

Estimating Stock Price Based on Information From Financial Statement Using Machine Learning Approach



CONTENTS

01 Introduction

02 Estimating Stock Price Based on Information From Financial Statement Using Machine Learning Approach

Introduction

01

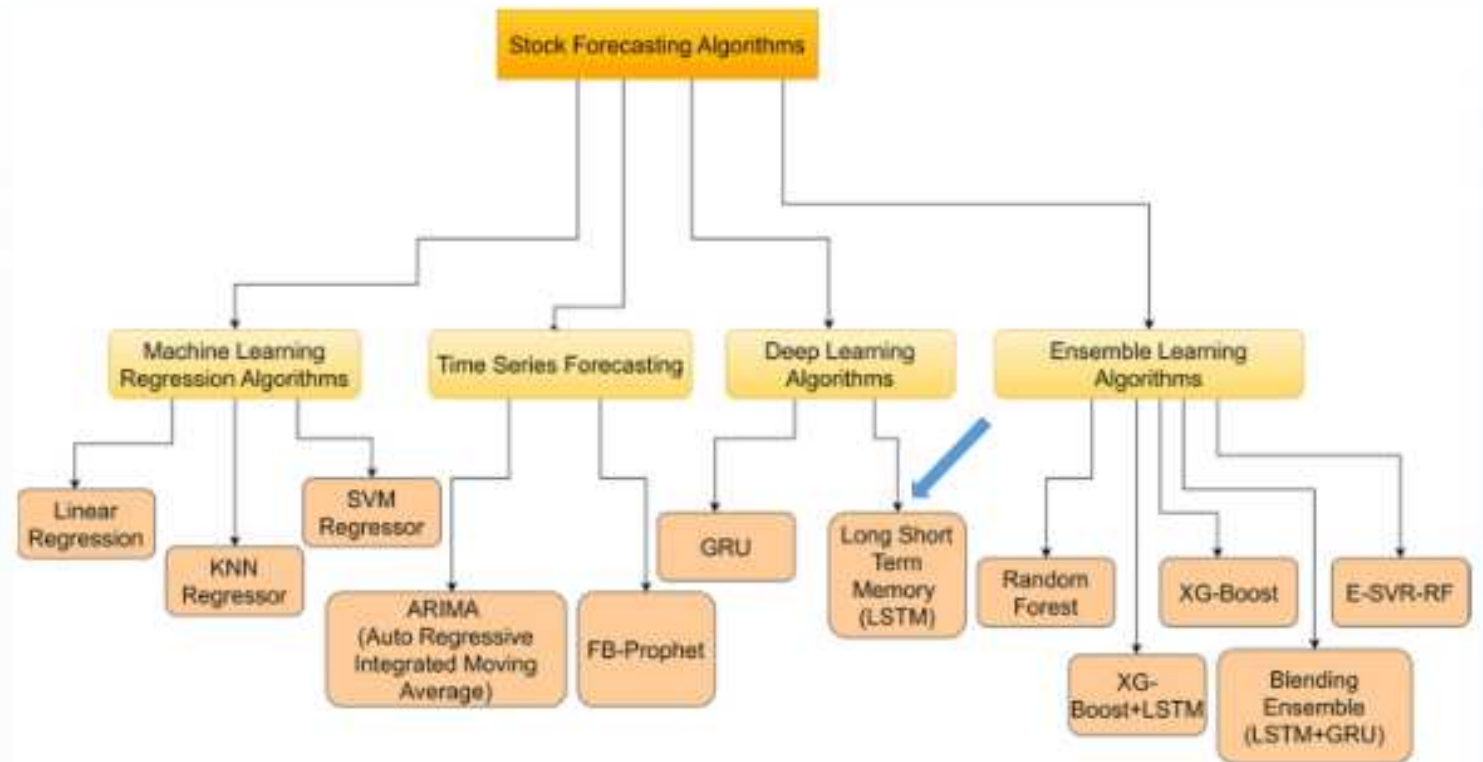




Background

Stock Forecasting Algorithm

Based on the studies presented so far, the best stock market forecasting solution is derived from fundamental analysis and technical analysis with sentiment analysis and deep learning models. Ensemble techniques provide especially promising forecasting outcomes. The algorithms that are used for stock market prediction by considering research papers are given in this figure.



**Estimating Stock Price
Based on Information
From Financial
Statement Using
Machine Learning
Approach**

02





About Research Paper



Title

Estimating stock price based on information from financial statement using machine learning approach



Publisher & Publish Date

A Thesis Submitted for the Degree of Master of Science in Computer Science
Department of Computer Engineering
FACULTY OF ENGINEERING - Chulalongkorn University in Thailand - 2022



Author

Thitikun Kunathananon



Methodology

The Proposed Model

The Long Short-Term Memory (LSTM) model is a Recurrent Neural Network (RNN) designed to handle time-series data.

Model Inputs and Features

The LSTM model takes both financial and macroeconomic indicators as input features

Model Output & Evaluation

The LSTM model predicts the percentage change in stock price from the previous quarter.

MAE (Mean Absolute Error) is used as the cost function to measure prediction accuracy

The Long Short-Term Memory (LSTM)

The Long Short-Term Memory (LSTM) model is a Recurrent Neural Network (RNN) designed to handle time-series data.

In this study, LSTM processes financial statement data at quarterly intervals to capture long-term dependencies and trends in stock prices.

The model's output is the predicted percentage change in stock price compared to the previous quarter.

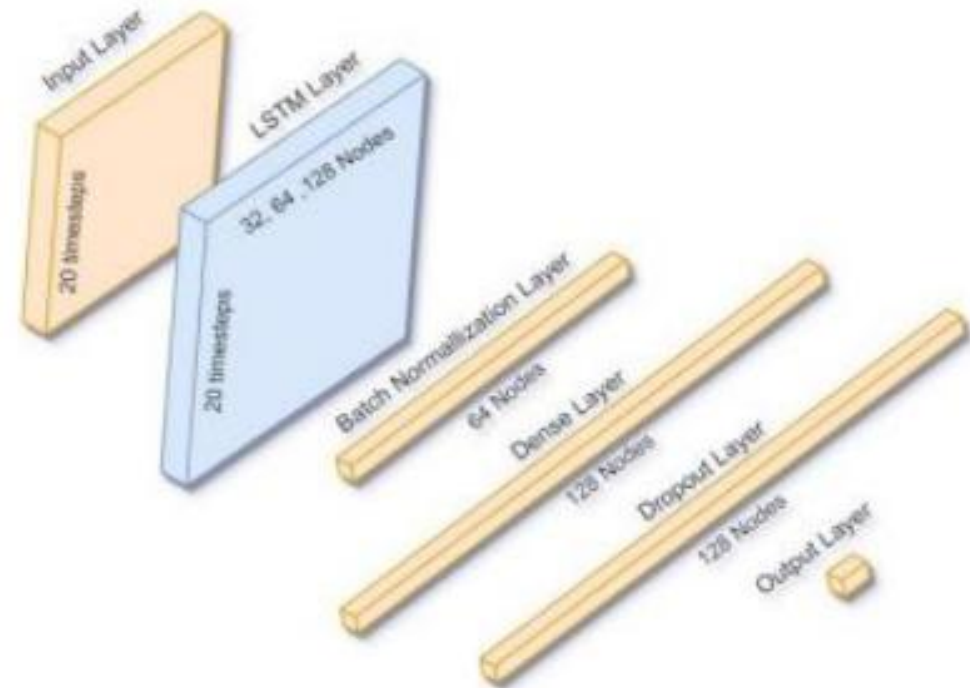


Figure 8 Our proposed LSTM architecture



Model Inputs and Features



Financial Statement Data Features

Total Assets	Total Equity	Total Liability
Earnings Per Share (EPS)	Closing Price	Market Capitalization
Price-to-Earnings Ratio (P/E Ratio)		



Macroeconomic Indicators

SET Index	Interest Rate	Inflation Rate
Gross Domestic Product (GDP)		





Data Description

Part 01

The selected data is a stock with data prior to 2005 from The Securities Exchange of Thailand (SET) and the Market for Alternative Investment (MAI).

Part 02

The Bank of Thailand

The National Economic and Social
Development Board

Part 03

The model utilized in this study outputs the daily adjusted close price, which is adjusted for stock splits, dividends, and capital gain distributions, and obtained from Yahoo Finance



Data Description

The input factors and descriptions are shown in Table

Factor	Description
M_TOTAL_ASSET	Total assets of each stock
M_TOTAL_EQUITY	Total equity of each stock
M_TOTAL_LIABILITY	Total liability of each stock
Z_PAR	Par of each stock
M_NET_PROFIT	Net profit of each stock
R_EPS	Earning per share of each stock
R_PE	Price to Earning Per Share of each stock
R_PB	Price to Book Value Ratio of each stock
M_BOOK_VALUE	Book value of each stock
R_TURNOVER	Turn over rate of each stock
M_MKT_CAP	Market Capitalization of each stock
R_INDEX_CLOSE	Closing SET index
M_MKT_PE	Market price to earnings
M_MKT_PBV	Market price to book value
M_MKT_YIELD	Market yield
M_MKT_CAP_SET	Market Capitalization of SET
Infection rate	Inflection Rate
Interest	Interest Rate
GDP	Gross Domestic Product



Data Splitting

The data has been separated into twenty quarters (five years) of training data, four quarters (one year following the training period) of validation data, and four quarters (one year following the validation period) of test data, as indicated in the figure

Year	2013				2014				2015				2016				2017				2018				2019				2020				2021			
Quarter	1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4
Ex. set 1																																				
Ex. set 2																																				
Ex. set 3																																				
Ex. set 4																																				
Ex. set 5																																				
Ex. set 6																																				
Ex. set 7																																				
Ex. set 8																																				



Training data



Validation data



Test data



Model Output & Evaluation



Model Output

For the stock price, the study uses an adjusted close price which changed for splits, dividends, and capital gain distributions. The closed daily adjusted price is calculated to find the average price by quarter as

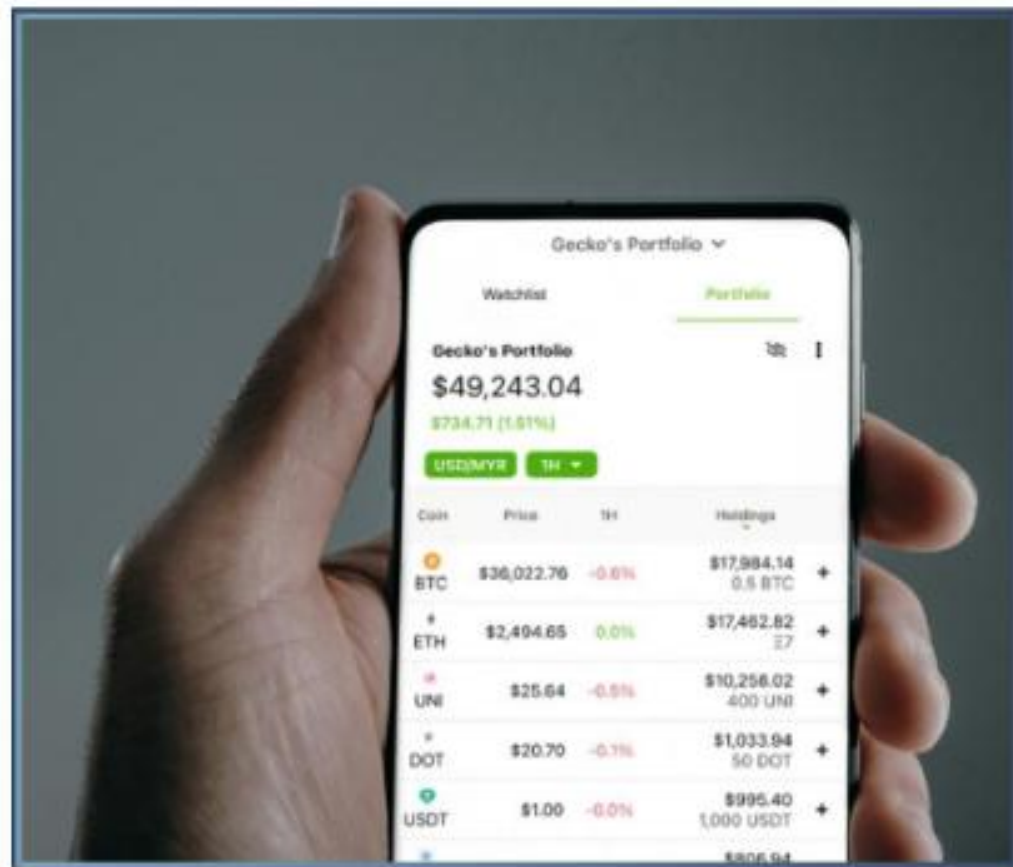
$$\text{Avg} = \frac{\sum_{i=1}^n x_i}{n}$$



Model Evaluation

evaluate the performance of the model using the MAE. This evaluation metric allows us to determine the percentage error of the model and provides valuable information on its accuracy and reliability.

$$\text{Output} = \frac{\text{Stock Price}_{t+1} - \text{Stock Price}_t}{\text{Stock Price}_t}$$





Experiment Result

01

Correlation of Factors

showing the relation between two variables: , the correlation between the various factors and the stock price in SET50 is depicted. the stock price is positively correlated with variables such as Equity, Net profit, Earning per Share (EPS), Market Capitalization, and SET Index.



02

The Prediction Performance Comparison Between Models

using three models: linear regression, Deep neural network (DNN), and LSTM. The result shows the LSTM model performed the best for predicting stock prices in SET50



03

The Prediction Performance by Industry

the LSTM model is considered to be the optimal choice for stock price prediction





Conclusion



Advantages

LSTM has the ability to capture long-term dependencies, handle noisy data, reduce human bias, and provide low-cost solutions for investors and institutions to increase profits and reduce losses.



Disadvantages

the performance of LSTM may vary among different industries, with some industries exhibiting higher prediction performance than others.

Thank You