Magic Stocks Model

Tim, Jade, & Jodi

Can machine learning be used to accurately predict the stock market?

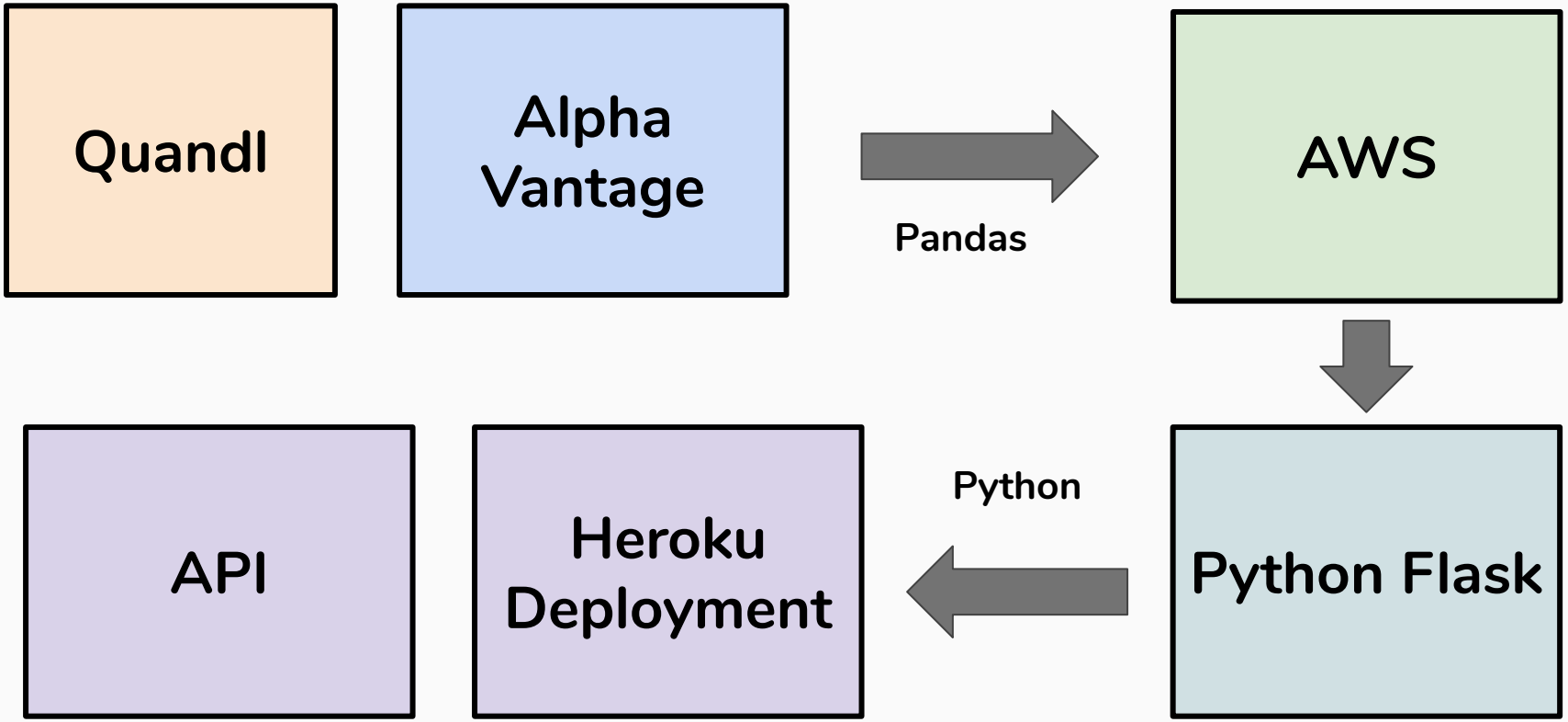
Project Overview

GOALS
Predict Future Prices
Visualize Findings
Serve Web Application

TOOLS
Alpha Vantage API
AWS - SQL
Keras - Tensorflow
Python - Flask
JavaScript - D3 - Plotly
Heroku

PRODUCT
Ten-Year Dataset
Data Visualization
Future Price Prediction

Dataset Selection & Creation



Project Overview

- **Extract** 10 years of stock data using the Alpha Vantage's API
- **Transform** data in Jupyter Notebook
- **Load** data into AWS Postgresql database
- Build and test **ML Models** in Jupyter Notebook and Google Colab (PySpark)
- Create data **Visualizations** with JavaScript, D3, and Plotly
- **Deploy** Python-Flask web application on Heroku

ML Models

Different machine learning models can be applied to stock market prediction task trying to predict the future stock price movement for different time frames and using different features that would affect stock prices.

- **Model Output:** the future direction of the stock price (up or down) in next 3, 7, 10, 14 days
    1. **Multilayer Perceptrons (MLP, sklearn.MLPClassifier)**
    2. **Deep Artificial Neural Networks (DNN, tensorflow.keras.models)**
    3. **Support Vector Machine (SVM)**

- **Model Output:** the future stock price, next day of closing price

  Long Short-Term Memory (LSTM), an artificial recurrent neural network (RNN) architecture used in the field of deep learning

Features Selection for DNN, MLP, SVM Models

These features were considered as technical analysis features for our stock prediction models, number of full features that are included in our project is 36.

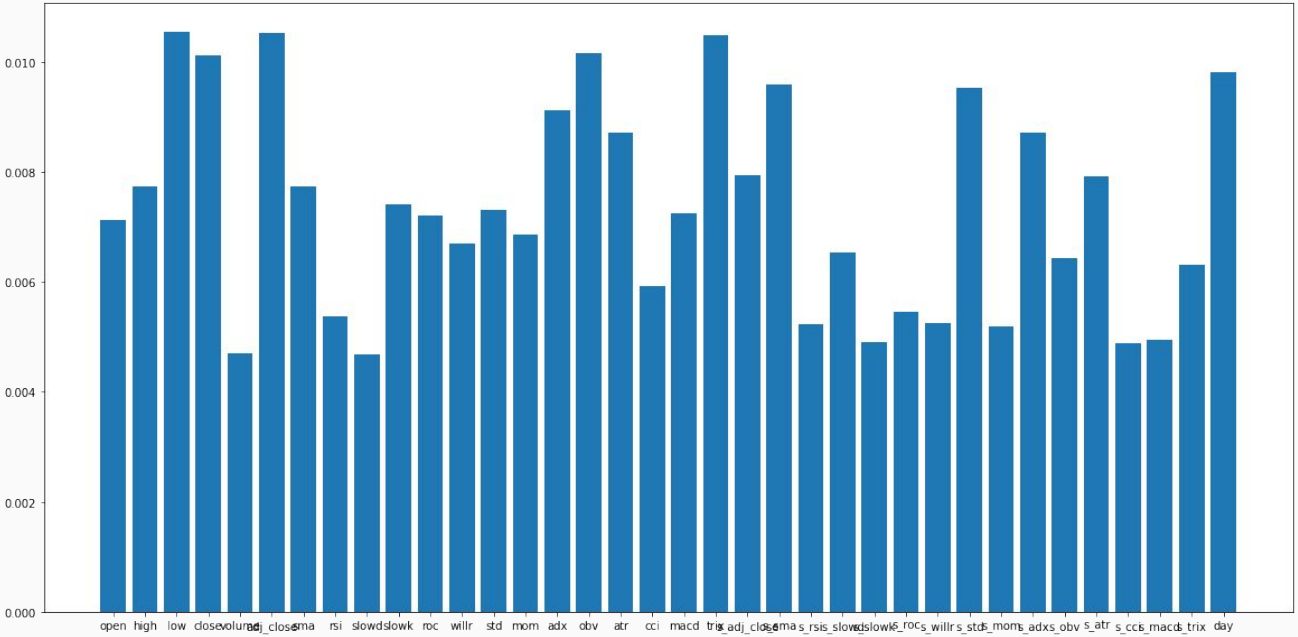
- **Daily Stock Price**: time, open, close, high, low, volume
- **Stock Technical Indicators**: sma, rsi, slowd, slowk, roc, willr, std, mom, adx, obv, atr, cci, macd, trix.
- **SPY&500 Technical Indicators**: sma, rsi, slowd, slowk, roc, willr, std, mom, adx, obv, atr, cci, macd, trix.
- **Technical Indicators:** are based on mathematical calculations, 14 days of window size.

Prediction accuracy on DNN, MLP,SVM model using full features

Company/ Accuracy	Next 3-day	Next 7-day	Next 10-day	Next 14-day
	DNN MLP SVM	DNN MLP SVM	DNN MLP SVM	DNN MLP SVM
AAPL	62%, 60%,51%	75%,69%,58%	79%,70%,61%	81%,73%,65%
KO	62%,54%,53%	71%,64%,56%	76%,67%,63%	79%,69%,67%
JPM	58%,53%,51%	68%,58%,53%	75%,63%,56%	81%,68%,61%
AMZN	59%,54%,54%	70%,60%,56%	79%,65%,59%	81%,69%,61%

Features Selection

The Extremely Randomized Trees algorithm is applied to feature selection and it suggests a subset of stock technical indicators are critical for predicting the stock trend, 10 features were selected based on their Gini importances. ('low', 'close','adx', 'obv', 'atr', 'trix', 's_adj_close','s_std', 's_adx', 's_atr')



Prediction accuracy on DNN, MLP,SVM model using selected features

Company/ Accuracy	Next 3-day	Next 7-day	Next 10-day	Next 14-day
	DNN, MLP, SVM	DNN, MLP, SVM	DNN, MLP, SVM	DNN, MLP, SVM
AAPL	59%,54%,57%	68%,67%,61%	71%,71%,66%	73%,75%,71%
KO	57%,57%,55%	67%,63%,69%	73%,64%,68%	76%,69%,73%
JPM	58%,57%,55%	67%,65%,63%	69%,67%,66%	73%,65%,68%
BAC	56%,55%,53%	70%,61%,62%	72%,65%,67%	70%,74%,71%

LSTM RNN applied for prediction of next day of closing price

Input data:

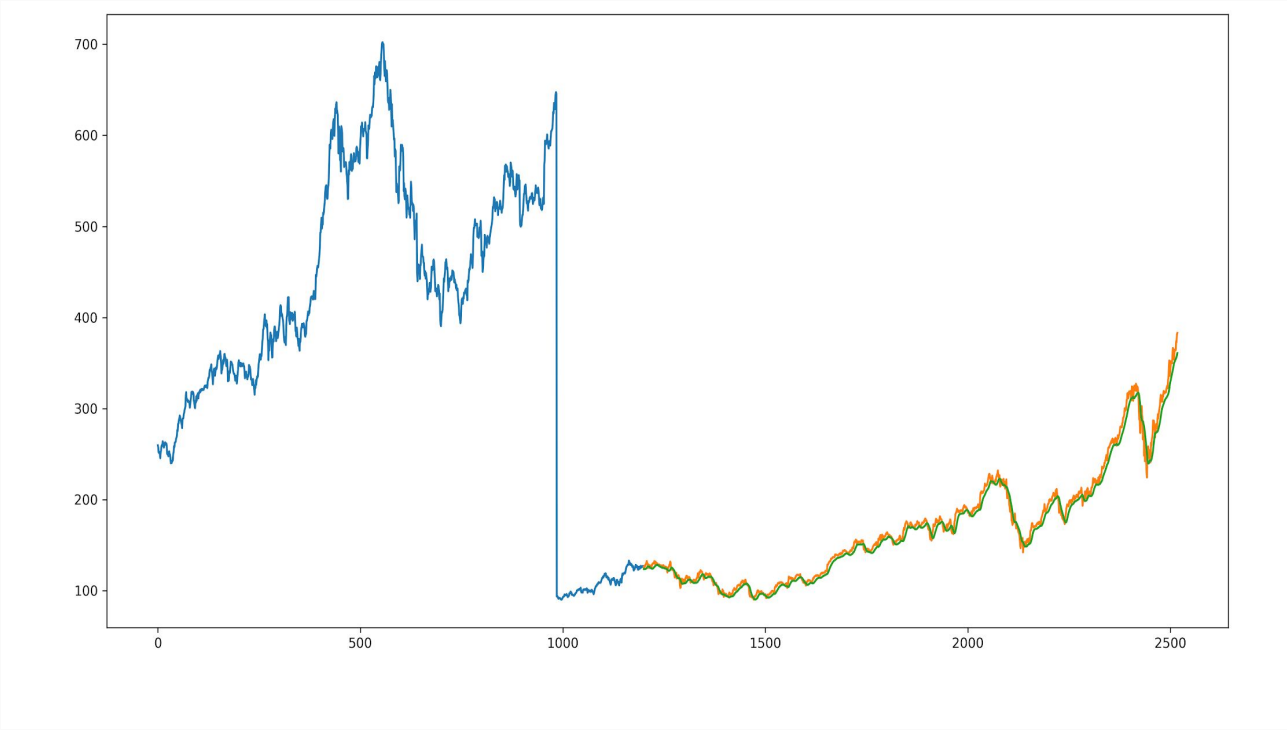
- 10 years historical stock price of AAPL

Input window size:

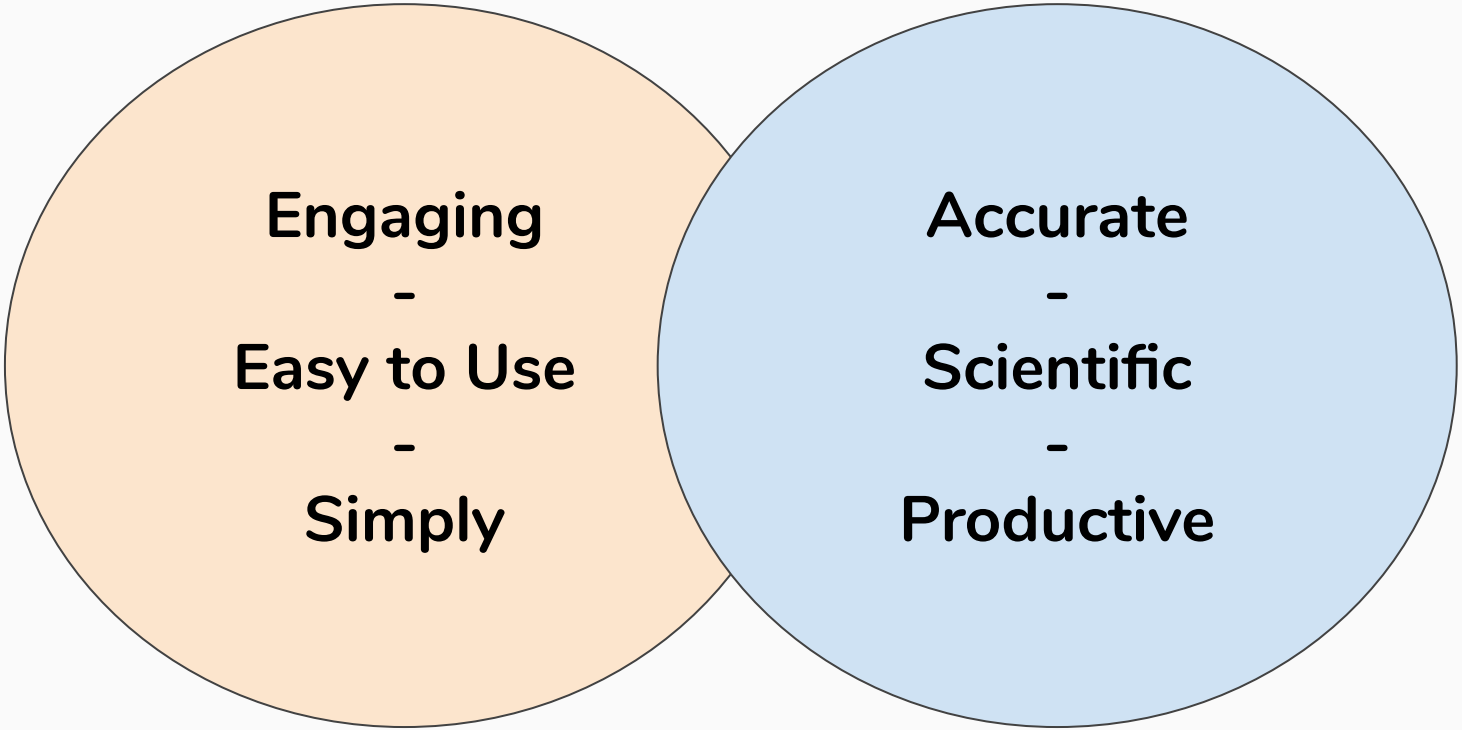
- 60 days

Output :

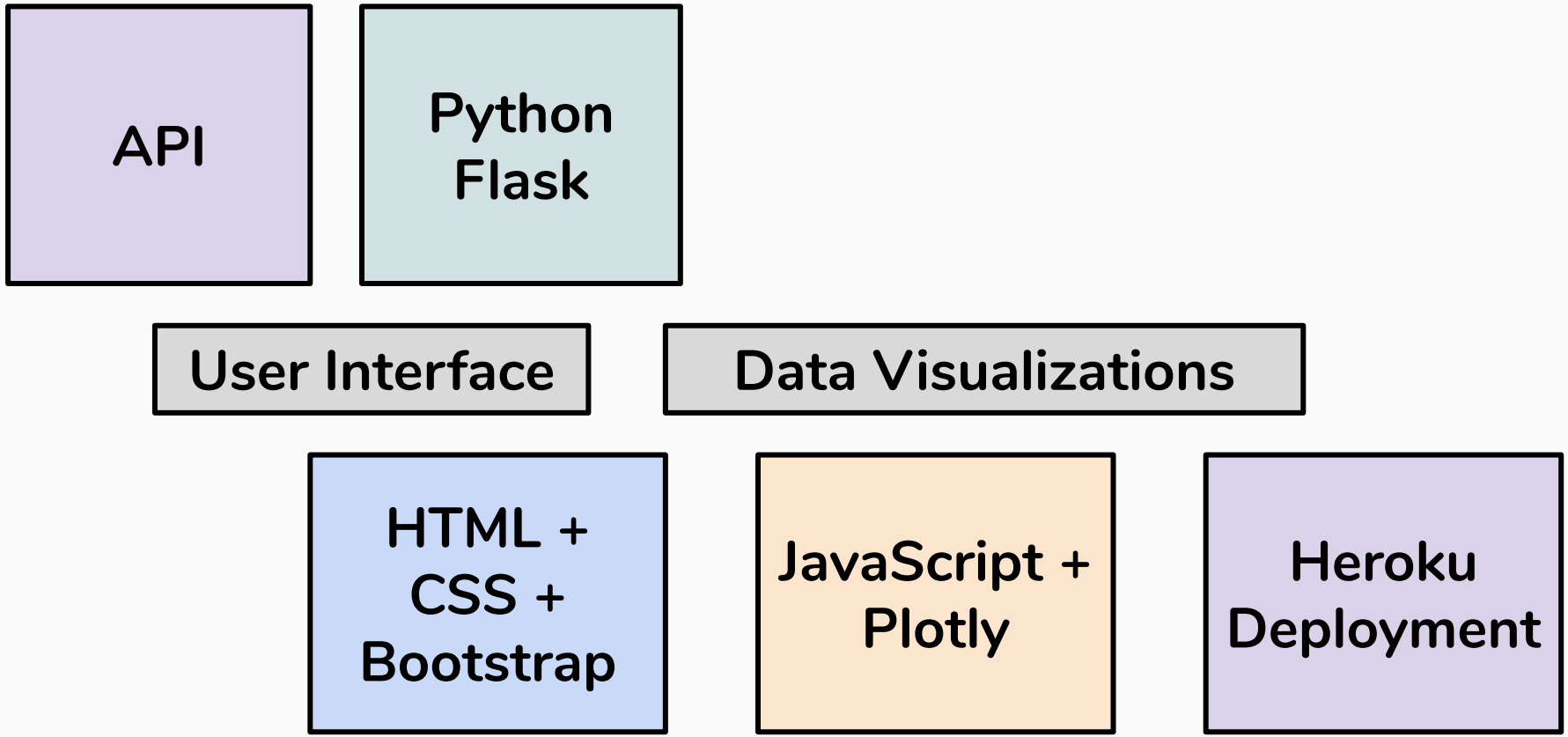
- next day of closing price



User Interface



App Design


API
Python
Flask
User Interface
Data Visualizations
HTML +
CSS +
Bootstrap
JavaScript +
Plotly
Heroku
Deployment

Continuing Improvements

User Education

- Elementary stock information, especially to coincide with vizzes
  - Ex. on our “data” page, give some definitions of those terms

ML Model & Investment Tools

- Grid Search with Scikit-Learn for tuning parameters
- Incorporate the sentiment information into model, explore different input window size
- On the “team” page, share some of our favorite tools and strategies
  - High-level investment books, apps, websites, & methodologies
  - Machine Learning references (articles & videos used for development)

The Joy of Tidying Up

- Consistent fonts and sizes across all pages
- A little less cheek from the website copywriter

Lessons Learned

Always the data mining...

- Good quality stock APIs exist
- Took time on the front end creating the database and deciding what stock features to include

Models take time!

- Thankfully, lots of people have already tried this, so we were able to learn from their trials
- Small, meaningful code that makes a huge difference