International Journal on Semantic Web and Information Systems
Volume 21 • Issue 1 • January-December 2025

Optimizing Stock Predictions With Bi-Directional LSTM and Levy Flight Fuzzy Social Spider Optimization (LFFSSO): LSTM Model

Deepa Raghunathan
https://orcid.org/0000-0003-2434-2198
Anna University, India & Dr. N.G.P. Institute of Technology, India

M. Krishnamoorthi
Dr. N.G.P. Institute of Technology, India

ABSTRACT

Stock Market Prediction (SMP) has developed into a significant area of research, especially in recent decades. Major novelty of the work is to develop an Evolutionary Bidirectional Long Short-Term Memory (EBi-LSTM) framework depends on investors' sentiment in tweets to Stock Market (SM). In addition, three feature selectors: the Chi-Square Test (CST), Analysis Of VAriance (ANOVA) technique and Mutual Information (MI) method are introduced for the selecting most important features. Levy Flight Fuzzy Social Spider Optimization (LFFSSO) algorithm is used for optimal tuning of parameters in the Bi-LSTM classifier. EBi-LSTM algorithm has been worked on datasets like Twitter, Stock, Weather, and Coronavirus disease (COVID-19). The proposed model extends the Valence Aware Dictionary and sEntiment Reasoner (VADER), TextBlob, and robustly optimized Bidirectional Encoder Representations from Transformers Retraining Approach (RoBERTa) for sentiment analysis. Highest results of 88.26%, 90.43%, 89.33% and 92.63% for precision, recall, F1-score and accuracy has been attained by proposed system.

KEYWORDS

Evolutionary Bidirectional Long Short-Term Memory (EBi-LSTM), Housing Development Finance Corporation (HDFC), Valence Aware Dictionary and Sentiment Reasoner (VADER)

INTRODUCTION

The COVID-19 virus of 2020 shook the world and raised concerns about the economy and worldwide health. The unpredictability of COVID-19 caused a worldwide decline in the stock market (SM), and the H1N1 pandemic had a similar effect. Numerous studies were conducted to comprehend the impacts of pandemics in various contexts. Specifically, the opinions shared on X are highly correlated with the SM. The sentiment, feelings, or meaning of big data can be identified by analyzing many forms of data, including text, photos, audio, and video, using artificial intelligence (AI) methods.

Because of the dynamic and unpredictable behavior of the SM, SMP has been regarded as an important and difficult task (Li & Wang, 2020; Arun et al., 2024). Since the SM is frequently seen

DOI: 10.4018/IJSWIS.367280

as erratic and volatile, maximizing profits and minimizing losses can be difficult, particularly in the banking industry. Implementing a useful strategy for "beating the market" has required substantial research. Considerable attention has been devoted in recent years to sentiment analysis (SA)—the study of classifying individual views and expressions as positive, negative, or neutral —utilizing natural language processing for social media channels and to examine the impact on the SM of opinions expressed on social media, examining previous information and identifying trends.

The social media network X is the ninth most popular website worldwide, with 300 million active users per month. Short messaging and content creation are combined in micro-blogging (MB). Common sites for MB about the SM include Facebook, Instagram, Pinterest, and X. Users of assured MB platforms can limit who can access their MB and publish entries using methods other than the web-based user interface. Among these, X is the most popular. Initially, it was intended to function as a messaging app allowing individuals to monitor one another. Real-time SM data can be accessed by a source through the stock application programming interface (API), a programming interface.

Researchers have been prompted to investigate X's potential impact beyond social networks because of its easily accessible API, which preserves posts that researchers may find useful, and its user-friendly features, which allow for filtering based on location and keywords (Arun et al., 2024; Nisar & Yeung, 2018). Research has consistently focused on its users in determining X's structural characteristics as a social channel. SMP is one of the hardest tasks and has occupied financial researchers and analysts for over 50 years (Zhao et al., 2023; Neelakandan & Paulraj, 2020). There has been almost no research on this subject during the COVID-19 pandemic.

Though the literature offers several definitions of SA, the best way to characterize SA in data extraction is as an application of SA. Opinion mining analyzes opinions, emotions, experiences, feelings, and actions in text form, emphasizing sentiment and emotions (Xing et al., 2018). Several investigations have focused on investigating SA's impact on SMP. There is a significant relationship between article publishing in SA and SM volatility.

Machine Learning (ML) are presented for sentiment prediction. The dynamic and ever-changing nature of SM, health concerns, and climate change affect the accuracy of stock prediction. Utilizing methods like Naive Bayes (NB), Support Vector Machine (SVM), and Deep Learning (DL), numerous SA research has been attempted at different levels. According to recent studies, the Recurrent Neural Network (RNN) version LSTM can extract stock time series data and has high SMP capabilities (Van Houdt et al.,2020). For DL approaches to work, a lot of training data is required.

However, previous investigations have provided unreliable predictions owing to the limited quantity of textual data that are received and processed. Because of the unpredictability of both the SM and the COVID-19 pandemic, particularly in relation to the banking industry, SMP is difficult in the field of economics (Al-Awadhi et al., 2020; Cepoi, 2020; Buckman et al., 2020).

This study's contribution is the use of SA on X and stock data to create an evolutionary bi-directional long short-term memory (EBi-LSTM) framework for SMP volatility. The Housing Development Finance Corporation (HDFC) Bank assessed and validated this strategy utilizing stock movement and sentiment data. Collected stock data, bank information, and tweets from Twitter. LFFSSO has been utilized to optimize the parameters of the Bi-LSTM classifier. Accuracy, precision, recall, and F1-score are some of the metrics used to assess the performance of the proposed classifier and the current classifiers.

LITERATURE REVIEW

Investors can quickly provide information about the stock's performance via Bardhan and Vaghela's (2021) consideration of two important factors: previous costs and social media data. Thus, to predict the stock performance, they integrate the opinions of various stakeholders from the internet with the stock's previous price. A decision tree is introduced for both prediction and classification to integrate