**Introduction:**

Sentiment analysis or opinion mining is a process that uses natural language processing (NLP) and machine learning (ML) to determine the emotional tone of a text. Sentiment analysis is used to analyze raw text to drive objective quantitative results using natural language processing, machine learning, and other data analytics techniques. It is used to detect positive or negative sentiment in text, and often businesses use it to gauge branded reputation among their customers.

In the current digital era, social media has become an increasingly important platform for facilitating interaction and information exchange among users worldwide. Twitter, being one of the most popular social media platforms, provides a data-rich environment that reflects users' opinions, trends, and activities in real-time [1]. Twitter data, with its continuous nature and sequential collection over periods, can be a valuable source for understanding and predicting consumer behavior and market trends in the future[2]

A **Time Series** is defined as a series of data points recorded at different time intervals.

Time Series forecasting is the process of using a statistical model to predict future values of a time series based on past results. Forecasting is the step where we want to predict the future values the series is going to take. Forecasting a time series is often of tremendous commercial value.

ARIMA stands for Autoregressive Integrated Moving Average Model. It belongs to a class of models that explains a given time series based on its own past values -i.e. its own lags and the lagged forecast errors. The equation can be used to forecast future values. Any ‘non-seasonal’ time series that exhibits patterns and is not a random white noise can be modelled with ARIMA models.

ARIMA Models are specified by three order parameters: (p, d, q), where, p is the order of the AR term, q is the order of the MA term and d is the number of differencing required to make the time series stationary.

In the competitive business world, accurate understanding of market segments and potential changes in the future is crucial for entrepreneurs and marketers. In this context, using text analysis and time series data methods to predict market segments based on Twitter data has become an intriguing research subject. In previous research, time series methods such as Autoregressive Integrated Moving Average (ARIMA) have been successfully applied in predictive analysis for various domains [3], including economics, finance, and sales.

However, despite the great potential of Twitter data for predicting market segments, the optimal utilization of such data is still unrealized. The overall volume of tweets generated on Twitter is massive, reaching an average of 493.354.38 tweets per year over the past eight years. Yet, there are still challenges in transforming this large and diverse volume of Twitter data into reliable and useful insights for business decision-makers.

despite the great potential of Twitter data for predicting market segments, the optimal utilization of such data is still unrealized. The overall volume of tweets generated on Twitter is massive, reaching an average of 493.354.38 tweets per year over the past eight years. Yet, there are still challenges in transforming this large and diverse volume of Twitter data into reliable and useful insights for business decision-makers.

our study aims to build a model for collecting and analyzing Twitter data [4, 5, 6, 7, 8, 9]. However, based on previous research, it is acknowledged that the forecasting or prediction results achieved are not fully aligned with actual data, with accuracy rates varying at 96.76%..

objectives:

Objective:

The objective of our project is to build a model that analyses the sentiment of social media posts(e.g. tweets) Overtime and predicts trends .this could help businesses, researchers or analysts understand public opinion about products, events or topics.

problem definition and challenges:

The main problem addressed in this project is to classify the polarity of a given text as positive, negative, or neutral. The goal is to understand how people feel about a product, service, or topic.

.

REFERENCE

[1] Zhao, X., Li, W., & Liu, B. (2023). Twitter Data Analysis for Market Segment Prediction Using ARIMA Time Series Analysis and Network Analysis. Information Processing & Management, 60(6), 102571.

[2] Yu, J., Zhang, L., & Li, C. (2023). Market Segmentation Forecasting Using Twitter Data and ARIMA Time Series Analysis: A Comparative Study of Different Sentiment Analysis Approaches. Journal of Interactive Marketing, 61, 15-29.

[3] Zhu, Y., Zhang, L., & Wu, J. (2019). Predicting Market Segments from Twitter Data Using ARIMA Time Series Analysis. International Journal of Information Management, 49, 13-24.

[4] Zhang, H., Zhao, M., & Li, Q. (2023). Predicting Market Segments from Twitter Data Using ARIMA Time Series Analysis: A Comparative Study. Journal of Marketing Analytics, 9(1), 45-59.

[5] Wang, L., Zhang, Y., & Chen, X. (2019). Market Segmentation Forecasting Based on Twitter Data Using ARIMA Time Series Analysis and Machine Learning Techniques. International Journal of Forecasting, 35(2), 537-549.

[6] Huang, L., Zhang, W., & Chen, H. (2020). Twitter Data Analysis for Market Segment Prediction Using ARIMA Time Series Analysis and Deep Learning Models. Information Sciences, 511, 1-15.

[7] Zhou, L., Wang, F., & Zhao, Y. (2021). Predicting Market Segments from Twitter Data Using ARIMA Time Series Analysis: A Long Short-Term Memory Approach. Expert Systems with Applications, 181, 115045.

[8] Li, J., Yang, Q., & Liu, Y. (2022). Market Segmentation Prediction Based on Twitter Data Using ARIMA Time Series Analysis and Topic Modeling. Decision Support Systems, 152, 113529.

[9] Zhang, T., Chen, Z., & Wang, H. (2022). Predicting Market Segments from Twitter Data Using ARIMA Time Series Analysis: A Hybrid Approach with Natural Language Processing. Journal of Business Research, 142, 373-385.