Define problem statement

Identify Data Sources

Clean Data

Discuss results and analysis

Implement ML algorithm(s)

Identify one or more ML algorithms

Feature selection

Conclusion

Data Cleaning steps:

S&P data:

* Take S&P 500 Historical data\_Start till 2018\_csv from datahub.io \cite{website12}

This gives S&P500 data till 2018. Has 1768 rows

Attributes: Data, SP500, CPI, Long Interest Rate

* Take S&P 500 Historical Data.xls file from Investing.com \cite{website11}

Sort the file by date from old to new. Took 59 records from this data set

Take S&P500 data from 5/1/2018 till 2023 from this file

Attributes: Price

* Random sampling to check the consistency of between data sets. The Price in the First data set lies within the High and Low prices in the second data set for each month. Hence good to use

GDP data:

* Take GDP data set from Fed \cite{website13}. This gives monthly GDP from 1992

Attribute: Nominal GDP

* Change all date formats on all Data sets to be consistent MM/DD/YYYY
* Take Pre 1992 GDP data from the website \cite{website14}

Change the Date format on this data set to match the date format on the Consolidated Data set. YYYY-MM-DD is changed to MM/DD/YYYY

Do Vlook up on date from the Consolidated Data set and obtain the GDP numbers from this data set

* Random sampling to check the consistency of between data sets. The Price in the First data set is pretty close to that from the second data set for each month. Hence good to use

Fed rate:

* Take the Fedfunds rate data from \cite{website06}

Do Vlook up on date from the Consolidated Data set and obtain the Fed rate numbers from this data set

Inflation data:

* Use the Historical Inflation data obtained from BLS
* This data needs significant transformation

Unemployment data:

* Use the Historical Inflation data obtained from BLS
* This data needs significant transformation

MACD:

* Plug in the Data and SP500 values from the Consolidated Data set into the MACD template
* Do Vlook up on date from the Consolidated Data set and obtain the MACD numbers from this data set

RSI:

* Plug in the Data and SP500 values from the Consolidated Data set into the RSI template
* Do Vlook up on date from the Consolidated Data set and obtain the RSI numbers from this data set