**Title:** Unveiling Sentiment Analysis Models in Stock Market: Exploring Feature Extraction Techniques and Diverse Models

**Abstract:**

Sentiment analysis in the stock market has become increasingly vital in today's financial landscape, offering insights into investor sentiment and market trends. Stock price prediction has been a topic of interest for researchers and analysts for a long time. But prediction of stock prices has always been difficult due to everchanging market. Previous studies in sentiment analysis have revealed that there is a strong correlation between stock prices and the published news articles.

This paper extracts features from time series data and social networks to investigates the efficiency of various feature extraction techniques and diverse modelling approaches in sentiment analysis for stock market prediction. This study explores feature extraction methods including TF-IDF Vectorizer, Bag-of-Words, Word2Vec, and Word Embedding, evaluating their ability to capture sentiment from textual data related to financial markets. Furthermore, a range of machine learning models such as MultinomialNB, Random Forest Classifier, Logistic Regression, Recurrent Neural Networks (RNN), and Long Short-Term Memory (LSTM) networks are employed to analyse the effectiveness of different modelling techniques in predicting stock market sentiment. Through comprehensive experimentation and evaluation, this research aims to find out what is the most accurate and precise way through which one can analyse stock market. Experimental results indicate that using Word Embedding with Long Short-Term Memory (LSTM) RNN outperforms any other combination in our experiment.

1. **Introduction**

In the world of financial market, Sentiment analysis as proven to be the most powerful tool for investors, analysts, and traders alike. By deciphering the sentiments embedded in various news articles, social network or any textual data market participants can gain valuable insight of stock market and its movement. In recent years, the application of sentiment analysis in the stock market has witnessed a surge, fuelled by advancements in natural language processing (NLP) techniques and machine learning algorithms.

This research paper delves into the intricate domain of stock market sentiment analysis, focusing on the exploration of various feature extraction techniques and diverse modelling approaches. Specifically, we investigate the efficacy of different feature extraction methods including TF-IDF Vectorizer, Bag-of-Words, Word2Vec, and Word Embedding, in capturing the nuanced sentiment expressed in textual data pertaining to financial markets.

Feature extraction serves as a critical preprocessing step in sentiment analysis, where raw textual data is transformed into numerical representations that can be fed into machine learning algorithms. TF-IDF Vectorizer computes the importance of each word in a document relative to a corpus, while Bag-of-Words represents text as a sparse matrix of word frequencies. On the other hand, Word2Vec and Word Embedding techniques aim to capture semantic relationships between words by mapping them to continuous vector space representations.

Furthermore, this paper explores the performance of diverse machine learning models in sentiment analysis tasks, leveraging the features extracted through the techniques. Various models such as Support Vector Machines (SVM), Random Forest, Recurrent Neural Networks (RNN), and Long Short-Term Memory (LSTM) networks are evaluated to discern their effectiveness in predicting stock market sentiment.

By comprehensively analysing the interplay between feature extraction methods and modelling techniques, this research aims to contribute to the advancement of sentiment analysis models in the context of financial markets. The insights garnered from this study not only hold implications for investors seeking to make informed decisions but also pave the way for the development of more robust and accurate sentiment analysis tools tailored for stock market forecasting.