ANALYTICAL STUDY OF MACHINE LEARNING MODELS FOR STOCK TRADING IN MALAYSIAN MARKET

HAZIRAH BINTI HALUL

UNIVERSITI SAINS ISLAM MALAYSIA

ANALYTICAL STUDY OF MACHINE LEARNING MODELS FOR STOCK TRADING IN MALAYSIAN MARKET

HAZIRAH BINTI HALUL
(Matric No : 3191167)

A thesis submitted in partial fulfilment for the degree of
MASTER OF SCENCE IN FINANCIAL MATHEMATICS

Faculty of Science and Technology
UNIVERSITI SAINS ISLAM MALAYSIA
Nilai, Negeri Sembilan

March 2021

BIODATA OF AUTHOR

Hazirah binti Halul (3191167) was born on 1st October 1996. She is currently residing at No 30, Jalan Mawar 10C, Taman Mawar 2, Bandar Baru Salak Tinggi, 43900 Sepang, Selangor Darul Ehsan. She was previously a student at Michigan State University and obtained Bachelor of Science in Astrophysics. Currently, she is a full time Master student of FST, USIM in Financial Mathematics. Before pursuing her master's degree, she worked at Michigan State University as Research Assistant in Physics and Astronomy Department and at Nuclear Agency Malaysia as Industrial Trainee in Material Analysis Department.

ACKNOWLEDGEMENT

First of all, I would like to thank Allah for giving me the chance and pleasure of doing this thesis in FST at USIM. I want to thank my supervisor, Dr. Karmila Hanim bt. Kamil for all the guidance and help during research. I would like to thank my coordinator, Dr. Siti Raihana bt. Hamzah for giving the opportunity to do the thesis. I want to also thank all the lecturers that have taught me in class for their insights and teaching. I would also like to thank my family and my friends for all the encouragement they gave me in doing the research and completing the thesis. I am happy that I was given this ability to share knowledge that I acquire in this thesis and that it would be beneficial to many others. Besides that, I get to learn other skills while studying in USIM. Without their help, I would probably not finish but crave for help to do my thesis.

Thank you.

ABSTRAK

Pada masa kini, *Machine Learning* (ML) boleh berfungsi sebagai salah satu cara penyelesaian untuk mempercepatkan proses membuat keputusan dalam peramalan pergerakan harga saham harian. Walaupun bagaimanapun, kerja-kerja penyelidikan dan analisis data yang menggunakan pelbagai model ML yang masih terhad dan tidak meluas telah membatasi para pelabur untuk menilai kecekapan dan kemampuan model-model ini. Kajian terdahulu biasanya memberi tumpuan kepada ramalan indeks saham atau pemilihan beberapa saham tertentu dengan ciri-ciri yang terhad. Oleh itu, sumbangan kajian ini lebih memberi tumpuan kepada penilaian model algoritma yang berbeza seperti model ML tradisional dan model *Deep Learning* dengan penggunaan data saham yang lebih besar dan parameter yang pelbagai dari syarikat-syarikat terpilih yang tersenarai di Bursa Malaysia. Tiga model ML tradisional yang digunakan ialah *Logistic Regression* (LR), *Support Vector Machine* (SVM), dan *Extreme Gradient Boosting* (XGB), manakala tiga lagi model *deep learning* yang terlibat ialah *Deep Belief Network* (DBN), *Multilayer Perception* (MLP), dan *Stacked Auto-Encoder* (SAE). Dengan menetapkan algoritma ML dan parameter-parameter tertentu berserta pengaplikasian kaedah *Walk-Forward Analysis* (WFA), reka bentuk algoritma bagi isyarat dagangan saham dapat dinilai berdasarkan dua kumpulan petunjuk penilaian, iaitu dari segi arah isyarat dan prestasi. Analisis perbandingan ke atas petunjuk penilaian bagi semua model algoritma dagangan saham telah dinilai dan dibincangkan. Bagi strategi perdagangan saham di Malaysia khususnya, dapatan kajian ini menunjukkan bahawa model *deep learning* menunjukkan prestasi yang lebih baik berbanding ML tradisional dari segi petunjuk penilaian arah isyarat di dalam ramalan siri masa. Namun begitu, model ML tradisional adalah lebih cekap berbanding model *deep learning* bagi petunjuk penilaian prestasi dari sudut keuntungan dan penilaian risiko.

Kata kunci: *Machine Learning*, *Walk-Forward Analysis*, ML tradisional, *deep learning*, ramalan siri masa

ABSTRACT

Nowadays, Machine Learning (ML) can serve as one of the solutions to accelerate the process of decision-making in forecasting daily stock market price movements. Nonetheless, inadequate number of research and lack of extensive data analysis using various ML models had limit the investors to appreciate the efficiency and capability of these models. Previous studies usually concentrate on the forecasting stock index or selecting a few stocks with restricted features. Therefore, this study focused to contribute on evaluating different algorithm models such as traditional ML and deep learning models with big stock data of multiple parameters from selected companies in Bursa Malaysia. The three traditional ML selected includes Logistic Regression (LR), Support Vector Machine (SVM), and Extreme Gradient Boosting (XGB), while another three deep learning models selected are Deep Belief Network (DBN), Multilayer Perception (MLP), and Stacked Auto-Encoder (SAE). By setting the ML algorithms and their parameter along with using Walk-Forward Analysis (WFA) method, the algorithm design of trading signal was evaluated based on two groups of evaluation indicators, namely directional and performance. Comparative analysis of evaluation indicators for all trading algorithms has been assessed and discussed. For stock trading in Malaysian stock market particularly, the experimental results of this study demonstrate that deep learning models have better performance in directional evaluation indicator compared to traditional ML in time series forecasting. However, traditional ML models are more efficient than deep learning in performance evaluation indicators in terms of profitability and risk assessment.

Keywords: Machine Learning, Walk-Forward Analysis, traditional ML, deep learning, time series forecasting

الملخص

في الوقت الحاضر ، يمكن أن يكون التعلم الآلي (ML) بمثابة أحد الحلول لتسريع عملية اتخاذ القرار في التنبؤ بحركات أسعار سوق الأسهم اليومية. ومع ذلك ، فإن العدد غير الكافي من الأبحاث والافتقار إلى تحليل البيانات الشامل باستخدام نماذج ML المختلفة قد حد من المستثمرين لتقدير كفاءة وقدرة هذه النماذج. تركز الدراسات السابقة عادةً على التنبؤ بمؤشر الأسهم أو اختيار عدد قليل من الأسهم ذات الميزات المقيدة. لذلك ، ركزت هذه الدراسة على المساهمة في تقييم نماذج الخوارزمية المختلفة مثل ML التقليدية ونماذج التعلم العميق مع بيانات الأسهم الكبيرة لمعايير متعددة من الشركات المختارة في بورصة ماليزيا. تشتمل نماذج ML التقليدية الثلاثة المحددة على الانحدار اللوجستي (LR)، وآلة المتجهات الداعمة (SVM)، وتعزيز التدرج الشديد (XGB)، في حين أن نماذج التعلم العميق الثلاثة الأخرى المحددة هي شبكة الإيمان العميق (DBN)، والإدراك متعدد الطبقات (MLP)، والمكدس التشفير التلقائي (SAE). من خلال تعيين خوارزميات ML والمعلمة الخاصة بها إلى جانب استخدام طريقة التحليل الفوري (WFA)، تم تقييم تصميم خوارزمية لإشارة التداول بناءً على مجموعتين من مؤشرات التقييم ، وهما الاتجاه والأداء. تم تقييم ومناقشة التحليل المقارن لمؤشرات التقييم لجميع خوارزميات التداول. بالنسبة لتداول الأسهم في سوق الأوراق المالية الماليزي على وجه الخصوص ، توضح النتائج التجريبية لهذه الدراسة أن نماذج التعلم العميق لها أداء أفضل في مؤشر التقييم الاتجاهي مقارنة بـ ML التقليدي في التنبؤ بالسلاسل الزمنية. ومع ذلك ، فإن نماذج ML التقليدية أكثر كفاءة من التعلم العميق في مؤشرات تقييم الأداء من حيث الربحية وتقييم المخاطر.

CONTENT PAGE

LIST OF TABLES

LIST OF FIGURES

ABBREVIATION

AE	Auto Encoder
ANN	Artificial Neural Network
AR	Accuracy Rate
AR*	Autoregression
ARR	Annualized Return Rate
ASR	Annualized Sharpe Ratio
B&H	Buy and Hold strategy
CART	Classification & Regression Tree
DBN	Deep Belief Network
DNN	Deep Neural Network
DT	Decision Tree
F1	F1 Score
FFNN	Feed-Forward Neural Network
FL	Functional Link
GARCH	Generalized Autoregressive Conditional Heteroskedasticity
GB	Gradient Boosting
GDA	Gradient Discriminant Analysis
GR	General Regression
GRU	Gated Recurrent unit
kNN	k-Nearest Neighbour
LR	Logistic Regression
LSTM	Long Short-Term Memory
MAE	Mean Absolute Error
MAPE	Mean Absolute Percentage Error
MDD	Maximum Drawdown
MLP	Multilayer Perception
MR	Mean Return
MSE	Mean Squared Error
NB	Naïve Bayes
NMSE	Normalized Mean Squared Error

PCA	Principal Component Analysis
PP	Profit Percentage
PR	Precision Rate
R	Return
RAE	Relative Absolute Error
RBF	Radial Basis Function
RBM	Restricted Boltzmann Machine
RF	Random Forest
RMSE	Root Mean Squared Error
RNN	Recurrent Neural Network
RR	Recall Rate
Rsqrt	R Squared
SAE	Stacked Auto-Encoder
SR	Sharpe Ratio
STD	Standard Deviation
SVM	Support Vector Machine
XGB	Extreme Gradient Boosting
WR	Winning Rate