Data Analytics and Its Applications

Data Analytics and Its Applications

Mohd Hilmi Hasan
Norshakirah Aziz
Emelia Akashah P Akhir
Izzatdin A. Aziz
Jafreezal Jaafar

UNIVERSITI TEKNOLOGI PETRONAS
Seri Iskandar ● 2023

Data Analytics and Its Applications

© Universiti Teknologi PETRONAS, 2020
Second Print, 2023

All rights reserved. No part of this publication may be reproduced or transmitted in any form or by any means, including photocopying and recording, without the written permission of the copyright holder, application for which should be addressed to Universiti Teknologi PETRONAS. Such written permission must also be obtained before any part of this publication is stored in a retrieval system of any nature.

The publisher, authors, contributors and endorsers of this publication each excludes liability for loss suffered by any person resulting in any way from the use of, or reliance on this publication.

Published by
UTP Press
Universiti Teknologi PETRONAS,
Owned by Institute of Technology PETRONAS Sdn. Bhd.
32610 Seri Iskandar, Perak Darul Ridzuan, Malaysia

Printed by
Reka Cetak Sdn. Bhd.
No. 14, Jalan Jemuju Empat 16/13D,
Seksyen 16, 40200 Shah Alam,
Selangor Darul Ehsan
Tel: +603-5510 4254 / +603-5510 4758
www.rekacetak.com

Perpustakaan Negara Malaysia Cataloguing-in-Publication Data

Mohd Hilmi Hasan
Data Analytics and Its Applications/Mohd Hilmi Hasan, Norshakirah Aziz, Emelia Akashah P Akhir, Izzatdin A. Aziz, Jafreezal Jaafar.
ISBN 978-967-2048-28-2
1. Data mining.
2. Big data.
I. Norshakirah Aziz. II. Emilia Akashah P Akhir
III. Izzatdin A. Aziz. IV. Jafreezal Jaafar.
V. Title.
006.312

CONTENTS

Preface vii
Introduction ix

1. **Hybrid Feature Selection Framework to Support Learning Analytics in Educational Data Mining** 1
Maryam Zaffar, Manzoor Ahmed Hashmani, Savita K. Sugathan

2. **Supervised Machine Learning: A Review of Training Algorithms for Financial Time-Series Forecasting** 23
Said Jadid Abdulkadir, Sameer Ahmad Khan

3. **Improved Latent Semantic Indexing-based Big Data Mining Methods for Customer Relationship Management** 47
Junzo Watada, Jianxiong Yang, Mohd Hilmi Hasan, Yukio Kodono

4. **Employment Analysis of Graduates Students Using K-Means** 73
Nurfarahani Mohd Sani, Shakirah Mohd Taib, Ahmad Izuddin Zainal Abidin, Syarifah Bahiyah Syed Mansoor

5. **A Comparison of Membership Function Types for Fuzzy-based Quality of Web Service Evaluation Model** 93
Mohd Hilmi Hasan, Izzatdin A. Aziz

6. Complaint Handling Algorithm Using Fuzzy Approach 115
Jafreezal Jaafar, Razulaimi Razali, Wan Isni Sofiah Wan Din

7. Data Analytics: Forecasting Asset Maintenance Failure using Holt Winter's Exponential Smoothing 129
Noor Hafila Syifaa Mohd Hatta, Norshakirah Aziz, Emelia Akashah P Akhir

8. Healthcare Assisted by Grid Computing and Fog Computing Using Healthcare IoTs: Practices and Challenges 137
Saddaf Rubab, Mohd Fadzil Hassan, Saurab Shukla

Authors' Biodata 155
Index 165

PREFACE

Data Analytics and Its Applications contains high quality papers from scientists in several disciplines of data analytics involving areas such as big data, analytics, algorithms and artificial intelligence.

The book promotes insights and understanding of the state of the art, and trends in big data analytics technology. Its contents cover original research as well as innovative applications that provide significant implications to the body of knowledge and real-world implementations. We hope that you enjoy this book and will take from it some inspirations for your future research.

Dr. Mohd Hilmi Hasan

Dr. Norshakirah Aziz

Dr. Emelia Akashah P Akhir

Dr. Izzatdin A Aziz

Dr. Jafreezal Jaafar

INTRODUCTION

Data analytics is an analysis process that determines the relationship between data to yield insights. Data analytics clusters and processes data and predicts future scenarios using statistical and mathematical analysis. Nowadays, data analytics has been a key solution for most organizations due to its benefits, namely cost reduction and faster yet better decision making. Data analytics also supports the exploration and innovation of new products and services through its ability to learn what customers want.

Data analytics is broken down into three basic types, descriptive, predictive and prescriptive analytics. Descriptive analytics describes the results of what has happened over a given time. Meanwhile, predictive analytics explains what is likely to occur in the future, and prescriptive analytics describes suggested actions for the predicted events.

This book provides a valuable window on data analytics and its applications. Chapter 1 entitled "Hybrid Feature Selection Framework to Support Learning Analytics in Educational Data Mining (EDM)", explores the hidden knowledge from educational settings. The knowledge extracted from the EDM technique can help the educational decision-making process, teaching process, and improvement in the academic performance of students. Feature selection techniques can enhance the performance of EDM tasks by selecting the most relevant features. These features can help in making strategies for improving the quality of education. The proposed hybrid feature selection framework is a step towards increasing the quality of educational data to guide decision making, and planning interventions to improve the academic performance of students. The proposed hybrid feature

selection framework shows better prediction accuracy as compared to the existing methods of feature selection.

Chapter 2 discusses challenges that need to be tackled when working with financial data as it poses different challenges due to its chaotic nature, which is volatile and non-linear, thus making the process of forecasting cumbersome. Various models have been proposed to tackle these challenges. A comparative analysis of different supervised machine learning training algorithms trained on recurrent neural networks is presented in this study. Ten different training algorithms have been used for comparative analysis. The experimental results show that the Bayesian regulation backpropagation algorithm on Bursa Malaysia KLCI data outperforms other training algorithms.

Chapter 3 entitled "Improved Latent Semantic Indexing-based Big Data Mining Methods for Customer Relationship Management", describes data analytics using the latent semantic indexing (LSI) method for data clustering analysis. It proves that LSI is useful in data clustering and matrix reduction process and cut a lot of computing time. The semantic and logical networks are not only compatible models, but also they have something to contribute to the others. The logic extends the expressive power of the semantic network, provides them with inference rules, procedural interpretation, and semantics. On the other hand, the semantic network draws attention to the advantages of using n-array relations. It provides a predicate logical inference system, with both a potentially useful path and an indexing abstract following strategy for guiding the mining for the data solution.

Chapter 4 entitled "Employment Analysis of Graduates Students Using K-Means" discusses the employment of fresh graduate students that needs to be given attention as it is one of the contributing factors to entire employment status in Malaysia. It is very important to analyse the employment trend to ensure the future generation is getting adequate job opportunities and industries are offering enough executive positions too. Hence, the analysis findings would help universities to improve and prepare its students to get a job placement. Thus, this study focuses on developing a career analytic by the industry for fresh graduate students using K-means. R tool is used for the data analysis process. The employment data set are taken from Universiti Teknologi PETRONAS (UTP) alumni database and

Malaysia government portal. The result of this research shows the relationship and pattern of employment trends by industry.

Chapter 5 comprises works on fuzzy logic as an artificial intelligence (AI) algorithm used in data analytics. The chapter "A Comparison of Membership Function Types for Fuzzy-based Quality of Web Service Evaluation Model" discusses the different implementation of fuzzy membership function in analysing the quality of web service.

Chapter 6 entitled "Complaint Handling Algorithm Using Fuzzy Approach" presents the application of fuzzy logic algorithm in text-based data analytics for the complaint handling system.

Chapter 7 entitled "Data Analytics: Forecasting Asset Maintenance Failure using Holt Winter's Exponential Smoothing" presents works on forecasting monthly machine downtime and studies the impact of data analytics in the automotive industry. A proposed asset maintenance management dashboard has been designed and developed by using Holt Winter's Exponential Smoothing. By having a comprehensive dashboard, the company can easily monitor its assets and machines in terms of performance, rejection ratio and downtime hours. Basically, it will help the company to reduce the loss caused by asset maintenance failure.

The final chapter of this book entitled "Healthcare Assisted by Grid Computing and IoT: Practices and Challenges" presents data analytics challenges regarding grid computation and internet of things (IoT). Data analytics should work together with IoT for the following reasons; 1) Data volume – IoT involves and produces huge amount of data that can be used for insights or decision making. 2) Market trend – IoT has become a buzzword in today's business market. The coupling between IoT and data analytics is a good investment by organizations as it offers a competitive advantage to them in competing in their market. This chapter discusses the implementation of grid computing and IoT, with support from data analytics, to provide new platforms for the healthcare industry. These platforms introduced smart devices to connect the patients and people with the health solution offering medical solutions required within less time as compared to the traditional ways of hospital visits and constant observation.

Overall, the data analytics techniques and algorithms (e.g. AI and machine learning) are interesting to learn, while the issues in data analytics are difficult and challenging. People are conducting works and research with enthusiasm, tenacity, and dedication to develop new data analytics techniques to provide new solutions to keep up with the ever-changing challenges and issues.