# Edris Safari

## DSC680 Project2, Milestone1

Abstract

In today's data-driven world, data streaming has emerged as an essential element for businesses worldwide.

Project White Paper

Real-Time Big Data Streaming with Kafka

Contents

[Business Problem 2](#_Toc131316585)

[Background/History 2](#_Toc131316586)

[Data Explanation 3](#_Toc131316587)

[Datasets 3](#_Toc131316588)

[Data Preparation 3](#_Toc131316589)

[Data Management 4](#_Toc131316590)

[Methods 4](#_Toc131316591)

[Analysis 4](#_Toc131316592)

[Technical Indicators 4](#_Toc131316593)

[RNN Model 6](#_Toc131316594)

[Conclusion 8](#_Toc131316595)

[Assumptions 9](#_Toc131316596)

[Limitations 9](#_Toc131316597)

[Challenges 9](#_Toc131316598)

[Future Uses & Product Roadmap 9](#_Toc131316599)

[Additional technical indicators 9](#_Toc131316600)

[New Model 9](#_Toc131316601)

[Graphical User Interface 10](#_Toc131316602)

[Other algorithms. 10](#_Toc131316603)

[Dataset 10](#_Toc131316604)

[Recommendations 10](#_Toc131316605)

[Implementation Plan 10](#_Toc131316606)

[Ethical Assessment 10](#_Toc131316607)

[References 11](#_Toc131316608)

# Business Problem

To stay competitive, companies must use data to gauge the health of their business and find ways to improve it. Business gain and lose clients due to various reasons, and

# Background/History

The goal of this project is to demonstrate a data streaming environment where data is live-fed to an application that will graph the data received.

# Data Explanation

For this phase of the project, we will use historical data augmented with live data from the stock market.

## Datasets

The dataset from yfinance will have values for Date, Open, High, Low, Close, Adjusted Close, and Volume.

## Data Preparation

Data is

## Data Management

# Methods

We will use Python as the primary programming language. We will use Matplotlib, Seaborn, Keras, Pandas, yfinance and other necessary libraries. Apache Kafka will beused to facilitate live data streaming, We will show technical indicators for individual stocks. We apply an AI model in Neural networks to predict the price of a given stock and show graphs and metrics to show confidence and accuracy of the prediction.

# Analysis

## Technical Indicators

Technical indicators are intrinsically pattern-based because they are based on time. They are signals that are produced by the price, volume, and other parameters. Shown graphically, these signals will show spike, declines, gradual/sharp decrease, or gradual/sharp increase, etc. These visuals will help technical analysist predict future price movements.

We created 5 technical indicators. The Bollinger band indicator is a popular one. Shown below, the black line is the 30-day average closing price of the stock, and the gray area is the band range. The upper part of this gray area is the 30-day average plus 3 standard deviations, and low part of the gray area in the average minus 3 standard deviation. The wider the band, the more volatile the stock is. There are

Chart, line chart

Description automatically generated

Relative Strength Index or RSI, measures speed and magnitude of a stock’s most recent price changes (up or down). This will help evaluate whether a stock is overvalued (don’t buy) or undervalued(buy). Shown below, the stock for AMAT was overvalued in 2016,2017 and undervalued since 2021-22.

Chart

Description automatically generated

Aroon Oscillators are technical indicators that are used to identify the strength and direction of a trend. The indicator consists of two lines: the Aroon-Up and the Aroon-Down. The Aroon-Up measures the number of periods since the highest high, while the Aroon-Down measures the number of periods since the lowest low. When the Aroon-Up is above the Aroon-Down, it indicates an uptrend, while when the Aroon-Down is above the Aroon-Up, it indicates a downtrend. Shown below, AMT is trending downward.

Chart, histogram

Description automatically generated

Acceleration Bands are a technical indicator that uses volatility to identify potential price trends in the market. The indicator consists of three lines: the upper band, the lower band, and the centerline. The upper and lower bands are based on the standard deviation of the price from the centerline, while the centerline is typically a moving average. Shown below, when the price touches or crosses the upper band it indicates that the trend is losing its momentum and due for a reversal.

Chart, line chart

Description automatically generated

## RNN Model

We used a RNN model and used two LSTM layers(shown below). LSTM is suitable for time series dataset. We used the closing price in the training data set to predict closing price in a test dataset.

Text

Description automatically generated with low confidence

The graph below shows the result. The disparity in the green and orange line shows the accuracy of the model.

Chart, line chart

Description automatically generated

Shown below, the disparity seems quite wide and persistent.

Chart, line chart

Description automatically generated

Using a different approach, we came up with this result:

Chart, line chart

Description automatically generated

No noticeable difference was found.

Chart, line chart, histogram

Description automatically generated

# Conclusion

We showed that technical indicators help investors make more accurate decisions about buying or selling stocks. They also help with the timing of the transaction, and the price target. Combined with fundamentals of the stock, and other factors such as interest rate, consumer confidence, overall health of the economy (both nationally and globally) and other external factors, investors of today can make educated and low risk decisions.

# Assumptions

We must assume that the investors are knowledgeable about the interpretation of the technical indicators, but not necessarily about machine learning. The machine learning aspect of this project is mainly the responsibility of the developers and data scientists.

# Limitations

One major limiting factor to this project is data quality. We must ensure that the quality of data is monitored and always maintained. Other limiting factor that we must be aware of is time to market constraint. Model and calculation accuracy is also a limitation that cannot be overlooked.

# Challenges

Challenges and issues that we could face in this project must be considered and addressed. This is part of risk assessment that all projects must go through. The table below lists the risks and their mitigation.

|  |  |
| --- | --- |
| Risk | Mitigation |
| Data Quality | Ensure data quality by performing a preliminary analysis |
| Data Security | Ensure data is secure both incoming and outgoing. Enable/utilize security measures. |
| Data Availability | Ensure data is available without interruption or delay that may affect the performance of the analysis |
| Technical Challenges | Enforce fault tolerance, redundancy, connection integrity |
| Ethical Violations | Ensure procedures are put in place that will reduce and remove risk of ethical violations by all parties involved. |

# Future Uses & Product Roadmap

The following improvements will be put on the roadmap.

## Additional technical indicators

These technical inidcators can be added to the portfolio:

1. Stochastic Oscillator(%K and %D)
2. Chaikin Money Flow
3. Parabolic SAR
4. Price Rate of Change
5. Volume Weighted Average Price

## New Model

Research and create classification model to predict up or down. Other RNN models and more advanced LSTM model can also help improve our portfolio,

## Graphical User Interface

A graphical user interface to allow users to select stocks, select graphs of different types (i.e., RSI, PVT, BBs, etc.) and be able to view the data daily, monthly, yearly, etc. The GUI will have to be web based and accessible by authorized personnel.

## Other algorithms.

Research and Implement new ML algorithms.

## Dataset

Augment data set with additional parameters to gauge the market fluctuation better.

# Recommendations

Market research at any level is an important asset. We recommend that the results of all the analysis done be used as a basis for future market analysis. This will also help in improvement of the process of stock analysis.

# Implementation Plan

To implement this project, we will perform the following main tasks:

1. Gather and prepare dataset.
2. Prepare Design Document.
3. Code, and test
4. Present
5. Deploy

# Ethical Assessment

Based on the ethical considerations we listed in our proposal; we have taken the following measures to address them.

Data privacy: Work with the legal team to ensure that the data collected is obtained in a legal and ethical manner. We will also work with the IT team to ensure that personal information for clients as well as employees is safe and secure.

Bias and fairness: Provide appropriate training about ethics involved in this field to people involved in all aspects of the product.

Transparency: Work with engineering to ensure that the data sources, analysis methods, and findings are transparent and easily understandable to all stakeholders.

Security: We will work with IT on security issues and data privacy as mentioned above.

Informed consent: Work with legal team to make sure that legal documents are in place to align with the customers on topics requiring consent.

Impact on society: Perform a thorough evaluation of the results, perform risk analysis, measure accuracy, and any metrics that can help ascertain minimal negative impact.

Questions and Answers

Q: What is technical analysis?

Q: What is machine learning?

Q: How do graphics help?

Q: Is technical analysis better than fundamentals?

Q: Are there other models or algorithms that can be used

Q: Are there other technical indicators?

Q: Can variables other than open, close, high, low be used, if so name 2-3.

Q. What are recurrent neural networks?

Q. Explain LSTM algorithm.

Q. Can use of ML adversely affect the stock market trading?

# References

1. 7 Technical Indicators to Build a Trading Toolkit <https://www.investopedia.com/top-7-technical-analysis-tools-4773275>
2. François Chollet. Deep Learning with Python (Kindle Locations 1504-1508). Manning Publications Co.. Kindle Edition.
3. What are Stock Fundamentals? <https://www.investopedia.com/articles/fundamental/03/022603.asp#:~:text=Stock%20fundamentals%20are%20key%20metrics,perceived%20value%20of%20a%20stock>
4. Technical Indicators Mathematical Description

<https://docs.anychart.com/Stock_Charts/Technical_Indicators/Mathematical_Description>

1. Understanding Indicators in Technical Analysis - <https://www.fidelity.com/bin-public/060_www_fidelity_com/documents/learning-center/Understanding-Indicators-TA.pdf>
2. Using Technical Indicators to Develop Trading Strategies <https://www.investopedia.com/articles/trading/11/indicators-and-strategies-explained.asp>
3. Technical Indicator: Definition, Analyst Uses, Types and Examples <https://www.investopedia.com/terms/t/technicalindicator.asp>
4. The basics of Bollinger Bands https://www.investopedia.com/articles/technical/102201.asp
5. Project Proposal-Project1\_Milestone1\_EdrisSafari.pdf