

Stock Sentiment Analysis Report

Introduction

The aim of this project is to perform stock sentiment analysis using news headlines and technical indicators to predict stock price movements. We will train multiple machine learning models, including Logistic Regression, Random Forest, Support Vector Classifier (SVC), and a Convolutional Neural Network (CNN), to identify buy and sell signals. The best performing model will be used to generate trading signals and evaluate the performance of a trading strategy.

Project Contains 2 jupyter Notebook

- 1.DataScrapp-Data Collecting of News and Stocks
2. SentimentAnalysis- Contains Complete Analysis

1.Data Collection

I used New York Times API to scrap the news headlines as it only allows to collect data before 2019 so I collected the news Data from 2016 to 2019 containing all types of daily news headlines , for the stocks data I collected it from yfinance website for the above range of time.

2.Code Explanation

Data Preparation

Step 1: Import Libraries I start by importing the necessary libraries for data manipulation, machine learning, and visualization.

Step 2: Data Loading load the stock price data and news headlines data into pandas DataFrames.

Step 3: Data Preprocessing Convert the 'Date' columns to datetime format and merge the stock data with news data based on the date.

Sentiment Analysis

Step 4: Sentiment Analysis I use VADER and TextBlob for sentiment analysis on the news headlines. The sentiment scores are added as new features in the combined DataFrame.

Technical Indicators

Step 5: Add Technical Indicators calculate the Simple Moving Average (SMA) and the Relative Strength Index (RSI) using the ta library and add these as features to the combined DataFrame.

Feature Engineering and Model Training

Step 6: Prepare Data for Training created a new column 'return' to calculate the stock's return and generate binary labels for upward (1) or downward (0) movement. then prepare the feature matrix (X) and the target vector (y) for training.

Step 7: Initialize and Train Models initialized multiple models (Logistic Regression, Random Forest, SVC, and CNN) and evaluate them using cross-validation to determine the best model based on accuracy.

Trading Signals and Backtesting

Step 8: Generate Trading Signals We use the best model to generate trading signals (buy/sell) for the stock data.

Step 9: Backtest Strategy We implement a backtesting function to simulate the trading strategy based on the generated signals and calculate the final portfolio value.

Step 10: Calculate Performance Metrics calculated various performance metrics such as Sharpe ratio, maximum drawdown, and win ratio to evaluate the trading strategy.

Also Plotted graph of ROC curve and Graph showing buy and sell trading signal according to data

Name-Aditya Kurne

Enroll-22123022

EPH