

KAEDAH PENYATUAN FITUR DAN PEMBELAJARAN MESIN BERGABUNG UNTUK MODEL KLASIFIKASI KEMISKINAN PELBAGAI DIMENSI

RUSNITA HAMDAN@ISNIN
AZURALIZA ABU BAKAR
NOR SAMSI AH SANI

PENERBIT UNIVERSITI KEBANGSAAN MALAYSIA
BANGI • 2024
www.ukm.my/penerbit

Cetakan Pertama / *First Printing*, 2024
Hak Cipta / *Copyright* Universiti Kebangsaan Malaysia, 2024

Hak cipta terpelihara. Tiada bahagian daripada terbitan ini boleh diterbitkan semula, disimpan untuk pengeluaran atau ditukarkan ke dalam sebarang bentuk atau dengan sebarang alat juga pun, sama ada dengan cara elektronik, gambar serta rakaman dan sebagainya tanpa kebenaran bertulis daripada Penerbit UKM terlebih dahulu.

All rights reserved. No part of this publication may be reproduced or transmitted in any form or by any means, electronic or mechanical including photocopy, recording, or any information storage and retrieval system, without permission in writing from Penerbit UKM.

Diterbitkan di Malaysia oleh / *Published in Malaysia by*
PENERBIT UNIVERSITI KEBANGSAAN MALAYSIA
43600 UKM Bangi, Selangor Darul Ehsan, MALAYSIA
www.ukm.my/penerbit
e-mel: penerbit@ukm.edu.my

Penerbit UKM adalah anggota / *is a member of the*
MAJLIS PENERBITAN ILMIAH MALAYSIA /
MALAYSIAN SCHOLARLY PUBLISHING COUNCIL
PERSATUAN PENERBIT BUKU MALAYSIA /
MALAYSIAN BOOK PUBLISHERS ASSOCIATION
No. Ahli / *Membership No.* 198302

Atur huruf oleh / *Typeset by*
PENERBIT UNIVERSITI KEBANGSAAN MALAYSIA
43600 UKM Bangi, Selangor Darul Ehsan, MALAYSIA

Dicetak di Malaysia oleh / *Printed in Malaysia by*
MASSA CREATIVE ENTERPRISE
Back side of GF-03A, Menara Paragon, Persiaran Bestari,
Cyber 11, 63000 Cyberjaya, Selangor Darul Ehsan, MALAYSIA



Data Pengkatalogan-dalam-Penerbitan
Perpustakaan Negara Malaysia
Rekod katalog untuk buku ini boleh didapati
dari Perpustakaan Negara Malaysia
ISBN 978-629-486-114-5

Kandungan

Senarai Jadual & Rajah ... 7

Senarai Singkatan ... 11

Prakata ... 13

BAB 1 Model Pengukuran Kemiskinan Pelbagai Dimensi ... 15

BAB 2 Kaedah Pembelajaran Mesin Bergabung dalam Pengelasan Kemiskinan ... 34

BAB 3 Data Sara Hidup dan Kemiskinan Malaysia ... 69

BAB 4 Kaedah Hutan Rawak dalam Penentuan Fitur Kemiskinan ... 90

BAB 5 Kaedah Penyatuan Fitur dari Pelbagai Sumber Data ... 103

BAB 6 Prestasi Model Klasifikasi Kemiskinan Pelbagai Dimensi ... 116

Rujukan ... 135

Indeks ... 145

Prakata

Dengan Nama Allah yang Maha Pemurah lagi Maha Penyayang.

Syukur Alhamdulillah kepada Allah SWT kerana dengan izin-Nya buku yang bertajuk *Kaedah Penyatuan Fitur dan Pembelajaran Bergabung untuk Model Klasifikasi Kemiskinan Pelbagai Dimensi* dapat disempurnakan. Pada tahun 2010, Malaysia telah memperkenalkan Indeks Kemiskinan Pelbagai Dimensi (Multidimensional Poverty Index, MPI) sebagai usaha untuk menggambarkan keadaan ketersisihan dalam pelbagai aspek kehidupan. Indeks MPI pada tahun 2019 telah dilaporkan oleh Jabatan Perangkaan Malaysia pada bacaan 1.1 peratus telah mendapat reaksi kurang baik daripada the *World Bank Group* dan *United Nations Special Rapporteur* berikutan daripada peratus yang terlalu rendah untuk sebuah negara yang mempunyai kadar pembangunan yang setara dengan Malaysia dan berpunca daripada reka bentuk pensampelan survei pendapatan isi rumah dan kemudahan asas (HEIS) yang digunakan sebagai asas pengukuran kemiskinan tidak mengambil kira pensampelan menyeluruh populasi rakyat Malaysia. Bagi mengatasi masalah reka bentuk pensampelan daripada satu sumber set data ini, buku ini telah menggunakan tiga sumber set data menggambarkan profil taraf hidup rakyat Malaysia, iaitu set data Banci Penduduk dan Perumahan Malaysia 2010, HEIS 2016 dan Bank Data Kemiskinan 2017 (eKasih). Set data yang heterogen dijana pada tahun berbeza tidak mempunyai kunci sepunya menyebabkan integrasi data tidak boleh dilaksanakan.

Buku ini mengupas kaedah mengenal pasti fitur penting dari sumber data berbeza dengan menggunakan kaedah hutan rawak dan membincangkan pendekatan penyatuan fitur dari pelbagai sumber untuk menghasilkan set data gabungan dan fitur penting untuk mengklasifikasi kemiskinan pelbagai dimensi. Buku ini membincangkan kaedah penyatuan fitur secara pembelajaran bergabung berasaskan kaedah pohon, iaitu pohon keputusan, hutan rawak dan XGBoost yang memfokuskan kepada pembinaan formula teknik gabungan fitur penting berdasarkan skor kedudukan. Selanjutnya, bagi memastikan janaan sampel data yang relevan dalam menghasilkan

fitur dan dimensi penting bagi mengklasifikasi kemiskinan dari pelbagai aspek ketersisihan, algoritma linear regresi digunakan.

Buku ini merupakan hasil penyelidikan pascasiswazah yang bertajuk *Pendekatan Penyatuan Fitur dalam Pembelajaran Bergabung untuk Model Klasifikasi Kemiskinan Pelbagai Dimensi*. Ucapan terima kasih juga ditujukan kepada Universiti Kebangsaan Malaysia (UKM) bagi penyelidikan ini melalui geran Cabaran Perdana CP-2017-015 melalui projek *LAB40: Enhancing the B40 Wellbeing Through Living Analytics*. Penghargaan buat Jabatan Perkhidmatan Awam (JPA) yang telah menaja pengajian ini serta Unit Penyelarasan Pelaksanaan, Jabatan Perdana Menteri dan Bank Data UKM yang telah bekerjasama berkongsi data dan maklumat bagi menghasilkan penyelidikan yang realistik. Penghargaan juga kepada Jabatan Statistik Negara untuk perolehan data asal melalui Bank Data UKM.

Rusnita Isnin@Hamdan
Azuraliza Abu Bakar
Nor Samsiah Sani