Stock Prediction Using KNN, Naive Bayes, and Logistic Regression

Project Proposal

Title: Stock Prediction Using KNN, Naive Bayes, and Logistic Regression

Submitted by: Salman Khan Se202007

Rana Zargham Ahmed Se202020

Fuzail Mughal Se202034

Class: BS-SE 7A

Instructor: Zahid Hussain

Date: 6 June, 2024

Abstract

This project aims to predict stock prices using three different machine learning models: K-Nearest Neighbors (KNN), Naive Bayes, and Logistic Regression. The goal is to compare the effectiveness of these models in the context of stock market prediction and determine which model provides the most accurate forecasts.

Introduction

The stock market is inherently volatile and unpredictable. Investors and financial analysts have long sought tools that can provide an edge in predicting stock movements. Machine learning offers powerful techniques to model and predict stock prices based on historical data. This proposal focuses on three algorithms: KNN, Naive Bayes, and Logistic Regression, each offering different approaches to the prediction task.

Stock Prediction Using KNN, Naive Bayes, and Logistic Regression

Objectives

- Implement and train KNN, Naive Bayes, and Logistic Regression models on historical stock data.
- Evaluate and compare the performance of each model in terms of accuracy and efficiency.
- Determine which model is most suitable for stock price prediction.

Methodology

Data Collection

Data is collected from Kaggle, focusing on relevant features such as price change, typical price, mean deviation, RSI, CCI, ROC, EMA12, EMA26, MACD, Bollinger Bands, and momentum.

Feature Selection

The following features are selected for model training:

- Price Change
- Typical Price
- Mean Deviation
- Avg Gain and Avg Loss
- RSI (Relative Strength Index)
- CCI (Commodity Channel Index)
- ROC (Rate of Change)
- EMA12 and EMA26 (Exponential Moving Averages)
- MACD (Moving Average Convergence Divergence)
- Bollinger Bands (BB_Middle, BB_Upper, BB_Lower)
- Momentum
- Label

Stock Prediction Using KNN, Naive Bayes, and Logistic Regression

Model Implementation

- K-Nearest Neighbors (KNN): Features are scaled using StandardScaler, and the KNN classifier istrained with k=5 neighbors.

- Naive Bayes: Gaussian Naive Bayes is used, assuming features follow a Gaussian distribution.- Logistic Regression: Relevant financial indicators are selected as features, followed by feature scaling for normalization.

Model Training

Each model follows a standard workflow that includes data splitting, feature scaling, and training on historical stock data.

Challenges and Learning Points

- Handling Data Imbalance: Mitigated through preprocessing techniques and model evaluations. - Model Selection and Optimization: Emphasized the trade-offs between model complexity and performance.

Future Directions

- Deep Learning: Exploring deep learning techniques like RNNs and LSTMs.
- Real-Time Data Application: Implementing models for real-time stock trading insights.
- Feature Engineering: Incorporating external data sources such as economic indicators or newssentiment analysis.

Stock Prediction Using KNN, Naive Bayes, and Logistic Regression

Conclusion

This project demonstrated the applicability of machine learning in financial analytics, emphasizing the importance of data analysis, model choice, and continuous improvement. Future efforts will focus on integrating sophisticated models and diverse data sources to enhance predictive accuracy.

References

- TradingView: https://www.tradingview.com/

- Binance: https://www.binance.com/