# Predicting S&P500 Daily Prices

By Niko Ganev

# Disclaimer

## NOT FINANCIAL ADVICE!

The content of this presentation and associated Github Repository and webpage are meant to showcase data science and programming skills, not to provide investment advice. The SPY ETF predictions in this project are only a sample of what kind of model and visualizations could be built with appropriate data, statistical analysis and programming tools. Keep in mind investing involves risk that can sometimes be significant. Investments fluctuate and you may potentially lose up to your total invested capital or even more. It is very important to do your own analysis based on your own circumstances and speak to a professional before making any investments. The underneath model is a student's project focusing on the technical aspects of the analysis and visualization and does not intend to achieve any returns. Any past performance is no guarantee of future return, nor is it necessarily indicative of future performance.

# Summary

## I- Assembling Dataset
-Correlated Assets
-News
-Technical indicators

## II- Exploratory Data Analysis
Get a modern PowerPoint Presentation that is beautifully designed.
I hope and I believe that this Template will your Time.

## III- Modelling
-ARIMA
-LSTM Neural Network
-KNN Cross Validation

## IV- Trading
Establishing a Trading Strategy and checking Returns

# I - Gathering the Data

# The Data Sources

**Reuters Headlines**

Historical financial headlines from January 2007 to December 2017 found on Harvard University Website. 16M headlines

**NewsApi**

Using NewsApi to stream daily news for free in production environment.

**Yahoo Finance**

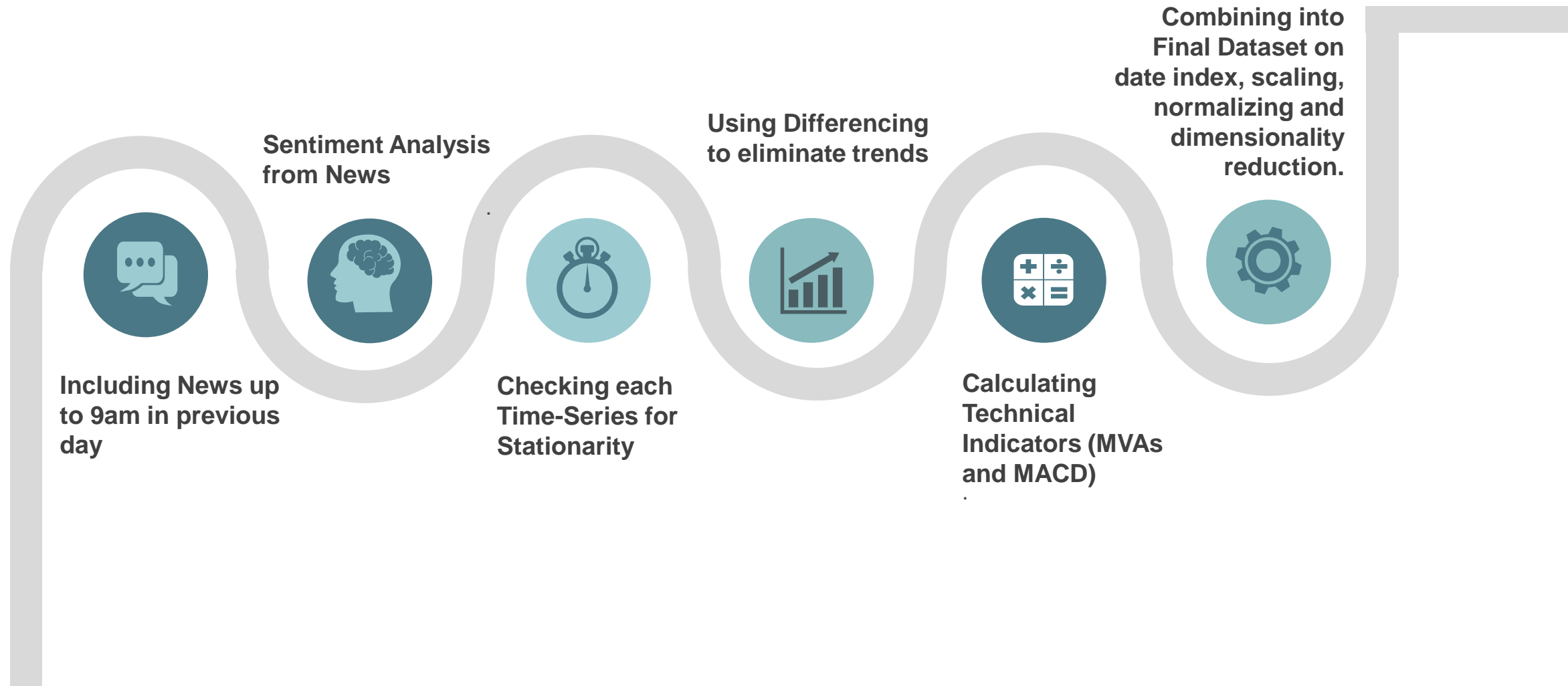Using Pandas Data Reader API to get live data from Yahoo finance.
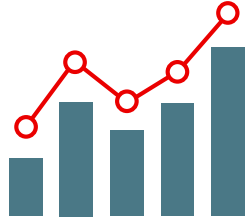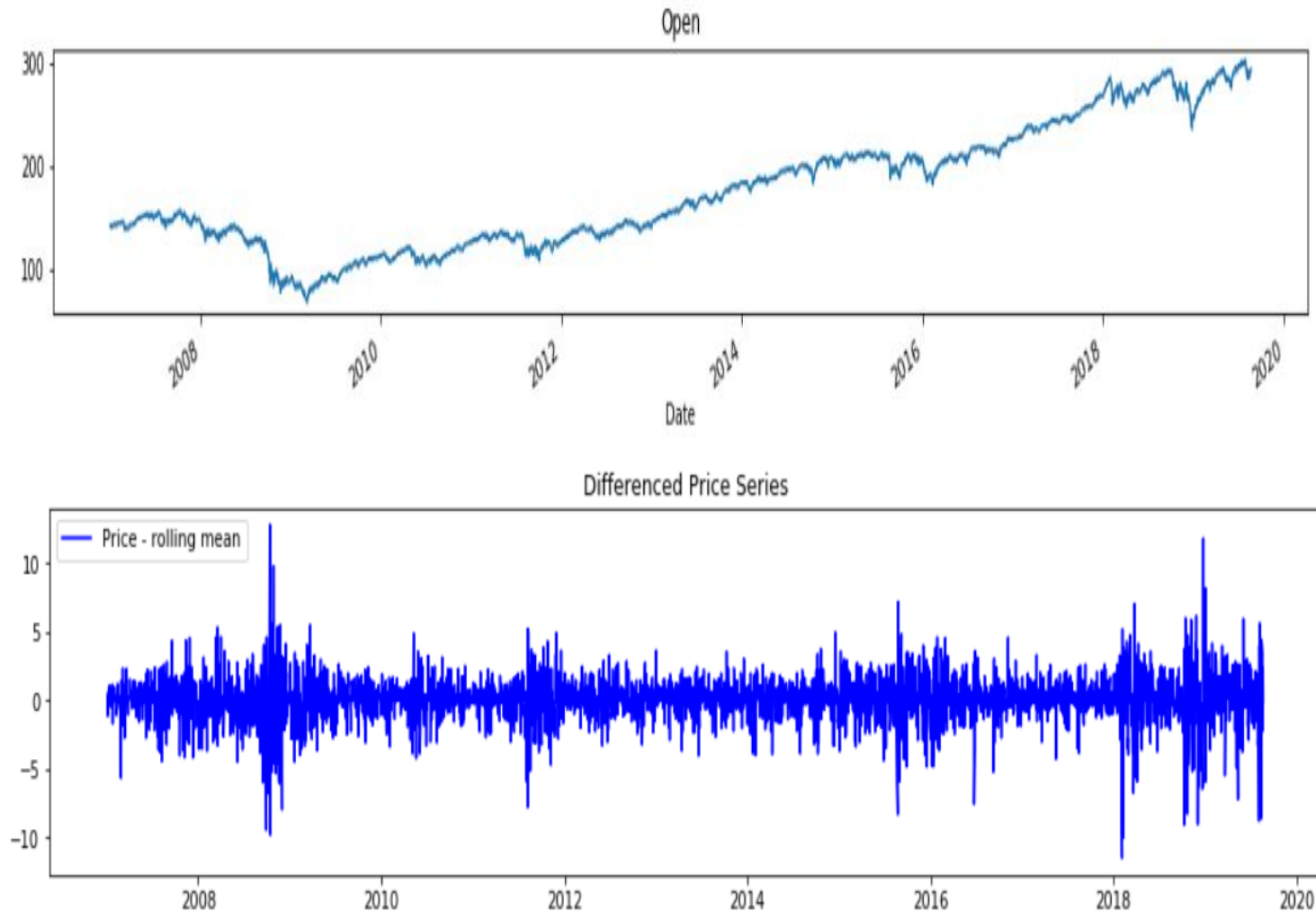
# The Features

**Correlated Assets**
ETF for major industries and foreign Exchanges

**News**
-
Sentiment Analysis

**Technical Indicators**

- Moving Averages
- MACD
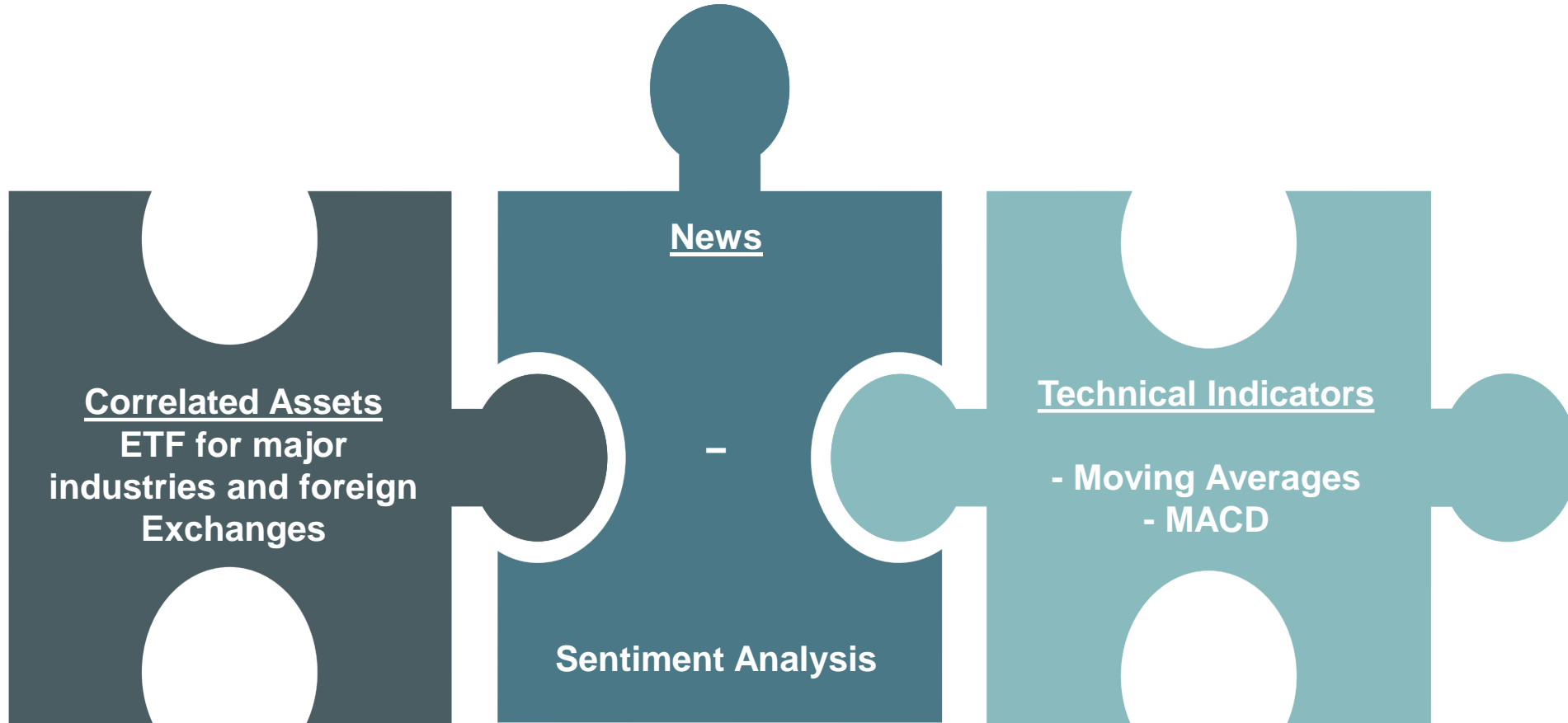
# II - Exploratory Data Analysis

# EDA Milestones



**Sentiment Analysis from News**

**Using Differencing to eliminate trends**

**Combining into Final Dataset on date index, scaling, normalizing and dimensionality reduction.**

**Including News up to 9am in previous day**

**Checking each Time-Series for Stationarity**

**Calculating Technical Indicators (MVAs and MACD)**

# Stationary Time Series

Differencing to remove trend

# III - Modelling

# ARIMA



SPY Prices Prediction
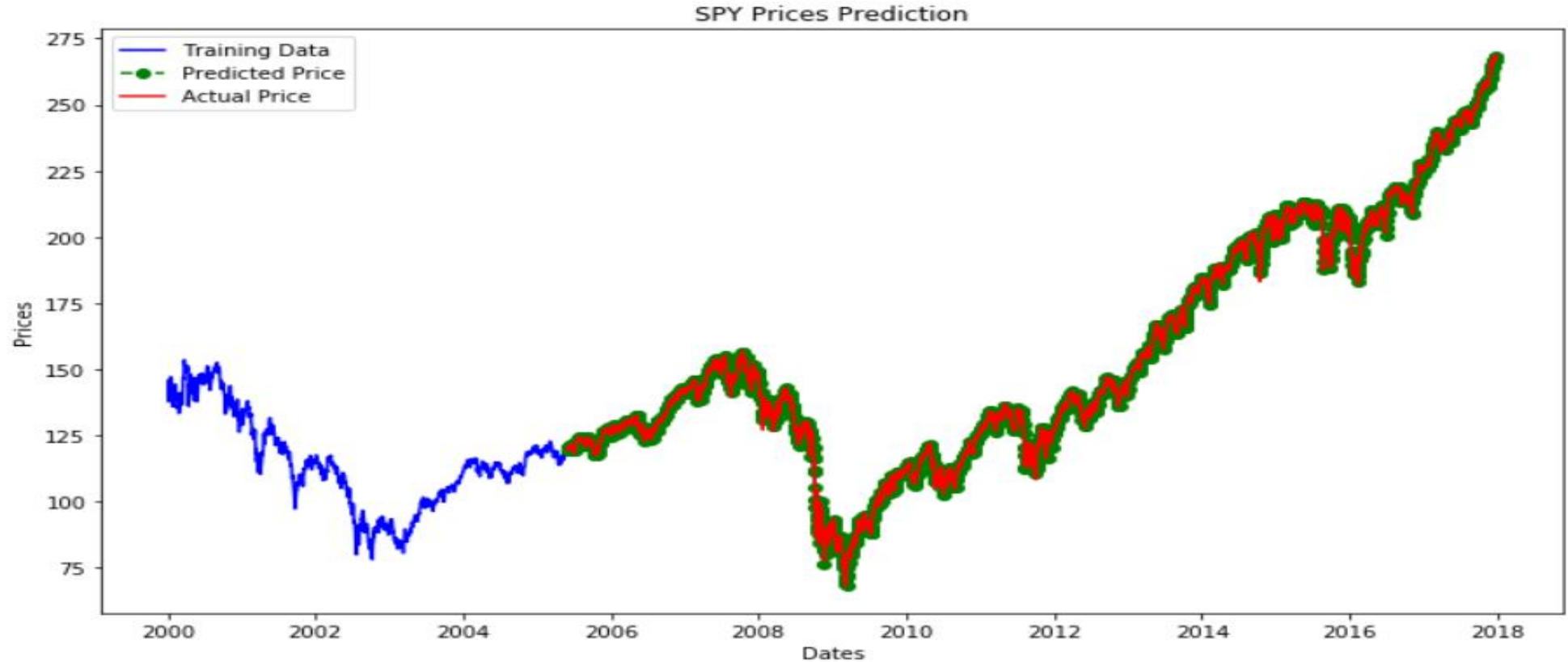
- Much easier to implement than a time series neural network
- Predictions are very close to actual prices

**UP** ⬆

- Not precise enough to build a trading algorithm
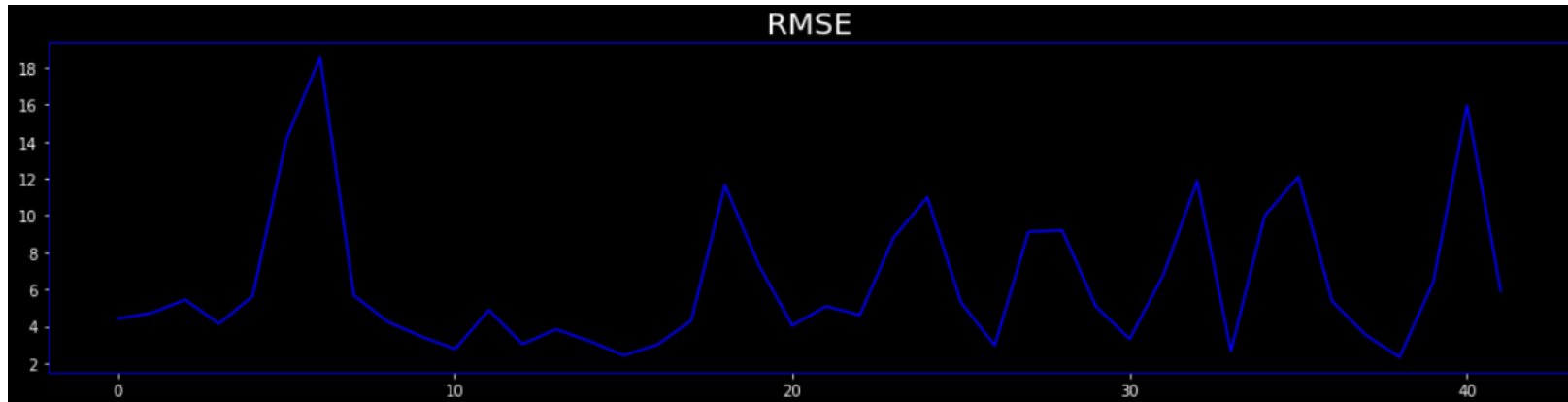- Hard to incorporate news into it

**DOWN** ⬇

# LSTM Neural Network with sliding window

- Low number of epoch to avoid over-fitting

- Five to seven days sliding window produced best results

- Model still showed lower performance on test set indicating over-fitting.
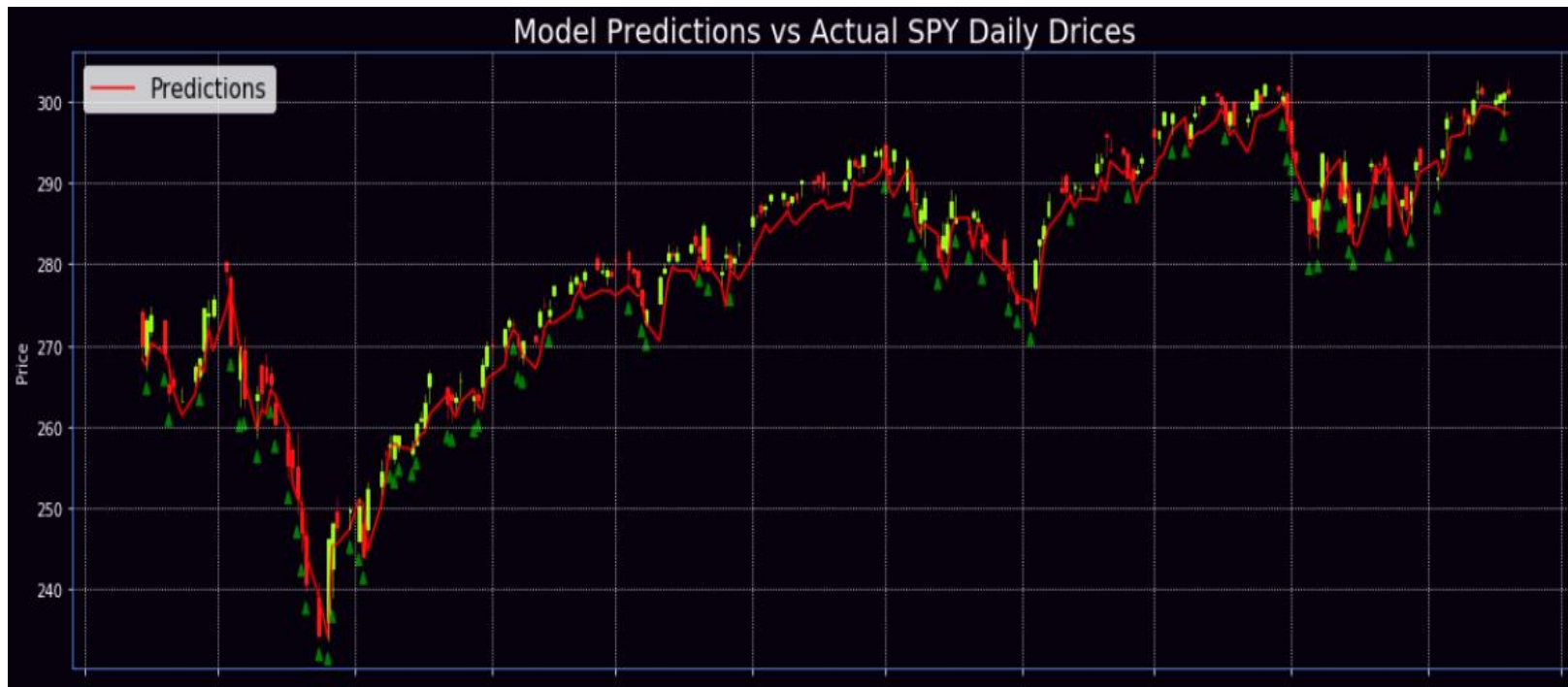
- -Necessity to chose optimal train set size.
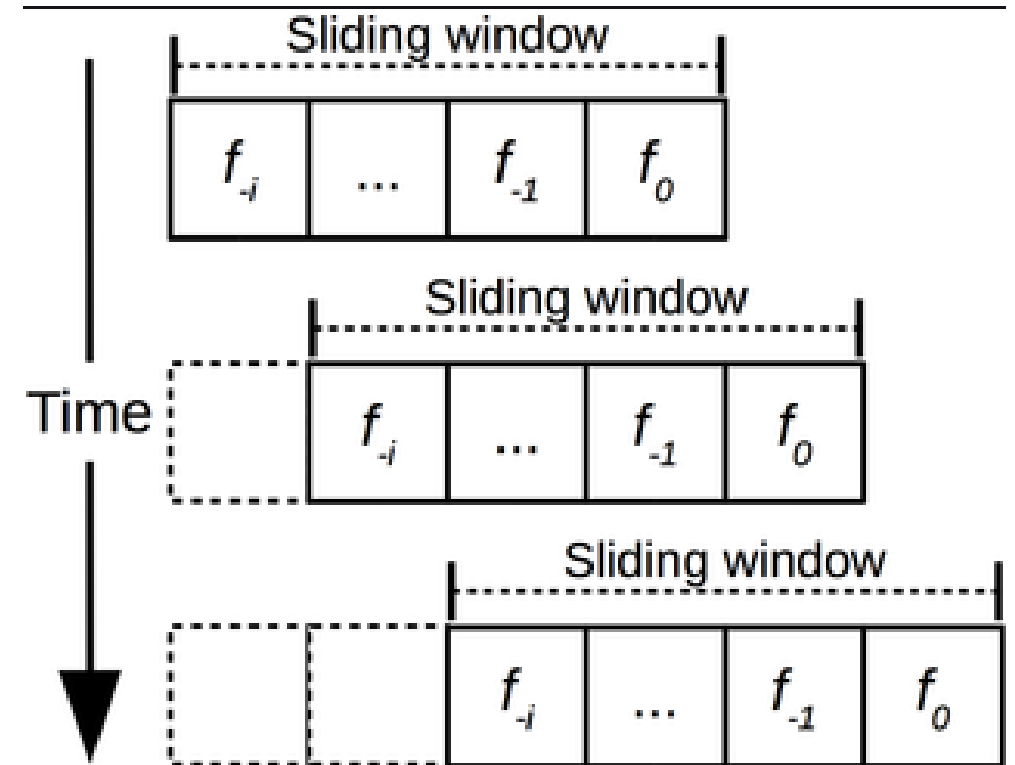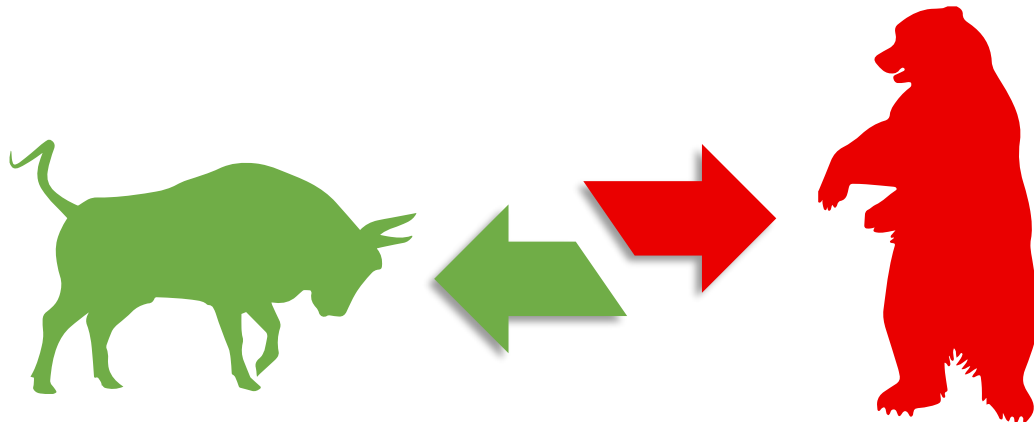
# K-Fold Cross-Validation



-RMSE lowest for train sets between 600 and 900 days

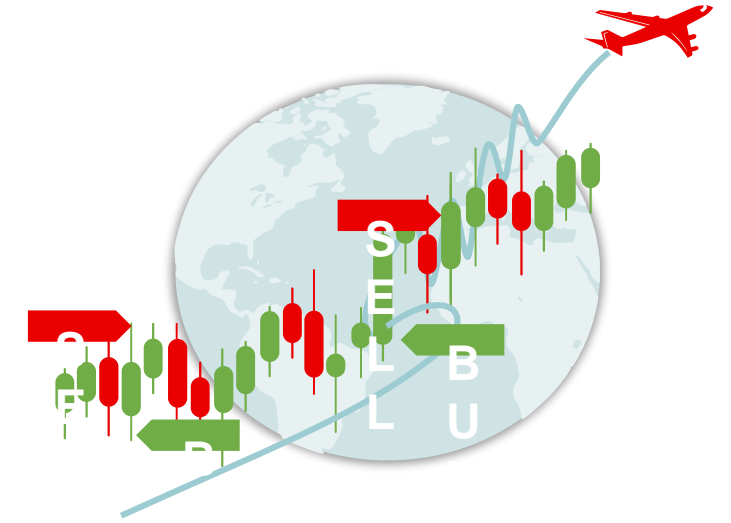-Rerunning Neural Network at 900 days significantly improves results.
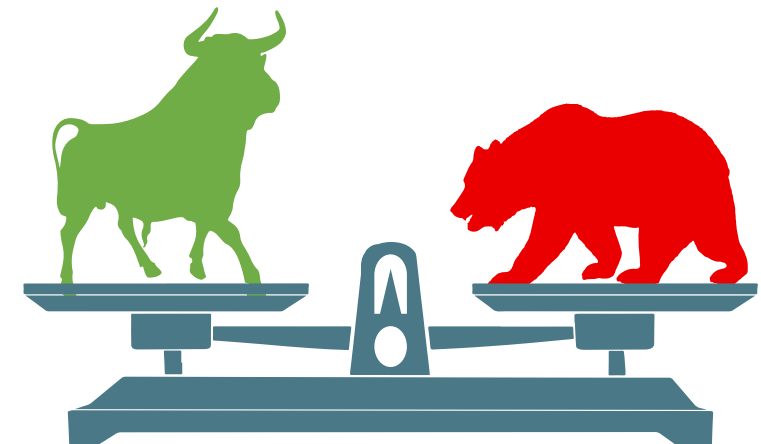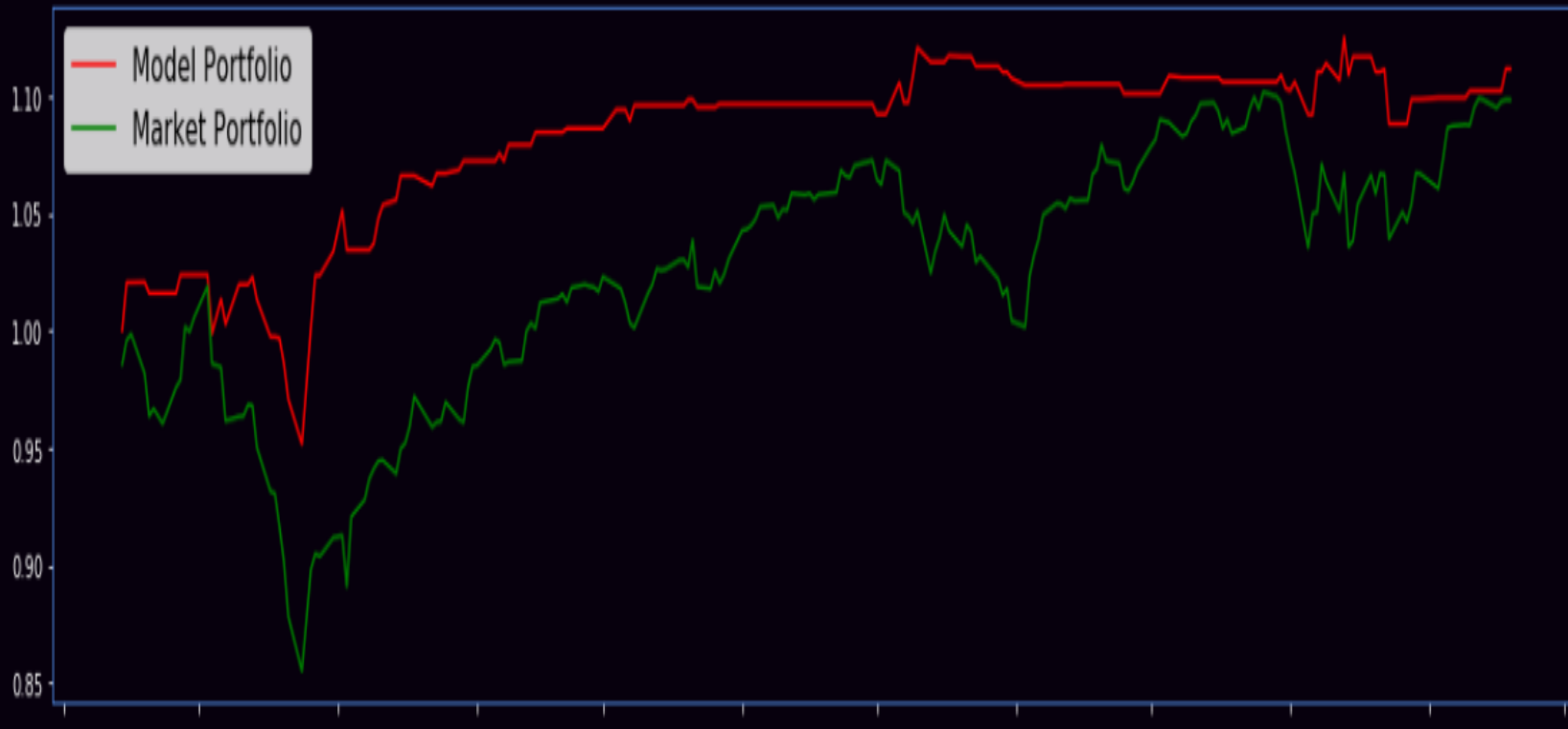
# IV - Trading Strategy

# Trading rules and r e t u r n s


Model Protfolio vs Market Portfolio Returns



✓ BUY as soon as prices is or below model predicted price

✓ SELL at day close

Thank You!