A Mini Project Synopsis

On

STOCK PRICE PREDICTION

Submitted in partial fulfillment of the requirement of
University of Mumbai for the Degree of

Bachelor of Engineering
In
Computer Engineering

Submitted By
Sudhanshu Chaurasia
Dhanesh .S
Jayesh Girdhar
KMR Dayasagar

Supervisor
Sujit Tilak

Department of Information Technology

PILLAI COLLEGE OF ENGINEERING

New Panvel – 410 206

UNIVERSITY OF MUMBAI

Academic Year 2020-21

DEPARTMENT OF INFORMATION TECHNOLOGY
Pillai College of Engineering
New Panvel – 410 206

CERTIFICATE

This is to certify that the requirements for the TE Mini Project Synopsis entitled '**Stock Price Prediction**' have been successfully completed by the following students:

Name	**Roll No.**
Sudhanshu Chaurasia	B 09
Dhanesh S	B 12
Jayesh Girdhar	B 20
KMR Dayasagar	B 25

in partial fulfillment of Bachelor of Engineering of Mumbai University in the Department of Computer Engineering, Pillai College of Engineering, New Panvel – 410 206 during the Academic Year 2018 – 2019.

_____

Supervisor
(Name of Project Guide)

_____ _____

Head of Department **Principal**
Dr. Sharvari Govilkar **Dr. Sandeep M. Joshi**

DEPARTMENT OF INFORMATION TECHNOLOGY
Pillai College of Engineering
New Panvel – 410 206

SYNOPSIS APPROVAL FOR B.E

This TE Mini Project Synopsis entitled "**Stock Price Prediction**" by **Sudhanshu , Dhanesh , Jayesh , Dayasagar** are approved for the degree of B.E. in **Computer Engineering**.

Examiners:

1. _____

2. _____

Supervisors:

1. _____

2. _____

Chairman:

1. _____

Date: 14/05/2021

Place:

Declaration

We declare that this written submission for TE Mini Project Synopsis Declaration entitled "**Project Title**" represent our ideas in our own words and where others' ideas or words have been included. We have adequately cited and referenced the original sources. We also declared that we have adhere to all principles of academic honesty and integrity and have not misrepresented or fabricated or falsified any ideas / data / fact / source in our submission. We understand that any violation of the above will cause for disciplinary action by institute and also evoke penal action from the sources which have thus not been properly cited or from whom paper permission have not been taken when needed.

Project Group Members:

Sudhanshu Chaurasia.: _____

Dhanesh S. : _____

Jayesh Girdhar : _____

KMR Dayasagar : _____

Date:

Place:

Table Of Contents

Chapter 01

Introduction

Today we live and breathe data. Forecasting the stock exchange data is an important financial subject which involves an assumption that the fundamental information publicly available in the past has some predictive relationships to the future stock returns. Stock market forecasting contains uncovering the market trends, planning investment tactics, identifying the best time to purchase the stocks and which stocks to purchase. A stock exchange or equity business sector is a non-direct, non-parametric framework that is difficult to model with any sensible exactness. It is the mix of speculators who need to purchase or offer or hold a share at a specific time. Prediction will continue to be an exciting locale of research, making scientists in the analytics field always desiring to enhance the existing forecasting models. The motivation is that companies and individuals are empowered to make investment decisions to develop viable system about their future endeavors.

Stock trend forecasting is considered as one of the most difficult tasks to achieve in money related gauging because of the difficulty in the multifaceted world of stock market. Many of the investors in the stock market are finding a technique that could guarantee easy profiting by forecasting the stock trends and minimize the risk of investing. This motivates the researchers in the domain field to delve and develop new forecasting models. Time series data analysis techniques use verifiable information as the premise for evaluating future results. Time series data can be defined as numerical data collected in a sequence over a period at regular intervals. The time series data can include the values collected at the end of every week, month, quarter, or year etc. The intention is to find if there is any link between the data collected so far and in what way does the data changes. To reduce the risk of investment, exchange of securities between the seller and buyer are facilitated by the stock exchanges. A stock exchange is an organization or a place where the stock traders or investors can deal with stocks. Some of the examples for stock market organizations include NASDAQ, NYSE, BSE, NSE etc.

Python is a programming language for statistical processing.

Chapter 02

Literature Review

Kannan, Sekar, Sathik and P. Arumugam in [13] used data mining technology to discover the hidden patterns from the historic data that have probable predictive capability in their investment decisions. The prediction of stock market is challenging task of financial time series predictions.

Jing Tao Yao and chew Lim tan in [14] used artificial neural networks for classification, prediction and recognition. Neural network training is an art. Trading based on neural network outputs, or trading strategy is also an art. Authors discuss a seven-step neural network prediction model building approach in this article. Pre and post data processing/analysis skills, data sampling, training criteria and model recommendation will also be covered in this article.

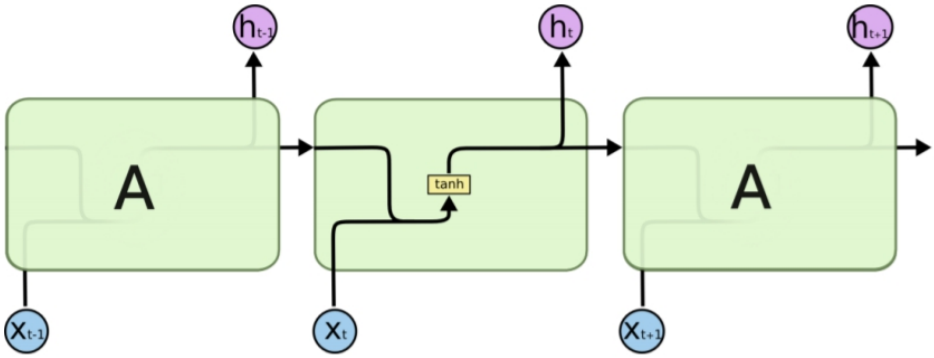
Tiffany Hui-Kuang and Kun-Huang Huarng in [15] used neural network because of their capabilities in handling nonlinear relationship and also implement a new fuzzy time series model to improve forecasting. The fuzzy relationship is used to forecast the Taiwan stock index. In the neural network fuzzy time series model where as insample observations are used for training and out-sample observations are used for forecasting. The drawback of taking all the degree of membership for training and forecasting may affect the performance of the neural network. To avoid this take the difference between observations. These reduce the range of the universe of discourse.

Md. Rafiul Hassan and Baikunthu Nath in [16] used Hidden Markov Models (HMM) approach to forecasting stock price for interrelated markets. HMM was used for pattern recognition and classification problems because of its proven suitability for modeling dynamic system. The author summarized the advantage of the HMM was strong statistical foundation. It‟s able to handle new data robustly and computationally efficient to develop and evaluate similar patterns. The author decides to develop hybrid system using AI paradigms with HMM improve the accuracy and efficiency of forecast the stock market.

Chapter 03

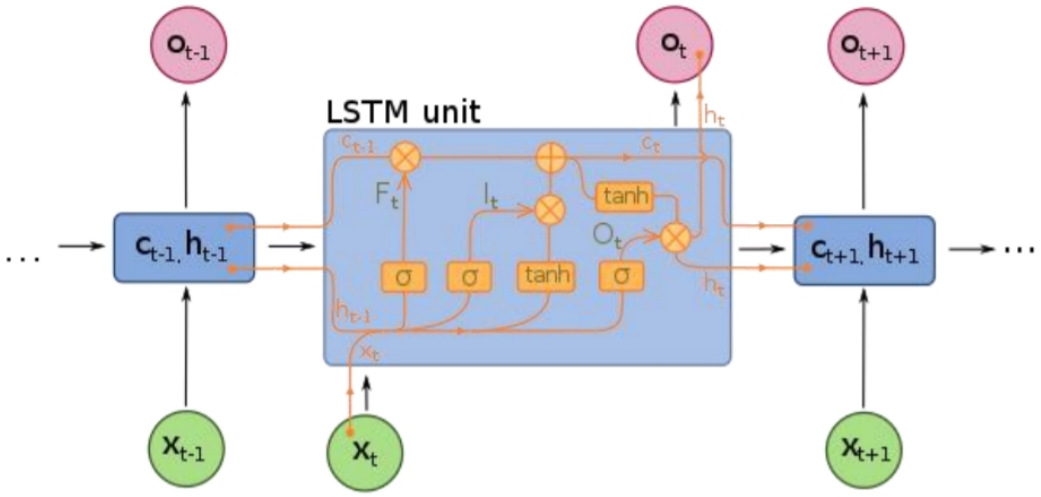
Methodology

Long short-term memory (LSTM) is a type of recurrent neural-network architecture in which the vanishing gradient problem is solved. LSTMs are capable of learning very long-term dependencies and they work tremendously well on a large variety of problems. LSTMs are first introduced by Hochreiter et al. in 1997 [10]. In addition to the original authors, many researchers contributed to the architecture of modern LSTM cells. Recurrent neural networks are generally designed in a chain like structure, looping back to the previous layers. In standard RNNs, this looping module will have a very simple structure, as shown in the Figure 3.1. This structure can be a simple tanh layer controlling the flow.



3.1 RNN looping module

Whereas in LSTMs, instead of this simple structure, they have four network layers interacting in a special way. General architecture of LSTMs is shown in Figure 5. Normally LSTM's are augmented by gates called "forget" gates. By controlling these gates, errors can be backpropagated through any number of virtual layers. This mechanism enables the network to learn tasks that depend on events that occurred millions of time steps ago.Fig 3.2 shows LSTM unit


LSTM unit
o_{t-1}
o_t
o_{t+1}
c_{t-1}, h_{t-1}
c_{t+1}, h_{t+1}
x_{t-1}
x_t
x_{t+1}
F_t
I_t
O_t
c_t
h_t
tanh
σ
σ
tanh
σ

3.1 LSTM UNIT

Chapter 04

System Analysis

Long Short-Term Memory models are extremely powerful time-series models. They can predict an arbitrary number of steps into the future. An LSTM module (or cell) has 5 essential components which allows it to model both long-term and short-term data.

- Cell state (c_t) - This represents the internal memory of the cell which stores both short term memory and long-term memories
- Hidden state (h_t) - This is output state information calculated w.r.t. current input, previous hidden state and current cell input which you eventually use to predict the future stock market prices. Additionally, the hidden state can decide to only retrieve the short or long-term or both types of memory stored in the cell state to make the next prediction.
- Input gate (i_t) - Decides how much information from current input flows to the cell state
- Forget gate (f_t) - Decides how much information from the current input and the previous cell state flows into the current cell state
- Output gate (o_t) - Decides how much information from the current cell state flows into the hidden state, so that if needed LSTM can only pick the long-term memories or short-term memories and long-term memories

The prediction of stock value is a complex task which needs a robust algorithm background in order to compute the longer term share prices. Stock prices are correlated within the nature of market; hence it will be difficult to predict the costs. The proposed algorithm using the market data to predict the share price using machine learning techniques like recurrent neural network named as Long Short Term Memory, in that process weights are corrected for each data points using stochastic gradient descent. This system will provide accurate outcomes in comparison to currently available stock price predictor algorithms.

Chapter 05

Software Description

Here is a detailed information and aspects of the softwares,API's and python libraries and modules used in this project.

1]WINDOWS 10:

Windows 10 is a Microsoft operating system for personal computers, tablets, embedded devices and internet of things devices.Windows 10 features
built-incapabilities that allow corporate IT departments to use mobile device
management (MDM) software to secure and control devices running the operating
system. In addition, organizations can use traditional desktop management software
such as Microsoft System Center Configuration Manager.

2] Visual Studio Code:

Visual Studio Code is a freeware source-code editor made by Microsoft for Windows, Linux and macOS. Features include support for debugging, syntax highlighting, intelligent code completion, snippets, code refactoring, and embedded Git.

3] Python 3.8.0:

Python is a multi-paradigm programming language. Object-oriented programming and structured programming are fully supported, and many of its features support functional programming and aspect-oriented programming (including by metaprogrammingand metaobjects (magic methods)).Many other paradigms are supported via extensions, including design by contract and logic programming.

4] NumPy:

NumPy is a Python library that provides a simple yet powerful data structure: the n-dimensional array. This is the foundation on which almost all the power of Python's data science toolkit is built, and learning NumPy is the first step on any Python data scientist's journey. This tutorial will provide you with the knowledge you need to use NumPy and the higher-level libraries that rely on it.

4] Pandas:

The Pandas datareader is a sub package that allows one to create a dataframe from various internet datasources, currently including:

- Yahoo! Finance
- Google Finance
- St.Louis FED (FRED)
- Kenneth French's data library
- World Bank
- Google Analytics

5] Scikit Learn:

Scikit-learn (formerly scikits.learn and also known as sklearn) is a free software machine learning library for the Python programming language.[3] It features various classification, regression and clustering algorithms including support vector machines, random forests, gradient boosting, *k*-means and DBSCAN, and is designed to interoperate with the Python numerical and scientific libraries NumPy and SciPy.

7] MatPlotlib:

Matplotlib is a comprehensive library for creating static, animated, and interactive visualizations in Python. Matplotlib makes easy things easy and hard things possible. Create. Develop publication quality plots with just a few lines of code.

Matplotlib is a cross-platform, data visualization and graphical plotting library for Python and its numerical extension NumPy. As such, it offers a viable open source alternative to MATLAB. Developers can also use matplotlib's APIs (Application Programming Interfaces) to embed plots in GUI applications.

A Python matplotlib script is structured so that a few lines of code are all that is required in most instances to generate a visual data plot.

Chapter 06

Need of Stock Prediction

Market prediction offers great profit avenues and is a fundamental stimulus for most researchers in this area. To predict the market, most researchers use either technical or fundamental analysis. Technical analysis focuses on analyzing the direction of prices to predict future prices, while fundamental analysis depends on analyzing unstructured textual information like financial news and earning reports. More and more valuable market information has now become publicly available online. This draws a picture of the significance of text mining strategies to extract significant information to analyze market behavior. While many papers reviewed the prediction techniques based on technical analysis methods, the papers that concentrate on the use of text mining methods were scarce. In contrast to the other current review articles that concentrate on discussing many methods used for forecasting the stock market, this study aims to compare many machine learning (ML) and deep learning (DL) methods used for sentiment analysis to find which method could be more effective in prediction and for which types and amount of data.

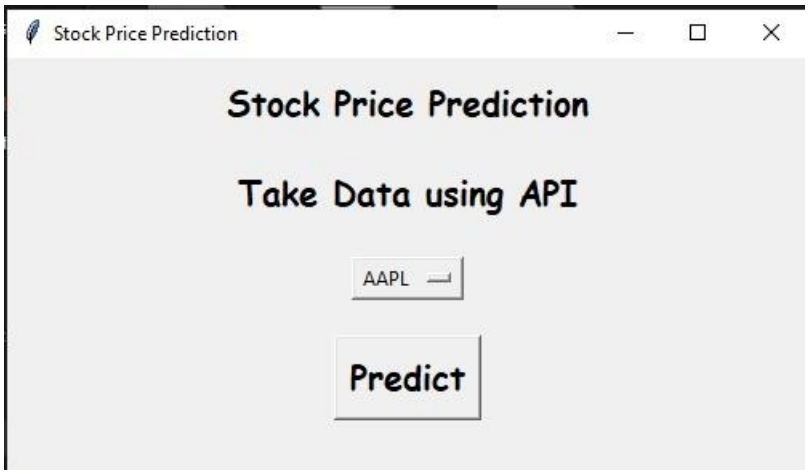
Chapter 7

Features

In this project we are trying to predict the highest price of the stocks of a particular company on everyday basis.

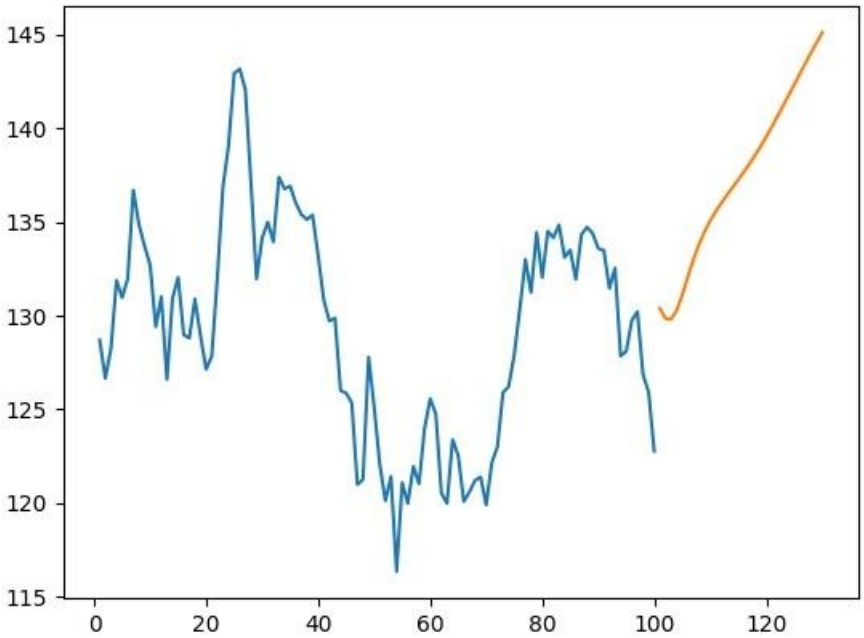
In this type of feature extraction we are predicting the stock price of each company by finding a relation between different companies. This inter-relation between these
companies will be used to predict the stock prices of a particular company in a better way. Each company's data will be individually used to predict each other company's stock price. The top three companies, which can predict a particular company with higher accuracy, will be used together to predict the stock price of a particular company.

We made a user friendly interface for ease.(Fig 7.1)

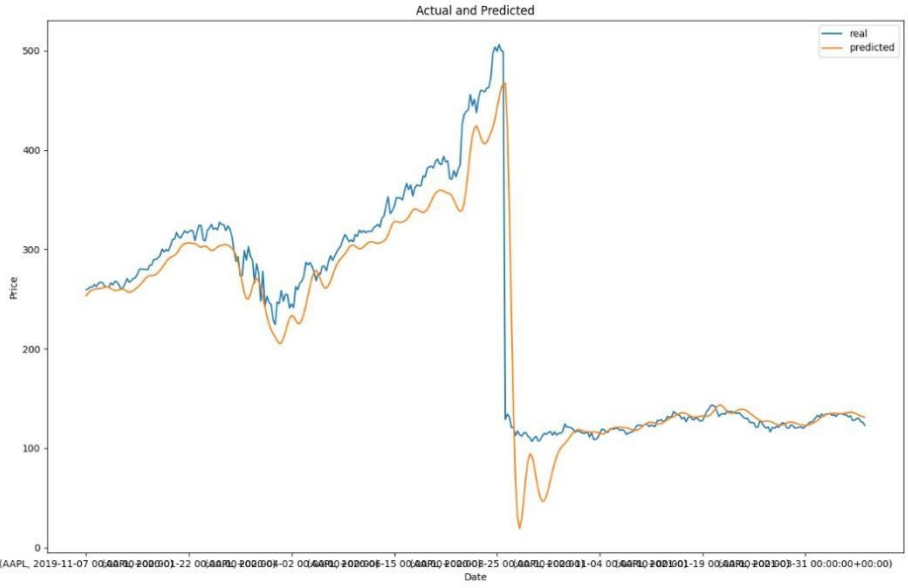


7.1 GUI Menu

The output shows us the graphical analysis



7.2 Thirty days ahead prediction using LSTM



7.3 Actual v/s Prediction

Chapter 8

Design and Procedure

Long-Term Dependencies problems in using RNN

RNN usually don't face any problems in connecting the past information to the present task because of its chain-like structure formed due to loops in the network but it is also possible that the gap between the relevant information in the past and the point in the present where it is to be needed becomes very large, then in such cases, it could become challenging for RNN to be able to learn to connect the information and finding patterns in the sequence of data. This is due to the Vanishing Gradient Problem.

What is Vanishing Gradient Problem?

In back propagation, the weight of the neural network is updated proportionally to the partial derivative of the error function for the current weights in each of the iterations of the training process.

But the problem arises when in some of the cases the gradients will be vanishingly small, that the value of the weight does not change at all and this may cause the neural network to completely stop the further training of the network.

This led to the invention of so-called LSTM.

LSTM Model in Python using TensorFlow and Keras

Now let us see how to implement an LSTM Model in Python using TensorFlow and Keras taking a very simple example.

Steps:

- Prepare the data
- Feature Scaling (Preprocessing of data)
- Split the dataset for train and test
- Converting features into NumPy array and reshaping the array into shape

accepted by LSTM model

- Build the architecture for LSTM network
- Compile and fit the model (Training)
- Evaluate the performance of model(Test)

Import all the required python packages and libraries

Chapter 09

Conclusion

In this project we have performed experiments on a novel approach to predict the stock prices using information data.The study of the share is carried out in this project and it can be carried out for several shares in the future.Prediction could be more reliable if the model trains a greater number of data sets using higher computing capacities, an increased number of layers, and LSTM modules.

Chapter 10

Future Scope

In future enhancement the inclusion of sentiment analysis from social media to understand what the market thinks about the price variation for a particular share and it can be implement this by adding twitter and Facebook API(Application Programming Interface) to our program as Facebook is a leading social media which has lots of market trend information posted by users.

References

[1] Wenping Zhang, Chunping Li, Yunming Ye, Wenjie Li and Eric W.T. Ngai "Dynamic Business Network Analysis for

Correlated Stock Price Movement Prediction", Published by the IEEE Computer Society 2015

[2] Chia-Hsuan Yeh and Chun-Yi Yang "Social Networks and Asset Price Dynamics", IEEE TRANSACTIONS ON EVOLUTIONARY COMPUTATION, , JUNE 2015

[3] Li-Xin Wang "Dynamical Models of Stock Prices Based on Technical Trading Rules Part I: The Models", IEEE TRANSACTIONS ON FUZZY SYSTEMS, AUGUST 2015

[4] Xiaodong Li a, Haoran Xie a,⇑, Li Chen b, Jianping Wanga, Xiaotie Deng "News impact on stock price return via sentiment analysis", ScienceDirect 2014

[5] Hong Keel Sul,Alan R. Dennis,Lingyao (Ivy) "Trading on Twitter: The Financial Information Content of Emotion in Social Media", 2014 47th Hawaii International Conference on System Science

[6] Abbasi, H. Chen, A. Salem, Sentiment analysis in multiple languages: feature selection for opinion classification in web forums, ACM Trans. Inform. Syst.(TOIS) 26 (2008) 12.

[7] S. Bao, S. Xu, L. Zhang, R. Yan, Z. Su, D. Han, Y. Yu, Mining social emotions from affective text, IEEE Trans. Knowl. Data Eng. 24 (2012) 1658–1670.

[8] M. Bautin, L. Vijayarenu, S. Skiena, International sentiment analysis for news and blogs, in: Proceedings of the International Conference on Weblogs and Social Media, 2008.

[9] F. Bießmann, J.M. Papaioannou, M. Braun, A. Harth, Canonical trends: detecting trend setters in web data, in: International Conference on Machine Learning, 2012.

[10] E. Cambria, C. Havasi, A. Hussain, Senticnet 2: a semantic and affective resource for opinion mining and sentiment analysis, in: FLAIRS Conference,

2012, pp. 202–207.

[11] E. Cambria, T. Mazzocco, A. Hussain, Application of multi-dimensional scaling and artificial neural networks for biologically inspired opinion mining, Biol. Inspired Cogn. Architec. 4 (2013) 41–53.

[12] E. Cambria, B. Schuller, Y. Xia, C. Havasi, New avenues in opinion mining and sentiment analysis, IEEE Intell. Syst. (2013).

[13] K. Senthamarai Kannan, P. Sailapathi Sekar, M.Mohamed Sathik and P. Arumugam, "Financial stock market forecast using data mining Techniques", 2010, Proceedings of the international multiconference of engineers and computer scientists.

[14] JingTao YAO and Chew Lim TAN , "Guidelines for Financial Prediction with Artificial neural networks".

[15] Tiffany Hui-Kuang yu and Kun-Huang Huarng, "A Neural network-based fuzzy time series model to improve forecasting", *Elsevier*, 2010, pp: 3366-3372. [16] Md. Rafiul Hassan and Baikunth Nath, "Stock Market forecasting using Hidden Markov Model: A New Approach", th

Proceeding of the 2005 5 international conference on intelligent Systems Design and Application 0-7695-2286-06/05, IEEE 2005.

[17] Ching-Hsue cheng, Tai-Liang Chen, Liang-Ying Wei, " A hybrid model based on rough set theory and genetic algorithms for stock price forecasting", 2010, pp. 1610-1629.

M.H. Fazel Zarandi, B. Rezaee, I.B. Turksen and E.Neshat, "A type-2 fuzzy rule-based experts system model for stock price analysis", Expert systems with Applications, 2009, pp. 139-154.

Appendix

The basic details to check the prediction:

-Take any company’s data using API

-Click the predict button

Index