UNIVERSITY OF TECHNOLOGY SYDNEY
Faculty of Engineering and Information Technology

MACHINE LEARNING ALGORITHMS FOR WEALTH DATA ANALYTICS

by

Ngoc Yen Nhi Vo

A Thesis Submitted
in Partial Fulfillment of the
Requirements for the Degree

Doctor of Philosophy

Sydney, Australia

2020

Certificate of Original Authorship

I, Ngoc Yen Nhi Vo, declare that this thesis, submitted in partial fulfilment of the requirements for the award of Doctor of Philosophy, in the School of Computer Science, Faculty of Engineering and IT at the University of Technology Sydney, Australia, is wholly my own work unless otherwise referenced or acknowledged. In addition, I certify that all information sources and literature used are indicated in the thesis. This document has not been submitted for qualifications at any other academic institution. This research is supported by the Australian Government Research Training Program.

Production Note:
Signature removed
prior to publication.

Ngoc Yen Nhi Vo

Sydney, Australia, 2020.

Acknowledgements

First of all, I express my sincere gratitude for my supervisor, Professor Guandong Xu, who had been a great mentor guiding me through my Ph.D. candidature stages. He had empowered and inspired me to work on multiple research projects which led to publications in top journals and conferences. I am grateful for the support of my co-supervisor, Shaowu Liu, who had been helpful with various technical difficulties. I am also greatly appreciate the contributions of all co-authours to my research publications that partially constructed this thesis: Professor Guandong Xu, Shaowu Liu Ph.D., Professor Xuezhong He, Professor Xitong Li, James Brownlow, Charles Chu, and Ben Culbert.

I thank Advanced Analytics Institute, School of Computer Science, Faculty of Engineering and IT, University of Technology Sydney for providing the infrastructures and computing power used in conducting empirical works. I am deeply grateful for the financial support received during my candidature time: UTS International Research Scholarship, FEIT Industry Scholarship, and Vice Chancellor Conference Fund scholarship. Part of my research was funded by Australian Research Council Linkage Project Scheme under LP170100891 and LP140100937 grants.

Finally, I am grateful for my dear family, especially my father Ngoc Bien Vo, my mother Thi Bao Di Phan, and my brother Ngoc Bao Vo, who had been supportive during my Ph.D. candidature stages, researching and writing my thesis. I am sincerely grateful for everything and especially the opportunities given in this life, from being born to the completion of this Doctor of Philosophy thesis.

List of Publications

Journal Papers

J-1. Vo, Nhi NY, Xuezhong He, Shaowu Liu, and Guandong Xu. "Deep Learning for Decision Making and the Optimization of Socially Responsible Investments and Portfolio." Decision Support Systems (2019): 113097.

J-2. Vo, Nhi NY, Xitong Li, Shaowu Liu, and Guandong Xu. "Leveraging Unstructured Call Log Data for Customer Churn Prediction." Knowledge-Based Systems (Under review).

Conference Papers

C-1. Vo, Nhi NY, Shaowu Liu, Xuezhong He, and Guandong Xu. "Multimodal mixture density boosting network for personality mining." In Proceedings of Pacific-Asia Conference on Knowledge Discovery and Data Mining, pp. 644-655. Springer, Cham, 2018. (PAKDD 2018 conference took place in Melbourne, Australia from 3^{rd} to 6^{th} June, 2018)

C-2. Vo, Nhi NY, Shaowu Liu, James Brownlow, Charles Chu, Ben Culbert, and Guandong Xu. "Client Churn Prediction with Call Log Analysis." In Proceedings of International Conference on Database Systems for Advanced Applications, pp. 752-763. Springer, Cham, 2018. (DASFAA 2018 conference took place in Queensland, Australia from 21^{st} to 24^{th} May, 2018)

C-3. Vo, Nhi NY, and Guandong Xu. "The volatility of Bitcoin returns and its correlation to financial markets." In Proceedings of International Conference on Behavioral, Economic, Socio-cultural Computing (BESC), pp. 1-6. IEEE, 2017. (BESC 2017 conference took place in Cracow, Poland from 16^{th} to 18^{th} October, 2017)

Contents

List of Figures

List of Tables

ABSTRACT

The thesis investigates multiple machine learning algorithms with big data approach and applies cutting-edge deep analytics to tackle the challenges in financial wealth management. In general, the existing research on wealth data analytics is limited with two main challenges. Firstly, the amount of quantitative research conducted is scarce and scattered across different approaches. Partially this is due to the lack of access to the data required for the research to use a quantitative approach. Secondly, the results are rudimentary and limited to a certain aspect of wealth data analytics. This lack of integration in existing research findings is a by-product of the simplistic approaches employed in lieu of big data analytics and deep learning techniques.

This research provides a broader and comprehensive approach for quantitative research within the wealth management field from both financial and customer aspects. Particularly, this research utilizes the big data of structured demographic, behavioral, communicational data, and unstructured textual information from wealth customers, plus additional financial market and corporate responsibility data from companies. This thesis exploits deep analytics techniques to provide a better framework for decision-making support based on the constructed mathematical and computational models, combined with customer segmentation modeling and quantitative finance approach.

From the customer aspect, the thesis applies big data analytics, text mining and interpretable machine learning in customer data analytics in wealth management. The proposed approaches and models are (1) MMDB for personality mining, (2) transfer learning for customer personality prediction, (3) ensemble model with text mining for churn prediction, (4) interpretable machine learning with SHAP-MRMR+ to extract customer insight, and (5) customer segmentation and managerial implications with personality and SOM.

From the financial aspect, this is one of the first research to utilize deep learning

for socially responsible investment. The proposed framework consists of (1) text mining of CSR reports for ESG ratings, (2) ESG-based quantitative models, (3) deep learning using Multivariate BiLSTM for stock return prediction, (4) MV-ESG for ESG-based portfolio optimization, and (5) reinforcement learning for socially responsible investment.

The empirical results show the advantages and effectiveness of deep learning algorithms and big data analytics in financial wealth data analytics. Through the completion of this thesis, various aspects of wealth data analytics have been researched and integrated into sophisticated frameworks, and the information systems can provide meaningful insights for multiple stakeholders, from researchers to individual investors and fund managers.

Abbreviation

2-D	Two dimensional
AGR	Agreeableness
ARMA	AutoRegressive Moving Average
AUC	Area Under the Curve
BG	Growth of Book Value per Share
Big Five	Big Five Personalities
BiLSTM	Bidirectional Long Short Term Memory network
Borderline-SMOTE	SMOTE and Borderline samples
CL	Call Logs data set
CON	Conscientiousness
CP	Customer Profiles
CRM	Customer Relationship Management
CSR	Corporate Social Responsibility
CSR-Sent	Corporate Social Responsibility Sentiment Dictionary
DCA	Discriminant Correlation Analysis
Doc2Vec	Document to Vector
DRIP	Deep Responsible Investment Portfolio
DT	Decision Tree
DY	Growth of Dividend Yield
EBITDA	Earning Before Interest, Tax, Depreciation and Amortization
EM	Expectation Maximization
ESG	Environmental, Social and Governance
EXT	Extraversion
FI	First Impression Dataset
GARCH	Generalized AutoRegressive Conditional Heteroskedasticity
GHG	Green House Gas

GMO	Genetically Modified Organism
GP	Gaussian Process
GRI	Global Reporting Initiative
GRU	Gated Recurrent Unit network
ID	Identification
IT	Information Technology
K	thousand (quantity)
LAD	Least Absolute Deviation
LI	Lexical Information
LIWC	Linguistic Inquiry and Word Count
LR	Logistic Regression
LSTM	Long Short Term Memory network
MA	Mean Accuracy
MAE	Mean Absolute Error
MAX-ESG	Maximum ESG portfolio
MAX-S	Maximum Sharpe portfolio
MAZ	Mean Absolute Z-Score
MDN	Mixture Density Network
MMD	Multimodal Mixture Density Network
MMDB	Multimodal Mixture Density BoostingNetwork
mRMR	Minimum Redundancy Maximum Relevance
MSE	Mean Squared Error
MV	Mean Variance model
MV-ESG	Mean Variance ESG model
MV-SGP	Mean Variance Stochasitc Goal Programming model
NB	Naïve Bayes
NEU	Neuroticism
NLP	Natural Language Processing
NN	Neural Network
NSGA-II	Non-Dominated Sorting Genetic Algorithm 2
OCSVM	One Class Support Vector Machine

OPN	Openness
P/B	Price per Book
P/ESG	Price per ESG
PE	Phrase Embedding
PT	Personality Traits
RF	Random Forest
RL	Reinforcement Learning
RMSE	Root Mean Squared Error
RNN	Recurrent Neural Network
ROC	Receiver Operating Characteristic
SDGs	United Nation Sustainable Development Goals
SHAP	Shapley Addictive Explanations
SHAP-MRMR+	Combined SHAP and positive mRMR method
SLSQP	Sequential Least Square Programming
SMO	Sequential Minimal Optimization
SMOTE	Synthetic Minority Over-sampling Technique
SMOTETomek	SMOTE and Tomek links
SMOTTEENN	SMOTE and Edited Nearest Neighbours
SOM	Self-Organizing Maps
SRI	Socially Responsible Investment
SSS	Small Sample Size
Super	Superannuation
SVD	Single Value Decomposition
SVM	Support Vector Machine
SVM-SMOTE	SMOTE and SVM
SVR	Support Vector Regression
TF-IDF	Term Frequency - Inverse Document Frequency
TI	Term Importance
Word2Vec	Word to Vector
XGB	XgBoost Extreme Gradient Boosting method
YT	Youtube Dataset