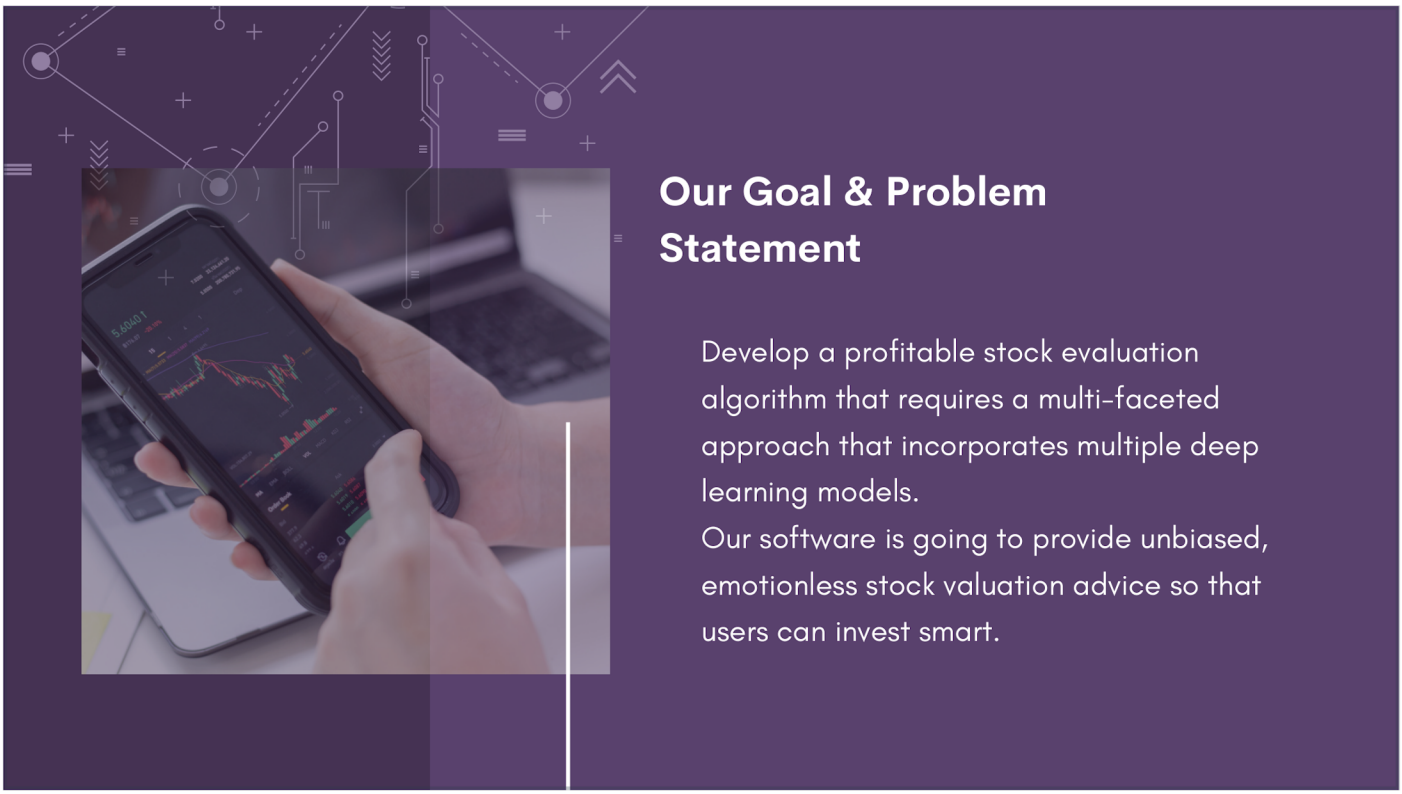
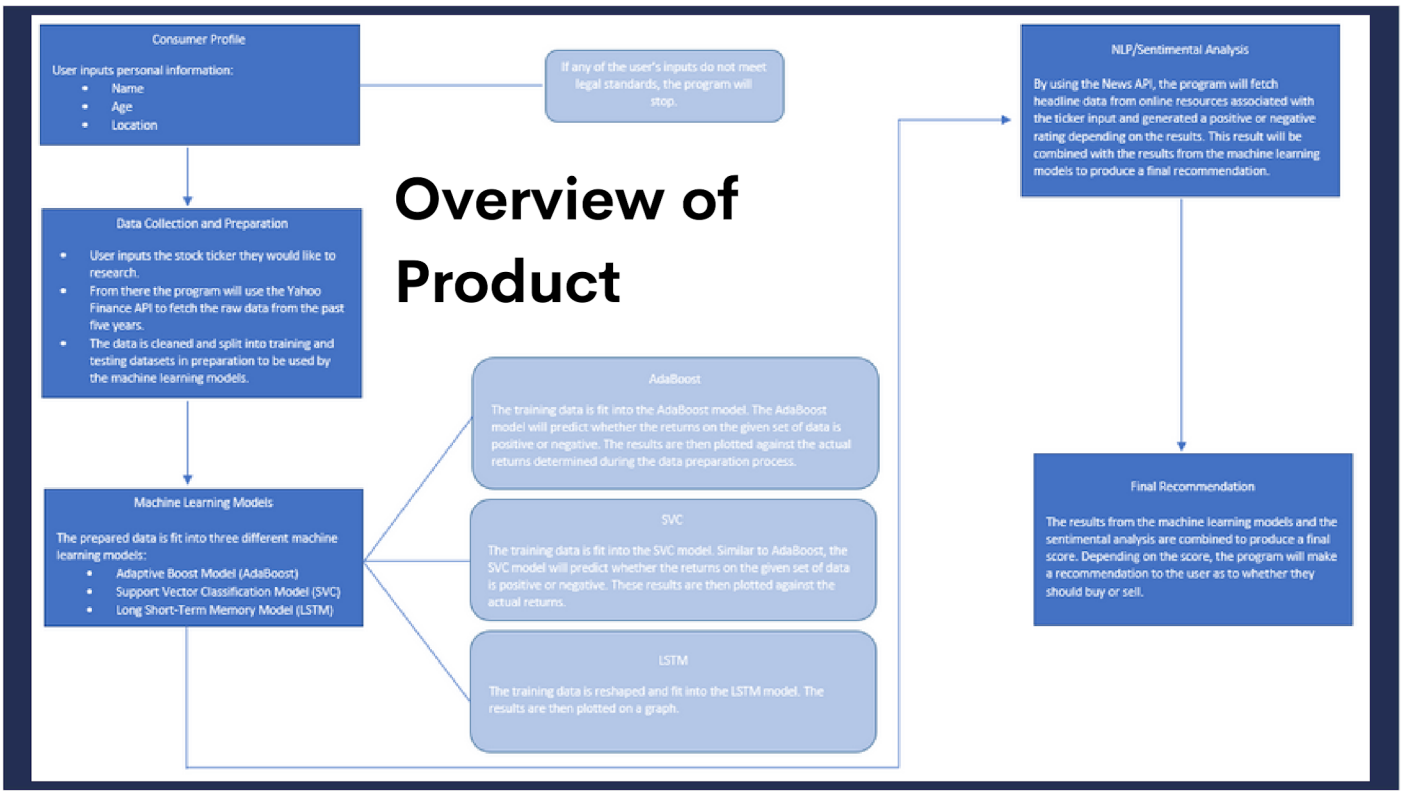
Algorithmic Evaluation in the Stock Market

AKSHAY VAS NYAMALA
EMMANUEL JOSEPH
JONATHAN COLEMAN
MOH JAISWAL
NOELIA NIEVAS



That assists on how profitable an asset is. By leveraging machine learning techniques investors can create evaluation algorithms that can analyze vast amounts of financial data, identify any patterns and make more accurate predictions about market trends.

User inputs their information to build a profile. The program will prompt the user with a set of queries and generate a structured data frame to store the acquired information.


Consumer Profile
User inputs personal information:
Name
Age
Location
If any of the user's inputs do not meet legal standards, the program will stop.
Overview of Product
NLP/Sentimental Analysis
By using the News API, the program will fetch headline data from online resources associated with the ticker input and generated a positive or negative rating depending on the results. This result will be combined with the results from the machine learning models to produce a final recommendation.
Data Collection and Preparation
User inputs the stock ticker they would like to research.
From there the program will use the Yahoo Finance API to fetch the raw data from the past five years.
The data is cleaned and split into training and testing datasets in preparation to be used by the machine learning models.
AdaBoost
The training data is fit into the AdaBoost model. The AdaBoost model will predict whether the returns on the given set of data is positive or negative. The results are then plotted against the actual returns determined during the data preparation process.
Machine Learning Models
The prepared data is fit into three different machine learning models:
Adaptive Boost Model (AdaBoost)
Support Vector Classification Model (SVC)
Long Short-Term Memory Model (LSTM)
SVC
The training data is fit into the SVC model. Similar to AdaBoost, the SVC model will predict whether the returns on the given set of data is positive or negative. These results are then plotted against the actual returns.
LSTM
The training data is reshaped and fit into the LSTM model. The results are then plotted on a graph.
Final Recommendation
The results from the machine learning models and the sentimental analysis are combined to produce a final score. Depending on the score, the program will make a recommendation to the user as to whether they should buy or sell.

ADA Boost

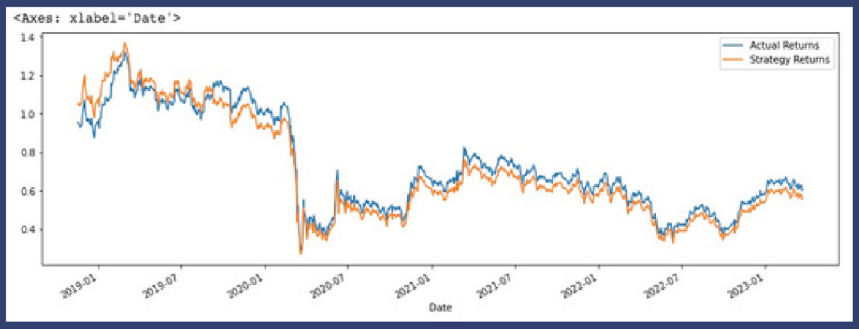
AdaBoost, is a machine learning ensemble algorithm that combines multiple weak learners to create a more robust and accurate model. The idea behind AdaBoost is to iteratively train multiple weak learners, which are models that are only slightly better than random guessing, on different samples of the training data.

In each iteration, the algorithm assigns more weight to the misclassified samples from the previous iteration, which forces the next weak learner to focus on the harder-to-classify data points. AdaBoost then combines the predictions from all weak learners, assigning higher weight to the more accurate models, to make the final prediction.

The strength of AdaBoost lies in its ability to improve the accuracy of weak learners by combining them into a stronger model. AdaBoost is particularly useful for classification problems, and it has been applied to various domains, such as computer vision, speech recognition, and financial analysis. However, it can be sensitive to noisy data and outliers.

SVM

In machine learning, SVMs are supervised learning models with associated learning algorithms that analyze data for classification and regression analysis. Developed at AT&T Bell Laboratories, SVMs are based on statistical learning frameworks and are one of the most robust prediction methods. Given a set of training examples, each marked as belonging to one of two categories, an SVM training algorithm builds a model that assigns new examples to one category or the other, making it a non-probabilistic binary linear classifier. An SVM maps training examples to points in space so as to maximise the width of the gap between the two categories. New examples are then mapped into that same space and predicted to belong to a category based on which side of the gap they fall.

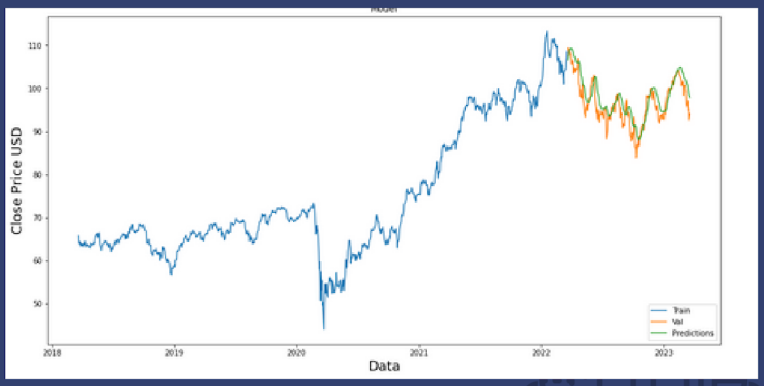


LSTM

LSTM is a type of recurrent neural network (RNN) architecture. LSTM networks are designed to address the vanishing gradient problem that occurs in standard RNNs, which can make it difficult for the network to learn long-term dependencies.

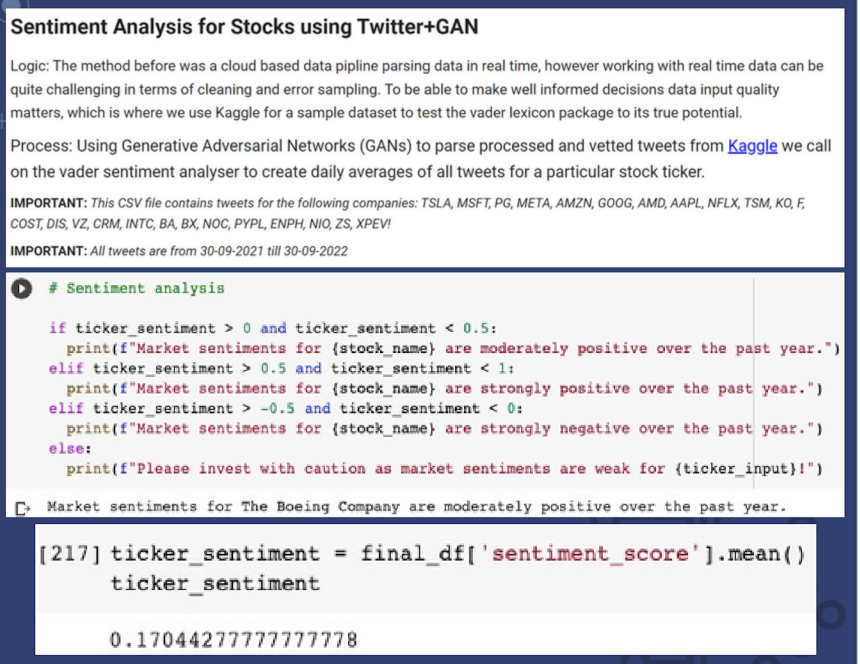
In an LSTM network, each neuron has a memory cell and three gates: an input gate, an output gate, and a forget gate. The input gate determines which information should be added to the memory cell, the output gate determines which information should be output from the memory cell, and the forget gate determines which information should be removed from the memory cell.

By selectively adding, removing, and outputting information from the memory cell, LSTM networks are able to effectively remember important information over long sequences of data. This makes them useful in applications such as natural language processing, speech recognition, and time series prediction.



NLP/Sentimental Analysis

NLP stands for Natural Language Processing, which is a field of computer science and artificial intelligence that focuses on the interaction between computers and humans using natural language. NLP is concerned with tasks such as language understanding, language generation, machine translation, sentiment analysis, and text summarization. It involves applying computational techniques to analyze, understand, and generate human language in a way that is both accurate and efficient. NLP has become increasingly important in today's world as more and more data is generated in the form of text, and there is a growing need to automate tasks that were previously performed by humans, such as customer service and content curation.



Sentiment Analysis for Stocks using Twitter+GAN

Logic: The method before was a cloud based data pipline parsing data in real time, however working with real time data can be quite challenging in terms of cleaning and error sampling. To be able to make well informed decisions data input quality matters, which is where we use Kaggle for a sample dataset to test the vader lexicon package to its true potential.

Process: Using Generative Adversarial Networks (GANs) to parse processed and vetted tweets from Kaggle we call on the vader sentiment analyser to create daily averages of all tweets for a particular stock ticker.

IMPORTANT: *This CSV file contains tweets for the following companies: TSLA, MSFT, PG, META, AMZN, GOOG, AMD, AAPL, NFLX, TSM, KO, F, COST, DIS, VZ, CRM, INTC, BA, BX, NOC, PYPL, ENPH, NIO, ZS, XPEV!*

IMPORTANT: *All tweets are from 30-09-2021 till 30-09-2022*

```
# Sentiment analysis

if ticker_sentiment > 0 and ticker_sentiment < 0.5:
  print(f"Market sentiments for {stock_name} are moderately positive over the past year.")
elif ticker_sentiment > 0.5 and ticker_sentiment < 1:
  print(f"Market sentiments for {stock_name} are strongly positive over the past year.")
elif ticker_sentiment > -0.5 and ticker_sentiment < 0:
  print(f"Market sentiments for {stock_name} are strongly negative over the past year.")
else:
  print(f"Please invest with caution as market sentiments are weak for {ticker_input}!")
```

```
Market sentiments for The Boeing Company are moderately positive over the past year.
```

```
[217] ticker_sentiment = final_df['sentiment_score'].mean()
      ticker_sentiment

      0.17044277777777778
```

In conclusion, our software solves a large problem in retail trading which is a lack of unbiased trading advice. To do this we are using machine learning algorithms and historical data which allows for users to eliminate fear and greed based on emotions and trade smarter.

CONCLUSION


Q and A
ASK AWAY!
SELL
BUY