SENTIMENT-BASED SUPPORT VECTOR MACHINE OPTIMIZED BY METAHEURISTIC ALGORITHMS FOR CRYPTOCURRENCY FORECASTING

BY

NOR AZIZAH BINTI HITAM

A thesis submitted in fulfilment of the requirement for the degree of Doctor of Philosophy in Computer Science

Kulliyyah of Information and Communication Technology
International Islamic University Malaysia

AUGUST 2022

ABSTRACT

Time series are used to model a variety of financial phenomena. The cryptocurrency forecasting problem is the focus of this thesis, which investigates time series forecasting challenges in finance. However, earlier research has neglected to consider the importance of sentiment and public opinion in today's market. The Commodity Channel Index (CCI), historical data and a machine learning algorithm are also employed in this study to improve the accuracy of time series forecasting. By employing hyperparameter optimization, this thesis intends to offer a novel sentiment-based support vector machine optimised by particle swarm and moth-flame optimization algorithms (SVMPSOMFO). PSO, GA, WOA, GOA, GWO, HS and MFO are compared against the proposed algorithm's performance for predicting cryptocurrency prices. A thorough investigation and discussion of all experimental results on different datasets are performed. From the findings, SVMPSOMFO outperforms other optimization methods in terms of accuracy rate when compared to a prediction model that excludes sentiment information. In addition, statistical tests are performed to validate the outcomes of the study.

خلاصة البحث

تُستخدم السلاسل الزمنية لنمذجة ظواهر مالية متنوعة. ومشكلة التنبؤ بأسعار العملات المشفرة هي محور هذه الأطروحة، التي تبحث في تحديات التنبؤ بالسلسلة الزمنية في الأمور المالية. و مع ذلك، فقد أهملت الأبحاث السابقة النظر في أهمية المشاعر والرأي العام في السوق اليوم. وقد تم استخدام مؤشر قناة السلع (CCI) والبيانات التاريخية وخوارزمية التعلم الآلي في هذه الدراسة لتحسين دقة تنبؤ السلاسل الزمنية. من خلال استخدام تحسين المعلمات الفائقة، تهدف هذه الأطروحة إلى تقديم آلة المتجه الداعم بصورة جديدة قائمة على المشاعر و محسّنة بواسطة خوارزميات تحسين عناصر السرب ولهب العثة (SVMPSOMFO). تم مقارنة PSO و GA و WOA و GOA و GWO و HS و MFO مع أداء الخوارزمية المقترحة للتنبؤ بأسعار العملات المشفرة. كما تم إجراء تحقيق و مناقشة شاملين لجميع النتائج التجريبية على مجموعات البيانات المختلفة. يظهر من النتائج، تفوق SVMPSOMFO على طرق التحسين الأخرى من حيث معدل الدقة عند مقارنته بنموذج التنبؤ الذي يستبعد معلومات المشاعر. بالإضافة إلى ذلك، تم إجراء الاختبارات الإحصائية للتحقق من صحة نتائج الدراسة.

APPROVAL PAGE

The thesis of Nor Azizah binti Hitam has been approved by the following:

_____
Amelia Ritahani binti Ismail
Supervisor

_____
Normaziah binti Abdul Aziz
Co-Supervisor

_____
Rizal Mohd Nor
Internal Examiner

_____
Nazlia Omar
External Examiner

_____
Mohammed Elwathig Saeed Mirghani
Chairman

DECLARATION

I hereby declare that this thesis is the result of my own investigations, except where otherwise stated. I also declare that it has not been previously or concurrently submitted as a whole for any other degrees at IIUM or other institutions.

Nor Azizah binti Hitam

Signature ..... Date.....

INTERNATIONAL ISLAMIC UNIVERSITY MALAYSIA

DECLARATION OF COPYRIGHT AND AFFIRMATION OF FAIR USE OF UNPUBLISHED RESEARCH

SENTIMENT-BASED SUPPORT VECTOR MACHINE OPTIMIZED BY METAHEURISTIC ALGORITHMS FOR CRYPTOCURRENCY FORECASTING

I declare that the copyright holders of this thesis are jointly owned by the student and IIUM.

By signing this form, I acknowledged that I have read and understand the IIUM Intellectual Property Right and Commercialization policy.

Affirmed by Nor Azizah binti Hitam

……..…….. ……..

Signature Date

INTERNATIONAL ISLAMIC UNIVERSITY MALAYSIA

DECLARATION OF COPYRIGHT AND AFFIRMATION OF FAIR USE OF UNPUBLISHED RESEARCH

SENTIMENT-BASED SUPPORT VECTOR MACHINE OPTIMIZED BY METAHEURISTIC ALGORITHMS FOR CRYPTOCURRENCY FORECASTING

I declare that the copyright holder of this thesis is International Islamic University Malaysia.

No part of this unpublished research may be reproduced, stored in a retrieval system, or transmitted, in any form or by any means, electronic, mechanical, photocopying, recording or otherwise without prior written permission of the copyright holder except as provided below

4. Any material contained in or derived from this unpublished research may be used by others in their writing with due acknowledgement.

5. IIUM or its library will have the right to make and transmit copies (print or electronic) for institutional and academic purposes.

6. The IIUM library will have the right to make, store in a retrieved system and supply copies of this unpublished research if requested by other universities and research libraries.

By signing this form, I acknowledged that I have read and understand the IIUM Intellectual Property Right and Commercialization policy.

Affirmed by Nor Azizah binti Hitam

……..…….. ……..

Signature Date

INTERNATIONAL ISLAMIC UNIVERSITY MALAYSIA

DECLARATION OF COPYRIGHT AND AFFIRMATION OF FAIR USE OF UNPUBLISHED RESEARCH

SENTIMENT-BASED SUPPORT VECTOR MACHINE OPTIMIZED BY METAHEURISTIC ALGORITHMS FOR CRYPTOCURRENCY FORECASTING

I declare that the copyright holder of this thesis is Nor Azizah Binti Hitam

By signing this form, I acknowledged that I have read and understand the IIUM Intellectual Property Right and Commercialization policy.

Affirmed by Nor Azizah binti Hitam

……..…….. ……..

Signature Date

This thesis is dedicated to my parents for laying the foundation of what I turned out to be in life.

ACKNOWLEDGEMENTS

All praises to Allah for the strengths and His blessings for the completion of this research and throughout the duration of my research.

I submit my heartiest gratitude to my respected supervisor I, Assoc. Prof. Dr Amelia Ritahani Ismail and supervisor II, Assoc. Prof. Dr. Noormaziah binti Abdul Aziz, for giving me the opportunity to and providing me invaluable guidance throughout this research. Not only that, I appreciate their emotional support, sincerity, understanding, patience, encouragement and for pushing me farther than I thought I could go. I am extremely grateful for what they have offered me. May Allah rewards them abundantly

My deepest gratitude goes to my beloved husband, Syed Kairil Musairi Syed Abdul Mutalib Jamalullail for his patience and companionship, and to my five wonderful children: Syed Ariq Darwish, Syed Aisy Fareeq, Syed Adel Harraz, Syed Aqil Jahran and the one and only girl, Sharifah Arwa Luthfia; for their love, understanding, prayers, and endless support throughout the process of working on this research. My parents, Hitam bin Muhammad, and Halimah binti Abdul Ghani, for their love, prayers, and passionate encouragement, made it possible for me to complete this research. Thank you to my siblings and in-laws for being there thick and thin..

Above ground, I am deeply indebted to my friends for all the moral support and encouragement through this PhD journey. Thank you for everything. This piece of work I dedicate to my parents, siblings, relatives, and friends for their emotional support and for making me believe in myself to keep going and complete this thesis. Thank you for reminding me the importance of working hard and working sincerely.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ALGORITHMS

LIST OF ABBREVIATIONS

ALO	Ant Lion Optimization
FFA	Firefly Algorithm
Forex	Foreign Exchange
GA	Genetic Algorithm
GOA	Grasshopper Optimization Algorithm
GWO	Grey Wolf Optimization
HS	Harmony Search
MFO	Moth Flame Optimization
NLP	Natural Language Processing
PSO	Particle Swarm Optimization
SVM	Support Vector Machine
SVMALO	Support Vector Machine with Ant Lion Optimization
SVMFFA	Support Vector Machine with Firefly Algorithm
SVMGA	Support Vector Machine with Genetic Algorithm
SVMGOA	Support Vector Machine with Grasshopper Optimization Algorithm
SVMGWO	Support Vector Machine with Grey Wolf Optimization
SVMHS	Support Vector Machine with Harmony Search
SVMMFO	Support Vector Machine with Moth Flame Optimization
SVMPSO	Support Vector Machine with Particle Swarm Optimization
SVMPSOMFO	Support Vector Machine with Particle Swarm Optimization and Moth Flame Optimization
SVMWOA	Support Vector Machine with Whale Optimization Algorithm
WOA	Whale Optimization Algorithm

CHAPTER ONE

INTRODUCTION

1.1 OVERVIEW

Forecasting plays a vital role in decision making process and planning in any organizations. It can be described as a prediction of a future events based on the historical data. This includes time series forecasting. As economic trends and market trends are inseparable, economists considered bond and stock markets anticipate economic trends. Market analysts look at the market trends more than the economic trends. Fundamental analysts must be specialized on fewer market sectors and need to understand the nature of industries and thus demand more work and concentration including the sentiment of investors. Investors sentiment do influence by the textual analysis that are available on most of the social media such as mood or sentiment information about stock market price (Bollen et al., 2011; Brendan O'Connor et al., 2010a), public opinions and sales forecast (Brendan O'Connor et al., 2010b).

Profit of all investors and traders are depending on the level of predictability, predicting future market will help them in making corrective measures in their forecasting. A huge growth in computing power have revolutionized virtually every angle of modern live, and financial market are no exception. Financial market is known as one of the major contributors to the capital budget (Baker & Wurgler, 2013) and many researchers show that its development give direct impact to the financial growth (Ake & Ognaligui, 2010). Various models have been introduced by previous researchers in conducting time series forecasting including the development of the Artificial Intelligence (AI) models that leads to an efficient stock market and give more confidence to investors to allocate more capital, finance and accommodate the products exchange (Ake & Ognaligui, 2010). This research aims to design novel algorithms for cryptocurrency future price forecasting that is capable to deal with

huge datasets. To achieve this aim, the proposed algorithm is inspired from sentiment-based Support Vector Machines (SVM) and optimized by particle swarm and moth flame as a methodology in developing the model. This thesis review most of the machine learning algorithms that are broadly used by the researchers in forecasting. There are many types of machine learning algorithms including artificial neural networks, support vector machines, deep learning and other approaches. This chapter will explain about some motivation of this thesis with a brief explanation on technical and fundamental analysis, financial technology along with the background of the study. Following this, the structure and content of the thesis is developed, stating the objective, research questions, significance and scope including limitation while working towards the completion.

1.2 PROBLEM BACKGROUND

Internet users may use social media to pursue common interests and to connect with others, sharing opinions, allow communication as well as finding the existing acquaintances (Nisar & Yeung, 2018; Syed Zulkarnain & Hitam, 2014; Wikipedia Contributors, 2017). To date, many studies being done in cryptocurrency market forecasting due to its popularity (Indera et al., 2017; Lee Kuo Chuen, 2015; Peng et al., 2018; Roche & Mcnally, 2016; Uzeki, 2018). A study by Parikh et al. (2021) discovers that the sentiment does play an important role in the cryptocurrency market as the high engagement of the investors to the social media do influence the transaction of the cryptocurrency market during the particular period.

However, apart from the engagement of the investors to the social media, there are other factors that can be considered in forecasting the cryptocurrency prices. First is the development of an efficient model (Kumar Dash et al., 2021). The development of an efficient model to forecast future cryptocurrency prices is still difficult due to its complexity in development and operations (Indera et al., 2017; Kara et al., 2011; Kumar Dash et al., 2021, Roche & Mcnally, 2016). The issue with the development of the model are related to the hyperparameter

optimization and the feature selection as mentioned by (Azmira et al., 2017; Buslim et al., 2021; Huerta et al., 2013; Kara et al., 2011). Huerta et al., (2013), has designed a model and simulate series of experiments to test the proposed model on the selected classifiers. In accordance with the proposed model, smallest observation set of data contribute to better performance as the trainings in SVM relies on the available data, but it draws an attention on the number of trainings needed by the classifiers to is left unknowns. This is also explored by Kara et al. (2011) where the proposed model is to forecast a movement of stock price by applying machine learning algorithms which are ANN and SVM in Istanbul Stock Exchange. The model has outperformed other models; however, the shortcomings are the model disregards the hyperparameter optimization and feature selection. This is followed by another research in financial forecasting presents that even though SVM does reduce the local minima issues and over-fitting, high complexity in its programming that pays more attention to inequality compared to the equality constraints (Azmira et al., 2017) becomes another deficiency to the SVM. (J. Wang, 2017) finds that the constraints of SVM in terms of parameter optimization and feature selection must be examined while developing a forecasting model, as revealed in the research of (Buslim et al., 2021). A study of Uras and Ortu (2021) discovers that additional of technical indicators to the classic microeconomic variables leads to an effective improvement in the prediction of cryptocurrency performances. However, no one knows which indicators can perform best when implemented in the machine learning algorithm.

1.3 PROBLEM STATEMENT

The Support Vector Machine (SVM) demonstrates an increase in forecasting the future price for it has the potential to lessen the local minima problem and over-fitting issue, nevertheless, it disregards the hyperparameter optimization and feature selection (Huerta et al., 2013), demand high complexity in its programming which demands more attention to inequality compared to the equality constraints (Azmira et al., 2017). A recent study of (Uras & Ortu, 2021) proved that the SVM is still a relevant machine learning algorithm for