

BAB II

TINJAUAN PUSTAKA

2.1 Studi Literatur

Penelitian terdahulu tentang MBTI (Myers-Briggs Type Indicator) dengan metode boosting dengan berbagai penelitian sebelumnya yang menggunakan MBTI sebagai basis data dan membandingkan efektifitas atau kinerja metode boosting. Penelitian – penelitian tersebut menjadi acuan penting dalam menentukan langkah – langkah sistematis penyusunan penelitian ini. Pada **Tabel 2.1** menjelaskan penelitian – penelitian tersebut.

Tabel 2. 1 Rincian Penelitian Terdahulu

No	Penulis	Judul	Dataset	Metode	Hasil
1	Gregorius Ryan, Pricillia Katarina , Derwin Suhartono	MBTI Personality Prediction Using Machine Learning and SMOTE for Balancing Data Based on Statement Senteces	(MBTI) Myers-Briggs Personality Type Dataset. Available online: https://www.kaggle.com/datasets/datasnaek/mbti-type (accessed on 20 November 2022).	Extreme Gradient Boosting classifier, Cat Boosting Classifier dengan model Word2Vec	79% untuk CatBoost dan 78% untuk XGBoost
2	Sharmeen Binti Syazwan Lai , Nur Huda Nabihan Binti Md Shahri , Mazni Binti	Comparing the Performance of AdaBoost, XGBoost, and Logistic Regression for Imbalanced Data	UCI Machine Learning Repository (https://archive.ics.uci.edu/ml/datasets.html), which contains numerous collections of	AdaBoost, XGBoost, Logistic Regression	Rata – rata akurasi berdasarkan Metrik : 99% untuk XGBoost, 98% untuk Adaboost, dan 84%

	Mohamad , Hezlin Aryani Binti Abdul Rahman, Adzhar Bin Rambli		datasets created by students of UC Irvine.		untuk Logistic
3	Jian Yang, Jinhan Guan	A Heart Disease Prediction Model Based on Feature Optimization and Smote- Xgboost Algorithm	Dataset dari direkam pada pasien nyata pada rumah sakit. Dan dataset tersebut dinamakan Heart Disease Dataset (HDD)	Metode XGBoost dengan 5 metode baseline menggunakan confusion matrix	93% untuk XGBoost
4	Mohammad Hossein Amirhossein i, Hassan Kazemian	Machine Learning Approach to Personality Type Prediction Based on the Myers– Briggs Type Indicator	(MBTI) Myers-Briggs Personality Type Dataset. Available online: https://www.kaggle.com/datasets/datasnaek/mbti-type	Extreme Gradient Boosting (XGBoost) dan Recurrent Neural Network (RNN)	Rata – rata akurasi dalam setiap tipe kepribadian untuk XGBoost di angka 75% , dan RNN di angka 67,2 %

Berdasarkan pada tabel 2.1 diatas terdapat beberapa penelitian terdahulu yang menggunakan dataset, metode, dan model yang sama. Tetapi tidak ada kasus yang serupa seperti penelitian ini. Hanya 2 penelitian yang hampir serupa. Seperti pada penelitian terdahulu nomor 1 dilakukan klasifikasi dengan kasus yang sama akan tetapi metode yang

dibandingkan yaitu XGBoost dan CatBoost lalu disempurnakan menggunakan SMOTE dengan akurasi metode XGBoost mencapai 78% , metode CatBoost 79%[2]. Hasil dari penelitian ini menjelaskan bahwa dataset ini cenderung kurang seimbang, maka dari itu penulis menggunakan SMOTE agar menghasilkan data yang kurang dan meningkatkan performa dari data tersebut. Menggunakan 2 fitur untuk representasi teks dan menunjukkan bahwa menggunakan SMOTE dengan algoritma ini meningkatkan akurasi klasifikasi dan skor F1.

Pada penelitian nomor 2 menggunakan dataset yang diambil dari UCI Machine Learning perbandingan menggunakan AdaBoost dan XGBoost namun tidak disempurnakan oleh smote. Yang menyentuh akurasi hingga 99% untuk XGBoost dan 98% untuk AdaBoost[11]. Penelitian yang membandingkan kinerja antara XGBoost dan AdaBoost dalam data yang tidak seimbang. Penulis juga menyarankan metode XGBoost yang secara umum memiliki kinerja terbaik dalam menawarkan akurasi yang lebih tinggi. Dan lebih efektif menangani ketidakseimbangan data dibandingkan dengan Adaboost.

Pada penelitian nomor 3 meneliti tentang penyakit jantung model ini menggabungkan pemilihan fitur dan metode SMOTE – XGBoost untuk meningkatkan akurasi prediksi pada kumpulan data medis yang tidak seimbang[12]. Menggunakan Metode XGBoost dan mendapat akurasi sebesar 93% dibandingkan dengan beberapa algoritma dasar. Menunjukkan kinerja yang unggul dalam akurasi, skor F1.

Untuk penelitian terakhir di nomor 4 penelitian ini lebih menekankan pentingnya optimalan fitur dalam peningkatan model prediksi menggunakan algoritma XGBoost dengan menggunakan dataset MBTI yang mendapat akurasi sebesar 75%. Hasilnya dibandingkan dengan metode yang digunakan menunjukkan peningkatan akurasi[3].

Dengan demikian, berdasarkan penjelasan di atas, penelitian ini bertujuan untuk mencapai hasil yang tinggi dalam mengoptimisasi dataset MBTI dengan membandingkan dua teknik boosting dan penyeimbangan dengan metode SMOTE.