DOI 10.36074/grail-of-science.07.06.2024.012

PREDICTING STOCK MARKET TRENDS WITH PYTHON

Mariia Kapinus
1-st Year of Study Master's Student
Faculty of Information Technologies
Taras Shevchenko National University of Kyiv, Ukraine

Kateryna Liashenko
Final Graduate of the Faculty of Mechanics & Mathematics
Taras Shevchenko National University of Kyiv, Ukraine

Scientific supervisors: Nikolina Ljepava
Dr , Department Chair (Management)
American University in Emirates UAE

Larysa Liashenko
Dr , Assistant Professor of the Department of Foreign Languages
Taras Shevchenko National University of Kyiv, Ukraine

Valerii Danylov
Dr , Professor of Institute of Applied System Analysis
Igor Sikorsky Kyiv Polytechnic Institute, Ukraine

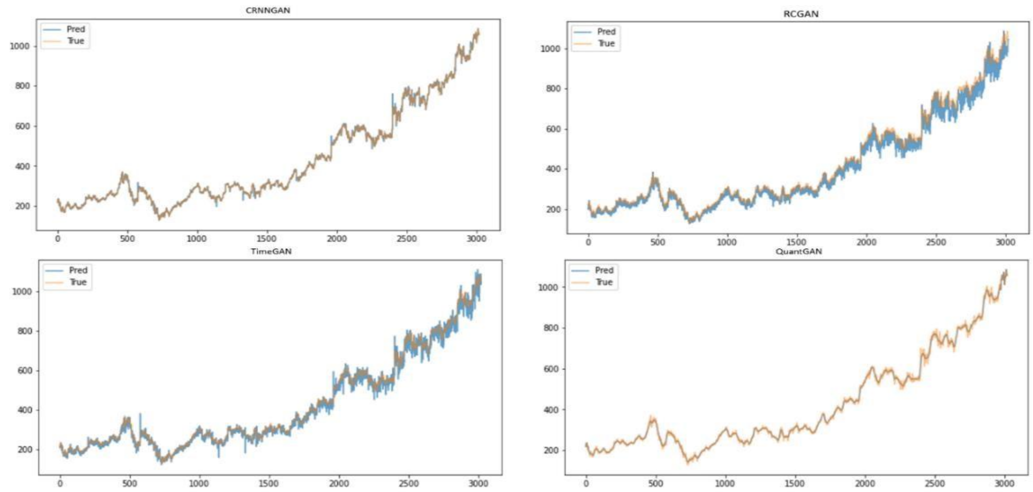
Summary. *Artificial Intelligence (AI) and machine learning (ML) have revolutionized the realm of stock market prediction, offering sophisticated tools to analyze vast volumes of data and anticipate market trends. This article provides a comprehensive overview of AI techniques, focusing on Python as the preferred platform for implementation. Beginning with an exploration of AI fundamentals, including machine learning and deep learning, it delves into various techniques employed for stock market prediction. Traditional statistical models such as linear regression and ARIMA are under scientific discussion alongside advanced ML algorithms like random forests and support vector machines. Moreover, the article highlights the efficacy of deep learning methodologies, particularly recurrent neural networks (RNNs) and long & short-term memory (LSTM) networks, in capturing temporal dependencies within stock market data. We also explored innovative developments such as Generative Adversarial Networks (GANs) for their potential in revealing hidden patterns influencing price movements. Throughout the discussion, we concluded that Python emerges as the preferred programming language due to its simplicity, extensive libraries, and versatility. Key libraries such as Pandas, NumPy, Scikit-learn, and TensorFlow play a pivotal role in data manipulation, preprocessing, and model development. The article outlines a structured approach to building predictive models, encompassing data collection, preprocessing, feature engineering, model selection, training, evaluation, and prediction. Despite the advancements in AI, challenges persist in stock market prediction, including market volatility, data quality issues, complexity of influencing factors, and risks of overfitting. Ultimately, we may witness AI and Python synergy, which*

empowers analysts and investors with deeper insights, enabling informed decision-making amidst the complexities of financial markets.
Keywords: artificial intelligence, machine learning, statistical models, deep learning, algorithm, programming language, stock market trends.

Artificial Intelligence (AI) is a multifaceted discipline that currently takes a remarkable place at the forefront of technological innovation, empowering machines to replicate human cognitive processes. Within the expansive realm of AI, Machine Learning (ML) emerges as a pivotal subset, employing algorithms and statistical models to execute tasks devoid of explicit programming instructions. Deep - Learning, a sophisticated branch of machine learning that uses neural networks to extract information directly from data sets, is developing this field. This hierarchical structure from AI's overarching scope down to the nuanced methodologies of ML and Deep Learning - serves as the cornerstone for developing intelligent systems capable of autonomous learning and decision-making. Through the amalgamation of cognitive modeling, algorithmic prowess, and data-driven insights, AI not only mimics but also extends the boundaries of human intelligence, revolutionizing industries and fostering groundbreaking advancements in technology. Now, let us delve into AI techniques for predicting stock market trends:

- *Statistical Models:* traders and analysts use traditional statistical methods such as linear regression, ARIMA (AutoRegressive Integrated Moving Average), and GARCH (Generalized Autoregressive Conditional Heteroskedasticity) to analyze time-series data and make forecasts. These models scrutinize historical stock prices and trading volumes to predict future price fluctuations based on observed patterns within the data [1] , [8].
- *Machine Learning Algorithms:* ML algorithms such as decision trees, random forests, support vector machines (SVM), and neural networks play a crucial role in stock market prediction. These algorithms are adept at handling intricate relationships among variables and capturing nonlinear patterns in the data. Ensemble methods like random forests amalgamate multiple models to enhance predictive accuracy [8].
- *Deep Learning*: Deep learning techniques, particularly recurrent neural networks (RNNs) and long short-term memory (LSTM) networks excel in analyzing sequential data. They excel in capturing temporal dependencies within stock market data, making them ideal for predicting short-term price movements and volatility [1], [8], [9]. For those delving into sequential data analysis, RNNs provide a potent tool. Unlike conventional models treating each data point in isolation, RNNs possess internal memory enabling them to learn from past sequences. This feature renders them particularly proficient in tasks such as stock price prediction, where comprehending historical trends and dependencies is pivotal. LSTM networks, a specific type of RNN, are especially suited for financial data due to their capability to handle long-term dependencies within sequences [8], [9], [10]. Generative Adversarial Networks (GANs), a groundbreaking development in deep learning, offer an innovative approach. These algorithms pit two neural networks against each other in an adversarial training process. The generator network endeavors to

generate synthetic data mimicking real market data seamlessly, while the discriminator network acts as a discerning critic, striving to identify synthetic data as counterfeit. This ongoing battle sharpens the generator's ability to capture intricate relationships within the data, potentially uncovering concealed patterns influencing price movements [7].



Pic. 1 Results achieved using GANs with RNNs, data unknown to models

However, mastery of these advanced tools requires a solid grounding in statistics and machine learning. Python stands out as the *favored platform* for deploying these techniques due to its wide array of libraries and adaptability. Although, the learning curve may be steep, the potential rewards are considerable. Proficiency in these methodologies enables investors to access a heightened predictive capacity, facilitating more informed investment choices and potentially yielding superior returns. Python is a popular programming language for AI and ML due to its simplicity and a vast array of libraries. Python provides a rich ecosystem of libraries for data analysis, machine learning, and deep learning, making it a preferred choice for stock market prediction tasks. Researchers commonly use several libraries in conjunction with each other for building and training deep learning models for stock market prediction:

1. *TensorFlow* or *PyTorch* as deep learning frameworks provide the foundation for building neural network architectures for stock market prediction. Both TensorFlow and PyTorch offer flexible APIs for defining and training deep learning models, allowing researchers and practitioners to implement various architectures such as convolutional neural networks (CNNs), recurrent neural networks (RNNs), long short-term memory networks (LSTMs), and transformer-based models like the Transformer and BERT [5], [6].
2. *Keras:* While TensorFlow has now integrated Keras as its high-level API, it's worth mentioning Keras separately because it provides a user-friendly interface for building and training deep learning models. Keras abstracts away the complexities of TensorFlow's low-level API, making it easier to prototype and experiment with different architectures and hyperparameters [5], [6].

3. *NumPy:* Users utilize NumPy for numerical computations and array operations, crucial for preprocessing input data and preparing it for training deep learning models. NumPy arrays serve common roles in representing stock market data, conducting normalization, scaling, and feature engineering, and constructing input tensors for training deep learning models [5].

4. *Pandas*: People use Pandas for data manipulation and analysis, especially for managing structured data like stock market datasets. The Pandas DataFrame is incredibly useful for tasks like loading, cleaning, and preprocessing historical stock price data, financial indicators, and other relevant features before using them as inputs for deep learning models [6].

❖ *Scikit-learn*: Although scikit-learn primarily focuses on classical machine learning algorithms, it also provides utilities for data preprocessing, feature selection, and model evaluation that can enhance deep learning approaches for stock market prediction. For example, you can utilize scikit-learn's preprocessing functions like StandardScaler or MinMaxScaler to scale input features, and its model evaluation tools to evaluate the performance of deep learning models [5], [6]. By leveraging these libraries in combination, researchers and practitioners can build and train deep learning models for stock market prediction, from preprocessing and feature engineering to model training and evaluation. Predicting stock market trends involves analyzing historical data to forecast future stock prices. This task is challenging due to the market's complexity and various influencing factors [6]. Now let us distinguish the main steps to build a predictive model:

❖ *Data Collection*: Gather historical stock market data, including prices, volumes, and other relevant metrics. Libraries like Pandas can help in data collection and manipulation [3], [4].

❖ *Data Preprocessing*: Clean the data by handling missing values, normalizing data, and encoding categorical variables. Pandas and Scikit-learn offer functions for data preprocessing [8].

❖ Feature Engineering: Create meaningful features from raw data to improve model performance. Features like moving averages, price-to-earnings ratios, and technical indicators can be useful.

❖ *Model Selection*: Choose an appropriate machine-learning algorithm based on the problem. For stock market prediction, algorithms like linear regression, decision trees, and ensemble methods (e.g., random forests) are commonly used [3], [4], [8].

❖ *Model Training*: Split the data into training and testing sets. Train the selected model on the training data using Scikit-learn's fit () function [3].

❖ *Model Evaluation*: Evaluate the model's performance using metrics like mean squared error (MSE), root mean squared error (RMSE), and R-squared. Scikitlearn provides functions for model evaluation [8].

❖ *Prediction*: Use the trained model to make predictions on new data. Visualize the predicted values alongside actual values using Matplotlib for analysis [3], [8].

Python Code Example

```
import pandas as pd
from sklearn.model_selection import train_test_split
from sklearn.ensemble import RandomForestRegressor
from sklearn.metrics import mean_squared_error
import matplotlib.pyplot as plt

# Load data
data = pd.read_csv('stock_data.csv')

# Data preprocessing and feature engineering
# ...

# Split data into features and target variable
X = data.drop('Close', axis=1)
y = data['Close']

# Split data into training and testing sets
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2, random_state=42)

# Initialize and train the model
model = RandomForestRegressor(n_estimators=100, random_state=42)
model.fit(X_train, y_train)

# Make predictions
predictions = model.predict(X_test)

# Evaluate model performance
mse = mean_squared_error(y_test, predictions)
print('Mean Squared Error:', mse)

# Visualize predictions vs actual values
plt.plot(y_test.index, y_test.values, label='Actual')
plt.plot(y_test.index, predictions, label='Predicted')
plt.xlabel('Date')
plt.ylabel('Stock Price')
plt.title('Stock Price Prediction')
plt.legend()
plt.show()
```

Challenges in Stock Market Prediction

It is crucial to acknowledge that even the most sophisticated models cannot guarantee perfect foresight. The inherent volatility of the stock market, coupled with the ever-present possibility of unforeseen events, necessitates a prudent approach. We should view AI models as invaluable aids for financial analysis, empowering investors with deeper insights to navigate the complex world of the stock market. Predicting stock market trends remains challenging due to several *factors:*

❖ *Market Volatility*: the stock market is inherently volatile, making it difficult to predict sudden price changes and market reactions to unforeseen events [10].

❖ *Data Quality:* stock market data can be noisy, incomplete, or subject to biases. Ensuring data quality and preprocessing are crucial for building accurate predictive models [2].

❖ *Complexity of Factors*: numerous factors influence stock prices, including economic indicators, company performance, market sentiment, and external events. Capturing all relevant factors in a model can be complex [1], [10].

❖ *Overfitting and Generalization*: ML models may overfit to historical data, capturing noise rather than true patterns. Ensuring model generalization and robustness to unseen data is essential [2].

In conclusion, Artificial Intelligence (AI) has emerged as a transformative force at the forefront of technological innovation, particularly through its subsets such as Machine Learning (ML) and Deep Learning. These methodologies, rooted in cognitive modeling and data-driven insights, have revolutionized various industries. Within stock market prediction, traditional statistical models, ML algorithms, and advanced deep learning techniques like recurrent neural networks (RNNs) and long short-term memory (LSTM) networks have shown promise in capturing complex market dynamics and forecasting trends. However, the journey towards mastering these techniques demands a solid foundation in statistics and machine learning, with Python emerging as the preferred programming language due to its rich ecosystem of libraries. By leveraging tools like TensorFlow, PyTorch, Keras, NumPy, Pandas, and scikit-learn, researchers and practitioners can build sophisticated predictive models that harness the power of AI for financial analysis. Despite the strides made in predictive modeling, challenges persist in accurately forecasting stock market trends. Market volatility, data quality issues, the complexity of influencing factors, and the risks of overfitting are among the hurdles that necessitate a cautious and pragmatic approach.

References:

[1] Akmaludheen K. K., D. D. (2023, May 19-21). Forecasting the Market: A Survey of Techniques.*2023 International Conference on Control, Communication and Computing.* P.6. URL: https://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=10164971

[2] Avci, R. (2023, March 27). *Challenges With Stock Price Prediction*. URL: https://medium.com/@rasim.avci/challenges-with-stock-price-prediction17e151bd79a7#:~:text=Data%20availability%20is%20a%20significant,used%20for%20tock%20price%20prediction

[3] Blumenfeld, Z. (2023, January 16). *The Definitive Guide to Building a Predictive Model in Python*. URL: https://neo4j.com/blog/build-predictive-model-python/

[4] Cousins, J. (2022, April 7). *6 steps to build a predictive model*. URL: https://eab.com/resources/blog/data-analytics-blog/predictive-modeling-101/

[5] Evans, D. (2023, December 13). *Hubspot*. Retrieved from The Complete Guide to Python AI Libraries.URL: https://blog.hubspot.com/website/python-ai-libraries

[6] Gravelle, R. (2023, July 14). *Top Python AI and Machine Learning Libraries*.URL: https://www.techrepublic.com/article/python-ai-libraries/

[7] Ian J. Goodfellow, J. (2014, June 10). Generative Adversarial Networks. *ArXiv*, p. 9. URL: https://arxiv.org/abs/1406.2661

[8] Kiprotich, B. (2023, March 31). *How to Build a Predictive Model Using Machine Learning with example. URL:* https://medium.com/@kiprotich01/how-to-builda-predictive-model-using-machine-learning-500be787127d
[9] Michael Biehl, N. P. (2023, January 12). *Survey of feature selection and extraction techniques for stock market prediction*. URL: https://jfinswufe.springeropen.com/articles/10.1186/s40854-022-00441-7
[10] Sekhon, G. (2023, January 27). URL: https://medium.com/@gianetan/challenges-in-predicting-stock-markets-in-general-andm-l-in-particular-6bff63c56787

ПРОГНОЗУВАННЯ ТЕНДЕНЦІЙ НА ФОНДОВОМУ РИНКУ ЗА ДОПОМОГОЮ PYTHON

Марія Капінус
студентка 1 курсу магістратури Факультету інформаційних технологій
Київський національний університет імені Тараса Шевченка, Україна
Катерина Ляшенко
студентка 4 курсу Механіко-математичного факультету
Київський національний університет імені Тараса Шевченка, Україна
Наукові керівники: Ніколіна Л'єпава
Д-р, Завідувачка кафедри менеджменту
Американський університет в Еміратах, ОАЕ
Лариса Ляшенко
Д-р, Асистент кафедри іноземних мов математичних факультетів
Київський національний університет імені Тараса Шевченка, Україна
Валерій Данилов
Д-р, Професор кафедри математичних методів системного аналізу
Національний технічний університет України «Київський політехнічний інститут імені Ігоря Сікорського», Україна

Анотація. *Штучний інтелект (AI) і машинне навчання (ML) зробили революцію у сфері прогнозування фондового ринку, запропонувавши складні інструменти для аналізу величезних обсягів даних і передбачення ринкових тенденцій. Ця стаття містить вичерпний огляд методів штучного інтелекту, зосереджуючись на Python як найкращій платформі для впровадження. Починаючи з вивчення основ штучного інтелекту, включаючи машинне та глибоке навчання, в статті поглиблено описуються різні методи, які використовуються для прогнозування фондового ринку. Традиційні статистичні моделі, такі як лінійна регресія та ARIMA, перебувають у стадії наукового обговорення разом із вдосконаленими алгоритмами машинного навчання, такими як випадкові ліси та опорні векторні машини. Крім того, у статті підкреслюється ефективність методологій глибокого навчання, зокрема рекурентних нейронних мереж (RNN) і мереж довготривалої та короткочасної пам'яті (LSTM), у фіксації тимчасової залежності в даних фондового ринку. Ми також досліджували інноваційні розробки, такі як Generative Adversarial Networks (GAN), на предмет їх потенціалу у виявленні прихованих закономірностей, що впливають на рух цін. Під час обговорення ми прийшли до висновку, що Python стає переважною мовою програмування завдяки своїй простоті, великим бібліотекам і універсальності. Ключові бібліотеки, такі як Pandas, NumPy, Scikit-learn і TensorFlow, відіграють ключову роль у маніпулюванні даними, попередній обробці та розробці моделей. У статті представлений структурований підхід до побудови прогнозних моделей, що включає збір даних, попередню обробку, розробку функцій, вибір* *моделі, навчання, оцінку та прогнозування. Незважаючи на досягнення у галузі ШІ, існують виклики у прогнозуванні фондового ринку, зокрема нестабільність ринку, проблеми з якістю даних, складність факторів впливу та ризики надмірного оснащення. Зрештою, ми можемо стати свідками синергії штучного інтелекту та Python, яка надає аналітикам та інвесторам глибше розуміння, дозволяючи приймати обґрунтовані рішення в умовах складності фінансових ринків.*

Ключові слова: *штучний інтелект, машинне навчання, статистичні моделі, глибоке навчання, алгоритм, програмна мова, тенденції на фондовому ринку.*

