PROJECT REPORT

AN EVALUATION OF HEART DISEASE PREDICTION USING AN EXTREME GRADIENT BOOSTING ALGORITHM

DAVIN CHANG
19.K1.0005

Faculty of Computer Science
Soegijapranata Catholic University
2022

HALAMAN PENGESAHAN

Judul Tugas Akhir:	:	An Evaluation Of Heart Disease Prediction Using An Extreme Gradient Boosting Algorithm
Diajukan oleh	:	Davin Chang
NIM	:	19.K1.0005
Tanggal disetujui	:	21 Desember 2022
Telah setujui oleh		
Pembimbing	:	Y.b. Dwi Setianto S.T., M.Cs.
Penguji 1	:	Yonathan Purbo Santosa S.Kom., M.Sc
Penguji 2	:	Hironimus Leong S.Kom., M.Kom.
Penguji 3	:	R. Setiawan Aji Nugroho S.T., MCompIT., Ph.D
Penguji 4	:	Rosita Herawati S.T., M.I.T.
Penguji 5	:	Y.b. Dwi Setianto S.T., M.Cs.
Penguji 6	:	Yulianto Tejo Putranto S.T., M.T.
Ketua Program Studi	:	Rosita Herawati S.T., M.I.T.
Dekan	:	Dr. Bernardinus Harnadi S.T., M.T.

Halaman ini merupakan halaman yang sah dan dapat diverifikasi melalui alamat di bawah ini.

sintak.unika.ac.id/skripsi/verifikasi/?id=19.K1.0005

DECLARATION OF AUTHORSHIP

I, the undersigned:

Name : Davin Chang

ID : 19.K1.0005

declare that this work, titled "An Evaluation Of Heart Disease Prediction Using An Extreme Gradient Boosting Algorithm", and the work presented in it is my own. I confirm that:

1. This work was done wholly or mainly while in candidature for a research degree at Soegijapranata Catholic University
2. Where any part of this thesis has previously been submitted for a degree or any other qualification at this University or any other institution, this has been clearly stated.
3. Where I have consulted the published work of others, this is always clearly attributed.
4. Where I have quoted from the work of others, the source is always given.
5. Except for such quotations, this work is entirely my own work.
6. I have acknowledged all main sources of help.
7. Where the work is based on work done by myself jointly with others, I have made clear exactly what was done by others and what I have contributed myself.

Semarang, December, 29, 2022

DAVIN CHANG

19.K1.0005

HALAMAN PERNYATAAN PUBLIKASI KARYA ILMIAH UNTUK KEPENTINGAN AKADEMIS

Yang bertanda tangan dibawah ini:

Nama	: Davin Chang
Program Studi	: Teknik Informatika
Fakultas	: Ilmu Komputer
Jenis Karya	: Skripsi

Menyetujui untuk memberikan kepada Universitas Katolik Soegijapranata Semarang Hak Bebas Royalti Nonekslusif atas karya ilmiah yang berjudul "An Evaluation Of Heart Disease Prediction Using An Extreme Gradient Boosting Algorithm". Dengan Hak Bebas Royalti Nonekslusif ini Universitas Katolik Soegijapranata berhak menyimpan, mengalihkan media/formatkan, mengelola dalam bentuk pangkalan data (database), merawat, dan mempublikasikan tugas akhir ini selama tetap mencantumkan nama saya sebagai penulis / pencipta dan sebagai pemilik Hak Cipta.

Demikian pernyataan ini saya buat dengan sebenarnya.

Semarang, 29 Desember 2022

Yang menyatakan

DAVIN CHANG

19.K1.0005

ACKNOWLEDGMENT

I have received a myriad of support, advice, and assistance throughout this document writing. I would like to thank my supervisors Y.b. Dwi Setianto S.T., M.Cs. for formulating this topic. I would also like to thank my friend Samuel Kurniawan Santoso, Jevon Carla, Andre Nugroho Pranoto, and Tan, Yudistira Suryakencana Adisatria for guiding with advice to finish this document.

I would like to thank my family and friends for giving me ceaseless love, support, and advices throughout my study in Soegijapranata Catholic University. You gave me great escape to rest my mind from my thesis.

ABSTRACT

Heart disease has recorded the most death cause in the world. A lot of researchers are trying to find better and more reliable machine learning to diagnose heart disease. Accuracy and the speed of computation become the main concern when classifying heart disease at its early stages related to human life.

This paper researched about Extreme Gradient Boosting (XGBoost) as an ensemble learning with boosting method to predict heart disease. The data will be preprocessed using handling missing value and removing outliers. The algorithm will be compared with 2 different datasets (with feature selection and without feature selection).

The outcome of this research hopefully can present the performance result of the Extreme Gradient Boosting algorithm using tenfold cross-validation and performance measures (Precision, Recall, F1-score, ROC Area, and Accuracy) when using feature selection and without using feature selection.

Keywords: heart disease, machine learning, ensemble learning, xgboost, feature selection

TABLE OF CONTENTS

LIST OF FIGURE

LIST OF TABLE