

BAB 1

PENDAHULUAN

1.1. Latar Belakang Masalah

Penyakit stroke yang terjadi akibat gangguan aliran darah ke otak. merupakan penyakit dengan peringkat kedua di dunia yang paling banyak menyebabkan kematian dan kecacatan [1]. Diagnosa penyakit stroke pun menjadi hal yang penting untuk dilakukan sebelum penyakit ini menjadi semakin parah dan mengakibatkan resiko kecacatan hingga kematian. Teknologi Machine Learning pun telah banyak diteliti untuk memprediksi penyakit stroke. Algoritma yang sering digunakan adalah Random Forest, Support Vector Machine (SVM), Artificial Neural Network (ANN), dan Decision Tree (DT).

Sebuah studi Systematic Review yang dilakukan oleh W.Wang et al [2] menemukan bahwa diantara keempat algoritma tersebut yang paling banyak memiliki performa terbaik adalah SVM, sehingga penulis mengangkat topik penelitian dengan algoritma tersebut. Namun demikian, W. Wang et al juga menyimpulkan bahwa penelitian terkait tidak memaparkan Hyper Parameter dan Hypertuning (Hyper Parameter Tuning). Hyper Parameter merupakan konfigurasi parameter yang mempengaruhi cara kerja algoritma Machine Learning. Sedangkan Hypertuning merupakan teknik yang berguna untuk meningkatkan performa Machine Learning dalam melakukan klasifikasi ataupun regresi.

Dari segi dataset, terdapat dua dataset medis yang telah diteliti untuk membandingkan tiga teknik Hypertuning yang cukup banyak dipakai yakni Randomized Search, Grid Search, dan Tree Base Pipeline Optimization Tool [3]. Kedua dataset tersebut adalah Cleveland heart disease (CHD) dataset dan Z-Alizadeh Sani (ZS) Dataset. Penelitian tersebut menemukan bahwa Teknik Randomized Search menghasilkan performa terbaik dibandingkan dengan yang lainnya. Di sisi lain, terdapat dataset Stroke dari Medical Clinic of Bangladesh (MCB) yang berjumlah 5110 data yang telah banyak diteliti dengan algoritma SVM dan metode Synthetic Minority Over-sampling Technique (SMOTE) untuk mengatasi masalah Data Imbalance [4]–[8],

namun setelah penulis melakukan penelusuran dari berbagai literatur ditemukan bahwa dataset tersebut belum diteliti untuk menjelaskan manfaat Randomized Search dengan algoritma Support Vector Machine.

Meskipun metode Hypertuning Grid Search Cross-validation dan Random Search digunakan, namun manfaat dan perbandingan metode tidak ditelusuri pada penelitian tersebut. Oleh karena itu, penelitian ini berjudul ‘PENERAPAN ALGORITMA SUPPORT VECTOR MACHINE DENGAN HYPERTUNING RANDOMIZED SEARCH PADA PREDIKSI STROKE’.

1.2. Rumusan Masalah

Rumusan masalah berdasarkan saran dan temuan yang ditemukan dari penelitian sebelumnya antara lain:

1. Data prediksi stroke yang perlu dianalisa sehingga dapat mendukung teknologi untuk mengantisipasi resiko penyakit stroke yang telah menjadi penyakit kedua yang paling banyak menyebabkan kematian dan menyebabkan kecacatan di dunia
2. Penelitian sebelumnya menemukan kurangnya analisa hypertuning maupun hyperparameter pada penelitian-penelitian terkait prediksi stroke dengan Machine Learning
3. Algoritma Support Vector Machine telah digunakan pada dataset pasien dari Clinic of Bangladesh namun tidak secara khusus memaparkan tentang manfaat dari hypertuning yang dilakukan

1.3. Tujuan

Penelitian ini bertujuan untuk menghasilkan analisa manfaat hypertuning Randomized Search terhadap prediksi stroke dengan algoritma Support Vector Machine pada dataset pasien dari Clinic of Bangladesh

1.4. Manfaat

Manfaat yang diharapkan dari dilakukannya penelitian ini, antara lain:

1. Penerapan algoritma Support Vector Machine dapat memberi kontribusi dalam dunia medis dengan menganalisa data pasien

2. Menghasilkan laporan mengenai manfaat metode Randomized Search pada algoritma Support Vector Machine dalam melakukan prediksi stroke
3. Memberikan kontribusi terkait Hypertuning yang dilaporkan kurang dilakukan oleh beberapa penelitian sebelumnya terkait prediksi stroke dengan Machine Learning

1.5. Batasan Masalah

Ruang lingkup penelitian ini, adalah sebagai berikut:

1. Jumlah dataset stroke yang diteliti memiliki jumlah data sebanyak 5110 data
2. Penelitian ini melakukan analisa komparatif antara performa prediksi stroke dengan hypertuning Randomized Search dan tanpa Randomized Search
3. Program yang digunakan adalah Jupyter Notebook dengan bahasa pemrograman Python 3

1.6. Keterbaruan

Berikut ini merupakan penelitian-penelitian sebelumnya yang telah melakukan prediksi stroke menggunakan dataset MCB yang diteliti pada penelitian ini namun terdapat kelemahan dari pemaparan tentang Hypertuning ataupun Hyperparameter seperti yang dijelaskan oleh W.Wang et al [2]:

1. Akbar et al. dalam penelitiannya yang berjudul **“An Ensemble Machine Learning Approach for Stroke Prediction”** [9] telah menggunakan algoritma Support Vector Machine dan algoritma-algoritma lainnya untuk memprediksi stroke menggunakan dataset Stroke dari MCB namun tidak menggunakan hyperparameter ataupun hypertuning.
2. Z. Li dalam penelitiannya yang berjudul **“Study of ICU Mortality Prediction and Analysis based on Random Forest”** [6] telah menggunakan dataset Stroke dari MCB dengan algoritma Random Forest untuk memprediksi stroke namun tidak memaparkan secara jelas terkait hypertuning dan hyperparameter
3. A Gaffney dalam penelitiannya yang berjudul **“An Ensemble Learning Algorithm for ICU Patient Mortality Prediction”** [7] telah menggunakan dataset Stroke dari MCB dengan Support Vector Machine dan tiga algoritma Machine Learning lainnya untuk memprediksi stroke. Hyperparameter yang

dijelaskan hanya berupa penggunaan 10-fold cross validation dan tidak memaparkan secara jelas terkait hypertuning yang digunakan meskipun penelitian tersebut menjelaskan bahwa “modelling and tuning of various classification models” yang mengindikasikan bahwa terdapat proses hypertuning untuk memaksimalkan performa model tersebut.

4. Cohen et al. dalam penelitiannya yang berjudul **“ICU Survival Prediction Incorporating Test-Time Augmentation to Improve the Accuracy of Ensemble-Based Models”** [8] telah menggunakan dataset Stroke dari MCB dengan H2O AutoML yang menggunakan berbagai algoritma Machine Learning untuk memprediksi stroke dan memaparkan secara jelas terkait hyperparameter namun tidak memaparkan teknik hypertuning yang digunakan.
5. R. Mitraa dan T. Rajendranb dalam penelitiannya yang berjudul **“Efficient Prediction of Stroke Patients Using Random Forest Algorithm in Comparison to Support Vector Machine,”** membandingkan algoritma Support Vector Machine dengan Random Forest untuk memprediksi stroke pada dataset MCB namun tidak memaparkan secara jelas mengenai Hypertuning dan Hyperparameter yang digunakan.