5: Public sentiment evolution over time**



Source @CentralBankRw



**Figure 7: Receiver operating curve for sentiment prediction**



The Logistic Regression's accuracy in the test data set is 96.3% with an F-measure (F1 score) of 98.7%, Cohen's Kappa statistics of 90.2%, and area under AUR curve of 98.8%. The CART showed an accuracy of 89.18% in prediction of the sentiment polarity of the test observations, with an F-measure of 85.004%, Cohen's Kappa statistics of 63.27%, and AUC of 85.37%.

The Random Forest algorithm attained a relatively higher accuracy of 90.88% compared to CART with the precision-recall proportion of 88.62%; Cohen's Kappa of 74.26% and relatively higher AUC of 88.99%, whereas Adaptive Boosting Machine lies in-between the Logistic Regression and Artificial Neural Network with an accuracy of 91.64% and an F-measure of 87.98%.

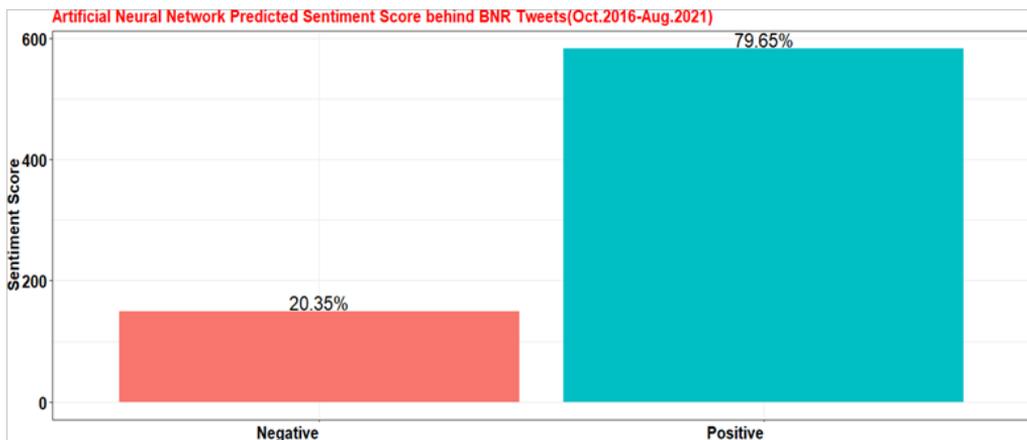
Among the five-trained classifiers, the greatest results have been achieved by the Artificial Neural Network (ANN) algorithm on all measures having the highest accuracy of 91.64%, and F-measure of 87.98%, a Cohen's Kappa statistics of 91.72% that suggests the perfect agreement between actual and predicted sentiment polarities. The ANN also is having the greater discriminating power of positive from negative sentiments since it covers 99.2% of the area under the ROC curve, as shown in Figure 8. Thus, the identification of optimal out-performed model is crucial in machine learning, especially when it comes to the effective and accurate prediction.

There are several existing criteria to select the optimal model taking into consideration the problem at hand. In the case of our study, the issue of imbalanced sentiment polarities is observed with 38.2% being negative and 61.8% being positive. To overcome the issue of imbalanced data set, Alghamdi and colleagues suggest that the Cohen's Kappa statistics and AUC are the best model performance evaluation metrics to be used for optimal model selection when dealing with binary classification (Hassan, 2012). With this in mind, the Artificial Neural Network (ANN) has been selected as the best performer model since it displays the best results (Table 2 and Figure 8)

compared to other models. Therefore, the empirical results suggest that the ANN should be used in predicting sentiment polarities.

For the near future, the predictive results show that positive sentiment will continue to dominate at 79.65%. This was expected since it validates the descriptive statistics and corroborates with previous studies done in Austin (US), where the model indicated positive outlook from the past data (Hsuanwei et al., 2016). Their findings revealed that the citizens' trust in their city has been largely positive and continues to remain positive. The assumption made was that, when the public trusts the institution's actions it takes time to wane. In addition, these findings allowed us to validate our third objective, that ANN in the case of NBR outperforms other machine learning classifiers in predicting the public sentiment. Figure 8 is a visual representation of the ANN predictive results, providing insights on how ANN almost agrees with the actual sentiment polarities prevailing in the extracted tweets.

**Figure 8: ANN predicted sentiment score**



With high results of positive emotions and sentiments from our sentiment score, the National Bank of Rwanda is challenged to keep building trust and be a reliable source for Twitter users. Trust in the central bank is a very valuable resource in the conduct of monetary policy. It is, therefore, very critical in today's environment for central banks to safeguard and build positive feelings and sentiments in economic agents as part of their communication strategy to drive the economy in the expected trajectory.

## 5. Conclusion, policy implications and limitations

In this study, we conducted a socio media sentiment analysis in the context of public feedback to the NBR Twitter posts. Two techniques that include a lexicon-based approach and a machine learning-based approach were used to perform text classification and prediction. The selected data provided one of the first attempts to explore and compare sentiment analysis techniques in the context of a central bank in Africa in use of social media. Our paper contributes to the body of literature to understand how sentiment analysis techniques can perform in the same way or differently in the context of central bank practices of social media. Our study also proved that the understandings of central bank social media data analysis on public feedback can significantly affect policy making and improve trust and relationship with the public. The analysis results revealed ex-ante reactions of the public towards the economic situations in the country and ex-post feedbacks towards the NBR events, communications, and activities.

Our paper also denotes how sentiment analysis results are useful to identify trends and patterns of public sentiment driven by unique events and announcements. The paper's findings have two implications. First, the public sentiment towards the central bank, in our case the NBR, can be influenced by its events, announcements, activities to name a few, so it is of paramount importance for central banks to consider public's opinions expressed via social media in the process of policy and decision making. Second, sentiment analysis has demonstrated to be an important tool for both recognizing current sentiments and predicting future sentiments.

Social media sentiment analysis using novel machine learning and lexicon-based approaches should be integrated into central banks strategy to mine the public's sentiment, in effort to hear their voice, promote public trust, accountability, and increase transparency of the central bank's activities. This will make the public more engaged and attentive with eagerness to contribute to the effectiveness of monetary policy implementation.

Also, social media presents an alternative source of data to tap into understanding economic agents' behaviours, expectations and perceptions that can replace costly economic surveys targeting businesses and individuals.

There are some limitations to this study. First, while we managed to carry a sentiment analysis using machine learning and lexicon-based approaches, there are other alternatives available that could provide variations in algorithms and accuracy.

How to choose and develop a benchmark model to provide a meaningful comparison will be an important matter. Second, the sentiment analysis techniques examined in this paper did not take into consideration the uses of Kinyarwanda, French, and Swahili languages as feedback on the NBR's posts. Therefore, sentiments dictionaries for low resources languages and natural language processing algorithms that take those languages into consideration are needed to capture these opinions and sentiments.

The future directions of our study will include an incorporation of French feedbacks in text analysis and development of a social media index. This index can be studied to investigate its relationship with other macroeconomic indicators, such as exchange rate, consumer price index, GDP, amongst others, to improve economic analysis and monetary policy analysis in central banks.

## References

- Angeletos, G.-M. and L. Jennifer. 2013. "Sentiments". *Econometrica*, 2: 739–80. Récupéré sur <https://economics.mit.edu/files/8394>
- Apoorv, A., X. Boyi, V. Ilia, R. Owen and P. Rebecca. 2010. "Sentiment analysis of Twitter data". Columbia University, 1-9. At <http://www.cs.columbia.edu/~julia/papers/Agarwaletal11.pdf>
- Ausserhofer, J. and A. Maireder. 2013. "National politics on Twitter: Structures and topics of a networked public sphere". . *Information, Communication & Society*. Volume 16; Pages 291–314
- Azar, P.D. 2016. "The wisdom of twitter crowds: Predicting stock market reactions to FOMC meetings via Twitter feeds". *Journal of Portfolio Management*, 113. Vol.42 (5) 123–134; DOI: <https://doi.org/10.3905/jpm.2016.42.5.123>
- Babatunde, O. and T. Mohammed. 2019. A Text Mining Analysis of Central Bank Monetary Policy Communication in Nigeria.
- Baharudin, B. 2010. "A review of machine learning algorithms for text-documents classification". *Journal of Advances in Information Technology*. doi:10.4304/jait.1.1.4–20
- Bannach-Brown, A. 2018. "The use of text-mining and machine learning algorithms in systematic reviews: Reducing workload in preclinical biomedical sciences and reducing human screening error". *BioRxiv*. doi:<https://doi.org/10.1101/255760>
- Blinder, A.S., M. Ehrmann, M. Fratzscher, J.D. Haan and D.-J. Jansen. 2008. "Central bank communication and monetary policy: A survey of theory and evidence". *Journal of Economic Literature*, 46(4): 910–45. Récupéré sur <https://www.jstor.org/stable/27647085>
- Blot, C. and P. Hubert. 2018. "Central bank communication during normal and crisis times". Sciences Po publications, info:hdl:2441/52p48pif509, Sciences Po.
- Bollen, J.H. 2011. "Twitter mood predicts the stock market". *Journal of computational science*, 2(1): 1–8.
- Brechler, J., V. Hausenblas, Z. Komárková and M. Plašil. 2019. Similarity and Clustering of Banks: Application to the Credit Exposures of the Czech Banking Sector. Prague: Czech National Bank.
- Brendan, T., B. Ohana and S. Jane Delany. 2016. "Sentiment classification using negation as a proxy for negative sentiment". The Twenty-Ninth International Flairs Conference. Florida Artificial Intelligence Research Society. 03–30
- David Bholat & Stephen Hans & Pedro Santos & Cheryl Schonhardt-Bailey, 2015. "Text mining for central banks," Handbooks, Centre for Central Banking Studies, Bank of England, number 33
- Dmitriy, P. 2020. "Social media adoption and usage in central banking". *Institute for Complex Strategic Studies*, 413–24. doi:10.1007/978-3-030-39296-3\_30

- Donato , M. and R. Davide. 2016. "From silence to voice: Monetary policy, central bank governance and communication". BAFFI CAREFIN Centre, 19. doi:<http://dx.doi.org/10.2139/ssrn.2815779>
- Emily A. Vogels, Brooke Auxier & Monica Anderson. 2020. Partisan differences in social media use show up for some platforms, but not Facebook. Pew Research Center, 2–19.
- Eriksson, M. 2018. "Lessons for crisis communication on social media: A systematic review of what research tells the practice". *International Journal of Strategic Communication*, 526–51. doi:<https://doi.org/10.1080/1553118X.2018.1510405>
- Farhadloo, M. and E. Rolland. 2016. "Fundamentals of sentiment analysis and its applications. *Sentiment Analysis and Ontology Engineering*, 1–24). doi:10.1007/978-3-319-30319-2\_1
- Gholampour, V. and E. van Wincoop. 2017. "What can we learn from euro-dollar tweets?" NBER Working Paper No. 23293. National Bureau of Economic Research, Cambridge, Massachusetts, March.
- González-Bailón, Sandra & Wang, Ning & Rivero, Alejandro & Borge-Holthoefer, Javier & Moreno, Yamir. (2014). Assessing the Bias in Samples of Large Online Networks. *Social Networks*. 38. 16–27. 10.1016/j.socnet.2014.01.004.
- Hansen, S.M. 2014. "Transparency and deliberation within the FOMC: A computational linguistics approach". *The Quarterly Journal of Economics*, 133(2): 801–70.
- Hassan, S., Y. He and H. Alani. 2012. *Semantic Sentiment Analysis of Twitter*. Berlin/Heidelberg: Springer.
- Hassan, S., Y. He, M. Fernandez and H. Alani. 2015. "Contextual semantics for sentiment analysis of Twitter". *Information Processing & Management*. doi:10.1016/j.ipm.2015.01.005
- Hsuanwei, M.C., C.F. Patricia and E. Lois. 2016. "Exploring government uses of social media through Twitter Sentiment". *Journal of Digital Information Management*, 14(4): 290–301. doi:10.6025/jdim/2016/14/5/290-301
- Janikowski, L. and A. Rzonca. 2018. *Central Bank Communication at Times of Non-Standard Monetary Policies*.
- Latha, I.H. 2019. "Sentiment analysis tool using machine learning algorithms". *Data Mining Project*, 14791–794.
- Masciandaro, D., D. Romelli and G. Rubera. 2020. *Tweeting on Monetary Policy and Market Sentiments: The Central Bank Surprise Index*.
- Meinusch, A. and P. Tillmann. 2017. "Quantitative easing and tapering uncertainty: Evidence from Twitter". *International Journal of Central Banking*, 227–58. At <https://www.ijcb.org/journal/ijcb17q4a7.htm>
- Miller, G.A. 1995. "WordNet: A lexical database for English". *ACM Digital Library*.
- Mohammad, S.M. 2017. *Challenges in Sentiment Analysis*. Berlin/Heidelberg: Springer.
- Nkuttner, K. 2001. *Monetary Policy Surprises and Interest Rates: Evidence from the Fed Funds Futures Market*. Amsterdam: Elsevier.
- Olteanu, A., C. Castillo, F. Diaz, E. Kiciman. 2016. "Social data: Biases, methodological pitfalls, and ethical boundaries. *Frontiers in Big Data* 2:13. <https://doi.org/10.3389/fdata.2019.00013>
- Orphanides, A. 2019. *Monetary Policy Strategy and its Communication*. Cambridge: Cambridge, Massachusetts.
- Pfeffer, J.M. 2018. "Tampering with Twitter's sample API". *EPJ Data Sci.*, 7: 1–3.

- PwC. 2020. Central Banking 2020: Ahead of the Curve. PwC. At <https://www.pwc.com/gx/en/financialservices/pdf/cb-2020-ahead-of-the-curve.pdf>
- Rachael, K. and Š. Karolina. 2020. Amplifying the Soundwaves: The Evolution of Social media Communications. Consulté le July 17, 2021. At <https://www.centralbanking.com/central-banks/monetary-policy/communication/7517146/amplifying-the-soundwaves-the-evolution-of-social-media-communications>
- Rasche, W.P. 2000. Perfecting the Market's Knowledge of Monetary Policy. Berlin/Heidelberg: Springer.
- Sergio R. Munoz & Shrikant I. Bangdiwala (1997) Interpretation of Kappa and B statistics measures of agreement, *Journal of Applied Statistics*, 24:1, 105-112, DOI: 10.1080/02664769723918
- Sreeja, I & Sunny, Joel & Jatian, Loveneet. (2020). Twitter Sentiment Analysis on Airline Tweets in India Using R Language. *Journal of Physics: Conference Series*. 1427. 012003. 10.1088/1742-6596/1427/1/012003.
- Labondance, P.H. 2018. "Central bank sentiment". OFCE, 35.
- StatsCounter. 2021. <https://gs.statcounter.com>. Consulté le October 13, 2020. At <https://gs.statcounter.com/social-media-stats/all/rwanda>
- Stefano Baccianella, A.E. 2006. SENTIWORDNET 3.0: An Enhanced Lexical Resource for Sentiment Analysis and Opinion Mining.
- Varol, O., E. Ferrara, F. Menczer and A. Flammini. 2017. "Early detection of promoted campaigns on social media". *EPJ Data Sci*, 6(1): 16.
- Youngjoon Lee, S., Kim and K.Y. Park. 2019. "Measuring monetary policy surprises using text mining: The case of Korea". Bank of Korea Working Paper No. 2019-11. Bank of Korea, April.
- Zhang, X., H. Fuehres and P.A.Gloor. 2011. "Predicting stock market indicators through Twitter 'I hope it is not as bad as I fear'". *Procedia – Social and Behavioral Sciences*, 26: 55–62.



## Mission

To strengthen local capacity for conducting independent, rigorous inquiry into the problems facing the management of economies in sub-Saharan Africa.

The mission rests on two basic premises: that development is more likely to occur where there is sustained sound management of the economy, and that such management is more likely to happen where there is an active, well-informed group of locally based professional economists to conduct policy-relevant research.

[www.aercafrica.org](http://www.aercafrica.org)

## Learn More



[www.facebook.com/aercafrica](https://www.facebook.com/aercafrica)



[www.instagram.com/aercafrica\\_official/](https://www.instagram.com/aercafrica_official/)



[twitter.com/aercafrica](https://twitter.com/aercafrica)



[www.linkedin.com/school/aercafrica/](https://www.linkedin.com/school/aercafrica/)

## Contact Us

African Economic Research Consortium  
Consortium pour la Recherche Economique en Afrique  
Middle East Bank Towers,  
3rd Floor, Jakaya Kikwete Road  
Nairobi 00200, Kenya  
Tel: +254 (0) 20 273 4150  
[communications@ercafrica.org](mailto:communications@ercafrica.org)