SPY Predictions with Advanced ML Techniques

Kyle Rodriguez

University of Houston

karodr20@cougarnet.uh.edu

Ece Şanal

University of Houston

esanal@cougarnet.uh.edu

**Keywords**

LSTM; Stock prediction; Machine Learning

# INTRODUCTION

The stock market has long fascinated investors, economists, and researchers due to its complexity, unpredictability, and potential for high returns. With the rise of accessible financial data and machine learning (ML) tools, predicting the stock prices using historical data has become a widely explored problem in artificial intelligence. However, market behavior is influenced by countless variables – from macroeconomic indicators to investor sentiment – making it inherently noisy and non-deterministic.

This project focuses on predicting short-term stock price movements using historical price data, specifically the **SPY ETF**, which tracks the performance of the S&P 500. Our goal is to apply and evaluate **Long Short-Term Memory (LSTM)** neural networks – a type of recurrent neural network well-suited for time series forecasting – to model sequential patterns in past prices and predict future trends. While our models currently rely only on historical price data, this work lays the foundation for integrating richer features like technical indicators or sentiment data in future versions.

# LITERATURE REVIEW

Forecasting stock prices is a classic yet complex challenge due to the market’s dynamic and stochastic nature. Over time, researchers have proposed a range of techniques, beginning with statistical models and evolving into more sophisticated machine learning and deep learning models.

Traditional models like **Autoregressive Integrate Moving Average (ARIMA)** have been used to analyze and forecast financial time series due to their simplicity and interpretability. However, ARIMA and similar models can assume stationarity and linearity – assumptions that often do not hold in financial markets where trends can shift suddenly.

With the rise of data-driven methods, machine learning algorithms such as **Support Vector Machines (SVMs)** and **Random Forests** have become popular. These models are capable of capturing non-linear relationships between variables and have shown improvements over classical techniques. For example, Patel et al. (2015) tested several ML models including Random Forest and found it to outperform SVM and artificial neural networks on Indian stock data [1]. Kara et al. (2011) also demonstrated promising results using SVMs and technical indicators on the Istanbul Stock Exchange [2].

The most significant progress in recent years has come from deep learning, especially the use of **Long Short-Term Memory (LSTM)** networks. Unlike traditional models, LSTMs are designed to capture long-term dependencies and temporal dynamics in sequential data. This makes them particularly well-suited for time series tasks like stock price prediction. Fischer and Krauss (2018) trained LSTM models on daily S&P 500 data and reported strong predictive performance compared to conventional models [3].

However, literature also highlights that financial markets are highly sensitive to external, non-quantifiable events such as political decisions, earnings reports, or global crises. As a result, even LSTMs can struggle with generalization and are often best suited as part of larger hybrid systems.

This project builds on these insights by implementing an LSTM-based model trained on historical price data for the SPY ETF. Through this, we aim to evaluate how well LSTMs can learn temporal patterns in market behavior and how they compare to baseline methods.

# TYPESET TEXT

## Normal or Body Text

SAMPLE: Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.

*Conference’10*, Month 1–2, 2010, City, State, Country.

Copyright 2010 ACM 1-58113-000-0/00/0010 …$15.00.

DOI: http://dx.doi.org/10.1145/12345.67890

Please use a 9-point Times Roman font, or other Roman font with serifs, as close as possible in appearance to Times Roman in which these guidelines have been set. The goal is to have a 9-point text, as you see here. Please use sans-serif or non-proportional fonts only for special purposes, such as distinguishing source code text. If Times Roman is not available, try the font named Computer Modern Roman. On a Macintosh, use the font named Times. Right margins should be justified, not ragged.

## Title and Authors

The title (Helvetica 18-point bold), authors' names (Helvetica 12-point) and affiliations (Helvetica 10-point) run across the full width of the page – one column wide. We also recommend phone number (Helvetica 10-point) and e-mail address (Helvetica 12-point). See the top of this page for three addresses. If only one address is needed, center all address text. For two addresses, use two centered tabs, and so on. For more than three authors, you may have to improvise.[[1]](#footnote-1)

## First Page Copyright Notice

Please leave 3.81 cm (1.5") of blank text box at the bottom of the left column of the first page for the copyright notice.

## Subsequent Pages

For pages other than the first page, start at the top of the page, and continue in double-column format. The two columns on the last page should be as close to equal length as possible.

Table . Table captions should be placed above the table

|  |  |  |  |
| --- | --- | --- | --- |
| **Graphics** | **Top** | **In-between** | **Bottom** |
| Tables | End | Last | First |
| Figures | Good | Similar | Very well |

## References and Citations

Footnotes should be Times New Roman 9-point, and justified to the full width of the column.

Use the “ACM Reference format” for references – that is, a numbered list at the end of the article, ordered alphabetically and formatted accordingly. See examples of some typical reference types, in the new “ACM Reference format”, at the end of this document. Within this template, use the style named *references* for the text. Acceptable abbreviations, for journal names, can be found here: <http://library.caltech.edu/reference/abbreviations/>. Word may try to automatically ‘underline’ hotlinks in your references, the correct style is NO underlining.

The references are also in 9 pt., but that section (see Section 7) is ragged right. References should be published materials accessible to the public. Internal technical reports may be cited only if they are easily accessible (i.e. you can give the address to obtain the report within your citation) and may be obtained by any reader. Proprietary information may not be cited. Private communications should be acknowledged, not referenced (e.g., “[Robertson, personal communication]”).

## Page Numbering, Headers and Footers

Do not include headers, footers or page numbers in your submission. These will be added when the publications are assembled.

# FIGURES/CAPTIONS

Place Tables/Figures/Images in text as close to the reference as possible (see Figure 1). It may extend across both columns to a maximum width of 17.78 cm (7”).

Captions should be Times New Roman 9-point bold. They should be numbered (e.g., “Table 1” or “Figure 2”), please note that the word for Table and Figure are spelled out. Figure’s captions should be centered beneath the image or picture, and Table captions should be centered above the table body.

# SECTIONS

The heading of a section should be in Times New Roman 12-point bold in all-capitals flush left with an additional 6-points of white space above the section head. Sections and subsequent sub- sections should be numbered and flush left. For a section head and a subsection head together (such as Section 3 and subsection 3.1), use no additional space above the subsection head.



Figure 1. Insert caption to place caption below figure.

.

## Subsections

The heading of subsections should be in Times New Roman 12-point bold with only the initial letters capitalized. (Note: For subsections and subsubsections, a word like *the* or *a* is not capitalized unless it is the first word of the header.)

### Subsubsections

The heading for subsubsections should be in Times New Roman 11-point italic with initial letters capitalized and 6-points of white space above the subsubsection head.

#### Subsubsections

The heading for subsubsections should be in Times New Roman 11-point italic with initial letters capitalized.

#### Subsubsections

The heading for subsubsections should be in Times New Roman 11-point italic with initial letters capitalized.

# ACKNOWLEDGMENTS

Our thanks to ACM SIGCHI for allowing us to modify templates they had developed.

# REFERENCES

1. Patel, J., Shah, S., Thakkar, P., and Kotecha, K. 2015. Predicting stock and stock price index movement using trend deterministic data preparation and machine learning techniques. *Expert Syst. Appl.* 42, 1 (Jan. 2015), 259-268. DOI=https://doi.org/10.1016/j.eswa.2014.07.040.
2. Kara, Y., Boyacioglu, M. A., and Baykan, Ö. K. 2011. Predicting direction of stock price index movement using artificial neural networks and support vector machines: The sample of the Istanbul Stock Exchange. *Expert Syst. Appl.* 38, 5 (May 2011), 5311–5319. DOI= https://doi.org/10.1016/j.eswa.2010.10.027.
3. Fischer, T., and Krauss, C. 2018. Deep learning with long short-term memory networks for financial market predictions. *Eur. J. Oper. Res.* 270, 2 (Oct. 2018), 654–669. DOI= https://doi.org/10.1016/j.ejor.2017.11.054.

Columns on Last Page Should Be Made As Close As Possible to Equal Length

1. If necessary, you may place some address information in a footnote, or in a named section at the end of your paper. [↑](#footnote-ref-1)