PROJECT REPORT

COMPARISON OF EXTREME GRADIENT BOOSTING ALGORITHM AND ARTIFICIAL NEURAL NETWORK ON DIABETES PREDICTION

JEVON CARLA
19.K1.0017

Faculty of Computer Science
Soegijapranata Catholic University
2023

HALAMAN PENGESAHAN

Judul Tugas Akhir:	: Comparison of Extreme Gradient Boosting Algorithm and Artificial Neural Network on Diabetes Prediction.
Diajukan oleh	: JEVON CARLA
NIM	: 19.K1.0017
Tanggal disetujui	: 13 Januari 2023
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.K1.0017

DECLARATION OF AUTHORSHIP

I, the undersigned:

Name : JEVON CARLA

ID : 19.K1.0017

declare that this work, titled " Comparison of Extreme Gradient Boosting Algorithm and Artificial Neural Network on Diabetes Prediction", 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, January, 17, 2023

JEVON CARLA

19.K1.0017

HALAMAN PERNYATAAN PUBLIKASI KARYA ILMIAH UNTUK KEPENTINGAN AKADEMIS

Yang bertanda tangan dibawah ini:

Nama : Jevon Carla

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 "Comparison of Extreme Gradient Boosting Algorithm and Artificial Neural Network on Diabetes Prediction". 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, 17 Januari 2023

Yang menyatakan

JEVON CARLA

19.K1.0017

ACKNOWLEDGMENT

Silahkan tuliskan anda ingin mengucapkan terima kasih atau ucapan persembahan ke siapapun yang anda rasa perlu ditulis disini

I have received a myriad of support, advice, and assistance throughout this document writing. I would like to thank my supervisors Mr. Y.b. Dwi Setianto S.T., M.Cs for formulating this topic. I would also like to thank my friend Davin Chang, 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 (ABSTRACT TITLE)

Diabetes is one of the serious diseases and it causes the sufferer to have high blood sugar due to the body unable to produce the required amount of insulin to regulate glucose. It may cause complications or may increase the risk of developing another disease like heart disease, kidney disease, blindness, etc. One of the best ways to fight this disease is by early diagnosis. If there are a lot of patient records, the machine learning classification algorithms play a great role in predicting whether a person has diabetes or not. The used dataset is Diabetes UCI Dataset from kaggle which has been collected using direct questionnaires from the patients of Sylhet Diabetes Hospital in Sylhet, Bangladesh, and approved by a doctor. The dataset has 520 data and 17 attributes. Several studies have been made in the last few decades and some of them show that Artificial Neural Networks (ANN) are one of the best algorithms for diabetes predictions, Extreme Gradient Boosting (XGBoost) is one of the popular machine learning algorithms used for classification, because of that reason the writer wants to find out whether XGBoost can be used on diabetes prediction and compare it with ANN. Both algorithms models were trained with the same ratio 80:20, 75:25, 70:30. 60:40, and 50:50. There are four models for the ANN with 3 hidden layers, 4 hidden layers, 5 hidden layers, and 6 hidden layers, as for the XGBoost models there are the first model with default parameters and the second one with the hyperparameters tuning. The accuracy, precision, recall, and f1 score of the models will be compared to find out which one has better performance. XGBoost performance able to achieve better performance but the third ANN models able to achieve highest score on 80:20, with 75:25 XGBoost with hyperparameters tuning able to achieve highest score, but XGBoost with default parameters have the same score as the the third ANN model, with 70:30 ratio, the third ANN model and both XGBoost models have the same score and have the highest score among all ratio. with 60:40 ratio, the first to third ANN models and XGBoost with default parameters have the same accuracy score but the third ANN models have the highest recall but lower precision than the XGBoost models. And with 50:50 XGBoost 2 has the best overall performances than the other models.

Keyword: diabetes, ann, xgboost, prediction, comparison

TABLE OF CONTENTS

LIST OF FIGURE

LIST OF TABLE