

Hak Cipta Diindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

PREDIKSI KADAR POLUTAN INDIKATOR KUALITAS UDARA DI PEKANBARU MENGGUNAKAN CATBOOST

TUGAS AKHIR

Disusun Sebagai Salah Satu Syarat
Untuk Memperoleh Gelar Sarjana Teknik
Pada Jurusan Teknik Informatika

Oleh

AMORINA HANJANI

NIM. 12050120365



UIN SUSKA RIAU

FAKULTAS SAINS DAN TEKNOLOGI
UNIVERSITAS ISLAM NEGERI SULTAN SYARIF KASIM RIAU
PEKANBARU
2024

LEMBAR PENGESAHAN

PREDIKSI KADAR POLUTAN INDIKATOR KUALITAS UDARA DI PEKANBARU MENGGUNAKAN CATBOOST

Oleh

AMORINA HANJANI

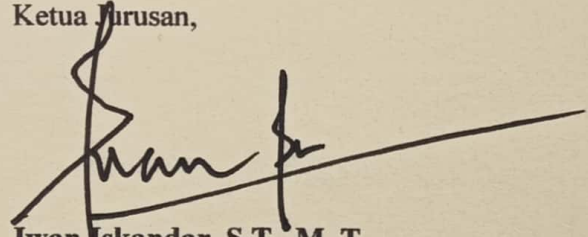
NIM. 12050120365

Telah dipertahankan di depan sidang dewan penguji
sebagai salah satu syarat untuk memperoleh gelar Sarjana Teknik
pada Universitas Islam Negeri Sultan Syarif Kasim Riau

Pekanbaru, 14 November 2024

Mengesahkan,

Ketua Jurusan,



Iwan Iskandar, S.T., M. T

NIP. 19821216 201503 1 003



Dr. Hartono, M. Pd

NIP. 19640301 199203 1 003

DEWAN PENGUJI

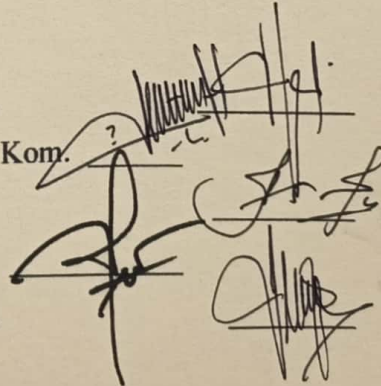
Ketua : Dr. Elin Haerani, S.T., M. Kom.

Pembimbing I : Dr. Lestari Handayani, S.T., M. Kom.

Pembimbing II : Iis Afrianty, S.T., M. Sc.

Penguji I : Muhammad Affandes, M. T.

Penguji II : Fadhilah Syafria, S.T., M. Kom.



LEMBAR PERSETUJUAN

PREDIKSI KADAR POLUTAN INDIKATOR KUALITAS UDARA DI PEKANBARU MENGGUNAKAN CATBOOST

TUGAS AKHIR

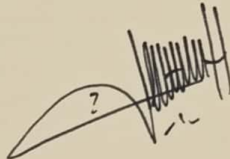
Oleh

AMORINA HANJANI

NIM. 12050120365

Telah diperiksa dan disetujui sebagai Laporan Tugas Akhir
di Pekanbaru, pada tanggal 14 November 2024

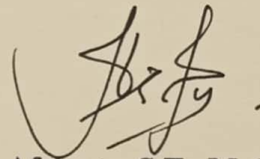
Pembimbing I,



Dr. Lestari Handayani, S.T., M. Kom

NIP. 19811113 200710 2 003

Pembimbing II



Iis Afrianty, S.T., M. Sc

NIP. 19880426 201903 2 009



LEMBAR HAK ATAS KEKAYAAN INTELEKTUAL

Tugas Akhir yang tidak diterbitkan ini terdaftar dan tersedia di Perpustakaan Universitas Islam Negeri Sultan Syarif Kasim Riau adalah terbuka untuk umum dengan ketentuan bahwa hak cipta pada penulis. Referensi kepustakaan diperkenankan dicatat, tetapi pengutipan atau ringkasan hanya dapat dilakukan seizin penulis dan harus disertai dengan kebiasaan ilmiah untuk menyebutkan sumbernya.

Penggandaan atau penerbitan sebagian atau seluruh Tugas Akhir ini harus memperoleh izin dari Dekan Fakultas Sains dan Teknologi Universitas Islam Negeri Sultan Syarif Kasim Riau. Perpustakaan yang meminjamkan Tugas Akhir ini untuk anggotanya diharapkan untuk mengisi nama, tanda peminjaman dan tanggal pinjam.

Hak Cipta Diindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

LEMBAR PERNYATAAN

Dengan ini saya menyatakan bahwa dalam Tugas Akhir ini tidak terdapat karya yang pernah diajukan untuk memperoleh gelar kesarjanaan di suatu Perguruan Tinggi, dan sepanjang pengetahuan saya juga tidak terdapat karya atau pendapat yang pernah ditulis atau diterbitkan oleh orang lain kecuali yang secara tertulis diacu dalam naskah ini dan disebutkan didalam daftar pustaka.

Pekanbaru, 18 November 2024

Yang membuat pernyataan,



AMORINA HANJANI

12050120365

LEMBAR PERSEMBAHAN

Tulisan pada lembar persembahan merupakan ekspresi bebas mahasiswa yang dimaksudkan sebagai ungkapan filosofis yang dapat memotivasi semua pihak, khususnya mahasiswa/penulis sendiri dalam penyelesaian tugas akhir.

Lembar persembahan format bebas, tidak berlebih-lebihan dan dalam komposisi warna grayscale atau hitam-putih.

Lembar persembahan dibuat secara wajar, tidak menyinggung, tidak mengandung hoaks dan tidak mengandung unsur SARA.

Lembar persembahan cukup dibuat 1 halaman ini saja.

Hak Cipta Diindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

Hak Cipta Diindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

ABSTRAK

Polusi udara akibat industri, transportasi, dan kebakaran hutan seperti yang terjadi di Pekanbaru pada 2019 semakin memperburuk kualitas udara hingga mencapai tingkat yang sangat berbahaya dan memicu penyakit pernapasan serta kematian. Penelitian ini bertujuan untuk menerapkan model Categorical Boosting (CatBoost) guna memprediksi kadar polutan indikator kualitas udara di Pekanbaru dan mengevaluasi pengaruh metode pre-processing dan pembagian data terhadap kinerja model. Penelitian ini berfokus pada prediksi polutan udara PM_{10} , SO_2 , CO , O_3 , NO_2 , $PM_{2.5}$. Hasilnya menunjukkan bahwa penggunaan metode pre-processing median dan pembagian data 80:20 dengan parameter $iterations=500$, $depth=6$, dan $learning_rate=0.01$ memberikan hasil evaluasi terendah pada $PM_{2.5}$ dengan RMSE sebesar 10.93 dan MAE terendah sebesar 8.65. Secara keseluruhan, model CatBoost tidak mampu menangkap pola dan tren data dengan baik. Penelitian lebih lanjut diperlukan untuk mengatasi keterbatasan model dalam menangkap pola tren data.

Kata kunci: kualitas udara, prediksi, catboost

ABSTRACT

Air pollution caused by industry, transportation, and forest fires as seen in Pekanbaru in 2019, further deteriorates air quality to extremely hazardous levels and triggers respiratory diseases and fatalities. This study aims to apply the Categorical Boosting (CatBoost) model to predict the Air Pollution Standard Index (ISPU) in Pekanbaru and evaluate the impact of pre-processing methods and data splitting on model performance. This research focuses on predicting the values of elements that serve as air quality indicators: PM10, SO₂, CO, O₃, NO₂, PM_{2,5}. The results show that using the median pre-processing method and 80:20 data split with parameters of iterations=500, depth=6, and learning_rate=0.01 yielded the lowest evaluation scores for PM_{2.5}, with an RMSE of 10.93 and the lowest MAE of 8.65. Overall, the CatBoost model is unable to effectively capture the patterns and trends in the data. Further research is needed to address the model's limitations in capturing data trend patterns.

Keywords: air quality, prediction, catboost

Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

KATA PENGANTAR

Assalammu'alaikum wa rohmatullohi wa barokatuh.

Alhamdulillah *robbil'amin*, tak henti-hentinya kami ucapkan kehadiran Allah *Subhanahu wa ta'ala*, yang dengan rahmat dan hidayah-Nya kami mampu menyelesaikan Tugas Akhir ini dengan baik. Tidak lupa bershalawat kepada Nabi dan Rasul-Nya, Nabi Muhammad *Sholallohu 'alaihi wa salam*, yang telah membimbing kita sebagai umatnya menuju jalan kebaikan.

Tugas Akhir ini disusun sebagai salah satu syarat untuk mendapatkan gelar sarjana pada jurusan Teknik Informatika Universitas Islam Negeri Sultan Syarif Kasim Riau. Banyak sekali pihak yang telah membantu kami dalam penyusunan laporan ini, baik berupa bantuan materi ataupun berupa motivasi dan dukungan kepada kami. Semua itu tentu terlalu banyak bagi kami untuk membalasnya, namun pada kesempatan ini kami hanya dapat mengucapkan terima kasih kepada:

1. Bapak Prof. Dr. Hairunas, M. Ag selaku Rektor Universitas Islam Negeri Sultan Syarif Kasim Riau.
2. Bapak Dr. Hartono, M. Pd. selaku Dekan Fakultas Sains dan Teknologi Universitas Islam Negeri Sultan Syarif Kasim Riau.
3. Bapak Iwan Iskandar, M. T. selaku Ketua Jurusan Teknik Informatika Fakultas Sains dan Teknologi Universitas Islam Negeri Sultan Syarif Kasim Riau.
4. Ibu Liza Afriyanti, M. Kom. selaku Penasehat Akademik Jurusan Teknik Informatika Fakultas Sains dan Teknologi Universitas Islam Negeri Sultan Syarif Kasim Riau.
5. Ibu Dr. Lestari Handayani, S.T., M. Kom. selaku Pembimbing I Laporan Tugas Akhir Jurusan Teknik Informatika Fakultas Sains dan Teknologi Universitas Islam Negeri Sultan Syarif Kasim Riau.



Hak Cipta Diindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

6. Ibu Iis Afrianty, S. T., M. Sc selaku Pembimbing II Laporan Tugas Akhir Jurusan Teknik Informatika Fakultas Sains dan Teknologi Universitas Islam Negeri Sultan Syarif Kasim Riau.
7. Bapak Muhammad Affandes, M.T selaku penguji I Laporan Tugas Akhir Jurusan Teknik Informatika Fakultas Sains dan Teknologi Universitas Islam Negeri Sultan Syarif Kasim Riau.
8. Ibu Fadhillah Syafria, S.T., M. Kom selaku Penguji II dan Koordinator Tugas Akhir Jurusan Teknik Informatika Fakultas Sains dan Teknologi Universitas Islam Negeri Sultan Syarif Kasim Riau.
9. Orang tua yaitu Bapak Amrun dan Ibu Erlina, serta adik Awraq dan kakak Veta yang selalu memberikan nasehat, dukungan, do'a, finansial dan motivasi saya dalam menyelesaikan tugas akhir ini.
10. Kepada teman-teman saya, khususnya Nurfazilla, Ayu Larasati, dan Rana Fatrika yang turut membantu dalam menyelesaikan tugas akhir ini.
11. Seluruh pihak yang belum kami cantumkan, terima kasih atas dukungannya, baik material maupun spiritual.

Kami menyadari bahwa dalam penulisan laporan ini masih banyak kesalahan dan kekurangan, oleh karena itu kritik dan saran yang sifatnya membangun sangat kami harapkan untuk kesempurnaan laporan ini. Akhirnya kami berharap semoga laporan ini dapat memberikan sesuatu yang bermanfaat bagi siapa saja yang membacanya.

Wassalamu'alaikum wa rohmatullohi wa barokatuh.

Pekanbaru, 14 November 2024

Amorina Hanjani

Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

DAFTAR ISI

| | |
|---|-----------|
| LEMBAR PERSEMBAHAN | i |
| ABSTRAK | ii |
| ABSTRACT | iii |
| KATA PENGANTAR | iv |
| DAFTAR ISI..... | vi |
| DAFTAR GAMBAR | viii |
| DAFTAR TABEL..... | x |
| DAFTAR RUMUS | xi |
| BAB 1 PENDAHULUAN | 1 |
| 1.1 Latar Belakang | 1 |
| 1.2 Rumusan Masalah | 5 |
| 1.3 Batasan Masalah..... | 5 |
| 1.4 Tujuan Penelitian..... | 6 |
| 1.5 Manfaat Penelitian..... | 6 |
| BAB 2 KAJIAN PUSTAKA..... | 7 |
| 2.1 Kajian Metode | 7 |
| 2.1.1 Kualitas Udara..... | 7 |
| 2.1.2 Data Mining Prediksi | 10 |
| 2.2 Penelitian Terkait | 17 |
| BAB 3 METODOLOGI PENELITIAN..... | 24 |
| 3.1 Tahapan Penelitian | 24 |
| 3.1.1 Pengumpulan Data | 24 |

Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

| | | |
|-----------------------------------|--|-----------|
| 3.1.2 | Data Mining untuk Prediksi | 25 |
| 3.1.3 | Skenario Uji | 30 |
| 3.1.4 | Prediksi Kadar Polutan Tahun 2024 | 31 |
| BAB 4 PEMBAHASAN | | 32 |
| 4.1 | Pengumpulan Data | 32 |
| 4.2 | <i>Preprocessing</i> | 34 |
| 4.3 | Pemodelan CatBoost | 41 |
| 4.4 | Evaluasi Model | 49 |
| 4.5 | Hasil dan Pembahasan | 50 |
| 4.5.1 | Hasil Prediksi Konsentrasi Polutan Udara | 50 |
| 4.5.2 | Hasil Skenario Pengujian | 54 |
| 4.5.3 | Hasil Prediksi Tahun 2024 | 57 |
| BAB 5 PENUTUP | | 62 |
| 5.1 | Kesimpulan | 62 |
| 5.2 | Saran | 62 |
| DAFTAR PUSTAKA | | 63 |
| LAMPIRAN A | | 73 |
| DAFTAR RIWAYAT HIDUP | | 76 |

DAFTAR GAMBAR

| | |
|---|----|
| Gambar 1 Flowchart Pelatihan CatBoost..... | 14 |
| Gambar 2 Flowchart Pengujian CatBoost..... | 15 |
| Gambar 3 Tahapan Penelitian | 24 |
| Gambar 4 Data Kualitas Udara | 32 |
| Gambar 5 Data Kualitas Udara Setelah Penggabungan Data | 33 |
| Gambar 6 Jumlah Nilai <i>Missing Value</i> | 34 |
| Gambar 7 Data Setelah Dilakukan Pengisian Nilai <i>Mean</i> | 35 |
| Gambar 8 Data Setelah Dilakukan Pengisian Nilai Median | 36 |
| Gambar 9 Data Setelah Dilakukan Penanganan Interpolasi Linear | 37 |
| Gambar 10 Kode Phyton untuk Mendeteksi Outlier..... | 37 |
| Gambar 11 Data PM10 Sebelum Dilakukan Penanganan Outlier | 40 |
| Gambar 12 Data PM10 Setelah Dilakukan Penanganan Outlier | 40 |
| Gambar 13 Hasil Pembagian Data | 41 |
| Gambar 14 Kode Phyton untuk Fitur Deret Waktu | 42 |
| Gambar 15 Kode Python untuk Memisahkan Fitur dan Target | 43 |
| Gambar 16 Kode Pyhton untuk Mencari Parameter Terbaik..... | 44 |
| Gambar 17 Contoh Pohon Keputusan PM10..... | 47 |
| Gambar 18 Pohon Keputusan Iterasi Pertama Semua Data PM10 | 47 |
| Gambar 19 Grafik Prediksi PM10..... | 51 |
| Gambar 20 Grafik Prediksi SO2 | 52 |
| Gambar 21 Grafik Prediksi CO..... | 52 |
| Gambar 22 Grafik Prediksi O3 | 53 |
| Gambar 23 Grafik Prediksi NO2..... | 53 |
| Gambar 24 Grafik Prediksi PM2,5..... | 54 |
| Gambar 25 Hasil Prediksi PM20 2024..... | 57 |
| Gambar 26 Hasil Prediksi SO2 2024 | 58 |
| Gambar 27 Hasil Prediksi CO 2024..... | 58 |
| Gambar 28 Hasil Prediksi O3 2024 | 59 |

Hak Cipta Dilindungi Undang-Undang

1. Diarangi mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

| | |
|--|----|
| Gambar 29 Hasil Prediksi NO2 2024..... | 59 |
| Gambar 30 Hasil Prediksi PM2,5 2024..... | 60 |





Hak Cipta Diindungi Undang-Undang

1. Diarangi mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

DAFTAR TABEL

| | |
|--|----|
| Tabel 1 Parameter untuk Model CatBoost | 14 |
| Tabel 2 Penelitian mengenai kasus pencemaran udara | 17 |
| Tabel 3 Penelitian prediksi kualitas udara menggunakan <i>machine learning</i> | 18 |
| Tabel 4 Penelitian perbandingan metode CatBoost dengan metode lainnya | 20 |
| Tabel 5 Penelitian terkait penggunaan CatBoost untuk prediksi | 22 |
| Tabel 6 Ratio Pembagian Data | 28 |
| Tabel 7 Daftar <i>Hyperparameter</i> yang digunakan | 29 |
| Tabel 8 Skema Skenario Uji yang Digunakan | 31 |
| Tabel 9 Hasil Outlier yang Ditemukan | 39 |
| Tabel 10 Hasil Data Setelah Penambahan Fitur | 42 |
| Tabel 11 Hasil Perolehan Parameter Terbaik | 44 |
| Tabel 12 Hasil Perhitungan Residu | 45 |
| Tabel 13 Hasil <i>Leaf</i> yang Diperoleh | 46 |
| Tabel 14 Hasil Prediksi Data PM10 | 48 |
| Tabel 15 Hasil Evaluasi model | 49 |
| Tabel 16 Hasil Nilai Prediksi Setiap Fitur | 50 |
| Tabel 17 Hasil Skenario Uji | 54 |
| Tabel 18 Hasil Skenario Uji Terbaik | 56 |
| Tabel 19 Hasil Evaluasi Prediksi 2024 | 60 |

DAFTAR RUMUS

| | |
|-------------|----|
| © Hak cipta | 16 |
| (1) | 16 |
| (2) | 26 |
| (3) | 26 |
| (4) | 26 |
| (5) | 26 |
| (6) | 27 |
| (7) | 27 |
| (8) | 28 |
| (9) | 29 |
| (10) | 30 |

Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

1. Diarangi mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

BAB 1

PENDAHULUAN

1.1 Latar Belakang

Kualitas udara memiliki peran krusial bagi kehidupan makhluk hidup di bumi, terutama bagi manusia (Prayudha et al., 2018). Ketika udara terkontaminasi oleh zat-zat yang tidak diinginkan yang memiliki efek berbahaya pada makhluk hidup dan lingkungan ini disebut sebagai polusi udara (Manisalidis et al., 2020). Zat-zat yang mencemari udara disebut polutan udara. Beberapa contoh polutan udara tersebut meliputi Nitrogen Dioksida (NO_2), Karbon Monoksida (CO), Ozon (O_3), Sulfur Dioksida (SO_2), dan Partikulat Materi (PM) (Singgih, 2020). Kadar unsur yang tinggi dalam senyawa ini dapat berdampak pada kesehatan masyarakat, khususnya memicu penyakit pernapasan, dan bahkan berpotensi menyebabkan kematian (Nababan et al., 2023).

Pencemaran udara saat ini telah mencapai tingkat yang mengkhawatirkan akibat perkembangan industri dan meningkatnya jumlah populasi manusia di dunia, yang menyebabkan peningkatan polusi udara (Raka Siwih, 2020). Lebih dari setengah dari dampak polusi udara terhadap kesehatan manusia terjadi di negara-negara yang sedang berkembang akibat kelebihan populasi dan urbanisasi yang tidak terkendali seiring dengan berkembangnya industrialisasi (Manisalidis et al., 2020). Berdasarkan laporan *Air Quality Life Index* (AQLI) tahun 2021, beberapa negara di Asia Selatan termasuk Indonesia, menyumbang lebih dari setengah total tahun hidup yang hilang secara global akibat dampak pencemaran udara terhadap harapan hidup.

Pekanbaru sebagai pusat pemerintahan dan Ibukota Provinsi Riau juga menjadi tujuan mobilitas terutama dari kabupaten dan kota di provinsi Riau. Meskipun meningkatnya jumlah kendaraan, kapasitas jalan tetap tidak berubah membuat kepadatan lalu lintas dapat terjadi yang berpotensi meningkatkan jumlah

Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:

- a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
- b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.

2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

pencemaran udara (Sasmita et al., 2022). Selain itu, pesatnya pembangunan Kota Pekanbaru dapat terlihat dari meningkatnya jumlah penduduk, kendaraan di jalan raya, dan jumlah industri (Photon et al., 2014). Studi sebelumnya mengenai hubungan antara kepadatan kendaraan dan polusi udara di Kota Pekanbaru khususnya di jalan Jendral Sudirman menunjukkan bahwa semakin banyak kendaraan di suatu lokasi, kecepatan kendaraan akan semakin menurun sehingga emisi CO yang dihasilkan meningkat. Hal ini menyebabkan kualitas udara di Kota Pekanbaru semakin memburuk (Sasmita et al., 2022).

Kebakaran hutan dan lahan berdampak besar terhadap lingkungan, kesehatan manusia, dan ekonomi masyarakat (Zahtamal et al., 2023). Kebakaran hutan dan lahan dapat meluas ke daerah di luar area yang terbakar karena menghasilkan kabut asap tebal dengan jangkauan yang luas. Sebaran kabut asap ini dapat mencapai daerah perkotaan dan pemukiman, menyebabkan penurunan kualitas udara yang berujung pada pencemaran udara (Putra Mulia et al., 2021). Pada tahun 2019, adanya kebakaran hutan dan lahan di Provinsi Riau memengaruhi kualitas udara di Kota Pekanbaru dengan Indeks Standar Pencemar Udara mencapai level tidak sehat sampai level berbahaya (Zahtamal et al., 2023). Kualitas udara yang buruk akibat kabut asap dapat memicu berbagai penyakit, seperti iritasi mata, batuk, flu, sakit kepala, sesak napas, dan alergi kulit (Putra Mulia et al., 2021). Berdasarkan data pemantauan kualitas udara otomatis yang diterbitkan oleh Kementerian Lingkungan Hidup dan Kehutanan tahun 2019, ditemukan bahwa rata-rata konsentrasi partikel $PM_{2.5}$ di stasiun *Air Quality Monitoring System* (AQMS) Kota Pekanbaru mengalami perbedaan sebelum dan setelah terjadinya kebakaran hutan dan lahan. Sebelum terjadinya kebakaran hutan dan lahan, rata-rata konsentrasi $PM_{2.5}$ adalah sebesar $23,64 \mu g/m^3$. Namun, ketika mempertimbangkan seluruh tahun, termasuk saat terjadinya kebakaran hutan dan lahan, rata-rata tahunan konsentrasi $PM_{2.5}$ mencapai $48,71 \mu g/m^3$.

Analisis polusi udara dianggap sebagai studi yang penting untuk memetakan tingkat polusi di berbagai kota (Maltare & Vahora, 2023). *Machine learning* memungkinkan untuk menghasilkan prakiraan *Air Quality Index* (AQI) yang lebih



Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:

- a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
- b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.

2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

tepat karena semakin banyaknya data historis yang dapat diakses untuk penelitian (Ravindiran et al., 2023). Sejumlah penelitian telah dilakukan dengan tujuan mengembangkan model prediksi yang optimal untuk memprediksi kualitas udara dengan memanfaatkan teknik dalam *machine learning* telah dilakukan oleh (Nababan et al., 2023) Penelitian ini menggunakan XGBoost yang dikombinasikan dengan Synthetic Minority Oversampling Technique (SMOTE). Hasilnya menunjukkan bahwa kombinasi SMOTE dan XGBoost memiliki kinerja yang sangat baik dalam memprediksi kualitas udara. Keduanya mencapai total akurasi sebesar 98,14%, presisi sebesar 78,94%, recall sebesar 79%, F1-score sebesar 98,14%, dan AUC-ROC sebesar 99,48%. Selanjutnya, penelitian yang dilakukan oleh (Putra et al., 2023) menerapkan *Support Vector Machine* (SVM) untuk memprediksi kualitas udara. Hasil penelitian menunjukkan bahwa kernel polinomial memberikan kinerja terbaik, dengan nilai *G-Mean* sebesar 0,98, *Specificity* 0,99, dan *sensitivity* 0,98. Selain itu, presisi, *recall*, *F1-Score*, dan akurasi juga mencapai nilai 0,98, menunjukkan ketepatan dan kehandalan model SVM dalam prediksi kualitas udara. Penelitian lain oleh (Murugan & Palanichamy, 2021) menggunakan *Multi-Layer Perceptron* (MLP) dan Random Forest. Hasil dari penelitian ini adalah Random Forest memberikan akurasi terbaik dalam prediksi Indeks Polusi Udara $PM_{2,5}$ di Malaysia dibandingkan MLP.

Adapun penelitian terkait penggunaan *machine learning* yang digunakan untuk prediksi dengan metode *Categorical Boosting* (CatBoost) juga telah dilakukan oleh (Ananda, 2023) untuk memprediksi gaya hidup berdasarkan *Socioeconomic Status* (SES). Penelitian ini berhasil prediksi kategori SES dengan menggunakan pemodelan algoritma CatBoost dengan hasil akurasi 85,94%. Selain itu, penelitian yang dilakukan oleh (Febriansyah Istianto et al., 2024) yang memprediksi curah hujan menggunakan metode CatBoost. Dari hasil evaluasi menggunakan *confusion matrix*, akurasi prediksi curah hujan cukup tinggi, yaitu mencapai 94,2%, berdasarkan atribut yang digunakan dalam data penelitian tersebut.

Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:

- a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
- b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.

2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

Penelitian mengenai perbandingan algoritma CatBoost dengan algoritma *machine learning* lainnya telah dilakukan oleh (Purbolingga et al., 2023) yang mengimplementasikan metode CatBoost dan XGBoost untuk memprediksi orang dengan risiko penyakit jantung. Hasil penelitian menunjukkan bahwa model CatBoost memperoleh skor yang sangat baik dan mencapai akurasi tertinggi. Setelah melalui proses penyetelan *hyperparameter* dan evaluasi model, terbukti bahwa CatBoost memberikan skor yang lebih tinggi daripada model XGBoost. Selanjutnya oleh (Huang et al., 2019) menggunakan Random Forest, SVM, CatBoost. Hasil penelitian menunjukkan CatBoost mengalami peningkatan signifikan dalam hal akurasi, stabilitas, dan biaya komputasi jika dibandingkan dengan Random Forest. Adapun penelitian yang dilakukan oleh (Ravindiran et al., 2023) yang membandingkan algoritma LightGBM (*Light Gradient Boosting Machine*), Random Forest, CatBoost, AdaBoost, dan XGBoost dalam memprediksi indeks kualitas udara. Hasil penelitian menunjukkan bahwa model CatBoost mengungguli model lainnya dengan koefisien korelasi R² sebesar 0,9998, *Mean Absolute Error* (MAE) sebesar 0,60, *Mean Square Error* (MSE) sebesar 0,58, dan *Root Mean Square Error* (RMSE) sebesar 0,76.

Categorical Boosting (CatBoost) adalah suatu implementasi dari *gradient boosting* yang memanfaatkan pohon keputusan biner sebagai model dasar untuk prediksi (Prokhorenkova et al., 2019). Ini adalah algoritma yang tangguh di antara pohon keputusan yang didorong gradien (GBDT) dan telah menghasilkan hasil yang bermanfaat dalam masalah terkait *machine learning* standar (Safaei et al., 2022). Menurut (Safaei et al., 2022), Model CatBoost dikembangkan untuk mengatasi kelemahan algoritma peningkatan gradien standar, seperti kebocoran target dan pergeseran prediksi. Model CatBoost mengubah algoritma *gradient boosting* lainnya melalui peningkatan yang teratur dan penanganan fitur kategoris yang fleksibel selama fase pelatihan. Hal ini dianggap sebagai salah satu kemajuan CatBoost yang paling berani dibandingkan dengan metode *gradient boosting* biasa (Safaei et al., 2022).

Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:

- a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
- b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.

2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

CatBoost juga telah banyak digunakan dalam prediksi indeks kualitas udara. Studi yang dilakukan oleh (Nurchaerani Kadir, 2024) menunjukkan bahwa dalam memprediksi nilai Air Quality Index (AQI), metode CatBoost memberikan hasil yang lebih baik dalam pemodelan kualitas udara di Kota Palembang dengan kesalahan prediksi yang lebih kecil. Penelitian lain juga dilakukan oleh (Ravindiran et al., 2023) menunjukkan bahwa metode CatBoost memiliki kinerja yang lebih unggul dibandingkan metode lainnya seperti Random Forest, AdaBoost, dan XGBoost dalam memprediksi nilai konsentrasi PM10 dan PM2,5 data kualitas udara. Berdasarkan hasil penelitian yang telah ditemukan, CatBoost telah terbukti menjadi metode yang efektif dalam memprediksi secara numerik. Oleh Karena itu, metode *Categorical Boosting* (CatBoost) digunakan dalam penelitian ini untuk memprediksi kadar polutan indikator kualitas udara karena telah terbukti berhasil dalam berbagai aplikasi prediksi. Melalui penggunaan algoritma ini, diharapkan penelitian ini dapat memberikan prediksi kadar polutan kualitas udara yang lebih akurat dan relevan serta dapat menjadi dasar untuk penelitian serupa.

1.2 Rumusan Masalah

Berdasarkan latar belakang tersebut, rumusan masalah dalam penelitian ini adalah bagaimana penerapan metode CatBoost untuk memprediksi kadar polutan indikator kualitas udara di Pekanbaru?

1.3 Batasan Masalah

Batasan masalah dalam penelitian ini adalah sebagai pedoman yang mempersempit cakupan penelitian ke dalam aspek-aspek khusus yang akan menjadi fokus dalam penyelidikan ini.

1. Penelitian ini berfokus pada prediksi kadar polutan udara yaitu PM₁₀, SO₂, CO, O₃, NO₂, PM_{2.5}.
2. Data yang digunakan terbatas pada wilayah Tenayan Raya dengan menggunakan data histori pola harian kualitas udara tahun 2019-2023 untuk pelatihan model.

Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:

- a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
- b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.

2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

1.4 Tujuan Penelitian

Tujuan penelitian ini adalah menerapkan metode CatBoost untuk memprediksi kadar polutan indikator kualitas udara di Pekanbaru.

1.5 Manfaat Penelitian

Adapun manfaat dari penelitian ini adalah sebagai berikut.

1. Memberikan informasi kepada masyarakat umum mengenai kinerja dari algoritma CatBoost yang diterapkan dalam memprediksi kualitas udara di kota Pekanbaru.
2. Hasil penelitian ini dapat menjadi landasan untuk penelitian selanjutnya di bidang ilmu data.



1. Diarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Diarang mengemukakan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

BAB 2

KAJIAN PUSTAKA

2.1 Kajian Metode

Kajian metode penelitian ini mencakup pemahaman mengenai kualitas udara, data mining prediksi, pengumpulan data, *preprocessing*, pembagian data, pemodelan menggunakan CatBoost, dan evaluasi model.

2.1.1 Kualitas Udara

Kualitas udara merujuk pada kondisi atau tingkat kemurnian udara di suatu area atau lokasi tertentu. Ini menunjukkan sejauh mana udara tersebut bersih atau terkontaminasi oleh berbagai polutan. Kualitas udara di dalam ruangan merupakan faktor penting yang dapat memengaruhi kesehatan tenaga kerja (Santoso et al., 2022). Kualitas udara yang baik menunjukkan bahwa udara memiliki sedikit atau tidak ada kontaminan yang berpotensi merugikan bagi kesehatan manusia dan lingkungan, sementara Kualitas udara yang buruk mencerminkan tingginya tingkat polusi yang dapat membahayakan kesehatan serta keberlanjutan lingkungan. Kualitas udara merujuk pada kondisi udara di suatu daerah yang mencakup berbagai parameter fisik, kimia, dan biologis yang berdampak pada kesehatan manusia, lingkungan, dan ekosistem secara keseluruhan (Ningrum, 2024).

Indeks Standar Pencemar Udara (ISPU) menjadi alat yang penting dalam mengukur tingkat pencemaran udara dan memberikan panduan kepada masyarakat untuk mengambil tindakan yang tepat demi menjaga kualitas udara yang sehat. Di Indonesia, Kementerian Lingkungan Hidup dan Kehutanan (KLHK) telah menerbitkan Peraturan Menteri Lingkungan Hidup dan Kehutanan Nomor 14 Tahun 2020 mengenai Indeks Standar Pencemar Udara (ISPU). Dalam peraturan tersebut, parameter yang digunakan untuk perhitungan ISPU meliputi Partikulat (PM_{10} dan $PM_{2.5}$), Karbon Monoksida (CO), Sulfur Dioksida (SO_2), Nitrogen Dioksida (NO_2), Ozon (O_3), dan Hidrokarbon (HC).

Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:

a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.

b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.

2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

1) *Particulate Matter* 10 (PM₁₀)

PM₁₀ merupakan partikel udara dalam wujud padat yang berdiameter kurang dari 10 µm. Partikel ini akan masuk ke paru-paru dan bertahan di dalam tubuh untuk waktu yang lama, sehingga dapat mengganggu sistem pernapasan manusia (Gunawan et al., 2018). PM₁₀ berasal dari debu jalan, debu konstruksi, pengangkutan material, emisi kendaraan, asap cerobong industri, dan ditambah dengan aktivitas penghancuran dan penggilingan (Jusuf et al., 2023).

2) *Particulate Matter* 2.5 (PM_{2.5})

PM_{2.5} merupakan partikel halus pada komponen udara yang berukuran lebih kecil dari 2,5 µm (mikrometer) dengan nilai ambang batas sebesar 65 µg/m³ yang dapat menimbulkan berbagai penyakit. Partikel yang ada di udara ambien biasanya berukuran 0,1 hingga 50 µm atau lebih (Serlina et al., 2023). PM_{2.5} berasal dari berbagai sumber seperti hasil pembakaran bahan bakar kendaraan bermotor, hingga kebakaran hutan (Ruchjana et al., 2021).

3) Karbon Monoksida (CO)

Karbon Monoksida (CO) adalah gas yang tidak berwarna dan tidak berbau, yang dapat berbahaya jika terhirup dalam jumlah besar. Gas CO dihasilkan dari proses pembakaran. Sumber utama gas CO di udara meliputi mobil, truk, dan kendaraan lainnya, serta mesin yang membakar bahan bakar fosil (Rizaldi et al., 2022).

4) Sulfur Dioksida (SO₂)

SO₂ merupakan polutan yang berasal dari sumber industri yang berperan sebagai prekursor untuk asam sulfat (H₂SO₄) dan merupakan komponen partikel aerosol yang berpengaruh pada deposisi asam, iklim global, serta lapisan ozon (Cahyono, 2011). Gas SO₂ di udara berasal dari berbagai sumber, termasuk pembangkit listrik, aktivitas pembakaran,

Hak Cipta Diindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:

- a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
- b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.

2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

pertambangan, dan pengolahan logam. Selain itu, sumber lokal seperti pemanasan rumah tangga dan distrik, serta sumber bergerak seperti mesin diesel juga menyumbang emisi gas ini (Amalia, 2022).

5) Nitrogen Dioksida (NO_2)

Gas nitrogen dioksida (NO_2) adalah polutan yang terdapat di udara ambien, bersama dengan nitrogen monoksida (NO). Gas-gas ini umumnya dihasilkan oleh aktivitas manusia, seperti pembakaran bahan bakar kendaraan, pembakaran sampah, pembakaran batubara, dan kegiatan industri (Masito, 2018). NO_2 dan nitrogen oksida (NO_x) lainnya juga berperan sebagai prekursor untuk sejumlah zat pencemar udara sekunder yang berbahaya, termasuk ozon (O_3) dan partikulat (PM) (Safira et al., 2021).

6) Ozon (O_3)

Ozon adalah molekul anorganik radikal yang terdiri dari tiga atom oksigen dan memiliki sifat sebagai oksidator yang kuat. Secara alami, ozon terbentuk dari molekul oksigen (O_2) di atmosfer bumi melalui interaksi dengan sinar ultraviolet atau aktivitas elektrik yang terjadi di atmosfer (Astriyani et al., 2023). Di atmosfer, pembentukan ozon berasal dari nitrogen oksida dan gas organik yang dihasilkan oleh emisi dari kendaraan dan industri (Lestari, 2010).

7) Hidrokarbon (HC)

Emisi Hidrokarbon (HC) yang dihasilkan oleh berbagai jenis mesin merupakan salah satu sumber pencemar udara. Hal ini disebabkan oleh pembakaran bahan bakar yang tidak sempurna dan ketidakcukupan pembakaran minyak pelumas silinder (Yusrianti, 2015). Sifat toksin hidrokarbon (HC) akan meningkat ketika berupa bahan pencemar dalam bentuk gas, cair, dan padat. Ini disebabkan oleh fakta bahwa padatan HC

Hak Cipta Diindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:

a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.

b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.

2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

Tahap ini mencakup pengidentifikasian dan penanganan terhadap data yang hilang (null), outlier, serta data yang tidak valid (Agung et al., 2023).

a. Penanganan *Missing Value*

Berikut ini adalah beberapa metode untuk menangani nilai yang hilang (Joshi & Patel, 2021).

- Menghilangkan baris dengan nilai yang hilang merupakan preferensi yang umum untuk data berukuran kecil, terutama jika persentase nilai yang hilang tidak melebihi 5%.
- Mengisi nilai yang hilang secara manual.
- Setiap nilai yang hilang digantikan dengan nilai yang tetap.
- Ukuran tendensi sentral (*mean, median, modus*) untuk data numerik.
- Mengganti nilai yang hilang dengan rata-rata berdasarkan kelas. Nilai yang paling mungkin dapat ditentukan menggunakan algoritma *machine learning* seperti *regresi linier* atau pohon keputusan.

b. Pemrosesan *Outlier*

Outlier adalah pola dalam data yang tidak sejalan dengan perilaku normal yang telah didefinisikan dengan jelas (Singh & Upadhyaya, 2012). Outlier dapat dihapus atau dikelola menggunakan metode statistik seperti trimming atau winsorizing (Agung et al., 2023). Teknik trimming yaitu menghapus data yang dianggap outlier pada persentil tertentu (Sihombing et al., 2023). Sedangkan, teknik winsorizing menggantikan setiap nilai variabel yang berada di atas atau di bawah persentil ke-k dengan nilai persentil ke-k itu sendiri (Dash et al., 2023). Sebelum dilakukan penanganan outlier, penting untuk mengetahui adanya outlier dalam dataset. Beberapa metode umum untuk mendeteksi outlier dalam observasi deret waktu univariat (Yaro et al., 2024).

- *Z-Score*

Hak Cipta Diindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:

- a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
- b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.

2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

Z-score adalah sistem yang merepresentasikan item perilaku abnormal dalam kaitannya dengan simpangan baku dan rata-rata kumpulan argumen (Anusha et al., 2019). Z-Score menggunakan nilai rata-rata dan standar deviasi untuk mendeteksi outlier dalam kumpulan data (Chikodili et al., 2021). Rentang Z-score dari -3 deviasi standar hingga +3 deviasi standar mencakup area dari sisi paling kiri hingga sisi paling kanan pada kurva distribusi normal standar. Biasanya nilai ± 3 ini digunakan sebagai ambang batas kritis dalam mendeteksi outlier pada data univariat (Yoseph et al., 2019).

- *Interquartile Range*

Interquartile Range (IQR) didefinisikan sebagai perbedaan antara persentil ke-25 dan ke-75 (juga disebut kuartil pertama dan ketiga). Oleh karena itu, jangkauan interkuartil menggambarkan 50% tengah pengamatan. Jika jangkauan interkuartil besar, itu berarti bahwa 50% tengah pengamatan berjarak lebar. Keuntungan penting dari jangkauan interkuartil adalah dapat digunakan sebagai ukuran variabilitas jika nilai ekstrem tidak dicatat secara tepat (Manikandan, 2011).

2) Integrasi Data

Integrasi data adalah proses yang menggabungkan dan menyatukan data. Proses ini dilakukan dengan mengumpulkan data dari berbagai sumber dan menyatukannya ke dalam penyimpanan data yang konsisten (Srivastava, 2014).

3) Transformasi Data

Transformasi data adalah proses yang melibatkan perubahan di mana atribut numerik diubah ke dalam rentang yang lebih kecil, seperti -1,0 hingga 1,0, atau 0,0 hingga 1,0 (Ulandari & Swanjaya, 2020). Normalisasi dan standarisasi adalah teknik yang sering digunakan untuk mengubah skala atribut (Agung et al., 2023).

Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:

- a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
- b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.

2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

4) Reduksi Data

Reduksi data merupakan teknik yang memampatkan data sedemikian rupa sehingga makna dari data tersebut tidak hilang (Joshi & Patel, 2021). Tujuannya adalah untuk menghilangkan redundansi, menghilangkan *noise*, dan memperbaiki efisiensi dalam pengolahan dan penyimpanan data.

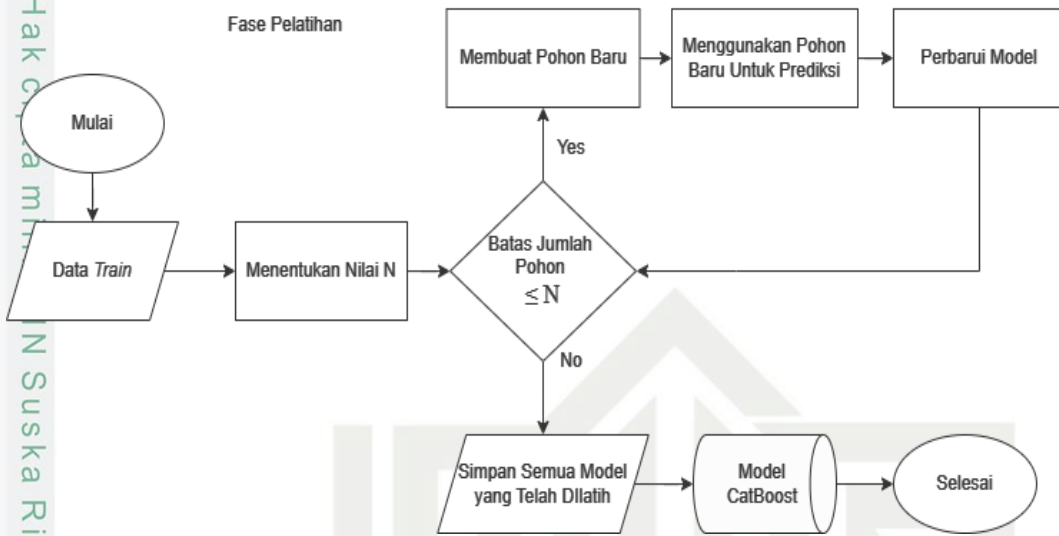
2.1.2.3 Pemodelan Menggunakan CatBoost

Pemodelan data adalah tahap dalam proses data mining yang menggunakan teknik dan algoritma untuk menciptakan model atau pola dari data yang telah dipilih dan diproses sebelumnya (Amalia Yunia Rahmawati, 2020). Sebelum melakukan pemodelan data, dataset dipisah menjadi dua bagian utama, yaitu data pelatihan dan data uji. Pembagian jumlah data latih dan data uji merupakan salah satu faktor yang memengaruhi akurasi, sehingga kesalahan dalam menentukan proporsi kedua jenis data tersebut dapat berdampak pada nilai akurasi dan presisi yang diperoleh (Baiq Nurul Azmi et al., 2023). Data *train* adalah data yang digunakan untuk melatih model CatBoost. Data *test* digunakan setelah proses pelatihan model selesai. Pembagian data yang sering dilakukan adalah 80:20, di mana 80% data digunakan untuk pelatihan dan 20% untuk pengujian. Rasio lain seperti 90:10, 70:30, 60:40, dan 50:50 juga biasa diterapkan dalam praktik (Joseph, 2022).

Pemodelan menggunakan CatBoost melibatkan penerapan algoritma *machine learning* yang merupakan bagian dari keluarga *Gradient Boosted Decision Trees* (GBDT) yang termasuk dalam teknik *ensemble learning* (Christian et al., 2022). Adapun tahapan pelatihan CatBoost dapat dilihat pada Gambar 1 (Nasution et al., 2023).

Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.



Gambar 1 Flowchart Pelatihan CatBoost

Prosesnya dimulai dengan menentukan nilai N yang mewakili jumlah iterasi atau jumlah pohon yang dilatih pada model CatBoost. Setiap pohon dibangun dengan pengurangan loss (kerugian) yang dibandingkan dengan pohon sebelumnya (Dewi, 2021). Adapun parameter-parameter yang digunakan untuk mengoptimalkan dan mengatur kinerja CatBoost agar sesuai dengan kebutuhan spesifik dapat dilihat pada Tabel 1 (Christian et al., 2022).

Tabel 1 Parameter untuk Model CatBoost

| Parameter | Keterangan |
|---------------|--------------------------------|
| Iterations | Jumlah pohon yang dapat dibuat |
| Depth | Jumlah kedalaman pohon |
| Learning Rate | Besaran tingkat pembelajaran |

Jika jumlah pohon yang diperbolehkan belum tercapai, model CatBoost akan terus menambahkan pohon baru pada setiap iterasi untuk meminimalkan kesalahan prediksi. Hal ini merupakan bagian dari proses *gradient boosting* yang memungkinkan model CatBoost untuk secara bertahap meningkatkan kinerjanya (Dewi, 2021). Tidak seperti algoritma GBDT lainnya, CatBoost menggunakan pohon simetri atau biasa disebut juga Oblivious Decision Tree (ODT) seimbang dan tidak terlalu rentan terhadap *overfitting*. Pada seluruh tingkat pohon

Hak Cipta Dilindungi Undang-Undang

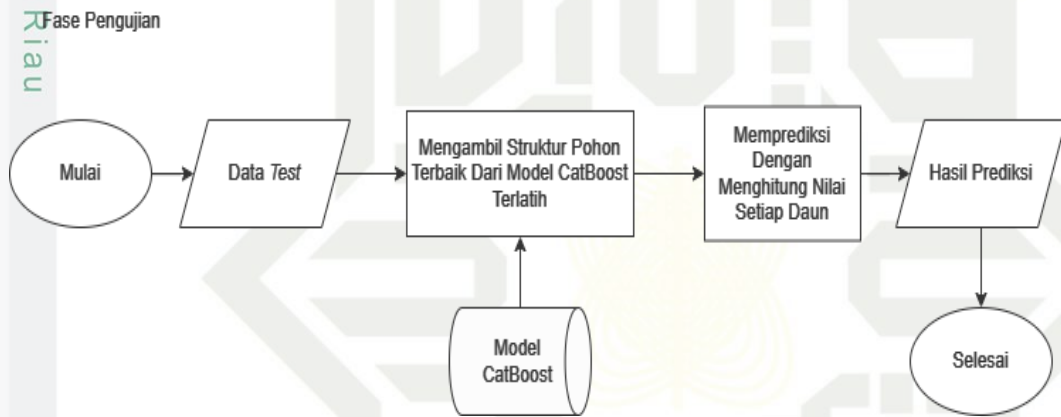
1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:

- a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
- b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.

2. Dilarang mengemukakan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

digunakan kriteria pemisahan yang sama, menggunakan fitur yang sama untuk membagi sampel menjadi partisi kiri dan kanan pada setiap tingkat pohon (Dewi, 2021). Struktur pohon yang digunakan di CatBoost juga berfungsi sebagai regularisasi, sehingga dapat memberikan manfaat yang berkualitas untuk banyak tugas (Ananda, 2023).

Setelah pelatihan CatBoost selesai, model akan dilanjutkan ke tahap pengujian. Langkah-langkah yang terlibat dalam proses pengujian CatBoost dapat dilihat pada Gambar 2 (Nasution et al., 2023).



Gambar 2 Flowchart Pengujian CatBoost

Selama tahap pengujian model, hasil prediksi akhir dihasilkan menggunakan pohon keputusan final yang merupakan hasil dari proses pelatihan model dengan parameter terbaik. Pohon keputusan final ini digunakan untuk memperoleh nilai target prediksi untuk setiap data uji. Selama tahap pengujian model, parameter diperoleh dari struktur pohon terbaik selama pelatihan digunakan oleh model CatBoost. Prediksi dibuat dengan menelusuri pohon dan menghitung jumlah daun (Nasution et al., 2023).

2.1.2.4 Evaluasi Model

Tahapan evaluasi hasil adalah langkah dalam proses data mining yang dilakukan untuk menilai kualitas dan akurasi model yang telah dikembangkan. (Amalia Yunia Rahmawati, 2020). Selain itu, evaluasi membantu dalam mengidentifikasi dan memperbaiki masalah yang mungkin muncul dalam model



Hak Cipta Diindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:

- a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
- b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.

2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

prediksi. Beberapa metrik yang dapat digunakan untuk mengevaluasi kinerja model prediksi sebagai berikut.

1) *Root Mean Squared Error* (RMSE)

Root Mean Squared Error (RMSE) adalah metrik yang umum digunakan untuk mengukur besaran rata-rata kesalahan antara nilai prediksi dan nilai observasi. RMSE dapat dihitung dengan Persamaan (1).

$$\text{RMSE} = \sqrt{\frac{1}{n} \sum_{i=1}^n y_i - \hat{y}_i^2} \quad (1)$$

Keterangan :

n : jumlah sampel atau observasi.

y_i : nilai sebenarnya dari observasi ke- i .

\hat{y}_i : nilai prediksi dari observasi ke- i .

2) *Mean Absolute Error* (MAE)

Dalam *Mean Absolute Error* (MAE), kesalahan yang berbeda tidak diberi bobot lebih atau kurang, tetapi skor meningkat secara linier dengan peningkatan kesalahan (Schneider & Xhafa, 2022). MAE berfungsi sebagai metrik untuk membandingkan kesalahan prediksi absolut. Ini lebih kuat terhadap outlier dibandingkan metrik lain seperti MSE atau RMSE (Lauble et al., 2023). Skor MAE dihitung dengan menghitung rata-rata dari selisih absolut antara nilai prediksi dan nilai sebenarnya. Nilai MAE dapat dihitung dengan Persamaan (2) sebagai berikut.

$$\text{MAE} = \frac{1}{n} \sum_{i=1}^n |y_i - \hat{y}_i| \quad (2)$$

Keterangan :

n : jumlah sampel atau observasi.

Hak Cipta Dilindungi Undang-Undang

1. Diarangi mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

y_i : nilai sebenarnya dari observasi ke-i.

\hat{y}_i : nilai prediksi dari observasi ke-i.

2.2 Penelitian Terkait

Dalam suatu penelitian, penting untuk memanfaatkan temuan dari penelitian sebelumnya yang relevan. Ini merujuk kepada hasil-hasil penelitian terdahulu yang berkaitan dengan topik penelitian serupa. Adapun temuan dan pembahasan penelitian sebelumnya yang relevan dengan penelitian ini dapat dilihat pada Tabel 2 sebagai berikut.

Tabel 2 Penelitian mengenai kasus pencemaran udara

| No | Penulis (Tahun) | Judul | Hasil Penelitian |
|----|-----------------------|--|---|
| 1 | (Keller et al., 2023) | <i>Air pollution impacts on in-hospital case-fatality rate of ischemic stroke patients</i> | Peningkatan karbon Konsentrasi monoksida, nitrogen dioksida, PM10, dan SO2 tidak berhubungan signifikan dengan kematian akibat stroke. Namun, konsentrasi SO2 secara signifikan berhubungan dengan tingkat kematian kasus stroke >8%, tidak bergantung pada tipe area tempat tinggal dan penggunaan area (OR 1,518 [95%CI 1,012–2,278], P = 0,044). |
| 2 | (SHW, 2019) | Analisis Pengaruh Faktor Meteorologi dan Kebakaran Hutan dan Lahan terhadap Penyakit Infeksi Saluran | Hasil penelitian ini menunjukkan bahwa pada bulan Maret 2014, yang merupakan bulan dengan jumlah hotspot terbanyak, arah angin bertiup dari lokasi kebakaran hutan dan lahan (Kabupaten Pelalawan dan Kabupaten Siak) menuju Kota Pekanbaru. Pada tahun 2014, kondisi kebakaran hutan |

Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

| | | |
|--|---|--|
| | <p>Pernapasan Akut (ISPA) di Kota Pekanbaru</p> | <p>dan lahan berdampak pada penurunan kualitas udara di Kota Pekanbaru. Kejadian infeksi saluran pernapasan akut (ISPA) pada tahun 2014 dipengaruhi oleh PM10 dan faktor curah hujan sebesar 54%. Sementara itu, pada tahun 2017, diasumsikan bahwa kebakaran hutan dan lahan tidak berdampak pada kualitas udara. Kejadian ISPA tahun 2017 dipengaruhi oleh PM10, curah hujan, dan suhu udara dengan kontribusi sebesar 25%. Penyebab utama terjadinya ISPA di Kota Pekanbaru pada tahun 2014 adalah kualitas udara yang buruk akibat kebakaran hutan dan lahan, sedangkan pada tahun 2017, penyebab utama kejadian ISPA adalah faktor meteorologi yang didukung oleh faktor lainnya.</p> |
|--|---|--|

Penelitian terkait prediksi kualitas udara menggunakan algoritma *machine learning* dapat dilihat pada Tabel 3.

Tabel 3 Penelitian prediksi kualitas udara menggunakan *machine learning*

| No | Penulis (Tahun) | Judul | Metode | Hasil Penelitian |
|----|------------------------|-------------------------------------|---------|--|
| 1 | (Nababan et al., 2023) | Prediksi Kualitas Udara Menggunakan | XGBoost | Hasil penelitian menggunakan 6 atribut yaitu PM10, SO2, CO, O3, NO2, |

Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

| | | | | |
|---|----------------------|---|-------------------------------------|--|
| | | Xgboost Dengan <i>Synthetic Minority Oversampling Technique</i> (Smote) Berdasarkan Indeks Standar Pencemaran Udara (ISPU) | | dan kategori menunjukkan bahwa SMOTE dan XGBoost memiliki kinerja yang lebih baik dalam memprediksi kualitas udara, dengan total akurasi sebesar 98,14%, presisi total sebesar 78,94%, recall sebesar 79%, F1-score sebesar 98,14%, dan nilai AUC-ROC sebesar 99,48%. |
| 2 | (Putra et al., 2023) | Klasifikasi prediksi kualitas udara Menggunakan metode <i>Support Vector Machine</i> (SVM) | <i>Support Vector Machine</i> (SVM) | Hasil penelitian ini menunjukkan bahwa kernel terbaik untuk sistem ini adalah kernel polinomial, dengan nilai G-Mean sebesar 0,98, Specificity 0,99, dan Sensitivity 0,98. Selain itu, nilai Presisi 0,98, Recall 0,98, F1-Score 0,98, serta akurasi sebesar 98% berdasarkan lima jenis zat. (PM10, SO2, CO, O3, dan NO2). |
| 3 | (Sang et al., 2021) | Prediksi Kualitas Udara Menggunakan Algoritma <i>K-Nearest Neighbor</i> | <i>K-Nearest Neighbor</i> | Hasil penelitian yang menggunakan parameter PM10, PM2.5, SO2, CO, O3, NO2, dan MAX menunjukkan bahwa saat model algoritma dievaluasi |

Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

| | | | |
|--|--|--|---|
| | | | dengan menggunakan confusion matrix, nilai K=7 memberikan kinerja terbaik, dengan akurasi mencapai 96%. Selain itu, nilai presisi sebesar 92%, recall 95%, dan f-measure 93%. |
|--|--|--|---|

Penelitian terkait perbandingan penggunaan metode CatBoost dengan metode lainnya dapat dilihat pada Tabel 4.

Tabel 4 Penelitian perbandingan metode CatBoost dengan metode lainnya

| No. | Penulis (Tahun) | Judul | Metode | Hasil Penelitian |
|-----|------------------------|--|---|---|
| 1 | (Ibrahim et al., 2020) | <i>Comparison of the CatBoost Classifier with other Machine Learning Methods</i> | Binary logistic regression, Random forest, Adaptive Boosting, Decision trees, Neural networks, gradient boost, XGBoost and Catboost | CatBoost dan XGBoost mencapai skor akurasi tertinggi dengan 94%. CatBoost berhasil mencapai skor (akurasi) tertinggi di kedua aplikasi (atau analisis). Metrik evaluasi lainnya juga mendukung kinerja algoritma ini. |

Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

| | | | | |
|---|---------------------------|--|--|--|
| 2 | (Banerjee et al., 2023) | <i>Prediction of transport behavior of nanoparticles using machine learning algorithm: Physical significance of important features</i> | Random forest, XGBoost, dan CatBoost | Kinerja terbaik dalam memprediksi Fraksi Retensi (RF) untuk seluruh rentang NP ($R^2 > 0,89$ dan $MSE < 0,007$) serta untuk masing-masing NP. |
| 3 | (Ravindiran et al., 2023) | <i>Air quality prediction by machine learning models: A predictive study on the Indian Coastal City of Visakhapatnam</i> | LightGBM (Light Gradient Boosting Machine), Random Forest, Catboost, Adaboost, dan XGboost | Hasil penelitian dengan menggunakan parameter CO, NO, NH3, NO2, NOX, dan SOX menunjukkan bahwa model CatBoost mengungguli model lainnya dengan koefisien korelasi R^2 sebesar 0,9998, MAE sebesar 0,60, MSE sebesar 0,58, dan RMSE sebesar 0,76. |
| 4 | (Nasional et al., 2024) | Perbandingan Performa Algoritma Metode Bagging dan Boosting pada Prediksi Konsentrasi PM10 di Jakarta Utara | Random Forest, CatBoost, XGBoost | Berdasarkan hasil pemodelan, algoritma Random Forest menghasilkan akurasi prediksi konsentrasi PM10 yang lebih tinggi ($R^2 = 0,6424$) dibandingkan XGBoost |

Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

| | | | |
|--|--|--|--|
| | | | ($R^2 = 0,6340$) dan CatBoost ($R^2 = 0,6294$) |
|--|--|--|--|

Penelitian terkait penggunaan metode CatBoost untuk kasus prediksi dapat dilihat pada Tabel 5.

Tabel 5 Penelitian terkait penggunaan CatBoost untuk prediksi

| No | Penulis (Tahun) | Judul | Hasil Penelitian |
|----|--------------------------|--|---|
| 1 | (Ananda, 2023) | <i>Machine learning</i> untuk prediksi gaya hidup berdasarkan <i>socioeconomic status</i> (SES) menggunakan algoritma CatBoost | Penelitian ini berhasil prediksi kategori SES dengan menggunakan pemodelan algoritma CatBoost dengan hasil akurasi 85,94% |
| 2 | (Christian et al., 2022) | Implementasi Penggunaan Algoritma <i>Categorical Boosting</i> (CatBoost) Dengan Optimisasi Hiperparameter Dalam Memprediksi Pembatalan Pesanan Kamar Hotel | Hasil evaluasi yang diperoleh dengan menggunakan algoritma CatBoost dan optimisasi hiperparameter menunjukkan performa yang lebih baik dibandingkan dengan algoritma yang digunakan dalam penelitian sebelumnya, yaitu Random Forest dengan optimisasi hiperparameter. Hiperparameter ini mampu menghasilkan model klasifikasi dengan akurasi yang dibulatkan sebesar 0,88 atau setara dengan 88%, serta nilai presisi sebesar 0,86 |

Hak Cipta Diindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

| | | | |
|---|-------------------------------------|--|--|
| | | | atau 86%. |
| 3 | (Febriansyah Istianto et al., 2024) | Prediksi Curah Hujan Menggunakan Metode <i>Categorical Boosting</i> (CatBoost) | Berdasarkan penelitian tersebut, evaluasi yang dilakukan dengan menggunakan <i>confusion matrix</i> menunjukkan bahwa model ini memiliki akurasi yang cukup tinggi dalam memprediksi curah hujan, dengan nilai akurasi mencapai 94,2%. |

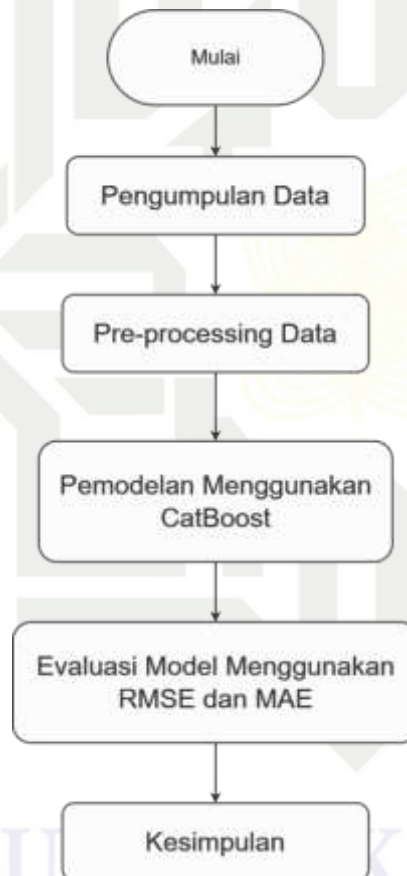
Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

BAB 3 METODOLOGI PENELITIAN

3.1 Tahapan Penelitian

Tahapan penelitian membantu untuk merencanakan, melaksanakan, dan mengevaluasi proyek prediksi dengan metode yang terstruktur. Berikut langkah-langkah dalam penelitian dapat dilihat pada Gambar 3.



Gambar 3 Tahapan Penelitian

3.1.1 Pengumpulan Data

Tahapan pengumpulan data dilakukan untuk mengumpulkan informasi yang diperlukan dalam penelitian. Informasi yang menjadi dasar penelitian ini bersumber dari studi literatur yang mencakup jurnal ilmiah, artikel, dan laporan



Hak Cipta Diindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:

- a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
- b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.

2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

penelitian terkait prediksi kualitas udara, *machine learning*, dan algoritma CatBoost yang akan digunakan untuk membangun dasar pengetahuan yang kuat dalam penelitian ini. Dataset yang digunakan dalam penelitian ini terdiri dari data kualitas udara Kota Pekanbaru, khususnya di wilayah Tenayan Raya yang diperoleh dari Dinas Lingkungan Hidup dan Kebersihan Kota Pekanbaru. Data harian kualitas udara untuk periode 1 Januari 2019 hingga 31 Desember 2023 digunakan dalam tahap pelatihan dan pengujian model, dengan parameter polutan utama seperti PM10, SO₂, CO, O₃, NO₂, dan PM_{2.5}. Selanjutnya, data harian dari 1 Januari hingga 18 Oktober 2024 digunakan dalam tahap implementasi model yang berfungsi sebagai acuan untuk membandingkan hasil prediksi model dengan kondisi aktual kualitas udara di tahun 2024.

3.1.2 Data Mining untuk Prediksi

Implementasi data mining untuk prediksi konsentrasi polutan udara melibatkan serangkaian proses yang sistematis dan terstruktur. Proses ini mencakup beberapa langkah, mulai dari pre-processing data, penerapan model prediksi yang dihasilkan, hingga evaluasi model.

3.1.2.1 Pre-Processing

Preprocessing merupakan tahap yang krusial dalam penelitian untuk memastikan bahwa data telah diproses dan siap digunakan dalam analisis. Pada tahap ini, dilakukan proses pembersihan data untuk mengatasi masalah nilai yang hilang dan *outlier* dalam dataset.

1) Penanganan *Missing Values*

Untuk mengatasi masalah ini, akan digunakan tiga metode imputasi, yaitu *mean*, *median*, dan interpolasi linear. Metode-metode ini dipilih berdasarkan referensi penelitian yang relevan, seperti penelitian (Noor et al., 2015) dan (Febiani et al., 2022) yang telah membuktikan efektivitas metode-metode tersebut dalam mengatasi masalah nilai yang hilang dalam data.

- a. Mean

Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

Imputasi rata-rata adalah metode yang mengisi nilai yang hilang dengan nilai rata-rata dari dataset. Rumus matematis untuk imputasi mean menggunakan Persamaan (3).

$$Mean = \frac{1}{n} \sum_{i=1}^{i=n} y_i \quad (3)$$

Dimana n adalah jumlah data yang tersedia dan y_i adalah jumlah nilai nilai dalam dataset.

b. Median

Median imputasi adalah metode statistik yang digunakan untuk memperkirakan nilai yang hilang dalam suatu dataset dengan menggantinya dengan nilai tengah suatu kumpulan data jika disusun dalam urutan menaik. Jika data dengan jumlah ganjil menggunakan Persamaan (4).

$$Median = \frac{(n+1)}{2} \quad (4)$$

Di mana n adalah jumlah nilai yang tidak hilang dalam kumpulan data.

c. Interpolasi Linear

Interpolasi linier digunakan untuk memperkirakan nilai yang tidak diketahui dalam rentang titik data yang diketahui. Ini menggunakan titik data yang ada untuk menyimpulkan dan mengisi nilai yang hilang atau tidak diketahui dalam kumpulan data. Rumus matematis untuk menghitung nilai yang hilang dapat menggunakan Persamaan (5).

$$f(x) = f(x_0) + \frac{f(x_1) - f(x_0)}{(x_1 - x_0)(x - x_0)} (x - x_0) \quad (5)$$

Keterangan

$f(x_0)$: Nilai fungsi pada titik x_0 .

Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

- $f(x_1)$: Nilai fungsi pada titik x_1 .
- x_0 dan x_1 : Dua titik data yang diketahui, dimana $x_0 < x_1$.
- x : Titik dimana nilai $f(x)$ akan dihitung.

Dimana x adalah variabel bebas, x_1 dan x_0 adalah nilai variabel bebas yang diketahui dan $f(x)$ adalah nilai variabel terikat untuk nilai x dari variabel bebas.

2) Pemrosesan *Outlier*

Tahap ini menggunakan metode IQR untuk mendeteksi *outlier* karena kemudahan implementasinya, keandalannya, dan ketahanannya terhadap outlier. Deteksi outlier menggunakan IQR dimulai dengan menghitung kuartil pertama (Q1) dan kuartil ketiga (Q3), Untuk menghitung IQR dapat dilakukan dengan Persamaan (6).

$$IQR = Q3 - Q1 \quad (6)$$

Keterangan :

Q1 : kuartil pertama (nilai persentil ke-25 dari data).

Q3 : kuartil ketiga (nilai persentil ke-75 dari data).

Nilai Q1 adalah nilai yang memisahkan 25% data terendah dari sisa data. Q3 adalah nilai yang memisahkan 75% data terendah dari 25% data tertinggi. langkah selanjutnya adalah menentukan batas atas dan batas bawah untuk mengidentifikasi nilai yang dianggap sebagai outlier dalam dataset. Untuk menghitung batas bawah dapat dihitung dengan Persamaan (7) sebagai berikut.

$$Batas\ Bawah = Q1 - 1.5 \times IQR \quad (7)$$

Sedangkan untuk menghitung batas atas menggunakan Persamaan (8) sebagai berikut.

Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

$$\text{Batas Atas} = Q3 + 1.5 \times IQR \quad (8)$$

Nilai yang berada di luar batas-batas ini akan dianggap sebagai outlier. Setelah outlier terdeteksi, teknik winsorizing digunakan untuk mengurangi efek outlier dalam set data. Semua nilai yang kurang dari Q1 dan lebih dari Q3 diganti menjadi Q1 dan Q3 secara berurutan.

3.1.2.2 Pemodelan CatBoost

Langkah berikutnya adalah tahap pemodelan. Model yang digunakan dalam penelitian ini adalah CatBoost. Untuk melakukan pemodelan, dilakukan pembagian data menggunakan teknik split antara data yang akan dilatih dan data yang akan diuji. Dalam penelitian ini, dataset akan dibagi dengan beberapa rasio berdasarkan penelitian (Aisyah et al., 2021). Rasio pembagian data pelatihan dan data uji dapat dilihat pada Tabel 6.

Tabel 6 Ratio Pembagian Data

| Data Latih | Data Uji |
|------------|----------|
| 90% | 10% |
| 80% | 20% |
| 70% | 30% |

Skenario ini dipilih untuk menilai kemampuan model CatBoost dalam mengidentifikasi pola pada data pelatihan dan menghasilkan hasil yang baik ketika diuji pada data pengujian. Setelah dilakukan pembagian data, tahapan pelatihan untuk model CatBoost dimulai. Pencarian parameter terbaik dilakukan dengan menggunakan GridSearchCV dengan cara mengulangi proses uji parameter untuk setiap parameter dalam daftar. Pada setiap iterasi, model akan dilatih menggunakan parameter saat ini. Beberapa *hyperparameter* yang diuji berdasarkan dari penelitian (Yang et al., 2024). Berikut adalah daftar parameter yang akan diuji dalam pemodelan menggunakan CatBoost ditunjukkan pada Tabel

Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

Tabel 7 Daftar Hyperparameter yang digunakan

| Parameter | Nilai |
|---------------|-----------------|
| iterations | 500, 1000 |
| depth | 6, 8, 10 |
| learning_rate | 0.01, 0.05, 0.1 |

1) Pelatihan Model

Model CatBoost memulai iterasi dengan mempertimbangkan prediksi sebelumnya serta memperhatikan informasi fitur yang tersedia. Dengan demikian, model secara adaptif memperbaiki prediksi sepanjang proses. Prediksi awal dapat ditetapkan sebagai nilai rata-rata dari target variabel untuk set pelatihan. Kemudian, pada setiap iterasi CatBoost, model akan memprediksi nilai *residual* dari selisih antara nilai target sebenarnya dan prediksi sebelumnya. Secara matematis, persamaan dapat ditulis dengan Persamaan (9).

$$r_i^{(t)} = y_i - \hat{y}_i^{(t-1)} \quad (9)$$

Keterangan:

$r_i^{(t)}$: Selisih antara nilai sebenarnya dan nilai prediksi pada iterasi ke- t .

y_i : Nilai sebenarnya untuk observasi ke- i .

$\hat{y}_i^{(t-1)}$: Nilai prediksi model pada iterasi sebelumnya.

Setelah residual dihitung, model kemudian membangun pohon baru dengan mengevaluasi kandidat split di setiap node. Split dipilih berdasarkan penurunan error atau gradien. Pemilihan struktur pohon terbaik dilakukan dengan perhitungan melalui pemisahan (split) yang berbeda, kemudian pohon dibangun dengan menggunakan split tersebut, menetapkan nilai pada daun yang diperoleh, menskor pohon, dan memilih

Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:

- a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
- b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.

2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

split terbaik. Selanjutnya, selama proses iterasi, model akan memperbaiki prediksi sebelumnya dengan menambahkan *residual* yang diprediksi pada iterasi tersebut. Proses ini akan berlanjut hingga jumlah iterasi maksimum tercapai.

2) Pengujian Model

Setelah model CatBoost dilatih menggunakan parameter uji dan menghasilkan prediksi pada setiap iterasi, langkah selanjutnya adalah menguji model menggunakan data yang belum pernah dilihat sebelumnya yaitu data *test*. Prediksi akhir merupakan kombinasi dari kontribusi setiap iterasi, di mana setiap kontribusi tersebut dikalikan dengan *learning rate*. Secara matematis, untuk menghitung prediksi akhir dapat ditulis dengan Persamaan (10).

$$\text{Prediksi Akhir} = \eta \times \sum_{i=1}^n \text{Output Leaf} \quad (10)$$

Keterangan:

η : Learning Rate

Output Leaf : Output dari daun pada pohon keputusan

3.1.2.3 Evaluasi Model

Tahapan evaluasi memastikan bahwa model yang digunakan dalam penelitian ini memiliki kinerja yang baik. Kinerja model dalam memprediksi kualitas udara diukur menggunakan metrik evaluasi RMSE dan MAE dengan Persamaan (1) dan (2) untuk menilai sejauh mana prediksi model mendekati nilai sebenarnya.

3.1.3 Skenario Uji

Skenario uji ini bertujuan untuk mengevaluasi kinerja CatBoost dalam memprediksi kualitas udara di Pekanbaru. Eksperimen ini akan melibatkan iterasi melalui beberapa model dengan hyperparameter yang berbeda, skema pembagian dan preprocessing data untuk mengoptimalkan akurasi prediksi. Skema skenario uji yang digunakan pada penelitian ini dapat dilihat pada Tabel 8.

Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

Tabel 8 Skema Skenario Uji yang Digunakan

| Tahapan | Metode | Parameter yang diuji |
|-------------------------------|-------------------------|---|
| <i>Preprocessing</i> | Imputasi | Mean |
| | | Median |
| | | Linear Interpolasi |
| Pembagian Data | <i>Train-Test Split</i> | 90:10 |
| | | 80:20 |
| | | 70:30 |
| <i>Hyperparameter Turning</i> | <i>Grid Search</i> | iterations : 500, 1000 depth : 6, 8, 10 learning_rate : 0.01, 0.05, 0.1 |
| Evaluasi | Metrik Evaluasi | RMSE |
| | | MAE |

3.1.4 Prediksi Kadar Polutan Tahun 2024

Model yang dilatih dengan data historis sebelumnya, digunakan untuk memprediksi kadar polutan kualitas udara untuk periode 1 Januari hingga 18 Oktober 2024. Hasil prediksi yang diperoleh nantinya akan dibandingkan dengan data aktual yang tersedia pada periode tersebut. Dengan membandingkan hasil prediksi dengan data aktual, evaluasi juga dilakukan untuk mengukur keakuratan model menggunakan RMSE dan MAE.

Hak Cipta Diindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:

a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.

b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.

2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

BAB 5

PENUTUP

5.1 Kesimpulan

Berdasarkan penelitian yang telah dilakukan, dapat diambil kesimpulan sebagai berikut.

1. Model CatBoost yang digunakan tidak berhasil menangkap pola dan tren data dengan baik. Hal ini terlihat dari perbedaan yang signifikan antara nilai aktual dan nilai prediksi yang dihasilkan model. Perbedaan besar tersebut disebabkan oleh fluktuasi data sepanjang tahun yang tidak berhasil ditangkap dengan baik oleh model.
2. Penggunaan metode preprocessing Median dan split data 70:30 atau 80:20 memberikan hasil evaluasi terbaik dengan nilai RMSE dan MAE terendah untuk sebagian besar fitur menunjukkan bahwa kombinasi ini cukup efektif dalam meningkatkan akurasi prediksi model CatBoost. Hal ini menunjukkan bahwa pemilihan metode preprocessing dan split data juga berperan penting dalam menentukan hasil prediksi pada model CatBoost.

5.2 Saran

Penyelidikan lebih lanjut diperlukan untuk mengidentifikasi akar penyebab buruknya kinerja model dan untuk mengeksplorasi pendekatan alternatif yang dapat menangkap pola dan tren mendasar dalam data dengan lebih baik. Berikut adalah beberapa saran yang dapat dipertimbangkan untuk penelitian selanjutnya.

1. Diperlukan upaya intensif dalam pembersihan, pengolahan, dan validasi data untuk memastikan kebersihan dan konsistensi data yang digunakan.
2. Eksplorasi lebih lanjut terhadap parameter pengujian untuk menemukan kombinasi yang optimal dengan karakteristik data yang spesifik.



Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:

- a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
- b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.

2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

DAFTAR PUSTAKA

- Agung, A., Daniswara, A., Kadek, I., & Nuryana, D. (2023). Data Preprocessing Pola Pada Penilaian Mahasiswa Program Profesi Guru. *Journal of Informatics and Computer Science*, 05, 97–100.
- Aisyah, S., Wahyuningsih, S., & Amijaya, F. (2021). Peramalan Jumlah Titik Panas Provinsi Kalimantan Timur Menggunakan Metode Radial Basis Function Neural Network. *Jambura Journal of Probability and Statistics*, 2(2), 64–74. <https://doi.org/10.34312/jjps.v2i2.10292>
- Amalia, S. (2022). Analisis Sulfur Dioksida (SO₂) Udara Ambient Menggunakan Metode Pararosanilin dengan Spektrofotometer UV-Visible Kabupaten Bandung , Jawa Barat. *Gunung Djati Conference Series*, 15(2774–6585), 11–15.
- Amalia Yunia Rahmawati. (2020). *Pengenalan Data Mining* (Issue July).
- Ananda, C. P. (2023). *Machine Learning untuk Prediksi Gaya Hidup Berdasarkan Socioeconomic Status (SES) Menggunakan Algoritma Catboost Studi Kasus: Mahasiswa UIN Jakarta*. 31–41. <https://doi.org/https://repository.uinjkt.ac.id/dspace/handle/123456789/73329>
- Anusha, P. V., Anuradha, C., Chandra Murty, P. S. R., & Kiran, C. S. (2019). Detecting outliers in high dimensional data sets using Z-score methodology. *International Journal of Innovative Technology and Exploring Engineering*, 9(1), 48–53. <https://doi.org/10.35940/ijitee.A3910.119119>
- Astriyani, M., Laela, I. N., Lestari, D. P., Anggraeni, L., & Astuti, T. (2023). Analisis Klasifikasi Data Kualitas Udara Dki Jakarta Menggunakan Algoritma C.45. *JuSiTik: Jurnal Sistem Dan Teknologi Informasi Komunikasi*, 6(1), 36–41. <https://doi.org/10.32524/jusitik.v6i1.790>
- Baiq Nurul Azmi, Arief Hermawan, & Donny Avianto. (2023). Analisis Pengaruh

Hak Cipta Diindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:

- a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
- b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.

2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

Komposisi Data Training dan Data Testing pada Penggunaan PCA dan Algoritma Decision Tree untuk Klasifikasi Penderita Penyakit Liver. *JTIM: Jurnal Teknologi Informasi Dan Multimedia*, 4(4), 281–290. <https://doi.org/10.35746/jtim.v4i4.298>

Banerjee, S., Bhavna, K., & Raychoudhury, T. (2023). Prediction of transport behavior of nanoparticles using machine learning algorithm: Physical significance of important features. *Journal of Contaminant Hydrology*, 258, 104237. <https://doi.org/10.1016/J.JCONHYD.2023.104237>

Cahyono, W. E. (2011). Kajian Tingkat Pencemaran Sulfur Dioksida Dari Industri Di Beberapa Daerah Di Indonesia. *Berita Dirgantara*, 12(4), 132–137. http://jurnal.lapan.go.id/index.php/berita_dirgantara/article/view/1661

Chikodili, N. B., Abdulmalik, M. D., Abisoye, O. A., & Bashir, S. A. (2021). Outlier Detection in Multivariate Time Series Data Using a Fusion of K-Medoid, Standardized Euclidean Distance and Z-Score. In *Communications in Computer and Information Science* (Vol. 1350, Issue November). Springer International Publishing. https://doi.org/10.1007/978-3-030-69143-1_21

Christian, J., Ernawati, I., & ... (2022). Implementasi Penggunaan Algoritma Categorical Boosting (Catboost) Dengan Optimisasi Hiperparameter Dalam Memprediksi Pembatalan Pesanan Kamar Hotel. ... *Mahasiswa Bidang Ilmu* ..., 2021, 641–651. <https://conference.upnvj.ac.id/index.php/senamika/article/view/2230>

Dash, C. S. K., Behera, A. K., Dehuri, S., & Ghosh, A. (2023). An outliers detection and elimination framework in classification task of data mining. *Decision Analytics Journal*, 6(January), 100164. <https://doi.org/10.1016/j.dajour.2023.100164>

Dewi, N. K. (2021). Deteksi Fake Follower Instagram menggunakan Catboost Classifier. In *UIN Syarif Hidayatullah Jakarta Fakultas Sains dan Teknologi*.

Febiani, N., Fauzan, A. C., & Huda, M. M. (2022). Implementasi Algoritma

Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:

- a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
- b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.

2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

Decision Tree C4.5 Dengan Improvisasi Mean Dan Median Pada Dataset Numerik. *Jurnal Teknik Informasi Dan Komputer (Tekinkom)*, 5(1), 105. <https://doi.org/10.37600/tekinkom.v5i1.435>

Febriansyah Istianto, A., Id Hadiana, A., & Rakhmat Umbara, F. (2024). Prediksi Curah Hujan Menggunakan Metode Categorical Boosting (Catboost). *JATI (Jurnal Mahasiswa Teknik Informatika)*, 7(4), 2930–2937. <https://doi.org/https://doi.org/10.36040/jati.v7i4.7304>

Gosavi, G., & Gawde, G. (2022). Prediction Techniques for Data mining. *National Conference on Emerging Trends in Computer Engineering and Technology, January*.

Gunawan, H., Ruslinda, Y., Bachtiar, V. S., & ... (2018). Model Hubungan Konsentrasi Particulate Matter 10 (PM10) di Udara Ambien Dengan Karakteristik Lalu Lintas di Jaringan Jalan Primer Kota Padang. *Jurnal UMJ Seminar Nasional Sains Dan Teknologi 2018*, 1–11. <https://jurnal.umj.ac.id/index.php/semnastek/article/view/3557%0Ahttps://jurnal.umj.ac.id/index.php/semnastek/article/download/3557/2661>

Huang, G., Wu, L., Ma, X., Zhang, W., Fan, J., Yu, X., Zeng, W., & Zhou, H. (2019). Evaluation of CatBoost method for prediction of reference evapotranspiration in humid regions. *Journal of Hydrology*, 574, 1029–1041. <https://doi.org/10.1016/J.JHYDROL.2019.04.085>

Ibrahim, A. A., Ridwan, R. L., Muhammed, M. M., Abdulaziz, R. O., & Saheed, G. A. (2020). Comparison of the CatBoost Classifier with other Machine Learning Methods. *International Journal of Advanced Computer Science and Applications*, 11(11), 738–748. <https://doi.org/10.14569/IJACSA.2020.0111190>

Joseph, V. R. (2022). Optimal ratio for data splitting. *Statistical Analysis and Data Mining*, 15(4), 531–538. <https://doi.org/10.1002/sam.11583>

Joshi, A. P., & Patel, B. V. (2021). Data Preprocessing: The Techniques for



Hak Cipta Diindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

Preparing Clean and Quality Data for Data Analytics Process. *Oriental Journal of Computer Science and Technology*, 13(0203), 78–81. <https://doi.org/10.13005/ojcs13.0203.03>

Jusuf, H., Prasetya, E., & Igirisa, N. (2023). Analisis Risiko Kesehatan Lingkungan Paparan Particulate Matter (Pm10) Dan Karbon Monoksida (Co) Pada Masyarakat Di Desa Buata Kecamatan Botupingge. *Sulolipu: Media Komunikasi Sivitas Akademika Dan Masyarakat*, 23(1), 187. <https://doi.org/10.32382/sulolipu.v23i1.3155>

Keller, K., Haghi, S. H. R., Hahad, O., Schmidtman, I., Chowdhury, S., Lelieveld, J., Münzel, T., & Hobohm, L. (2023). Air pollution impacts on in-hospital case-fatality rate of ischemic stroke patients. *Thrombosis Research*, 225(October 2022), 116–125. <https://doi.org/10.1016/j.thromres.2023.03.006>

Lauble, S., Steuer, D., & Haghsheno, S. (2023). a Comparative Evaluation of Catboost Decision Trees and Expert Intuition To Predict Durations in the Predesign Phase. *Proceedings of the European Conference on Computing in Construction*. <https://doi.org/10.35490/EC3.2023.161>

Lestari, E. R. C. (2010). Hubungan kebakaran hutan dan lahan di Provinsi Riau terhadap kualitas udara dan resiko kesehatan masyarakat di Kota Pekanbaru. *Universitas Indonesia*, i–90.

Mai Sarah Tarigan, P., Tata Hardinata, J., Qurniawan, H., Safii, M., & Winanjaya, R. (2022). Implementasi Data Mining Menggunakan Algoritma Apriori Dalam Menentukan Persediaan Barang (Studi Kasus : Toko Sinar Harahap). *Just IT: Jurnal Sistem Informasi, Teknologi Informasi Dan Komputer*, 12(2), 51–61. <https://doi.org/https://doi.org/10.24853/justit.12.2.%25p>

Malley, B., Ramazzotti, D., & Wu, J. T. (2016). Secondary Analysis of Electronic Health Records. *Secondary Analysis of Electronic Health Records*, 1–427. <https://doi.org/10.1007/978-3-319-43742-2>

Maltare, N. N., & Vahora, S. (2023). Air Quality Index prediction using machine



Hak Cipta Diindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:

a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.

b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.

2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

learning for Ahmedabad city. *Digital Chemical Engineering*, 7(November 2022), 100093. <https://doi.org/10.1016/j.dche.2023.100093>

Manikandan, S. (2011). Measures of dispersion. *Journal of Pharmacology and Pharmacotherapeutics*, 2(4), 315–316. <https://doi.org/10.4103/0976-500X.85931>

Manisalidis, I., Stavropoulou, E., Stavropoulos, A., & Bezirtzoglou, E. (2020). Environmental and Health Impacts of Air Pollution: A Review. *Frontiers in Public Health*, 8(February), 1–13. <https://doi.org/10.3389/fpubh.2020.00014>

Masito, A. (2018). Risk Assessment Ambient Air Quality (NO₂ And SO₂) and The Respiratory Disorders to Communities in the Kalianak Area of Surabaya. *Jurnal Kesehatan Lingkungan*, 10(4), 394. <https://doi.org/10.20473/jkl.v10i4.2018.394-401>

Murugan, R., & Palanichamy, N. (2021). Smart city air quality prediction using machine learning. *Proceedings - 5th International Conference on Intelligent Computing and Control Systems, ICICCS 2021*, 1048–1054. <https://doi.org/10.1109/ICICCS51141.2021.9432074>

Nababan, A. A., Jannah, M., Aulina, M., & Andrian, D. (2023). Prediksi Kualitas Udara Menggunakan Xgboost Dengan Synthetic Minority Oversampling Technique (Smote) Berdasarkan Indeks Standar Pencemaran Udara (Ispu). *JTIK (Jurnal Teknik Informatika Kaputama)*, 7(1), 214–219. <https://doi.org/10.59697/jtik.v7i1.66>

Nasional, J., Informasi, S., Rizkiani, E., & Brahma, D. (2024). Perbandingan Performa Algoritma Metode Bagging dan Boosting pada Prediksi Konsentrasi PM₁₀ di Jakarta Utara. 01, 74–81. <https://doi.org/https://doi.org/10.25077/TEKNOSI.v10i1.2024.74-80>

Nasution, F. A., Saadah, S., & Yunanto, P. E. (2023). Credit Risk Detection in Peer-to-Peer Lending Using CatBoost. *Jurnal RESTI (Rekayasa Sistem Dan Teknologi Informasi)*, 7(5), 1056–1062.

Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:

a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.

b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.

<https://doi.org/10.29207/resti.v7i5.5139>

Ningrum, F. (2024). *Identifikasi Jenis-Jenis Aeroplankton Sebagai Dasar Parameter Kualitas Udara di Wilayah Pekanbaru*. *11*(1), 152–161. <https://doi.org/https://doi.org/10.31849/bl.v11i1.19887>

Noor, N. M., Al Bakri Abdullah, M. M., Yahaya, A. S., & Ramli, N. A. (2015). Comparison of linear interpolation method and mean method to replace the missing values in environmental data set. *Materials Science Forum*, *803*(July 2015), 278–281. <https://doi.org/10.4028/www.scientific.net/MSF.803.278>

Nurchaerani Kadir. (2024). *Analisis Performa Metode Random Forest dan Catboost Dalam Pemodelan Kualitas Udara Kota Palembang*.

Prayudha, J., Pranata, A., & Al Hafiz, A. (2018). Implementasi Metode Fuzzy Logic Untuk Sistem Pengukuran Kualitas Udara Di Kota Medan Berbasis Internet of Things (Iot). *Jurteksi*, *4*(2), 141–148. <https://doi.org/10.33330/jurteksi.v4i2.57>

Prokhorenkova, L., Gusev, G., Vorobev, A., Dorogush, A. V., & Gulin, A. (2019). *CatBoost : unbiased boosting with categorical features. Section 4*, 1–23.

Purbolingga, Y., Marta, D., Rahmawatia, A., & Wajhi, B. (2023). Perbandingan Algoritma CatBoost dan XGBoost dalam Klasifikasi Penyakit Jantung. *Jurnal APTEK Vol. 15 No 2 (2023) 126-133*, *15*(2), 126–133. <http://journal.upp.ac.id/index.php/aptek/article/download/1930/1163/4970>

Putra Mulia, Nofrizal, & Dewi, W. N. (2021). Analisis Dampak Kabut Asap Karhutla Terhadap Gangguan Kesehatan Fisik Dan Mental. *Health Care : Jurnal Kesehatan*, *10*(1), 62–68. <https://doi.org/10.36763/healthcare.v10i1.103>

Putra, R. E., Elektro, F. T., Telkom, U., Kallista, M., Elektro, F. T., Telkom, U., Setianingsih, C., Elektro, F. T., Telkom, U., Udara, A. P., Udara, B. I. K., Pendahuluan, I., & Indonesia, I. (2023). *Klasifikasi prediksi kualitas udara Menggunakan metode Support Vector Machine (SVM)*. *10*(4), 3790–3796.

Hak Cipta Diindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:

- a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
- b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.

2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

- Raka Siwih. (2020). Peran Dinas Lingkungan Hidup dan Kehutanan Provinsi Riau Dalam Pengendalian Pencemaran Lingkungan Hidup (Studi Kasus Pencemaran Udara Akibat Kebakaran Hutan Di Kota Pekanbaru). *Electoral Governance Jurnal Tata Kelola Pemilu Indonesia*, 12(2), 6. <https://talenta.usu.ac.id/politeia/article/view/3955>
- Ravindiran, G., Hayder, G., Kanagarathinam, K., Alagumalai, A., & Sonne, C. (2023). Air quality prediction by machine learning models: A predictive study on the indian coastal city of Visakhapatnam. *Chemosphere*, 338. <https://doi.org/10.1016/j.chemosphere.2023.139518>
- Razali, A., Maksum, H., & Daswarman. (2014). Perbandingan Gas Karbon Monoksida (CO) Dan Hidrokarbon (HC) Yang Menggunakan Catalyst Kuningan Dengan Catalyst Tembaga Pada Motor Empat Langkah. *Automotive Engineering Education Journals*, 3(4), 1–9. <https://ejournal.unp.ac.id/students/index.php/poto/article/download/3162/2285>
- Rizaldi, M. A., Azizah, R., Latif, M. T., Sulistyorini, L., & Salindra, B. P. (2022). Dampak Paparan Gas Karbon Monoksida Terhadap Kesehatan Masyarakat yang Rentan dan Berisiko Tinggi. *Jurnal Kesehatan Lingkungan Indonesia*, 21(3), 253–265. <https://doi.org/10.14710/jkli.21.3.253-265>
- Ruchjana, B. N., Arianto, A. T., Parmikanti, K., & Suhandi, B. (2021). Peramalan Konsentrasi Particulate Matter 2.5 (PM2.5) menggunakan Model Vector Autoregressive dengan Metode Maximum Likelihood Estimation. *KUBIK: Jurnal Publikasi Ilmiah Matematika*, 6(1), 1–12. <https://doi.org/10.15575/kubik.v6i1.8046>
- Safaei, N., Safaei, B., Seyedekrami, S., Talafidaryani, M., Masoud, A., Wang, S., Li, Q., & Moqri, M. (2022). E-CatBoost: An efficient machine learning framework for predicting ICU mortality using the eICU Collaborative Research Database. In *PLoS ONE* (Vol. 17, Issue 5 May). <https://doi.org/10.1371/journal.pone.0262895>



Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:

- a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
- b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.

2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

- Safira, M. C., Fauzan, A., & Adhiwibawa, M. A. S. (2021). Interpolasi Polutan Nitrogen Dioksida (NO₂) Yogyakarta Dengan Pendekatan Ordinary Kriging Dan. *Jurnal Aplikasi Statistika & Komputasi Statistik*, 14(2), 55–66.
- Sang, A. I., Sutoyo, E., & Darmawan, I. (2021). Analisis Data Mining untuk Klasifikasi Data Kualitas Udara DKI Jakarta Menggunakan Algoritma Decision Tree dan Support Vector Machine. *E-Proceeding of Engineering*, 8(5), 8954–8963.
- Santoso, I., Juanda, & Dahlan, A. (2022). *Pengawasan Kualitas Udara Dalam Gedung*. 1–23.
- Schneider, P., & Xhafa, F. (2022). Anomaly detection. *Anomaly Detection and Complex Event Processing over IoT Data Streams*, 49–66. <https://doi.org/10.1016/B978-0-12-823818-9.00013-4>
- Serlina, Y., Bachtiar, V. S., & Putra, I. (2023). Analisis Konsentrasi Particulate Matter 2,5 di Udara Ambien dan Rekomendasi Tanaman Pereduksi PM_{2,5} di Perumahan Unand Blok B, Ulu Gadut, Kota Padang. *Jurnal Serambi Engineering*, 8(4), 7516–7524. <https://doi.org/10.32672/jse.v8i4.6863>
- Shankar, D. D., Azhakath, A. S., Khalil, N., Sajeev, J., Mahalakshmi, T., & Sheeba, K. (2023). Data Mining for Cyber Biosecurity Risk Management – a comprehensive review. *Computers & Security*, 103627. <https://doi.org/10.1016/j.cose.2023.103627>
- SHW, M. N. P. (2019). *Analisis Pengaruh Faktor Meteorologi dan Kebakaran Hutan dan Lahan terhadap Penyakit Infeksi Saluran Pernapasan Akut (ISPA) di Kota Pekanbaru*. <http://repository.ipb.ac.id/handle/123456789/100532>
- Shombing, P. R., Suryadiningrat, S., Sunarjo, D. A., & Yuda, Y. P. A. C. (2023). Identifikasi Data Outlier (Pencilan) dan Kenormalan Data Pada Data Univariat serta Alternatif Penyelesaiannya. *Jurnal Ekonomi Dan Statistik Indonesia*, 2(3), 307–316. <https://doi.org/10.11594/jesi.02.03.07>

Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:

- a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
- b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.

2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

- Singgih, I. K. (2020). Air Quality Prediction in Smart City's Information System
ARTICLE INFO. / *International Journal of Informatics Information System and Computer Engineering*, 1, 35–46.
<https://doi.org/https://doi.org/10.34010/injiiscom.v1i1.4020>
- Singh, K., & Upadhyaya, S. (2012). Outlier Detection: Applications And Techniques. *International Journal of Computer ...*, 9(1), 307–323.
<http://search.ebscohost.com/login.aspx?direct=true&profile=ehost&scope=sit e&authtype=crawler&jrnl=16940784&AN=73150560&h=i4TJm6g8nLsTJh cyMBIVsybVnJ9dDMRPUQ7ZLZ8IBk76dVDDRAgMCCc258zyyjrF/zu+Mv vsObGzF2pYu0H1DPg==&crl=c>
- Srivastava, S. (2014). Weka: A Tool for Data preprocessing, Classification, Ensemble, Clustering and Association Rule Mining. *International Journal of Computer Applications*, 88(10), 26–29. <https://doi.org/10.5120/15389-3809>
- Sugiyono, D. (2019). Metode penelitian kuantitatif kualitatif dan R&D. In *Penerbit Alfabeta*.
- Ulandari, D. A., & Swanjaya, D. (2020). Perbandingan Transformasi Data pada Penentuan Peserta Bimbingan Belajar Menggunakan Metode Perceptron. *Prosiding SEMNAS INOTEK ...*, 191–196.
- Yang, C. C., Wang, C. F., Lin, W. M., Chen, S. W., & Hu, H. W. (2024). Evaluating the performance of an AI-powered VBAC prediction system within a decision-aid birth choice platform for shared decision-making. *Digital Health*, 10(195). <https://doi.org/10.1177/20552076241257014>
- Yaro, A. S., Maly, F., Prazak, P., & Maly, K. (2024). Outlier Detection Performance of a Modified Z-Score Method in Time-Series RSS Observation With Hybrid Scale Estimators. *IEEE Access*, 12, 12785–12796. <https://doi.org/10.1109/ACCESS.2024.3356731>
- Yoseph, F., Heikkila, M., & Howard, D. (2019). Outliers identification model in point-of-sales data using enhanced normal distribution method. *Proceedings*

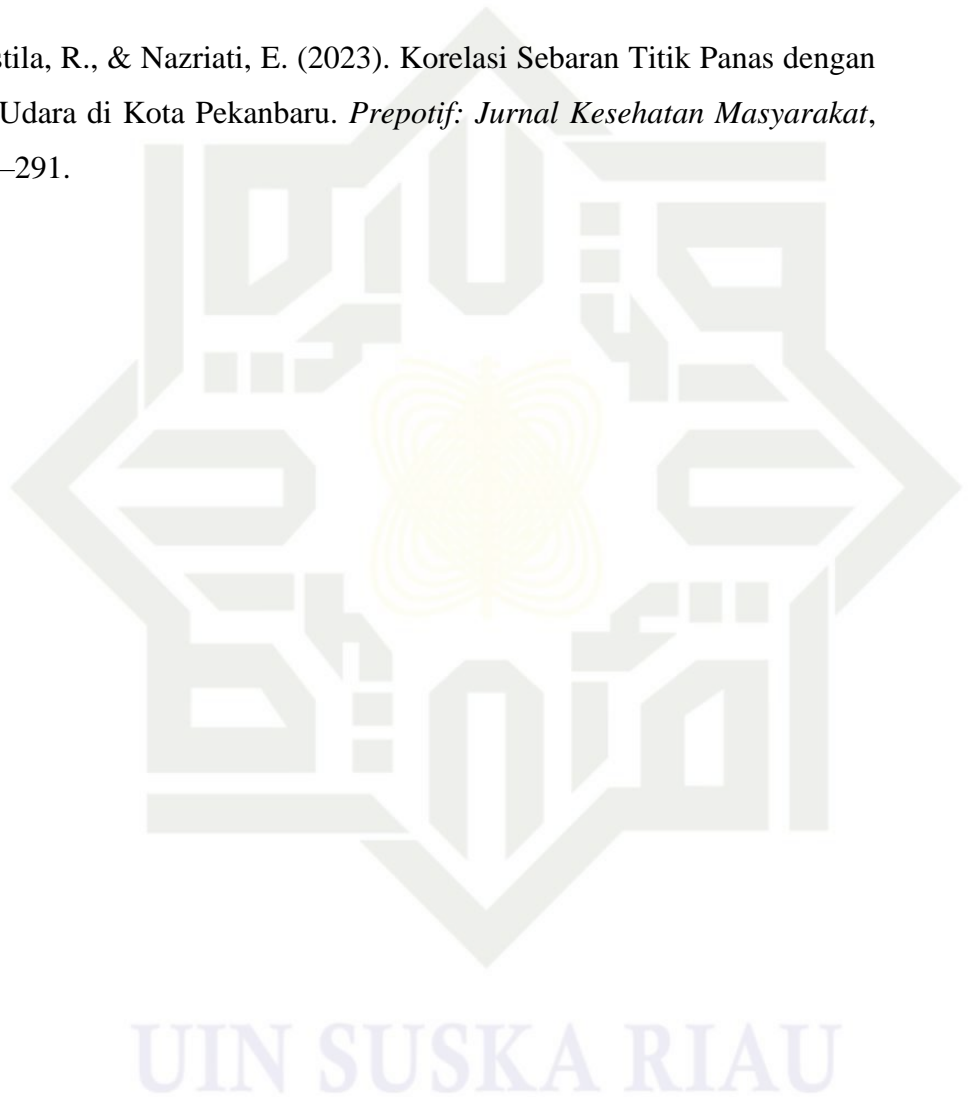
Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

- *International Conference on Machine Learning and Data Engineering, ICMLDE 2019*, 72–78. <https://doi.org/10.1109/iCMLDE49015.2019.00024>

Yusrianti, Y. (2015). Studi Literatur tentang Pencemaran Udara Akibat Aktivitas Kendaraan Bermotor di Jalan Kota Surabaya. *Al-Ard: Jurnal Teknik Lingkungan*, 1(1), 11–20. <https://doi.org/10.29080/alard.v1i1.29>

Zahntamal, Restila, R., & Nazriati, E. (2023). Korelasi Sebaran Titik Panas dengan Kualitas Udara di Kota Pekanbaru. *Prepotif: Jurnal Kesehatan Masyarakat*, 7(1), 283–291.





Hak Cipta Diindungi Undang-Undang

1. Diarangi mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan satu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Diarangi mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

LAMPIRAN A



SURAT KETERANGAN PENELITIAN

Nomor : BL.04.00/Kesbangpol/1363/2024



- a. Dasar :
1. Undang-Undang Republik Indonesia Nomor 14 Tahun 2008 Tentang Keterbukaan Informasi Publik.
 2. Undang-Undang Republik Indonesia Nomor 25 Tahun 2009 Tentang Pelayanan Publik.
 3. Peraturan Pemerintah Republik Indonesia Nomor 18 Tahun 2016 Tentang Perangkat Daerah.
 4. Peraturan Menteri Dalam Negeri Nomor 3 Tahun 2018 Tentang Penerbitan Surat Keterangan Penelitian.
 5. Peraturan Daerah Kota Pekanbaru Nomor 9 Tahun 2016 Tentang Pembentukan dan Susunan Perangkat Daerah Kota Pekanbaru.
- b. Menimbang :
- Surat dari Dekan Fakultas Sains dan Teknologi UIN SUSKA Riau, nomor B-3344/F.V/PP.00.9/04/2024 tanggal 30 April 2024, perihal pelaksanaan kegiatan Penelitian Riset/Pra Riset dan pengumpulan data untuk bahan Skripsi.

MEMBERITAHUKAN BAHWA :

1. Nama : AMORINA HANJANI
2. NIM : 12050120365
3. Fakultas : SAINS DAN TEKNOLOGI UIN SUSKA RIAU
4. Jurusan : TEKNIK INFORMATIKA
5. Jenjang : S1
6. Alamat : JL. HOS COKROAMINOTO GG. PRIBADI KEL. SEKIP HULU KEC. RENGAT-INDRAGIRI HULU
7. Judul Penelitian : PREDIKSI KUALITAS UDARA MENGGUNAKAN CATBOOST (STUDI KASUS : PEKANBARU)
8. Lokasi Penelitian : DINAS LINGKUNGAN HIDUP DAN KEBERSIHAN KOTA PEKANBARU

Untuk Melakukan Penelitian, dengan ketentuan sebagai berikut :

1. Tidak melakukan kegiatan yang menyimpang dari ketentuan yang telah ditetapkan yang tidak ada hubungan dengan kegiatan Riset/Pra Riset/ Penelitian dan pengumpulan data ini.
2. Pelaksanaan kegiatan Riset ini berlangsung selama 6 (enam) bulan terhitung mulai tanggal Surat Keterangan Penelitian ini diterbitkan.
3. Berpakaian sopan, mematuhi etika Kantor/Lokasi Penelitian, bersedia meninggalkan foto copy Kartu Tanda Pengenal.
4. Melaporkan hasil Penelitian kepada Walikota Pekanbaru c.q Kepala Badan Kesatuan Bangsa dan Politik Kota Pekanbaru, paling lambat 1 (satu) minggu setelah selesai.

Demikian Rekomendasi ini dibuat untuk dipergunakan sebagaimana mestinya.

Pekanbaru, 3 Mei 2024



Tembusan

- Yth :
1. Dekan Fakultas Sains dan Teknologi UIN SUSKA Riau di Pekanbaru.
 2. Yang Bersangkutan.



Hak Cipta Diindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.



UIN SUSKA RIAU

KEMENTERIAN AGAMA REPUBLIK INDONESIA
UNIVERSITAS ISLAM NEGERI SULTAN SYARIF KASIM RIAU
FAKULTAS SAINS DAN TEKNOLOGI

كلية العلوم و التكنولوجيا

FACULTY OF SCIENCES AND TECHNOLOGY

Jl. H.R. Soebrantas No. 155 KM. 15 Tuah Madani - Pekanbaru 28293 PO.Box. 1004 Telp.(0761) 589026 - 589027
Fax. (0761) 589 025 Web. www.uin-suska.ac.id, Email : faste@uin-suska.ac.id

Nomor : B- 3217 /F.V/PP.00.9/ 04/2024
Sifat : Penting
Hal : Mohon Izin Penelitian dan Pengambilan Data
Tugas Akhir/Skripsi

Pekanbaru, 23 April 2024

Kepada Yth.
Kepala Dinas Lingkungan Hidup Dan Kebersihan Pekanbaru
Jl. Datuk Setia Maharaja No.04, Simpang Tiga
Kec. Bukit Raya, Kota Pekanbaru, Riau 28125

Assalamu'alaikum Wr. Wb.

Dengan hormat, sehubungan telah dimulainya mata kuliah Tugas Akhir pada Fakultas Sains dan Teknologi UIN Suska Riau, Kami bermaksud mengirimkan mahasiswa :

| | |
|---------------------|------------------------|
| Nama | : Amorina Hanjani |
| NIM | : 12050120365 |
| Fakultas | : Sains dan Teknologi |
| Program Studi / Smt | : Teknik Informatika/8 |
| No. HP / E-mail | : 082284480310 |

untuk pengambilan data yang sangat dibutuhkan dalam Tugas Akhir mahasiswa tersebut yang berjudul "Prediksi Kualitas Udara di Pekanbaru Menggunakan CatBoost". Kami mohon kiranya Saudara berkenan memberikan izin dan fasilitas demi kelancaran Tugas Akhir mahasiswa yang bersangkutan.

Demikian surat ini Kami sampaikan, atas bantuan dan kerjasama Saudara kami ucapkan terima kasih.



Amorina Hanjani, ST, M.Eng., Ph.D
NIP. 19850616211011016

Tembusan :
Yth. Rektor UIN Suska Riau.



Hak Cipta Diindungi Undang-Undang

1. Diarangi mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.



SURAT IZIN TUGAS AKHIR
 PROGRAM STUDI TEKNIK INFORMATIKA
 FAKULTAS SAINS DAN TEKNOLOGI
 UNIVERSITAS ISLAM NEGERI SULTAN SYARIF KASIM RIAU

Ketua Program Studi Teknik Informatika Menerangkan Mahasiswa Di Bawah Ini :

Nama : Amorina Hanjani
 NIM / Semester : 12050120365 / 7
 Program Studi : Teknik Informatika
 Alamat : Jl. Melayu, Simpang Baru, Pekanbaru
 No. Telp/HP : 082284480310
 Email Student : 12050120365@students.uin-suska.ac.id

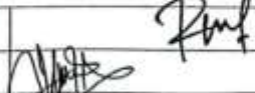
Lulus Verifikasi Serta Validasi Persyaratan Tugas Akhir dan Diberikan **Izin Melaksanakan Penelitian dan Pengumpulan Data** pada :

Perusahaan /Instansi : Dinas Lingkungan Hidup dan Kehutanan Provinsi Riau
 Alamat : Jl. Jendral Sudirman, No. 468, Pekanbaru, 28126, Jadirejo, Sukajadi, Pekanbaru, Riau 28121
 Bidang Kajian/Judul : Prediksi Kualitas Udara Menggunakan Algoritma CatBoost
 Waktu Pelaksanaan :

Demikianlah Surat Izin Ini Diberikan Untuk Keperluan Mahasiswa Melengkapi Persyaratan Pengajuan Surat Pengantar Tugas Akhir Fakultas Sains dan Teknologi, UIN Sultan Syarif Kasim Riau.

Pekanbaru, 17 Januari 2023
 Ketua Program Studi Teknik Informatika

Iwan Iskandar, S.T., M.T.
 NIP. 19821216 201503 1 003

| | | |
|--------------------------|---|--|
| Sekretaris Program Studi | : Reski Mai Candra, S.T., M.Sc. NIP. 19860505 201503 1 006 |  |
| Koordinator Tugas Akhir | : Fadhilah Syalria, S.T., M.Kom. NIK. 130 517 102 | |



DAFTAR RIWAYAT HIDUP

Data Pribadi

| | | |
|----------------------|--|---|
| Nama Lengkap | : Amorina Hanjani |  |
| Tempat/Tanggal Lahir | : Rengat, 23 Juli 2002 | |
| Jenis Kelamin | : Perempuan | |
| Agama | : Islam | |
| Kewarganegaraan | : Indonesia | |
| Alamat | : Jl. Hos Coktoaminoto, Gg. Pribadi, Kel. Sekip Hulu, Kec. Rengat | |
| No. Hp | : 0822 8448 0310 | |
| Email | : 12050120365@students.uin-suska.ac.id | |

Pendidikan

| | |
|------------|-----------------------------------|
| 2006-2007 | : TK Aisyah Indragiri Hulu Rengat |
| 2007-2014 | : SD N 016 Rengat |
| 20014-2017 | : SMP N 01 Rengat |
| 2017-2020 | : SMA N 1 Bantan |

UIN SUSKA RIAU

Hak Cipta Diindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.