IMPLEMENTASI SYNTHETIC MINORITY OVERSAMPLING TECHNIQUE (SMOTE) PADA ALGORITMA EXTREME GRADIENT BOOSTING (XGBOOST) UNTUK KLASIFIKASI INDEKS STANDAR PENCEMARAN UDARA (ISPU)

SKRIPSI

Oleh :

ACHMAD FAUZIHAN BAGUS SAJIWO

20081010069

PROGRAM STUDI INFORMATIKA

FAKULTAS ILMU KOMPUTER

UNIVERSITAS PEMBANGUNAN NASIONAL "VETERAN"

JAWA TIMUR

2024

LEMBAR PENGESAHAN SKRIPSI

Judul : **IMPLEMENTASI SYNTHETIC MINORITY OVERSAMPLING TECHNIQUE (SMOTE) PADA ALGORITMA EXTREME GRADIENT BOOSTING (XGBOOST) UNTUK KLASIFIKASI INDEKS STANDAR PENCEMARAN UDARA (ISPU)**

Oleh : **ACHMAD FAUZIHAN BAGUS SAJIWO**

NPM : **22081010069**

Telah Diseminarkan Dalam Ujian Skripsi Pada :

Hari Kamis, Tanggal 04 Juli 2024

Mengetahui

Dosen Pembimbing

1.

Dr. Basuki Rahmat, S.Si, MT.

NIP. 19690723 2021211 002

2.

Achmad Junaidi, S.Kom, M.Kom.

NPT. 3 7811 04 0199 1

Dosen Penguji

1.

Eva Yulia Puspaningrum, S.Kom., M.Kom

NIP. 19890705 2021212 002

2.

Retno Mumpuni, S.Kom., M.Sc

NPT. 172198 70 716054

Menyetujui

Dekan

Fakultas Ilmu Komputer

Prof. Dr. Ir. Novirina Hendrasarie, MT.

NIP. 19681126 199403 2 001

Koordinator Program Studi

Informatika

Fetty Tri Anggraeny, S.Kom, M.Kom

NIP .19820211 2021212 005

SURAT PERNYATAAN BEBAS DARI PLAGIASI

Saya, mahasiswa Program Studi Sarjana Informatika Universitas Pembangunan Nasional "Veteran" Jawa Timur, yang bertanda tangan di bawah ini:

Nama : Achmad Fauzihan Bagus Sajiwo

NPM : 20081010069

Menyatakan dengan sesungguhnya bahwa Skripsi/Tugas Akhir yang saya kerjakan berjudul:

"IMPLEMENTASI SYNTHETIC MINORITY OVERSAMPLING TECHNIQUE (SMOTE) PADA ALGORITMA EXTREME GRADIENT BOOSTING (XGBOOST) UNTUK KLASIFIKASI INDEKS STANDAR PENCEMARAN UDARA (ISPU)"

bukan merupakan plagiasi sebagian atau keseluruhan dari Skripsi/Tugas Akhir/Penelitian orang lain dari juga bukan merupakan produk dan software yang saya beli dari pihak lain. Saya juga menyatakan bahwa Skripsi/Tugas Akhir ini secara keseluruhan adalah pekerjaan Saya sendiri, kecuali yang dinyatakan dalam Daftar Pustaka dan tidak pernah diajukan untuk syarat memperoleh gelar di Universitas Pembangunan Nasional "Veteran" Jawa Timur maupun di Institut Pendidikan lain. Bukti hasil pengecekan plagiasi dokumen ini dapat ditelusuri melalui QR Code di bawah.

Apabila di kemudian hari terbukti bahwa dokumen ini merupakan plagiasi karya orang lain, saya sanggup menerima sanksi sesuai aturan yang berlaku.

Demikian atas perhatiannya disampaikan terima kasih.

Surabaya, 4 Juli 2024

Hormat saya,

Achmad Fauzihan Bagus Sajiwo

NPM. 20081010069

IMPLEMENTASI SYNTHETIC MINORITY OVERSAMPLING TECHNIQUE (SMOTE) PADA ALGORITMA EXTREME GRADIENT BOOSTING (XGBOOST) UNTUK KLASIFIKASI INDEKS STANDAR PENCEMARAN UDARA (ISPU)

Nama Mahasiswa	**: Achmad Fauzihan Bagus Sajiwo**
NPM	**: 20081010069**
Program Studi	**: Informatika**
Dosen Pembimbing	**: Dr. Basuki Rahmat, S.Si. MT.**
	Achmad Junaidi, S.Kom, M.Kom.

ABSTRAK

Polusi udara adalah masuknya zat-zat berbahaya ke atmosfer, yang dapat disebabkan oleh tindakan manusia, baik sengaja maupun tidak sengaja, serta oleh peristiwa alam. Menurut *Air Quality Live Index* (AQLI) pada bulan April 2021, DKI Jakarta sebagai ibu kota negara, menempati posisi keenam di dunia dengan kota tingkat kualitas udara yang paling buruk. Untuk menghadapi masalah polusi udara yang terus memburuk, perlu diambil tindakan yang tepat dan efektif. Satu diantaranya adalah melakukan penelitian klasifikasi indeks standar pencemaran udara (ISPU).

Penerapan klasifikasi ISPU membutuhkan metode yang dapat mengolah dan menganalisis pola data dari sensor-sensor yang mengukur tingkat polutan udara. Metode yang digunakan pada penelitian ini adalah *eXtreme Gradient Boosting* (XGBoost). Untuk membantu menyeimbangkan data, pada penelitian ini menggunakan *Synthetic Minority Over-sampling Technique* (SMOTE). Data yang digunakan adalah ISPU DKI Jakarta tahun 2022-2023 yang berasal dari website Satu Data Jakarta: https://satudata.jakarta.go.id/home.

Hasil klasifikasi indeks standar pencemaran udara menggunakan algoritma *eXtreme Gradient Boosting* dengan *Synthetic Minority Over-sampling Technique*, didapatkan *accuracy* sebesar 99.63%. Dari perhitungan *confusion matrix* didapatkan nilai *precision, recall* dan *f1-score*. Untuk kelas 0 pada *precision* didapatkan nilai sebesar 99%, pada *recall* sebesar 100% dan *f1-score* sebesar 100%. Untuk kelas 1 pada *precision* didapatkan nilai sebesar 100%, pada *recall*

sebesar 99% dan *f1-score* sebesar 100%. Untuk kelas 2 pada *precision* didapatkan nilai sebesar 100%, pada *recall* sebesar 100% dan *f1-score* sebesar 100%.

Kata kunci: *ISPU, Klasifikasi, XGBoost, Imbalanced Data, SMOTE*

IMPLEMENTATION OF SYNTHETIC MINORITY OVERSAMPLING TECHNIQUE (SMOTE) IN THE EXTREME GRADIENT BOOSTING ALGORITHM (XGBOOST) FOR CLASSIFICATION OF AIR POLLUTION STANDARD INDEX (ISPU)

Nama Mahasiswa	**: Achmad Fauzihan Bagus Sajiwo**
NPM	**: 20081010069**
Program Studi	**: Informatika**
Dosen Pembimbing	**: Dr. Basuki Rahmat, S.Si. MT.**
	Achmad Junaidi, S.Kom, M.Kom.

ABSTRACT

Air pollution is the entry of harmful substances into the atmosphere, which can be caused by human actions, whether intentional or unintentional, as well as by natural events. According to the Air Quality Live Index (AQLI) in April 2021, DKI Jakarta, as the nation's capital, is in sixth place in the world with the city with the worst air quality level. To deal with the problem of air pollution which continues to worsen, appropriate and effective action needs to be taken. One of them is conducting research on the classification of the air pollution standard index (ISPU).

Implementing the ISPU classification requires a method that can process and analyze data patterns from sensors that measure air pollutant levels. The method used in this research is eXtreme Gradient Boosting (XGBoost). To help balance the data, this research used Synthetic Minority Over-sampling Technique (SMOTE). The data used is the DKI Jakarta ISPU for 2022-2023 which comes from the Satu Data Jakarta website: https://satudata.jakarta.go.id/home.

The results of the standard air pollution index classification using the eXtreme Gradient Boosting algorithm with Synthetic Minority Over-sampling Technique, obtained an accuracy of 99.63%. From the confusion matrix calculations, precision, recall and f1-score values are obtained. For class 0, the precision value is 99%, the recall is 100% and the f1-score is 100%. For class 1, the precision score was 100%, the recall was 99% and the f1-score was 100%. For

class 2, the precision score was 100%, the recall was 100% and the f1-score was 100%.

Keywords: *ISPU, Classification, XGBoost, Imbalanced Data, SMOTE*

KATA PENGANTAR

Puji syukur kita panjatkan kehadirat Allah Swt. yang telah memberikan rahmat dan hidayah-Nya sehingga penulis dapat menyelesaikan laporan skripsi yang berjudul "Implementasi Synthetic Minority Oversampling Technique (SMOTE) Pada Algoritma Extreme Gradient Boosting (XGBoost) Untuk Klasifikasi Indeks Standar Pencemaran Udara (ISPU)" ini tepat pada waktunya. Penulis berusaha dengan semaksimal mungkin dalam penyusunan laporan ini baik dari hasil bimbingan bersama dosen pembimbing, maupun diskusi dengan teman-teman penulis.

Surabaya, 04 Juli 2024

Penulis

Achmad Fauzihan Bagus Sajiwo

UCAPAN TERIMA KASIH

Dalam kesempatan ini, penulis mengucapkan banyak terima kasih kepada semua pihak yang telah membantu dalam pelaksanaan kegiatan perkuliahan maupun penyusunan laporan ini kepada :

1. Orang tua penulis yang selalu memberikan doa serta dukungan baik secara moril ataupun materil kepada penulis
2. Bapak Prof. Dr. Ir. Akhmad Fauzi, M.MT., selaku Rektor Universitas Pembangunan Nasional "Veteran" Jawa Timur.
3. Ibu Prof. Dr. Ir. Novirina Hendrasarie, MT., selaku Dekan Fakultas Ilmu Komputer, Universitas Pembangunan Nasional "Veteran" Jawa Timur.
4. Ibu Fetty Tri Anggraeny, S.Kom, M.Kom. selaku Koordinator Program Studi Informatika Universitas Pembangunan Nasional "Veteran" Jawa Timur.
5. Bapak Andreas Nugroho Sihananto, S.Kom., M.Kom., selaku Koordinator Skripsi Program Studi Informatika Universitas Pembangunan "Veteran" Jawa Timur.
6. Bapak Dr. Basuki Rahmat, S.Si. MT., selaku dosen pembimbing 1 yang telah membimbing penulis dari awal penyusunan laporan sampai penanda tanganan laporan ini.
7. Bapak Achmad Junaidi, S.Kom, M.Kom., selaku dosen pembimbing 2 yang telah membimbing penulis dari awal penyusunan laporan sampai penanda tanganan laporan ini.
8. Ibu Henni Endah Wahanani, ST. M.Kom. selaku dosen wali Program Studi Informatika Fakultas Ilmu Komputer Universitas Pembangunan Nasional "Veteran" Jawa Timur.
9. Seluruh dosen dan staff Program Studi Informatika dan staff Fakultas Ilmu Komputer Universitas Pembangunan Nasional "Veteran" Jawa Timur.
10. Badan Legislatif Mahasiswa Fakultas Ilmu Komputer Universitas Pembangunan Nasional "Veteran" Jawa Timur (BLM FASILKOM UPN "Veteran" Jawa Timur) dan Badan Eksekutif Mahasiswa Fakultas Ilmu Komputer Universitas Pembangunan Nasional "Veteran" Jawa Timur (BEM FASILKOM UPN "Veteran" Jawa Timur) yang telah menjadi

wadah bagi penulis untuk belajar berorganisasi dan menambah relasi yang dapat berguna bagi masa depan penulis.

11. Teman-teman penulis yang senantiasa bersama dari semester 1 sampai sekarang.

Akhir kata, penulis berharap laporan ini dapat memberikan manfaat serta menjadi bahan referensi bagi penulis maupun pembaca. Penulis menyadari bahwa laporan yang ditulis ini masih jauh dari kata sempurna. Oleh karena itu, kritik dan saran yang membangun penulis butuhkan demi masa depan bersama.

DAFTAR ISI

DAFTAR TABEL

DAFTAR GAMBAR

DAFTAR KODE PROGRAM