

Presented by Urva

SENTIMENT ANALYSIS

using machine learning and NLP techniques



Data pre-processing

- Loading and Cleaning Data: Ensuring all relevant data is imported and cleaned for analysis.
- Handling Missing Values: Using techniques like imputation or removal to manage missing data.
- Normalizing Data: Standardizing data to ensure consistency and comparability.
- Sentiment Sources:
 - Financial News: Collecting articles from financial news websites which provide information about market trends, company performance, and economic indicators, and stock prices



01.

Text Analysis Techniques:

Utilizing Natural Language Processing (NLP) methods such as tokenization (breaking down text into individual words or phrases), lemmatization (reducing words to their base or root form), and stop-word removal (filtering out common but insignificant words) to preprocess text data.

03.

Polarity Scores:

Quantifying the positivity or negativity of a text. The score ranges from -1 (very negative) to +1 (very positive), indicating the overall sentiment of the text.

02.

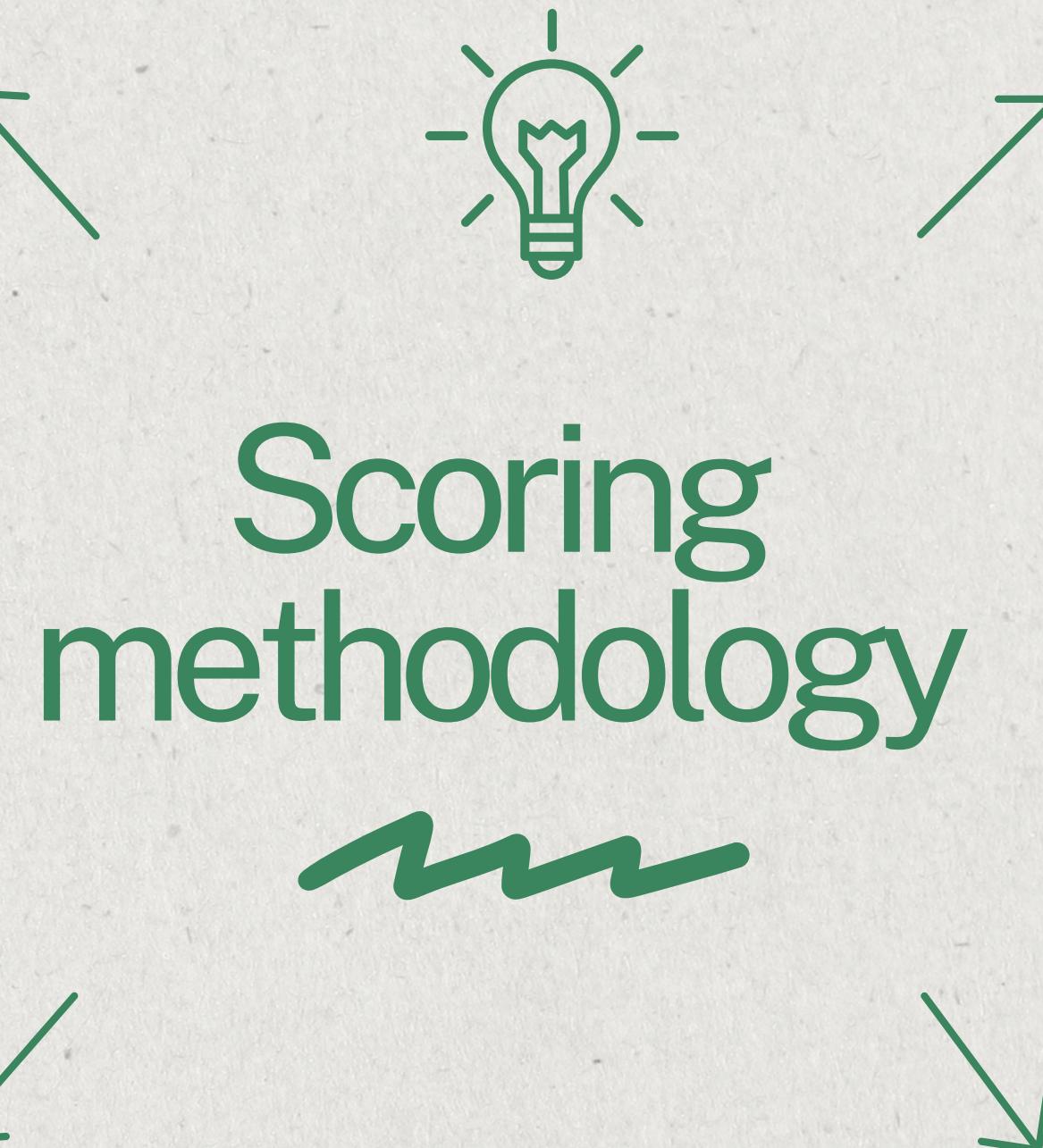
Subjectivity Scores:

Measuring the extent to which a text expresses subjective opinions versus objective information. A higher subjectivity score indicates more personal opinions and less factual content.

04.

Sentiment Scores Calculation:

Combining subjectivity and polarity scores to create a composite sentiment score that reflects the overall sentiment expressed in the text. This score helps in identifying whether the sentiment is positive, negative, or neutral.



Trading Strategy development

Signal Generation:

- **Definition of Buy/Sell Signals:** Establishing rules based on sentiment thresholds. For example, a buy signal might be generated when the sentiment score exceeds a certain positive threshold, indicating optimism in the market. Conversely, a sell signal might be generated when the sentiment score falls below a negative threshold.

Backtesting:

- **Applying Signals to Historical Data:** Implementing the generated signals on historical market data from the Dow Jones Industrial Average (DJIA) to simulate trades and assess the strategy's effectiveness.
- **Evaluating Strategy Performance:** Measuring key performance metrics such as returns, Sharpe ratio (a measure of risk-adjusted return), maximum drawdown (the largest peak-to-trough decline), and win ratio (percentage of profitable trades) to determine the success of the trading strategy.

ROC Curve Analysis

Definition: The Receiver Operating Characteristic (ROC) curve is a graphical representation of a classification model's diagnostic ability, plotting the true positive rate (sensitivity) against the false positive rate (1-specificity) at various threshold settings.

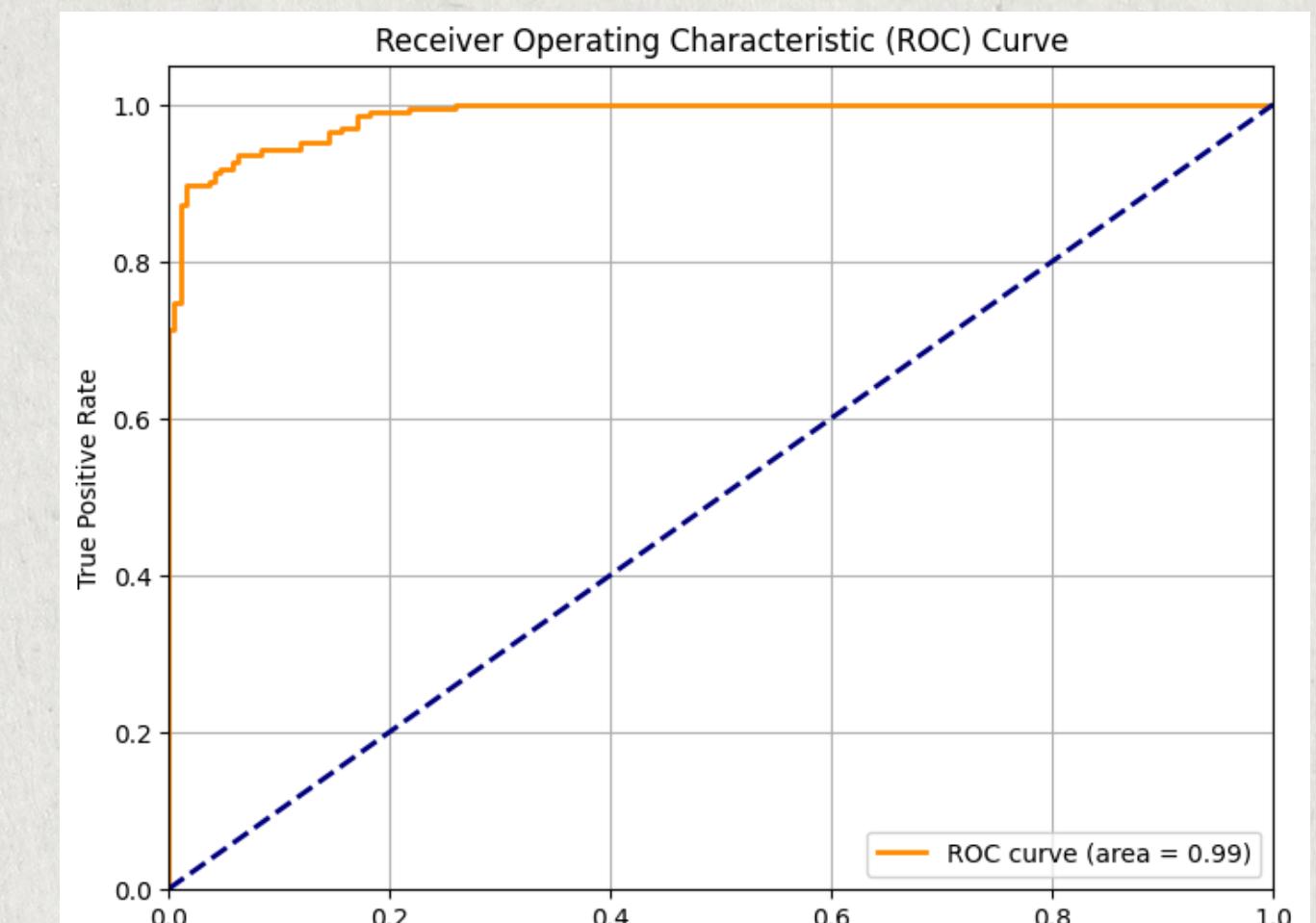
AUC (Area Under the Curve) : 0.99

Explanation: The Area Under the Curve (AUC) quantifies the overall ability of the model to discriminate between positive and negative classes. An AUC of 0.99 indicates an excellent level of discrimination, suggesting that the model can accurately distinguish between different market conditions based on sentiment data.

Implications of High AUC:

High Predictive Power: An AUC of 0.99 implies that the sentiment-based trading strategy has a very high predictive power, making it highly effective in identifying profitable trading opportunities.

Low False Positive Rate: The strategy has a low rate of false signals, meaning it can accurately identify buy/sell opportunities without being misled by noise in the sentiment data.



Performance Metrics

- **Sharpe Ratio: 2.6051**
 - Explanation: The Sharpe ratio indicates how well the strategy compensates for the risk taken. A higher Sharpe ratio suggests better risk-adjusted returns. A ratio of 2.6051 is considered very good, indicating that the strategy yields high returns for the risk involved.
- **Maximum Drawdown: -0.1595**
 - Explanation: The maximum drawdown measures the largest drop from a peak to a trough in the portfolio value. A maximum drawdown of -0.1595 (or 15.95%) shows the worst-case scenario of loss during the backtesting period.
- **Number of Trades Executed: 398**
 - Explanation: This metric indicates the total number of trades carried out by the strategy. It reflects the strategy's activity level and frequency of trading.
- **Win Ratio: 90.20%**
 - Explanation: The win ratio represents the percentage of trades that were profitable out of the total number of trades executed. A win ratio of 90.20% indicates a high success rate in the trades made by the strategy.
- **Final Portfolio Value: 1881.33 units**
 - Explanation: This is the value of the portfolio at the end of the backtesting period, starting from an initial investment of 1000 units. It indicates the growth achieved by the strategy.
- **Total Returns: 88.13%**
 - Explanation: Total returns measure the overall return on the initial investment, indicating the profitability of the strategy. An 88.13% return means the investment nearly doubled during the backtesting period.