STOCK MARKET PREDICTOR

USING HIDDEN MARKOV MODELS

Machine Intelligence

BACHELOR OF TECHNOLOGY- V Sem CSE

Department of Computer Science & Engineering

SUBMITTED BY

Batch No:

Student name 1: Rishikesh Sastri SRN: PES1UG20CS340
Student name 2: Rithvik Grandhi SRN: PES1UG20CS342
Student name 3: Rohan C SRN: PES1UG20CS345

PES UNIVERSITY

(Established under Karnataka Act No. 16 of 2013)

100 Feet Ring Road, BSK III Stage, Bengaluru-560085

Abstract and Scope

Hidden Markov Models are an incredibly interesting type of stochastic process that are under-utilized in the Machine Learning world. They are particularly useful for analyzing time series. This, combined with their ability to convert the observable outputs that are emitted by real-world processes into predictable and efficient models makes them a viable candidate to be used for stock market analysis. The stock market has several interesting properties that make modeling non-trivial, namely volatility, time dependence and other similar complex dependencies. HMMs are suited to dealing with these complications as the only information they require to generate a model is a set of observations (in this case historical stock market data). Our project emphasizes on using the HMM efficiently and tackle the properties mentioned above. It first builds a training model on data collected on the given set of days and uses the model created to predict the prices into N (it being the input given by user) days in the future. The accuracy of the model depends on the number of training cases run to create the model. The project used the following dependencies given below.

Dependencies

- Pandas_datareader - Allows one to download data directly from Yahoo finance
- NumPy - Required for fast manipulation of financial data (e.g., calculating fractional change)
- Matplotlib - Required for visualisation of results
- Hmmlearn - Open-source package that allows for creation and fitting of HMM's
- Sklearn - Used to calculate metrics to score the results and split the data, will be removed in future to reduce dependency
- Tqdm - Used to track progress whilst training
- Argparse - Required for console inputs

As the project depends on HMMs, it has quite the large scope. Basically, HMMs are used in fields where the dataset or sequence is not immediately observable. Some other fields where HMMs have exceptional applications are Neuroscience, Cryptanalysis, Machine translation, Gene Prediction and Virus Detection.

Feasibility Study:

Predicting the stock market is one of the most important applications of Machine Learning in finance. Predicting the stock market has been the bane and goal of investors since its inception. Every day billions of dollars are traded on the stock exchange, and behind every dollar is an investor hoping to make a profit in one way or another. Entire companies rise and fall daily depending on market behaviour. If an investor is able to accurately predict market movements, he offers a tantalizing promise of wealth and influence. Today, so many people are making money staying at home trading in the stock market. It is a plus point for you if you use your experience in the stock market and your machine learning skills for the task of stock price prediction.

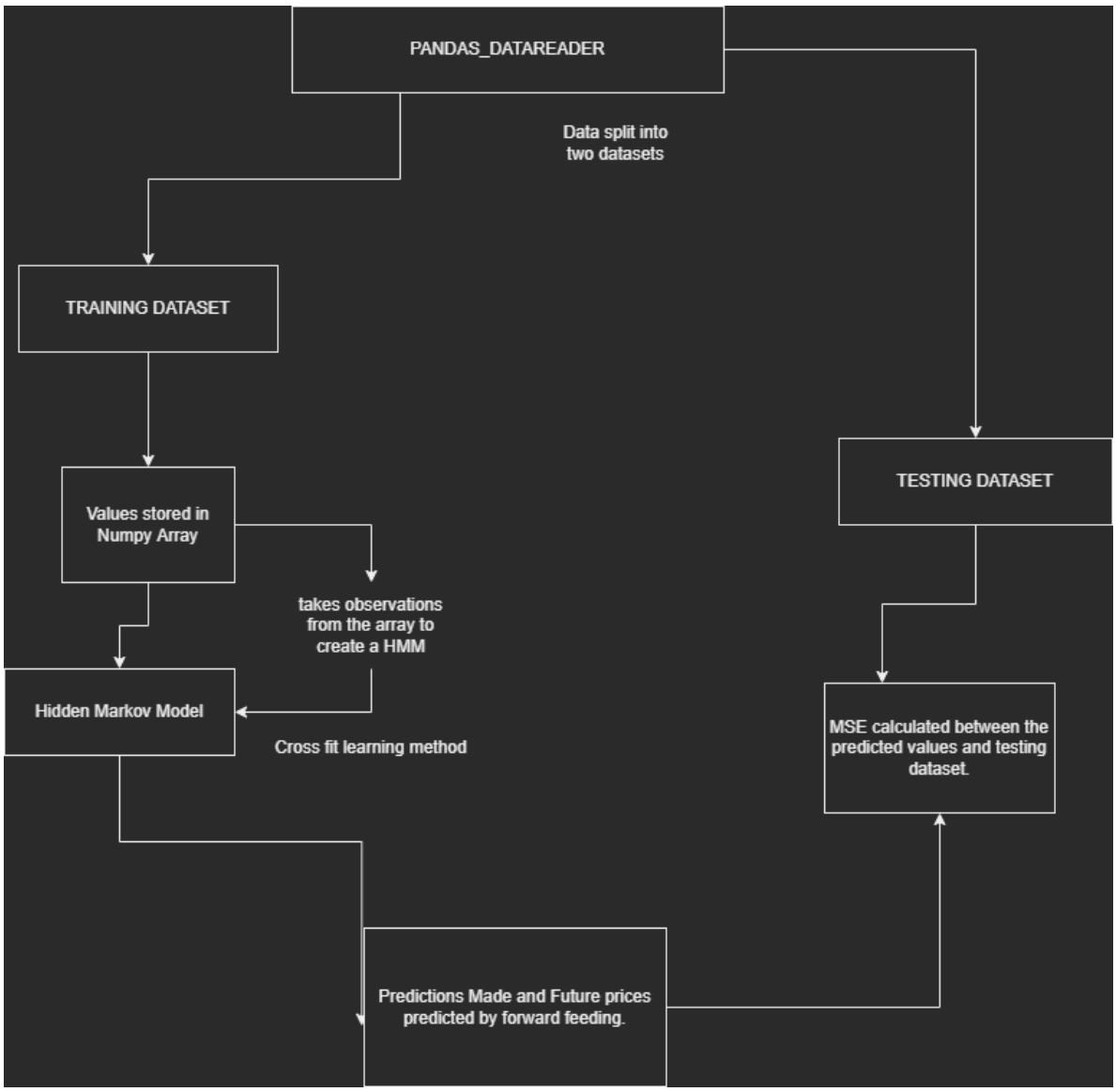
We are taking the apple stocks as an example for the project demo. The was due to recently conducted September event of Apple, one of the most favorite events for all Apple users. Every time a company comes up with an innovative product, it leads to an increase in its stock price.

Design Approach/ Methodology/ Planning of work

Stock market data is downloaded via pandas_datareader and the data is split into training and testing datasets. The fractional changes for any given day (from open to close, high to open, open to low) in the training dataset are computed and stored in a NumPy array. These fractional changes can be seen as the observations for the HMM and are used to train the continuous HMM with hmm learns fit method. The model then predicts the closing price for each day in the training dataset, based on the given days opening price. This prediction is found by calculating the highest scoring potential outcome out of a pre-determined set of outcomes (e.g., +0.001%, -0.001% etc.).

All predictions as well as the actual close prices for the testing period are stored in an excel file and the Mean Squared Error between the two is printed out. The MSE is also

included in the file name for future reference. Future days are predicted by feeding forward the prediction values. Unfortunately, at present there is no method in place to account for overnight/ weekend futures trading, and so for the future predictions the n+1 days open price is the same as the nth days closing price.



References

Historical Data –
https://quantquote.com/historical-stock-data
Description of Technical Analysis -
http://www.investopedia.com/terms/t/technicalanalysis.asp
Berkeley paper on ANN-based analysis -
http://www.cs.berkeley.edu/~akar/IITK_website/EE671/report_stock.pdf
Efficient Market Hypothesis -
http://www.e-m-h.org/Fama70.pdf
Finding anomalies in time series –
https://blog.twitter.com/2015/introducing-practical-and-robust-anomaly-detection-in-a-time-series
Scale: The universal laws of growth –
http://www.penguinrandomhouse.com/books/314049/scale-by-geoffrey-west/9781594205583/
Nonparametric Econometrics: A Primer –
http://socserv.mcmaster.ca/racine/ECO0301.pdf
An Investigation into the Use of Reinforcement Learning Techniques within the Algorithmic Trading Domain –
http://www.doc.ic.ac.uk/teaching/distinguished-projects/2015/j.cumming.pdf
Technical Analysis –
http://www.mrao.cam.ac.uk/~mph/Technical_Analysis.pdf
An empirical analysis of algorithmic trading on financial markets –
http://lib.ugent.be/fulltxt/RUG01/001/315/567/RUG01-001315567_2010_0001_AC.pdf