

**IMPLEMENTASI SUPERVISED MACHINE LEARNING
DENGAN METODE CRISP-DM DALAM PREDIKSI RATE
PEMUATAN PUPUK STUDI KASUS PT PETROKIMIA
GRESIK**

KERJA PRAKTIK



**BERLIAN SAFRI PRAKOSO
I0320017**

**PROGRAM STUDI SARJANA TEKNIK INDUSTRI
FAKULTAS TEKNIK
UNIVERSITAS SEBELAS MARET
SURAKARTA
2023**

**IMPLEMENTASI SUPERVISED MACHINE LEARNING
DENGAN PENDEKATAN CRISP-DM DALAM PREDIKSI
RATE PEMUATAN PUPUK STUDI KASUS PT
PETROKIMIA GRESIK**

KERJA PRAKTIK



**BERLIAN SAFRI PRAKOSO
I0320017**

**PROGRAM STUDI SARJANA TEKNIK INDUSTRI
FAKULTAS TEKNIK
UNIVERSITAS SEBELAS MARET
SURAKARTA
2023**

LEMBAR PENGESAHAN

Judul Laporan Kerja Praktik:

**IMPLEMENTASI SUPERVISED MACHINE LEARNING
DENGAN PENDEKATAN CRISP-DM DALAM PREDIKSI
RATE PEMUATAN PUPUK STUDI KASUS PT PETROKIMIA
GRESIK**

Disusun Oleh:

Berlian Safri Prakoso

I0320017

Mengesahkan
Kepala Program Studi Teknik
Industri
Fakultas Teknik

Disetujui

Dosen Pembimbing

Dr. Eko Liquiddanu, S.T., M.T.
NIP. 197101281998021001

Dr. Eko Liquiddanu, S.T., M.T.
NIP. 197101281998021001

SURAT KETERANGAN KERJA PRAKTIK

Yang bertanda tangan dibawah ini menerangkan bahwa mahasiswa:

Nama : Berlian Safri Prakoso

NIM : I0320017

Program Studi : Teknik Industri – Universitas Sebelas Maret

Telah melaksanakan Kerja Praktik di PT Petrokimia Gresik

Lama Kerja Praktik: 2 bulan (2 Januari 2023 – 28 Februari 2023)

Tanggal Penilaian : 17 Februari 2023

Nama Penilai : Ir. Ari Primantara, S.T., M.MT, IPM

Jabatan Penilai : VP Pergudangan dan Pengantongan

Tanda Tangan &

Stempel Perusahaan



FORM PENILAIAN KERJA PRAKTIK

Mohon diisi dan dicek seperlunya

Nama Mahasiswa : Berlian Safri Prakoso

NIM : I0320017

Program Studi : Teknik Industri – Universitas Sebelas Maret

Telah melaksanakan KERJA PRAKTIK di:

Nama Perusahaan : PT Petrokimia Gresik

Alamat Perusahaan : Jl. A. Yani, Kabupaten Gresik

Lama Kerja Praktik : 01/01/2023 sampai dengan 28/02/2023

Topik yang dibahas : Penerapan *Machine Learning* untuk Memprediksi *Rate* Pemuatan Pupuk *In Bag* pada Gudang Lini 1 PT Petrokimia Gresik

Nilai (sesuai kondite mahasiswa yang bersangkutan)

Sikap	:	<table><tbody><tr><td>50</td><td>60</td><td>70</td><td>80</td><td>90</td><td>100</td></tr></tbody></table>	50	60	70	80	90	100
50	60	70	80	90	100			
Kerajinan	:	<table><tbody><tr><td>50</td><td>60</td><td>70</td><td>80</td><td>90</td><td>100</td></tr></tbody></table>	50	60	70	80	90	100
50	60	70	80	90	100			
Prestasi	:	<table><tbody><tr><td>50</td><td>60</td><td>70</td><td>80</td><td>90</td><td>100</td></tr></tbody></table>	50	60	70	80	90	100
50	60	70	80	90	100			
Nilai rata-rata	:	<table border="1"><tbody><tr><td>100</td></tr></tbody></table>	100					
100								

Tanggal Penilaian : 17 Februari 2023

Nama Penilai : Ir. Ari Primantara, S.T., M.MT, IPM

Jabatan Penilai : VP Pergudangan dan Pengantongan

Tanda Tangan &

Stempel Perusahaan



KATA PENGANTAR

Puji dan syukur penulis panjatkan kepada Tuhan Yang Maha Esa atas segala rahmatnya penulis dapat menyelesaikan Laporan Kerja Praktik dengan judul “Implementasi Supervised Machine Learning dengan Metode CRISP-DM dalam Prediksi Rate Pemuatan Pupuk: Studi Kasus PT Petrokimia Gresik” dengan tepat waktu sebagai salah satu syarat bagi penulis dalam menyelesaikan pendidikan di Program Studi Sarjana Teknik Industri Universitas Sebelas Maret Surakarta. Laporan kerja praktik ini disusun setelah penulis melaksanakan kerja praktik di PT. Petrokimia Gresik dari tanggal 1 Januari 2023 sampai dengan 20 Februari 2023.

Laporan kerja praktik ini dapat terselesaikan dengan baik tentunya berkat bantuan dari berbagai pihak. Oleh karena itu, penulis mengucapkan terimakasih sebanyak-banyaknya kepada:

1. Tuhan Yang Maha Esa yang telah memberikan kemudahan dalam melaksanakan kerja praktik di PT Petrokimia Gresik dan juga menyelesaikan laporan kerja praktik dengan baik.
2. Orang tua dan saudara penulis yang selalu mendoakan serta memberi dukungan moril maupun material.
3. Bapak Dr. Eko Liquiddanu, S.T., M.T. selaku Kepala Program Studi Sarjana Teknik Industri, Fakultas Teknik, Universitas Sebelas Maret Surakarta, sekaligus sebagai dosen pembimbing kerja praktik yang telah memberikan arahan, dukungan, dan bimbingan bagi penulis sehingga laporan kerja praktik dapat terselesaikan dengan baik.
4. Bapak Taufiq Rochman S.TP., M.T. selaku koordinator kerja praktik Program Studi Sarjana Teknik Industri, Fakultas Teknik, Universitas Sebelas Maret Surakarta.
5. Bapak Ir. Ari Primantara S.T, M.MT, IPM, CSCA, CPLM. Selaku VP Departmen Pergudangan PT. Petrokimia Gresik sekaligus pembimbing lapangan yang senantiasa membantu dan memberikan *insight* dan dukungan selama kegiatan kerja praktik berlangsung.
6. Bapak Aminudin selaku AVP Gudang III Departmen Pergudangan yang selalu memberikan dukungan dan informasi mengenai pergudangan selama kegiatan kerja praktik berlangsung.
7. Seluruh karyawan PT Petrokimia Gresik terutama karyawan departemen pergudangan yang secara langsung maupun tidak langsung telah ikut andil dalam penyelesaian laporan ini.
8. Ervizar Buana dan keluarga yang telah membantu dan memfasilitasi penulis selama pelaksanaan kerja praktik di Kota Gresik.
9. Keluarga dan kerabat-kerabat Teknik Industri angkatan 2020 Universitas Sebelas Maret yang telah memberikan motivasi dalam pelaksanaan dan penyusunan Laporan Kerja Praktik.

10. Shafa, Tiara, dan Ervizar selaku teman kerja praktik satu departemen yang selalu memberikan dukungan dan sebagai *partner* diskusi selama kegiatan kerja praktik berlangsung.
11. Semua pihak lain yang telah membantu dalam penyusunan laporan ini yang tidak dapat disebutkan satu per satu.

Penulis menyadari bahwa dalam penyusunan laporan ini masih banyak terdapat kekurangan. Oleh karena itu, penulis mengharapkan kritik dan saran yang konstruktif sebagai masukan serta perbaikan dalam penulisan laporan kedepannya. Semoga laporan ini dapat bermanfaat bagi penulis dan pembaca sekalian.

Gresik, 17 Februari 2023

Penulis

Berlian Safri Prakoso

DAFTAR ISI

HALAMAN JUDUL.....	i
LEMBAR PENGESAHAN	ii
SURAT KETERANGAN KERJA PRAKTIK	ii
FORM PENILAIAN KERJA PRAKTIK	iii
KATA PENGANTAR	iv
DAFTAR ISI.....	vi
DAFTAR TABEL.....	viii
DAFTAR GAMBAR	ix
BAB I PENDAHULUAN.....	I-1
1.1 Latar Belakang	I-1
1.2 Rumusan Masalah	I-3
1.3 Tujuan Penelitian.....	I-3
1.4 Manfaat Penelitian.....	I-3
1.5 Batasan Masalah.....	I-4
1.6 Asumsi.....	I-4
1.7 Sistematika Penulisan.....	I-5
BAB II TINJAUAN PUSTAKA	II-1
2.1 Tinjauan Umum Perusahaan	II-1
2.1.1 Profil Perusahaan	II-1
2.1.2 Sejarah Perusahaan.....	II-2
2.1.3 Visi Misi Perusahaan	II-3
2.1.4 Tata Nilai Perusahaan	II-3
2.1.5 Struktur Organisasi	II-5
2.1.6 Alur Distribusi.....	II-6
2.1.7 Wilayah Distribusi	II-6
2.1.8 Ruang Lingkup Gudang Perusahaan.....	II-7
2.1.9 Transformasi Digital pada Pergudangan.....	II-7
2.1.10 Produk yang dihasilkan perusahaan	II-9
2.2 Tinjauan Pustaka	II-10
2.2.1 Data Mining	II-10
2.2.2 CRISP-DM.....	II-10
2.2.3 Deep Learning.....	II-11

2.2.4	<i>Artificial Neural network</i>	II-11
2.2.5	<i>Metrics Evaluation</i>	II-12
2.2.6	Outlier	II-13
2.2.7	<i>One-Hot Encoding</i>	II-14
2.2.8	Analisis Hubungan dan Hipotesis	II-14
BAB III	METODOLOGI PENELITIAN	III-1
3.1	Pemahaman Bisnis (<i>Business Understanding</i>)	III-2
3.2	Pemahaman Data (<i>Data Understanding</i>)	III-2
3.3	Persiapan Data (<i>Data Preparation</i>).....	III-2
3.4	Pemodelan (<i>Modelling</i>)	III-2
BAB IV	PENGUMPULAN DAN PENGOLAHAN DATA	IV-1
4.1	Pemahaman Terhadap Data (<i>Data Understanding</i>)	IV-1
4.1.1	Tinjauan Data yang Tersedia	IV-1
4.1.2	Kebutuhan Data untuk Analisis	IV-3
4.1.3	Eksplorasi Data Awal.....	IV-3
4.2	Persiapan Data (<i>Data Preparation</i>).....	IV-4
4.2.1	Pemilihan Dataset.....	IV-4
4.2.2	Pembersihan Data.....	IV-4
4.2.3	Transformasi dan Pemrosesan Data	IV-6
4.2.4	Penggabungan Data dari Sumber Berbeda.....	IV-8
4.3	Pengujian Hubungan dan Signifikansi	IV-9
4.4	Pengembangan Model Machine Learning.....	IV-11
4.5	Evaluasi Model Machine Learning	IV-13
4.6	<i>Hyperparameter Tuning</i>	IV-18
BAB V	ANALISIS DAN INTERPRETASI HASIL.....	V-1
5.1	Analisis Deskriptif.....	V-1
5.2	Analisis Hubungan dan Pengujian Hipotesis.	V-6
5.3	Evaluasi Model.....	V-9
5.4	Analisis <i>Hyperparameter Tuning</i>	V-11
BAB VI	KESIMPULAN DAN SARAN	VI-1
6.1	Kesimpulan.....	VI-1
6.2	Saran	VI-2
DAFTAR PUSTAKA		

DAFTAR TABEL

Tabel 4.1 Gabungan dataset pemuatan pupuk dan cuaca.....	IV-3
Table 4.2 Gabungan dataset pemuatan pupuk dan cuaca (lanjutan)	IV-4
Table 4.3 <i>Metrics evaluation</i> model	IV-13
Table 4.4 <i>Metrics evaluation</i> model dengan AutoKeras.....	IV-19
Table 4.5 <i>Metrics evaluation</i> model dengan penghapusan <i>outlier</i>	IV-22
Table 5.1 Data maksimum, minimum, dan <i>mean</i> rate	V-3
Table 5.2 Koefisien korelasi dan <i>p-value</i>	V-6

DAFTAR GAMBAR

Gambar 2.2.1 Logo Perusahaan PT Petrokimia Gresik	II-1
Gambar 2.2.2 Sejarah Perusahaan PT Petrokimia Gresik.....	II-2
Gambar 2.2.3 Struktur Perusahaan	II-5
Gambar 2.2.4 Alur Distribusi Perusahaan.....	II-6
Gambar 2.2.5 Ruang Lingkup Departemen Pergudangan	II-7
Gambar 2.2.6 Logo Aplikasi WMS Perusahaan	II-8
Gambar 2.2.7 Logo Aplikasi DTMS Perusahaan.....	II-8
Gambar 2.2.8 Logo Aplikasi Sistro Perusahaan	II-9
Gambar 2.2.9 Produk-produk Pupuk Perusahaan	II-9
Gambar 2.10 Produk-produk Non Pupuk Perusahaan	II-10
Gambar 3.3.1 <i>Flowchart</i> Metodologi Penelitian	III-1
Gambar 4.4.1 Data pemuatan pupuk WMS PT Petrokimia Gresik	IV-1
Gambar 4.4.2 Data cuaca historis Kabupaten Gresik	IV-2
Gambar 4.4.3 Data cuaca historis Kabupaten Gresik (lanjutan).....	IV-2
Gambar 4.4 Proses menyeleksi data yang akan dianalisis	IV-4
Gambar 4.5 Proses menyeleksi aktivitas Pengeluaran Pupuk / Non Pupuk Melalui JT.....	IV-4
Gambar 4.6 <i>Outlier</i> pada dataset pemuatan pupuk	IV-5
Gambar 4.7 Menghapus outlier pada atribut lama pemuatan	IV-5
Gambar 4.8 Menghapus outlier pada atribut lama <i>approval</i>	IV-6
Gambar 4.9 <i>Dataset</i> pemuatan pupuk yang sudah bersih dari <i>outlier</i>	IV-6
Gambar 4.10 Pengkodean ulang variabel dengan teknik <i>One-Hot Encoding</i> ...	IV-7
Gambar 4.11 Pengkodean ulang variabel dengan teknik <i>One-Hot Encoding</i> (lanjutan)	IV-7
Gambar 4.12 Penambahan atribut rate	IV-7
Gambar 4.13 Penggabungan dataset pemuatan dengan dataset cuaca	IV-8
Gambar 4.14 Penggabungan dataset pemuatan dengan dataset cuaca (lanjutan)	IV-9
Gambar 4.15 Hasil uji korelasi dan signifikansi variabel lama pemuatan terhadap rate.....	IV-9
Gambar 4.16 Hasil uji korelasi dan signifikansi variabel jumlah terhadap rate.....	IV-9
Gambar 4.17 Hasil uji korelasi dan signifikansi variabel <i>shift</i> kerja terhadap rate	IV-9
Gambar 4.18 Hasil uji korelasi dan signifikansi variabel jam terhadap rate ..	IV-10
Gambar 4.19 Hasil uji korelasi dan signifikansi variabel <i>temperature</i> terhadap rate	IV-10
Gambar 4.20 Hasil uji korelasi dan signifikansi variabel precipitation terhadap rate	IV-10
Gambar 4.21 Hasil uji korelasi dan signifikansi variabel <i>humidity</i> terhadap rate	IV-10
Gambar 4.22 Hasil uji korelasi dan signifikansi variabel <i>windspeed</i> terhadap rate	IV-10

Gambar 4.23 Model <i>summary</i> prediksi rate pemuatan pupuk	IV-12
Gambar 4.24 Grafik <i>learning curve loss</i>	IV-16
Gambar 4.25 Grafik <i>learning curve</i> MAE	IV-17
Gambar 4.26 Visualisasi <i>scatter plot</i> hasil prediksi dengan data aktual.....	IV-18
Gambar 4.27 Grafik <i>learning curve Hyperparameter</i> dengan <i>trial and error</i>	IV-19
Gambar 4.28 Grafik <i>learning curve Hyperparameter</i> dengan AutoKeras.....	IV-19
Gambar 4.29 Histogram distribusi atribut rate.....	IV-20
Gambar 4.30 Grafik <i>learning curve loss function Hyperparameter</i> dengan penghapusan <i>outlier</i>	IV-22
Gambar 4.31 Grafik <i>learning curve MAE Hyperparameter</i> dengan penghapusan <i>outlier</i>	IV-22
Gambar 4.32 Visualisasi <i>scatter plot</i> hasil prediksi dengan data actual Setelah <i>Hyperparameter</i> dengan metode penghapusan <i>outlier</i>	IV-23
Gambar 5.1 Jumlah pemuatan pupuk tiap Gudang	V-1
Gambar 5.2 Jumlah pupuk yang dimuat	V-2
Gambar 5.3 Jumlah pemuatan pupuk selama tahun 2022.....	V-3
Gambar 5.4 Histogram distribusi variabel numerik.....	V-5
Gambar 5.5 <i>Heatmap</i> untuk visualisasi korelasi antar variabel.....	V-9

BAB I

PENDAHULUAN

Bab ini menjelaskan mengenai latar belakang, rumusan masalah, tujuan penelitian, manfaat penelitian, batasan masalah, asumsi, serta sistematika penulisan laporan kerja praktik yang dilaksanakan di PT Petrokimia Gresik.

1.1 Latar Belakang

Indonesia merupakan negara agraris yang memiliki lahan pertanian yang luas. Sektor pertanian menjadi salah satu sektor yang paling diperhatikan oleh pemerintah karena sektor ini memiliki kontribusi terbesar kedua terhadap pencapaian program Sustainable Development Goals (SDG's) mengenai ketahanan pangan, tidak ada kelaparan, perbaikan nutrisi, serta mendorong budidaya pertanian yang berkelanjutan. Berdasarkan data dari Badan Pusat Statistik luas panen padi di Indonesia pada tahun 2022 diperkirakan sebesar 10,61 juta hektare dengan produksi sekitar 55,67 juta ton GKG yang mana angka ini mengalami peningkatan sebanyak 194,71 ribu hektare atau 1,87 persen dibandingkan luas panen padi di 2021 yang sebesar 10,41 juta hektar. Tak hanya menjadi penyumbang terbesar kedua terhadap pencapaian program SDG's mengenai ketahanan pangan, sektor pertanian juga memiliki peran strategis sebagai penggerak utama ekonomi nasional. Hal ini terbukti dari kontribusi sektor pertanian yang menjadi penyumbang terbesar kedua terhadap Produk Domestik Bruto (PDB) Indonesia yaitu sebesar 13,28 persen pada tahun 2021.

Dengan lahan pertanian yang luas dan potensi yang besar, sektor pertanian di Indonesia memiliki peluang besar untuk terus dikembangkan dan dijadikan sebagai sumber daya ekonomi yang berkelanjutan. Sebagai contoh, produksi padi yang terus meningkat setiap tahunnya menunjukkan bahwa sektor pertanian di Indonesia memiliki potensi untuk memberikan dampak positif bagi pertumbuhan ekonomi dan kesejahteraan masyarakat. Oleh karena itu pemerintah terus berupaya menjalankan sejumlah program guna meningkatkan kapasitas produksi padi/beras nasional.

Untuk mendukung upaya pemerintah mengenai peningkatan kapasitas produksi pertanian salah satu cara yang dapat dilakukan adalah dengan memaksimalkan penggunaan pupuk dalam proses budidaya. Pupuk merupakan salah satu faktor penting dalam meningkatkan produktivitas tanaman dan kualitas hasil panen karena pupuk merupakan sumber unsur hara yang dapat meningkatkan pertumbuhan dan produktivitas tanaman (Mansyur et al., 2021). Penggunaan pupuk yang tepat dan terukur dapat membantu memperbaiki kondisi tanah dan memberikan nutrisi yang cukup bagi pertumbuhan tanaman, sehingga dapat meningkatkan produksi dan kualitas hasil panen (Fatma Adelia et al., 2013).

Melihat pentingnya pupuk untuk tanaman khususnya sektor pertanian ditambah misi pemerintah dalam peningkatan produksi pertanian maka produksi dan distribusi pupuk juga perlu ditingkatkan. Berdasarkan data dari PT Pupuk Indonesia (Persero) tercatat total produksi sebesar 18.842.442 ton (*unaudited*) selama tahun 2022. Hal ini menunjukkan kuantitas produksi yang cukup besar sehingga untuk mencapai tujuan peningkatan produksi pertanian melalui penggunaan pupuk perlu ditopang dengan proses distribusi pupuk yang baik pula.

PT Petrokimia Gresik merupakan salah satu produsen pupuk terbesar di Indonesia yang menjadi bagian dari PT Pupuk Indonesia (Persero). PT Petrokimia Gresik memproduksi berbagai macam jenis pupuk maupun non pupuk seperti Urea, ZA, SP-36, NPK Phonska, DAP, NPK Kebomas, ZK dan Petroganik. Perusahaan ini menjadi salah satu bagian penting dari misi pemerintah dalam peningkatan kapasitas produksi pertanian. Selain melakukan produksi perusahaan juga perlu melakukan *warehousing* dan juga mendistribusikan produknya kepada konsumen secara cepat dan efisien.

Namun, proses distribusi pupuk PT Petrokimia Gresik melibatkan biaya yang signifikan, terutama dalam kegiatan bongkar muat produk dari dan ke gudang menggunakan jasa perusahaan rekanan tenaga kerja bongkar muat (TKBM). Untuk mengoptimalkan penggunaan sumber daya dan meminimalkan biaya, perusahaan perlu melakukan alokasi jumlah truk yang optimal dalam tiap *shift* distribusi.

Dalam rangka mencapai tujuan tersebut, informasi mengenai rate pemuatan pupuk oleh TKBM pada tiap *shift* menjadi penting. Dengan mengetahui informasi ini, PT Petrokimia Gresik dapat membuat keputusan strategis yang tepat dalam pengalokasian jumlah truk pada tiap *shift*, serta melakukan prediksi untuk masa depan. Salah satu pendekatan yang dapat digunakan untuk melakukan prediksi ini adalah dengan menggunakan teknik machine learning.

Machine learning adalah suatu cabang dari kecerdasan buatan yang memungkinkan komputer untuk belajar dari data dan pengalaman masa lalu. Dengan menggunakan algoritma machine learning, perusahaan dapat menganalisis data historis mengenai rate pemuatan pupuk, serta faktor-faktor yang mempengaruhinya seperti kondisi cuaca, waktu pemuatan, dan lainnya. Dengan demikian, perusahaan dapat membangun model prediksi yang dapat digunakan untuk memprediksi rate pemuatan pupuk di masa depan.

Dengan adanya prediksi yang akurat, PT Petrokimia Gresik dapat mengoptimalkan alokasi jumlah truk, mengurangi biaya operasional, serta meningkatkan efisiensi dalam proses distribusi pupuk. Selain itu, hasil analisis dan prediksi yang diperoleh dari model machine learning juga dapat memberikan wawasan dan pemahaman yang lebih mendalam mengenai faktor-faktor yang berpengaruh pada rate pemuatan pupuk, sehingga dapat membantu perusahaan dalam pengambilan keputusan strategis yang lebih baik.

Dalam rangka mencapai tujuan tersebut, penelitian ini akan mengadopsi kerangka CRISP-DM (Cross-Industry Standard Process for Data Mining) yang terdiri dari beberapa tahap, termasuk pemahaman bisnis, pemahaman data, pemodelan, evaluasi, dan penyebaran. Melalui penerapan kerangka CRISP-DM, diharapkan penelitian ini dapat memberikan pemahaman yang komprehensif mengenai faktor-faktor yang mempengaruhi rate pemuatan pupuk, serta menyediakan rekomendasi yang tepat bagi PT Petrokimia Gresik dalam mengoptimalkan distribusi pupuk secara efisien.

1.2 Rumusan Masalah

Rumusan masalah yang akan dibahas dalam penelitian yang dilakukan selama melaksanakan kerja praktik di PT Petrokimia Gresik adalah sebagai berikut:

1. Apa saja faktor-faktor yang berpotensi mempengaruhi rate pemuatan pupuk?
2. Bagaimana hubungan antara variabel independen dengan variabel dependen yaitu rate pemuatan pupuk?
3. Bagaimana performa model prediksi yang dikembangkan dalam memprediksi rate pemuatan pupuk?

1.3 Tujuan Penelitian

Tujuan penelitian yang dilakukan selama melaksanakan kerja praktik di PT Petrokimia Gresik adalah sebagai berikut:

1. Mengidentifikasi faktor-faktor yang berpotensi mempengaruhi rate pemuatan pupuk di PT Petrokimia Gresik..
2. Menganalisis hubungan antara variabel independen dengan variabel dependen yaitu rate pemuatan pupuk.
3. Membangun model prediksi dan mengevaluasi performa model yang dikembangkan dalam memprediksi rate pemuatan pupuk.

1.4 Manfaat Penelitian

Manfaat penelitian yang diperoleh selama melaksanakan kerja praktik di PT Petrokimia Gresik adalah sebagai berikut:

1. Bagi Perusahaan
 - a. Hasil penelitian ini dapat memberikan wawasan dan informasi yang lebih akurat kepada PT Petrokimia Gresik dalam pengambilan keputusan terkait pengelolaan distribusi pupuk.

- b. Hasil penelitian ini berguna untuk peningkatan efisiensi operasional, dengan adanya model prediksi yang dapat memperkirakan tingkat pemuatan pupuk di masa depan, perusahaan dapat mengoptimalkan alokasi jumlah truk secara lebih efisien, sehingga dapat mengurangi biaya operasional yang terkait dengan distribusi pupuk.
 - c. Peningkatan profitabilitas: Dengan mengurangi biaya operasional melalui pengoptimalan alokasi truk, PT Petrokimia Gresik dapat meningkatkan profitabilitasnya. Penggunaan yang lebih efisien dari sumber daya dan pengurangan biaya dapat berdampak positif pada margin keuntungan perusahaan.
2. Bagi Mahasiswa
- a. Penelitian ini memberikan kesempatan bagi peneliti untuk mengembangkan keterampilan dalam menganalisis data, membangun model prediksi, dan melakukan evaluasi terhadap hasil penelitian.
 - b. Peneliti mendapatkan pengalaman praktis dalam mengaplikasikan konsep-konsep teoritis ke dalam suatu studi kasus nyata dalam industri pupuk.

1.5 Batasan Masalah

Batasan masalah dalam penelitian yang dilakukan selama melaksanakan kerja praktik di PT Petrokimia Gresik adalah sebagai berikut:

- 1. Model prediksi hanya untuk produk *in-bag* yang dikerjakan oleh tenaga kerja bongkar muat (TKBM)
- 2. Penelitian tidak mempertimbangkan kondisi alat berat *forklift*
- 3. Penelitian hanya terbatas pada gudang lini 1 PT Petrokimia Gresik.

1.6 Asumsi

Asumsi dalam penelitian yang dilakukan selama melaksanakan kerja praktik di PT Petrokimia Gresik adalah sebagai berikut:

- 1. Pekerja TKBM bekerja secara wajar dan memiliki kemampuan yang normal.

2. Jumlah TKBM yang melayani pemuatan pupuk diasumsikan selalu sama

1.7 Sistematika Penulisan

Sistematika penulisan yang digunakan pada penulisan laporan penelitian selama melaksanakan kerja praktik di PT Petrokimia Gresik adalah sebagai berikut:

BAB I PENDAHULUAN

Bab ini membahas mengenai pengantar dari permasalahan yang akan dikaji dalam laporan kerja praktik ini. Bab 1 pendahuluan ini menjelaskan latar belakang, rumusan masalah, tujuan penelitian, manfaat penelitian, batasan masalah, asumsi, dan sistematika penulisan laporan kerja praktik.

BAB II TINJAUAN PUSTAKA

Bab ini menjelaskan mengenai tinjauan umum dari perusahaan yang menjadi tempat pelaksanaan kerja praktik serta membahas teori-teori yang berkaitan dengan penelitian untuk dijadikan acuan dalam menganalisis dan menyelesaikan masalah.

BAB III METODOLOGI PENELITIAN

Bab ini menjelaskan mengenai tahapan-tahapan terstruktur yang digunakan dalam penelitian dan disajikan bentuk flowchart beserta penjelasan mengenai setiap tahapan tersebut.

BAB IV PENGUMPULAN DAN PENGOLAHAN DATA

Bab ini membahas mengenai pengumpulan data-data perusahaan yang diperlukan untuk penelitian serta menjelaskan proses pengolahan data sesuai dengan metode yang ditentukan untuk menyelesaikan masalah yang dibahas.

BAB V ANALISIS DAN INTERPRETASI HASIL

Bab ini membahas mengenai analisis permasalahan yang dirumuskan dan interpretasi hasil pengolahan data yang telah dilakukan.

BAB VI KESIMPULAN DAN SARAN

Bab ini menjelaskan kesimpulan penelitian sesuai masalah yang dirumuskan serta berisi saran untuk perusahaan maupun penelitian selanjutnya sesuai topik yang dikaji.

BAB II

TINJAUAN PUSTAKA

Bab ini menjelaskan mengenai tinjauan umum perusahaan tempat kerja praktik yaitu PT. Petrokimia Gresik dan membahas landasan teori yang mengacu pada tema yang dibahas dalam laporan kerja praktik ini.

2.1 Tinjauan Umum Perusahaan

2.1.1 Profil Perusahaan

Berikut merupakan logo perusahaan PT Petrokimia Gresik:



Gambar 2.2.1 Logo Perusahaan PT Petrokimia Gresik

Nama Perusahaan : PT Petrokimia Gresik

Alamat Perusahaan : Jl. Jenderal Ahmad Yani, Ngipik, Kabupaten
Gresik, Jawa Timur, 61119

No.Telp / Fax : (031) 3981811

Tahun berdiri : 10 Juli 1972

Alamat *Website* : <https://petrokimia-gresik.com/>

Bidang Usaha : Industri Manufaktur Pupuk

Makna Logo :

- a. Inspirasi logo PT Petrokimia Gresik adalah seekor kerbau berwarna kuning keemasan yang berdiri tegak di atas kelopak daun yang berujung lima dengan tulisan warna putih di bagian tengah.

- b. Seekor kerbau berwarna kuning keemasan atau dalam bahasa Jawa sebagai Kebomas yaitu penghargaan perusahaan kepada daerah di mana PT Petrokimia Gresik berdomisili, yakni di Kecamatan Kebomas di Kabupaten Gresik. Kerbau ini juga melambangkan simbol sahabat petani yang memiliki sifat loyal, tidak bias, pemberaani, dan giat bekerja.
- c. Kelopak daun hijau berujung lima yang melambangkan kelima sila Pancasila, sedangkan tulisan PT Petrokimia Gresik merupakan singkatan dari Petrokimia Gresik.
- d. Warna kuning keemasan pada gambar kerbau merepresentasikan keagungan, kejayaan, dan keluhuran budi. Padan hijau pada kelopak daun berujung lima menggambarkan kesuburuan dan kesejahteraan.
- e. Tulisan PT Petrokimia Gresik berwarna putih mencerminkan kesucian, kejujuran, dan kemurnian, sedangkan batas hitam pada sekuruh komponen logo merepresentasikan kewibawaan dan elegan.
- f. Warna hitam pada penulisan nama perusahaan melambangkan kedalaman, stabilitas, dan keyakinan yang teguh. Nilai-nilai yang selalu mendukung seluruh proses kerja.

2.1.2 Sejarah Perusahaan

Berikut merupakan diagram sejarah perusahaan PT Petrokimia Gresik beserta penjelasannya:



Gambar 2.2.2 Sejarah Perusahaan PT Petrokimia Gresik

PT Petrokimia Gresik merupakan sebuah pabrik pupuk terlengkap di Indonesia yang awal berdirinya disebut “Proyek Petrokimia Surabaya”. Setelah itu, kontrak pembangunannya ditandatangani pada tanggal 8 Desember 1964. Proyek ini diresmikan oleh Presiden RI, HM. Soeharto tanggal 10 Juli 1972 yang kemudian pada tanggal tersebut ditetapkan sebagai hari jadi PT Petrokimia Gresik.

PT Petrokimia Gresik saat ini memiliki luas area lebih dari 450 hektar di Kabupaten Gresik, Jawa Timur. Total produksi per 31 Oktober 2022 mencapai 8,9 juta ton/tahun yang terdiri produk pupuk yang dihasilkan sebanyak 5 (lima) juta ton/tahun dan produk non pupuk sebanyak 3,9 juta ton/tahun. PT Petrokimia Gresik ini merupakan anak perusahaan dari PT Pupuk Indonesia (Persero) yang bertransformasi menuju perusahaan “Solusi Agroindustri” dalam mendukung tercapainya program ketahanan pangan nasional dan kemajuan dunia pertanian.

Struktur pemegang saham di PT Petrokimia Gresik adalah PT Pupuk Indonesia (Persero) yang memiliki lembar saham sebanyak 2.393.033 atau senilai dengan Rp2.393.033.000.000 (99,97550 dan Yayasan Petrokimia Gresik memiliki 60 lembar saham atau senilai dengan Rp60.000.000 (0,0025%). Jumlah karyawan di Petrokimia Gresik per 31 Oktober 2022 sebanyak 1.957 karyawan.

2.1.3 Visi Misi Perusahaan

Visi misi Petrokimia Gresik yaitu:

a. Visi:

Menjadi produsen pupuk dan produk kimia lainnya yang berdaya saing tinggi dan produknya paling diminati oleh konsumen.

g. Misi:

1. Mendukung penyediaan pupuk nasional untuk tercapainya prohran swasembada pangan
2. Meningkatkan hasil usaha untuk menunjang kelancaran kegiatan operasional dan pengembangan usaha perusahaan.
3. Mengembangkan potensi usaha untuk mendukung industri kimia nasional dan berperan aktif dalam *community development*

2.1.4 Tata Nilai Perusahaan

Berikut merupakan tata nilai PT Petrokimia Gresik:

1. Amanah

- Definisi: memegang teguh kepercayaan yang diberikan.
- Panduan perilaku:
 - Memenuhi janji dan komitmen
 - Bertanggung jawab atas tugas, keputusan, dan tindakan yang dilakukan
 - Berpegang teguh kepada nilai moral dan etika

2. Kompeten

- Definisi: terus belajar dan mengembangkan kapabilitas.
- Panduan perilaku:
 - Meningkatkan kompetensi diri untuk menjawab tantangan yang selalu berubah.
 - Membantu orang lain belajar.
 - Menyelesaikan tugas dengan kualitas terbaik.

3. Harmonis

- Definisi: saling peduli dan menghargai perbedaan.
- Panduan perilaku:
 - Menghargai setiap orang apapun latar belakangnya.
 - Suka menolong orang lain.
 - Membangun lingkungan kerja yang kondusif.

4. Loyal

- Definisi: berdedikasi dan mengutamakan kepentingan bangsa dan negara.
- Panduan perilaku:
 - Menjaga nama baik sesama karyawan, pimpinan BUMN, dan negara.
 - Rela berkorban untuk mencapai tujuan yang lebih besar.
 - Patuh kepada pimpinan sepanjang tidak bertentangan dengan hukum dan etika.

5. Adaptif

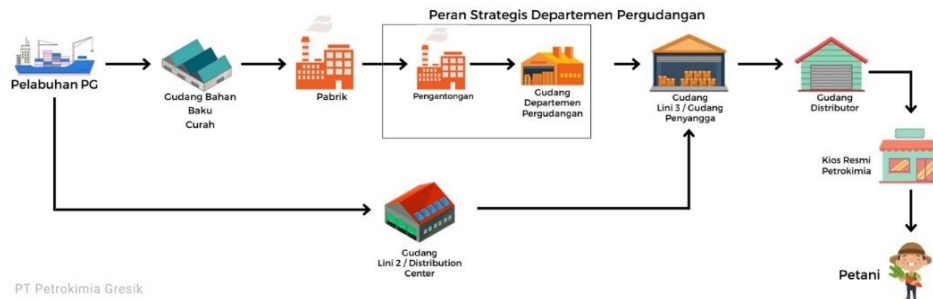
- Definisi: terus berinovasi dan antusias dalam menggerakkan ataupun menghadapi perubahan.
- Panduan perilaku:
 - Cepat menyesuaikan diri untuk menjadi lebih baik.
 - Terus-menerus melakukan perbaikan mengikuti perkembangan teknologi.
 - Bertindak proaktif.

6. Kolaboratif

- Definisi: membangun kerja sama yang sinergis.

2.1.6 Alur Distribusi

Berikut merupakan alur distribusi di PT Petrokimia Gresik dengan atau tanpa gudang penyangga:



Gambar 2.2.4 Alur Distribusi Perusahaan

Alur distribusi pengiriman pupuk yang berasal dari PT Petrokimia Gresik berasal dari Pelabuhan PG yang dikirim melalui kapal. Selanjutnya, akan dikirim melalui GBB atau Gudang Bahan Baku ataupun juga dapat langsung melalui DC (*Distribution Center*). Apabila kapal PG mengirimkan ke GBB, berarti kapal tersebut memuat bahan-bahan baku pupuk yang nantinya akan diproses melalui pabrik. Apabila pengiriman dari Pelabuhan PG langsung menuju DC, berarti kapal PG tersebut memuat pupuk jadi yang berasal dari anak-anak perusahaan Pupuk Indonesia. Setelah diproduksi di pabrik, selanjutnya akan dikirim ke UPP atau Unit Pengantongan Pupuk dimana pupuk-pupuk yang masih *fresh* berasal dari pabrik dalam kondisi curah. Setelah pupuk-pupuk tersebut dalam kondisi in bag, selanjutnya akan dikirim ke gudang departemen pergudangan. Tujuan pengiriman ke departemen pergudangan selain untuk proses penyimpanan juga untuk menunggu kendaraan yang digunakan untuk mengirim pupuk-pupuk. Transportasi yang digunakan untuk mengirim pupuk-pupuk ke gudang lini 3 atau gudang penyangga menggunakan truck. Pengiriman pupuk-pupuk ke gudang penyangga bertujuan agar pupuk-pupuk tersebut dikirim ke kabupaten-kabupaten yang berada di Indonesia, begitu juga dari pupuk-pupuk yang berasal dari DC akan langsung dikirimkan melalui gudang penyangga. Pupuk-pupuk yang telah berada di gudang penyangga, selanjutnya akan dikirim kepada gudang distributor yang akan dikirimkan ke retail atau kios-kios terdekat untuk nantinya petani atau konsumen dapat melakukan transaksi pembelian pupuk.

2.1.7 Wilayah Distribusi

Berikut merupakan wilayah distribusi untuk pupuk bersubsidi yang berada di PT Petrokimia Gresik:

1. Pupuk ZA: seluruh Indonesia
2. Pupuk SP- 36: seluruh Indonesia

3. Pupuk Urea: terletak di 28 kabupaten dan kota di Jawa Timur dan di seluruh kabupaten /kota di NTT, Maluku, Maluku Utara, Papua, Papua Barat
4. Pupuk Phonska: di seluruh Indonesia kecuali di Sumatera Selatan dan Lampung. Untuk daerah Jawa Barat tersebar di 10 kabupaten/kota.
5. Pupuk Organik Petroganik: di seluruh Indonesia kecuali di DKI Jakarta, Banten, Jawa Barat, dan 4 kabupaten/kota di Jawa Tengah.
6. Pupuk Phonska Oca: terletak di 11 provinsi di Indonesia

2.1.8 Ruang Lingkup Gudang Perusahaan

Ruang lingkup pergudangan yang berada di PT Petrokimia Gresik terbagi atas 4 (empat) gudang di antaranya:

Gudang I

1. BS Belerang
2. BS ZA 1/3
3. BS Urea 1A
4. Gudang Urea 1A
5. Pengantongan Urea 1A
6. BS ZA 2
7. Gudang ZA
8. Pengantongan ZA
9. Kaptan
10. Mixtro
11. Gudang Urea 1B
12. Pengantongan Urea 1B
13. BS Urea 1B
14. Gudang Sparepart PPBJ

Gudang III

1. GBB A
2. GBB B
3. GBB C
4. PetroCas
5. GMG 1
6. GMG
7. Jembatan Timbang IV
8. Jembatan Timbang V
9. Jembatan Timbang II
10. Jembatan Timbang I
11. Jembatan Timbang III
12. KIG Q
13. KIG Beton

Gudang II

1. Gudang PF 1
2. 02 400
3. Pengantongan PF I
4. Pengantongan Phonska 2&3
5. Conveyor System area Gudang II
6. Curing PF I
7. 02 A650
8. Gudang Phonska III
9. 09 400
10. Pengantongan Phonska I
11. 09-650

Gudang IV

1. Produk Pengembangan
2. Curing PF 2
3. 22-400
4. Gudang PF 2
5. Pengantongan Phonska 4
6. Conveyor System area Gudang IV
7. Alf 3
8. Curah 50.000
9. Dome
10. PA
11. Coal
12. Sulphur
13. Puri Gypsum 1
14. CR / Gypsum
15. Puri Gypsum 2

Gambar 2.2.5 Ruang Lingkup Departemen Pergudangan

2.1.9 Transformasi Digital pada Pergudangan

1. *Warehouse Management System (WMS)*

WMS atau Warehouse Management System merupakan sistem manajemen terintegrasi yang digunakan untuk pengelolaan stok, tata letak, alat berat, penerimaan produk, pengelolaan & pengeluaran

produk, utilitas pada proses penerimaan dan pengeluaran sebagai bagian dari pengelolaan dan pengendalian dengan penerapan SCA (*supply chain analytics*). WMS ini terdiri dari 2 versi dimana WMS 2CE ditujukan untuk sarana peningkatan kepuasan pelanggan dalam pelayanan muatan produk non subsidi di pergudangan. Fitur utama lainnya yaitu digitalisasi PBB, antrian / posisi truk, digitalisasi laporan & posisi gudang. Berikut merupakan gambar / logo dari aplikasi WMS:



Gambar 2.2.6 Logo Aplikasi WMS Perusahaan

2. *Digital Transport Management System (DTMS)*

DTMS merupakan sistem manajemen transportasi internal terintegrasi untuk pengelolaan order truk internal (curah dan *in bag*) dan manajemen kinerja berbasis aplikasi android dan website. DTMS ini mengambil dan mengolah data transaksi antar gawai DTMS dan bluetooth cubeacon yang dipasang pada setiap check point (gudang, jembatan timbang, dan pelabuhan) yang ditampilkan dalam dashboard kinerja sistem transport internal. Berikut merupakan logo aplikasi DMTS milik PT Petrokimia Gresik:



Gambar 2.2.7 Logo Aplikasi DTMS Perusahaan

3. *System Scheduling Truck Online (SISTRO)*

Sistem integrasi untuk penjadwalan truk, informasi kesediaan stok dan buruh di terminal PKG dimana pada aplikasi ini dilengkapi dengan pembagian alokasi kuota muat di setiap *shift* untuk mengatasi keterlambatan kedatangan truk. Aplikasi SISTRO ini mampu melacak keberadaan truk saat proses muat karena dilengkapi dengan *security*

pass yang dilengkapi dengan *barcode* untuk dapat discan diseluruh *check point*. Aplikasi SISTRO ini terinspirasi dari aplikasi TIX-ID di mana pada aplikasi TIX-ID, kita sebagai pengguna yang ingin menonton film apa yang kita ingin tonton dan juga hari, jam, dan tanggal berapa. Hal ini juga berlaku apa aplikasi SISTRO yang cara kerjanya hampir mirip seperti TIX-ID. Berikut merupakan logo dari aplikasi SISTRO:



Gambar 2.2.8 Logo Aplikasi Sistro Perusahaan

2.1.10 Produk yang dihasilkan perusahaan

Produk-produk yang dihasilkan oleh PT Petrokimia Gresik terbagi atas 2 jenis produk, yaitu pupuk dan non pupuk. Untuk produk pupuk terdapat 2 jenis pupuk yang disubsidi oleh pemerintah yaitu pupuk urea dan pupuk phonska. Berikut merupakan beberapa contoh produk pupuk yang dihasilkan oleh Petrokimia Gresik:



Gambar 2.2.9 Produk-produk Pupuk Perusahaan

(Sumber: <https://petrokimia-gresik.com/product-category/pupuk>)

Selanjutnya, yaitu untuk produk non pupuk yang dihasilkan oleh Petrokimia Gresik. Pada produk-produk non pupuk yang dihasilkan oleh PT Petrokimia Gresik, mereka menghasilkan beberapa produk, seperti bahan

kimia, makanan hewan ternak, dan juga menghasilkan jasa (*engineer*, pelabuhan, laboratorium). Berikut merupakan beberapa contoh produk non pupuk yang dihasilkan oleh Petrokimia Gresik:



Gambar 2.10 Produk-produk Non Pupuk Perusahaan

(Sumber: <https://petrokimia-gresik.com/product-category/pupuk/>)

2.2 Tinjauan Pustaka

2.2.1 Data Mining

Data mining adalah proses analisis data yang melibatkan penggalian informasi yang berguna dari kumpulan data besar (M et al., 2018). Tujuannya adalah untuk mengidentifikasi pola-pola dalam data input dan membangun hubungan yang dapat digunakan untuk memecahkan masalah yang ada. Dengan menggunakan berbagai metode analisis, data mining membantu menghasilkan pengetahuan yang berharga dari data yang ada di dalam database. Dalam banyak bidang seperti penelitian, pendidikan, rumah sakit, dan pemasaran, teknik data mining digunakan untuk mengungkapkan pola-pola tersembunyi yang ada dalam data dan mengkategorikan informasi menjadi wawasan yang bermanfaat. Contoh teknik *data mining* yang umum digunakan yaitu regresi, klasifikasi, dan *clustering*.

2.2.2 CRISP-DM

CRISP-DM (*Cross-Industry Standard Process for Data Mining*) adalah suatu metode atau kerangka kerja yang digunakan dalam proses data mining. Metode ini melibatkan penggalian informasi yang berguna dari kumpulan data besar dengan tujuan mengidentifikasi pola-pola dalam data input dan membangun hubungan yang dapat digunakan untuk memecahkan masalah yang ada. Dalam banyak bidang seperti penelitian, pendidikan, rumah sakit,

dan pemasaran, teknik data mining digunakan untuk mengungkapkan pola-pola tersembunyi dalam data dan mengkategorikan informasi menjadi wawasan yang bermanfaat. Contoh teknik data mining yang umum digunakan adalah regresi, klasifikasi, dan clustering. Melalui pendekatan yang sistematis, CRISP-DM membantu mengelola proyek *data mining* dari awal hingga akhir dengan enam tahap yang saling terkait yaitu *Business Understanding*, *Data Understanding*, *Data Preparation*, *Modeling*, *Evaluation*, dan *Deployment* (Abasova et al., 2021).

2.2.3 Deep Learning

Deep learning adalah sebuah metode dalam bidang kecerdasan buatan (*artificial intelligence*) yang menghasilkan hasil yang lebih bagus dalam pemrosesan data yang kompleks. Berbeda dengan teknik *machine learning* tradisional, *deep learning* menggunakan jaringan saraf tiruan yang dalam dan memiliki banyak lapisan (*layer*) untuk mempelajari representasi data secara otomatis dan lebih baik (Sakellariopoulos et al., 2019). *Deep learning* mampu mengekstraksi fitur-fitur yang lebih abstrak dan kompleks dari data, sehingga dapat menghasilkan prediksi yang lebih akurat dan wawasan yang lebih mendalam. *Deep learning* dapat diaplikasikan kedalam banyak kasus *machine learning* seperti pengenalan wajah, pengenalan suara, dan pemrosesan bahasa alami (*Natural Language Processing*). Dengan kemampuannya yang unggul, *deep learning* menjadi teknik yang sangat relevan dan berpotensi besar untuk mengatasi tantangan pemrosesan data yang kompleks.

2.2.4 Artificial Neural network

Artificial Neural network (ANN) adalah komponen dasar dalam *deep learning* yang digunakan untuk memecahkan masalah yang kompleks dengan memanfaatkan persamaan nonlinear (Tahwia et al., 2021). Metode ini terinspirasi dari sistem saraf biologis manusia dan telah menjadi salah satu pendekatan yang populer dalam bidang kecerdasan buatan. Dalam ANN, terdapat sejumlah unit pemrosesan sederhana yang disebut *neuron*, yang saling terhubung membentuk struktur jaringan. Melalui proses pembelajaran dengan menggunakan data, ANN dapat menyesuaikan bobot koneksi antar neuron untuk mengenali pola dan membuat prediksi berdasarkan data yang diberikan. Keunggulan utama ANN terletak pada kemampuannya dalam memodelkan kompleksitas masalah yang *nonlinier* dan adaptabilitasnya terhadap data yang diberikan. Dengan demikian, ANN telah digunakan dalam berbagai aplikasi seperti pengenalan pola, pengenalan suara, pengolahan citra, dan prediksi dalam berbagai domain. Sebagai metode yang canggih, ANN telah membuktikan keunggulannya dalam menyelesaikan masalah yang rumit dalam berbagai bidang ilmu pengetahuan dan teknologi.

2.2.5 Metrics Evaluation

Metrics evaluation adalah kriteria atau ukuran yang digunakan untuk mengevaluasi kinerja dan keakuratan model machine learning. *Metrics evaluation* penting karena dapat memberikan informasi tentang sejauh mana model mampu melakukan prediksi yang akurat dan sejauh mana model tersebut dapat menjelaskan variasi dalam data. Selain itu *metrics evaluation* juga dapat memberikan umpan balik tentang kelemahan model yang dikembangkan sehingga dapat dianalisis penyebabnya dan dilakukan perbaikan. Pada penelitian ini *metrics evaluation* yang digunakan adalah *Mean Squared Error*, *R-Squared*, *Mean Absolute Error*, *Root Mean Squared Error*, dan *Adjusted R-Squared*.

2.2.5.1 Mean Squared Error

Mean Squared Error (MSE) mengukur rata-rata dari perbedaan kuadrat antara nilai prediksi dan nilai sebenarnya. MSE memberikan informasi tentang tingkat kesalahan prediksi secara keseluruhan. MSE mencerminkan keakuratan keseluruhan model dengan semakin rendah nilai MSE menunjukkan prediksi yang lebih baik. Berikut adalah rumus untuk menghitung MSE.

$$MSE = \frac{1}{N} \sum_{i=1}^N (y_i - \hat{y}_i)^2$$

N = jumlah data

y_i = nilai aktual data ke-i

\hat{y}_i = nilai prediksi data ke-i

2.2.5.2 R-Squared

R-Squared memberikan gambaran tentang sejauh mana variabilitas variabel dependen dapat dijelaskan oleh variabel independen yang ada dalam model. Nilai R^2 berkisar antara 0 hingga 1, di mana nilai yang lebih tinggi menunjukkan model yang dapat menjelaskan variasi data dengan lebih baik. Berikut adalah rumus untuk menghitung *R-Squared*

$$R^2 = 1 - \frac{\text{sum of squared prediction}}{\text{sum of squared total}} = \frac{\sum_i (y_i - \hat{y}_i)^2}{\sum_i (y_i - \bar{y}_i)^2}$$

\bar{y}_i = nilai rata-rata

y_i = nilai aktual data ke-i

\hat{y}_i = nilai prediksi data ke-i

2.2.5.3 Mean Absolute Error

Mean Absolute Error (MAE) mengukur rata-rata perbedaan absolut antara nilai prediksi dan nilai sebenarnya. Metrik ini memberikan gambaran tentang tingkat kesalahan prediksi secara keseluruhan tanpa

memperhatikan arah perbedaan. Semakin rendah nilai MAE maka semakin baik model dalam melakukan prediksi. Berikut adalah rumus untuk menghitung MAE.

$$MAE = \frac{1}{N} \sum_{i=1}^N |y_i - \hat{y}_i|$$

N = jumlah data

y_i = nilai aktual data ke- i

\hat{y}_i = nilai prediksi data ke- i

2.2.5.4 *Root Mean Squared Error*

Root Mean Squared Error (RMSE) merupakan akar kuadrat dari MSE dan memberikan gambaran tentang tingkat kesalahan prediksi dalam satuan yang sama dengan variabel dependen. RMSE memberikan informasi yang lebih intuitif tentang tingkat kesalahan prediksi. Nilai RMSE yang lebih rendah menunjukkan bahwa model memiliki tingkat kesalahan prediksi yang lebih kecil. Model dengan RMSE yang rendah lebih akurat dalam melakukan prediksi.

$$RMSE = \sqrt{MSE}$$

MSE = *mean squared error*

2.2.5.5 *Adjusted R-Squared*

Adjusted R-Squared adalah variasi dari *R-Squared* yang memperhitungkan jumlah variabel independen dalam model. Metrik ini membantu mengurangi penyesuaian yang tidak signifikan ketika menambahkan variabel independen tambahan ke dalam model. Nilai *Adjusted R-Squared* berkisar antara 0 hingga 1. Nilai yang lebih tinggi menunjukkan bahwa model secara lebih baik dapat menjelaskan variasi dalam data.

$$Adjusted R^2 = \left\{ 1 - \left[\frac{(1 - R^2)(n - 1)}{n - k - 1} \right] \right\}$$

R^2 = *R-Squared*

n = jumlah total observasi.

k = jumlah variabel independen dalam model.

2.2.6 *Outlier*

Outlier merujuk pada data atau observasi yang memiliki nilai yang jauh berbeda dengan mayoritas data yang ada. Kehadiran outlier dalam kumpulan data dapat menyebabkan peningkatan terjadinya alarm positif palsu dan mempengaruhi hasil analisis data yang dilakukan (Vinutha et al., 2018). Salah satu metode yang sering digunakan untuk mendeteksi outlier

adalah metode IQR (Interquartile Range). Metode ini mengukur sebaran data dengan memanfaatkan konsep kuartil, data yang dianggap outlier adalah data yang berada diluar batas atas ataupun batas bawah. Rumus untuk mencari outlier dalam data menggunakan metode IQR adalah sebagai berikut:

$$IQR = Q3 - Q1$$

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

$$Batas\ Atas = Q3 + (1.5 \times IQR)$$

Keterangan:

Q3 = kuartil atas

Q1 = kuartil bawah

IQR = *Interquartile Range*

2.2.7 One-Hot Encoding

Mesin atau komputer hanya memiliki kemampuan untuk memproses dan memahami data numerik, sehingga data yang masih dalam bentuk kategorikal perlu dilakukan pengkodean variabel agar dapat dipahami algoritma machine learning sehingga meningkatkan kualitas model (De La Bourdonnaye & Daniel, 2021). Salah satu metode yang digunakan untuk pengkodean variabel kategorikal adalah One-Hot Encoding. One-Hot Encoding adalah teknik yang mengubah variabel kategorikal menjadi representasi numerik biner. Dalam metode ini, setiap kategori pada variabel kategorikal diubah menjadi vektor biner dengan panjang yang sama dengan jumlah kategori yang ada. Nilai 1 ditempatkan pada posisi yang sesuai untuk menunjukkan kehadiran kategori tersebut, sementara nilai 0 digunakan untuk menunjukkan ketidakhadirannya. Kelebihan One-Hot Encoding terletak pada fakta bahwa metode ini tidak menghasilkan urutan numerik yang dapat mempengaruhi model dalam menarik kesimpulan yang salah. Selain itu, One-Hot Encoding juga memungkinkan model untuk mempelajari hubungan yang setara antara kategori-kategori tersebut tanpa memberikan bobot lebih pada satu kategori dibandingkan yang lain.

2.2.8 Analisis Hubungan dan Hipotesis

Sebelum melakukan *modelling* perlu dilakukan analisis korelasi dan signifikansi antara variabel independen dan variabel dependen, untuk menguji nilai korelasi dan signifikansi menggunakan uji korelasi pearson. Rumus untuk menghitung koefisien korelasi Pearson (r) adalah sebagai berikut:

$$r = \frac{\sum(X_i - \bar{X})(Y_i - \bar{Y})}{\sqrt{\sum(X_i - \bar{X})^2 \sum(Y_i - \bar{Y})^2}}$$

x_i = nilai actual data ke-i pada variabel independen

y_i = nilai aktual data ke-I pada variabel dependen

\bar{x} = nilai rata-rata variabel independen

\bar{y} = nilai rata-rata variabel dependen

Koefisien korelasi Pearson berkisar antara -1 hingga 1. Nilai positif menunjukkan korelasi positif, nilai negatif menunjukkan korelasi negatif, dan nilai mendekati 0 menunjukkan tidak adanya korelasi linier antara variabel.

Setelah dilakukan uji korelasi selanjutnya adalah uji signifikansi untuk melihat apakah hubungan yang diamati antara variabel independen dan variabel dependen adalah signifikan secara statistik dalam sampel yang dianalisis. Dalam konteks pengujian signifikansi hubungan korelasi menggunakan korelasi Pearson, hipotesis nol (H_0) menyatakan bahwa tidak ada hubungan antara dua variabel yang diuji. Hipotesis alternatif (H_1) menyatakan bahwa terdapat hubungan antara kedua variabel. Pada tahap pengujian, dilakukan perhitungan statistik uji (statistik uji-t) berdasarkan sampel yang digunakan. Nilai *p-value* kemudian dihitung, Berdasarkan nilai *p-value* yang diperoleh dapat diambil keputusan untuk menerima atau menolak hipotesis nol. Jika nilai *p-value* lebih kecil dari tingkat signifikansi yang ditetapkan sebelumnya ($\alpha = 0,05$), maka hipotesis nol dapat ditolak dan disimpulkan bahwa terdapat hubungan linier yang signifikan antara kedua variabel dalam populasi. Sebaliknya, jika nilai *p-value* lebih besar dari tingkat signifikansi, maka tidak ada cukup bukti statistik untuk menolak hipotesis nol dan tidak dapat disimpulkan adanya hubungan linier yang signifikan. Berikut adalah rumus untuk menghitung uji-t

$$t = \frac{r \times \sqrt{n-2}}{\sqrt{1-r^2}}$$

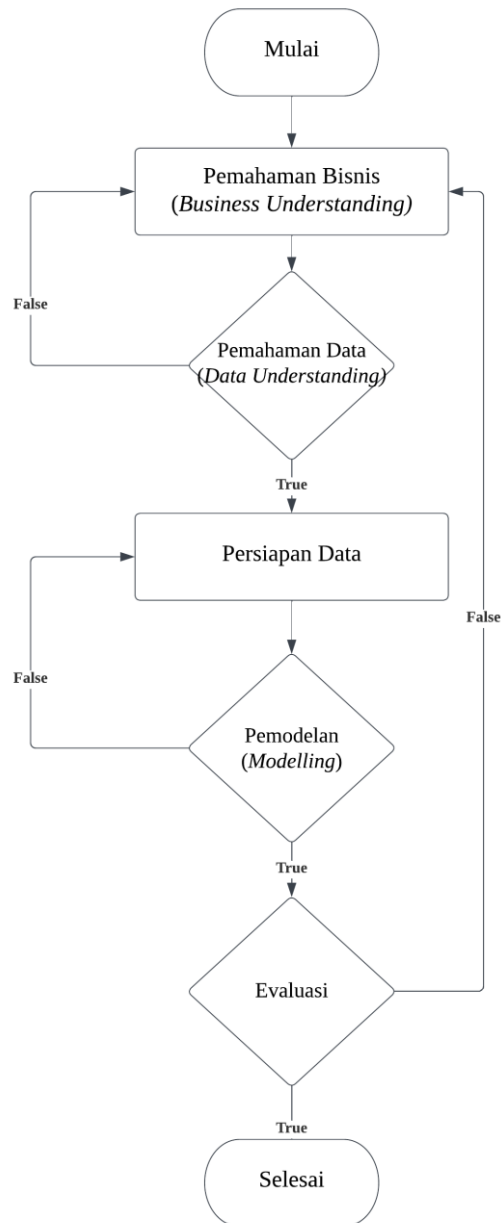
r = koefisien korelasi Pearson

n = jumlah sampel observasi

BAB III

METODOLOGI PENELITIAN

Bab ini menjelaskan mengenai tahapan metodologi penelitian yang dilakukan di PT Petrokimia Gresik yang terdiri dari *flowchart* penelitian beserta penjelasannya. Berikut *flowchart* metodologi penelitian yang dilakukan.



Gambar 3.3.1 *Flowchart* Metodologi Penelitian

3.1 Pemahaman Bisnis (*Business Understanding*)

Pada tahap ini fokus utama adalah untuk memperoleh pemahaman yang komprehensif tentang tujuan dan persyaratan bisnis yang terlibat. Informasi yang diperoleh kemudian digunakan untuk merumuskan definisi masalah dalam data mining serta menyusun rencana awal yang bertujuan untuk mencapai tujuan bisnis tersebut. Pemahaman bisnis ini sangat bergantung pada komunikasi yang efektif dan kejelasan yang tinggi. Tantangannya adalah bahwa setiap pemangku kepentingan mungkin memiliki tujuan, perspektif, dan preferensi yang berbeda dalam menghubungkan informasi yang relevan. Tanpa memiliki pemahaman yang jelas, ringkas, dan komprehensif tentang tujuan proyek, risiko pemborosan sumber daya menjadi semakin tinggi. Oleh karena itu, penting untuk memiliki pandangan yang jelas dan menyeluruh tentang tujuan proyek untuk memastikan efisiensi dan efektivitas dalam penggunaan sumber daya yang tersedia.

3.2 Pemahaman Data (*Data Understanding*)

Pemahaman terhadap data melibatkan serangkaian kegiatan untuk mempersiapkan dan mengevaluasi data yang diperlukan, termasuk pengumpulan data itu sendiri. Terdapat hubungan erat antara pemahaman bisnis dan pemahaman terhadap data, karena dalam tahap ini perumusan masalah data mining dan pemahaman tentang data yang tersedia dilakukan. Data dikumpulkan sebagai bagian dari proses ini. Pemahaman tentang kebutuhan dan persyaratan bisnis akan menentukan jenis data yang dikumpulkan, sumber data yang digunakan, dan metode pengumpulan data yang digunakan.

3.3 Persiapan Data (*Data Preparation*)

Persiapan data dilakukan setelah proses pengumpulan data, di mana data perlu diidentifikasi, dipilih, dibersihkan, dan diorganisir ke dalam format yang diinginkan. Setelah dataset dipilih, langkah selanjutnya adalah memeriksa adanya kasus yang meragukan, data yang hilang, atau data yang ambigu. Proses persiapan data dapat melibatkan iterasi dan tidak harus dilakukan dalam urutan tertentu. Aktivitas persiapan data dapat meliputi transformasi variabel, penghapusan data yang tidak relevan atau duplikat, pengisian nilai yang hilang, dan normalisasi atau pengkodean ulang data.

Tujuan dari persiapan data adalah untuk memastikan bahwa data yang digunakan dalam analisis sesuai dengan kebutuhan penelitian atau tujuan bisnis, serta bersih dan terstruktur dengan baik untuk analisis lanjutan.

3.4 Pemodelan (*Modelling*)

Pada tahap pemodelan, algoritma diterapkan untuk menemukan pola dalam data. Model dipilih berdasarkan jenis data dan tujuan penelitian. Dalam penelitian ini, menggunakan *Artificial Neural networks* (ANN) untuk melakukan prediksi. Model ini membantu mengidentifikasi faktor yang berpengaruh dan memberikan estimasi nilai target. Diharapkan penelitian ini memberikan wawasan dan prediksi yang berguna bagi perusahaan atau kepentingan bisnis yang relevan.

BAB IV

PENGUMPULAN DAN PENGOLAHAN DATA

Bab ini membahas tahapan pengumpulan dan pengolahan data yang dilakukan dalam penelitian di PT Petrokimia Gresik. Pada bab ini, akan dijelaskan langkah-langkah yang diambil untuk memperoleh dan mempersiapkan data yang relevan guna mendukung analisis yang dilakukan. pada penelitian ini data pemuatan pupuk diperoleh dari program *wareouse management system* PT Petrokimia dan data historis cuaca dari proses *scraping* data di internet.

4.1 Pemahaman Terhadap Data (*Data Understanding*)

Subbab ini menjelaskan mengenai kebutuhan data untuk analisis, tinjauan data yang tersedia, dan eksplorasi data awal.

4.1.1 Tinjauan Data yang Tersedia

Pada penelitian ini menggunakan data pemuatan pupuk gudang lini 1 yang didapat dari aplikasi *warehouse management system* (WMS) internal perusahaan PT Petrokimia yang diambil pada periode bulan Juni 2022 hingga bulan Desember 2022 yang terdiri dari 77748 baris dan 12 kolom. Selain itu, untuk mendukung hipotesis juga digunakan data historis cuaca Kabupaten Gresik bulan Juni hingga Desember 2022 yg didapat dari penyedia data cuaca Visual Crossing. Untuk data pemuatan pupuk terdiri dari 12 kolom atau atribut seperti pada gambar berikut:

Id	Nama_gudang	Nama_aktivitas	Nama_karu	Nama_checker	Nama_produk	Jumlah	Waktu_pemBua	Waktu_pengijau	Waktu_approva	Lama_pem	Lama_appro
61605	Gudang Curah 5	Pengeluaran Pup	Agung Mahardh	Ach. Muchid	ASAM SULFAT (H	20.58	20/07/2022 11:0	20/07/2022 11:0	20/07/2022 11:0	0	0
66181	Gudang Curah 5	Pengeluaran Pup	Agung Mahardh	Ach. Muchid	ASAM SULFAT (H	7.57	28/07/2022 10:0	28/07/2022 10:0	28/07/2022 10:0	1	0
61037	Gudang Curah 5	Pengeluaran Pup	Agung Mahardh	Fitriyanto	ALF3 @. 50 Kg	20	19/07/2022 14:1	19/07/2022 14:2	19/07/2022 14:2	3	1
61045	Gudang Curah 5	Pengeluaran Pup	Agung Mahardh	Fitriyanto	ALF3 @. 50 Kg	20	19/07/2022 14:1	19/07/2022 14:1	19/07/2022 14:2	6	4
77215	Gudang Curah 5	Pengeluaran Pup	Agung Mahardh	Fitriyanto	ALF3 @1000KG	1	18/08/2022 09:0	18/08/2022 09:1	18/08/2022 09:1	11	1
61044	Gudang Curah 5	Pengeluaran Pup	Agung Mahardh	Fitriyanto	ALF3 @. 50 Kg	20	19/07/2022 14:0	19/07/2022 14:0	19/07/2022 14:1	12	10
61593	Gudang Curah 5	Pengeluaran Pup	Agung Mahardh	Muhammad Riza	Petrocas @50KG	10	20/07/2022 11:0	20/07/2022 11:0	20/07/2022 11:1	13	4
61198	Gudang Curah 5	Pengeluaran Pup	Agung Mahardh	Fitriyanto	ALF3 @. 50 Kg	20	20/07/2022 10:4	20/07/2022 10:5	20/07/2022 11:0	17	15
61038	Gudang Curah 5	Pengeluaran Pup	Agung Mahardh	Fitriyanto	ALF3 @. 50 Kg	20	19/07/2022 13:5	19/07/2022 14:0	19/07/2022 14:1	20	16

Gambar 4.4.1 Data pemuatan pupuk WMS PT Petrokimia Gresik

Kolom 'Id' merupakan atribut untuk id aktivitas pemuatan pada suatu waktu, kolom 'Nama_gudang' merupakan atribut yang menampilkan nama gudang tempat terjadinya aktivitas pemuatan pupuk, kolom 'Nama_Aktivitas' merupakan atribut yang menampilkan nama aktivitas atau kegiatan dari id yang bersangkutan, kolom 'Nama_karu' merupakan atribut yang menampilkan nama kepala regu yang bertugas pada aktivitas tersebut, kolom 'Nama_checker' adalah atribut yang menampilkan checker yang bertugas atau menangani aktivitas dari id tersebut, checker memiliki tugas mengecek jumlah dan posisi pemuatan dari gudang kedalam truk dan melaporkannya pada kepala regu, kolom 'Nama_produk' adalah atribut yang menampilkan produk yang dimuat pada suatu aktivitas, kolom

‘jumlah’ menampilkan jumlah produk yang dimuat dalam satuan Ton, kolom ‘waktu_pembuatan’ adalah atribut yang menampilkan waktu dan tanggal dari id aktivitas pertama kali dibuat, kolom ‘Waktu_pengajuan’ adalah atribut yang menampilkan waktu dan tanggal ketika checker mengajukan laporan pemuatan kepada kepala regu, kolom ‘Waktu_approval’ adalah atribut yang menampilkan waktu dan tanggal saat kepala regu menyetujui laporan dari checker sehingga truk bisa masuk proses selanjutnya yaitu ke jembatan timbang untuk penimbangan muatan, kolom ‘Lama_pemuatan’ adalah atribut yang menampilkan durasi proses pemuatan dari awal hingga selesai yang mana selisih dari waktu approval dan waktu pembuatan, dan kolom ‘Lama_approval’ adalah atribut yang menampilkan durasi proses persetujuan laporan dari checker kepada kepala regu yang mana merupakan selisih dari waktu approval dan waktu pengajuan.

Untuk data historis cuaca terdiri dari 24 kolom atau atribut seperti pada gambar berikut:

Name	Datetime	Temp	Feelslike	Dew	Humidity	Precip	Precipprob	Precip type	Snow	Snowdepth	Windgust	Windspeed	Winddir
gresik	2023-02-01T00:00:00	25	25	23	88.68	0	0		0	0	4.7	7.6	230
gresik	2023-02-01T01:00:00	25	25	23.1	89.21	0	0		0	0	4.7	5.1	270
gresik	2023-02-01T02:00:00	25	25	23	88.68	0	0		0	0	2.2	9.4	300
gresik	2023-02-01T03:00:00	25	25	24	94.19	0	0		0	0	3.6	3.6	260
gresik	2023-02-01T05:00:00	25.5	25.5	23	86.12	0	0		0	0	13	7.6	315

Gambar 4.4.2 Data cuaca historis Kabupaten Gresik

Sealevelpressure	Cloudcover	Visibility	Solarradiation	Solarenergy	Uvindex	Severerisk	Conditions	Icon	Stations
1007	27.5	5	0		0	10	Partially cloudy	partly-cloudy-night	WARR,96935599
1007.2	48.1	5	0		0	10	Partially cloudy	partly-cloudy-night	WARR,96935099
1006	27.5	5	0		0	10	Partially cloudy	partly-cloudy-night	WARR,96935599
1006	27.5	6	0		0	10	Partially cloudy	partly-cloudy-night	WARR,96935599
1006	27.5	6	0		0	10	Partially cloudy	partly-cloudy-night	WARR,96935599

Gambar 4.4.3 Data cuaca historis Kabupaten Gresik (lanjutan)

Kolom ‘Name’ menunjukkan lokasi data cuaca yaitu Kabupaten Gresik, kolom ‘datetime’ menunjukkan tanggal dan waktu pengamatan cuaca. ‘temp’ menggambarkan suhu udara rata-rata, sedangkan ‘feelslike’ mencerminkan suhu yang dirasakan oleh manusia. ‘dew’ mengacu pada titik embun, yaitu suhu di mana udara jenuh dengan uap air. Kolom ‘humidity’ menunjukkan tingkat kelembaban relatif udara dalam persen. ‘precip’ menunjukkan curah hujan dalam satuan milimeter, sedangkan ‘precipprob’ memberikan peluang terjadinya curah hujan dalam persentase. Informasi mengenai jenis curah hujan tidak tersedia dalam dataset ini. Kolom ‘snow’ menampilkan jumlah salju yang terjadi dalam satuan sentimeter, sedangkan ‘snowdepth’ mengindikasikan kedalaman salju dalam sentimeter. ‘windspeed’ menyajikan kecepatan angin dalam kilometer per jam, dan ‘windgust’ menunjukkan kecepatan angin puncak dalam kilometer per jam. ‘winddir’ memberikan informasi mengenai arah angin dalam derajat. ‘sealevelpressure’ mengacu pada tekanan udara pada permukaan laut dalam

milibar. 'cloudcover' menunjukkan persentase cakupan awan di langit. 'visibility' menggambarkan jarak pandang dalam kilometer. 'solarradiation' menyajikan radiasi matahari dalam watt per meter persegi, sedangkan 'solarenergy' menunjukkan jumlah energi matahari yang diterima dalam megajoule per meter persegi. 'uvindex' mengindikasikan tingkat indeks sinar ultraviolet (UV). Informasi mengenai risiko cuaca ekstrem tidak disediakan dalam dataset ini. Kolom 'conditions' memberikan deskripsi singkat tentang kondisi cuaca pada waktu yang bersangkutan. 'icon' adalah gambar simbol yang menggambarkan kondisi cuaca secara visual. Terakhir, kolom 'stations' mencantumkan daftar sumber stasiun cuaca terdekat yang mengamati kondisi cuaca di daerah tersebut.

4.1.2 Kebutuhan Data untuk Analisis

Untuk melaksanakan analisis dalam penelitian ini, diperlukan data yang mencakup atribut-atribut penting terkait pemuatan pupuk. Data pemuatan pupuk yang diperlukan mencakup informasi tentang jenis produk pupuk, jumlah pupuk yang dimuat, lama pemuatan, dan atribut lain yang relevan. Sedangkan untuk memperkuat prediksi juga diperlukan cuaca untuk melihat seberapa besar pengaruh variabel cuaca terhadap produktivitas pemuatan pupuk. Data tentang cuaca yang diperlukan mencakup informasi suhu, kelembaban udara, jenis presipitasi, curah hujan, dan atribut lain yang mungkin berpengaruh. Kombinasi dari kedua jenis data ini diharapkan memberikan wawasan yang lebih komprehensif mengenai faktor-faktor yang memengaruhi proses pemuatan pupuk dan efisiensinya.

4.1.3 Eksplorasi Data Awal

Salah satu tahapan pemahaman data adalah melalui eksplorasi data, eksplorasi data penting untuk dilakukan untuk menemukan fakta atau informasi terhadap data sehingga bisa dilakukan analisis lebih lanjut. Pada penelitian ini data yang digunakan adalah gabungan dari dataset pemuatan pupuk dan dataset cuaca pada kota Gresik. Dataset gabungan memiliki 47150 baris dan 13 kolom, berikut adalah 5 baris awal dari dataset.

Tabel IV.1 Gabungan dataset pemuatan pupuk dan cuaca

Nama Gudang	Nama Produk	Lama Approval	Shift Kerja	Jumlah
Gudang ZA UREA	ZA NS @50 KG	1	3	29
Gudang Phonska Area III	NPK 15-10-12 Sub @50KG	30	3	26.22
Gudang Phonska Area III	NPK 15-10-12 Sub @50KG	9	3	26.22
Gudang Urea I B	Urea SUB @50 KG	4	3	39
Gudang Fosfat & Phonska Area I	NPK 15-10-12 Sub @50KG	18	3	31.86

Table IV.2 Gabungan dataset pemuatan pupuk dan cuaca (lanjutan)

Lama Pemuatan	Rate	Temp	Precip	Humidity	Windspeed	Conditions	Jam
54	0.54	25	0	94.19	9.4	0	0
49.33	0.53	25	0	94.19	9.4	0	0
49.33	0.53	25	0	94.19	9.4	0	0
54.67	0.71	25	0	94.19	9.4	0	0
84.43	0.38	25	0	94.19	9.4	0	0

Dari dataset dapat diketahui bahwa pada gudang ZA UREA produk ZA NS @50 KG dimuat pada *shift* kerja ke-3 dengan jumlah pemuatan 29 Ton dan lama pemuatan 54 menit dan rate dari pemuatan ini adalah 0,54 yang didapat dari pembagian antara jumlah pemuatan dengan lama pemuatan. Pada atribut nama gudang dan nama produk akan diubah menjadi data kategorikan dengan cara *encoding* menggunakan teknik *One-Hot Encoding* supaya bisa dimasukkan kedalam model ANN.

4.2 Persiapan Data (*Data Preparation*)

Subbab ini menjelaskan mengenai pemilihan dataset, pembersihan data, transformasi dan pemrosesan data, pemilihan fitur, dan ekstraksi fitur.

4.2.1 Pemilihan Dataset

Pada penelitian ini, fokus analisis dilakukan terhadap produktivitas pemuatan pupuk khususnya pada jenis pupuk *in-bag*. Oleh karena itu, diperlukan seleksi data untuk menghilangkan data yang tidak relevan atau diluar lingkup pengamatan. Pada dataset ini, ciri-ciri pupuk *in-bag* adalah yang mengandung satuan @xx Kg kecuali untuk produk Alf3.

```
1 # filter data yang hanya mengandung nama produk yang dianalisis
2 data = data[data['nama_produk'].str.contains('@50KG|@25KG|@1KG|@5KG|@40Kg|@20KG|@50 KG')]
```

Gambar IV.4 Proses menyeleksi data yang akan dianalisis

Selain itu, pada penelitian ini hanya aktivitas pemuatan pupuk dengan kategori 'Pengeluaran Pupuk / Non Pupuk Melalui JT' yang akan dipilih sebagai data yang akan digunakan dalam analisis lebih lanjut.

```
1 # filter data yang hanya mengandung jenis aktivitas 'Pengeluaran Pupuk / Non Pupuk Melalui JT'
2 data = data[data['nama_aktivitas'] == 'Pengeluaran Pupuk / Non Pupuk Melalui JT']
```

Gambar IV.5 Proses menyeleksi aktivitas Pengeluaran Pupuk / Non Pupuk Melalui JT

4.2.2 Pembersihan Data

Sebelum data bisa diolah dan dianalisis data perlu dibersihkan terlebih dahulu karena pembersihan data merupakan tahap penting dalam proses

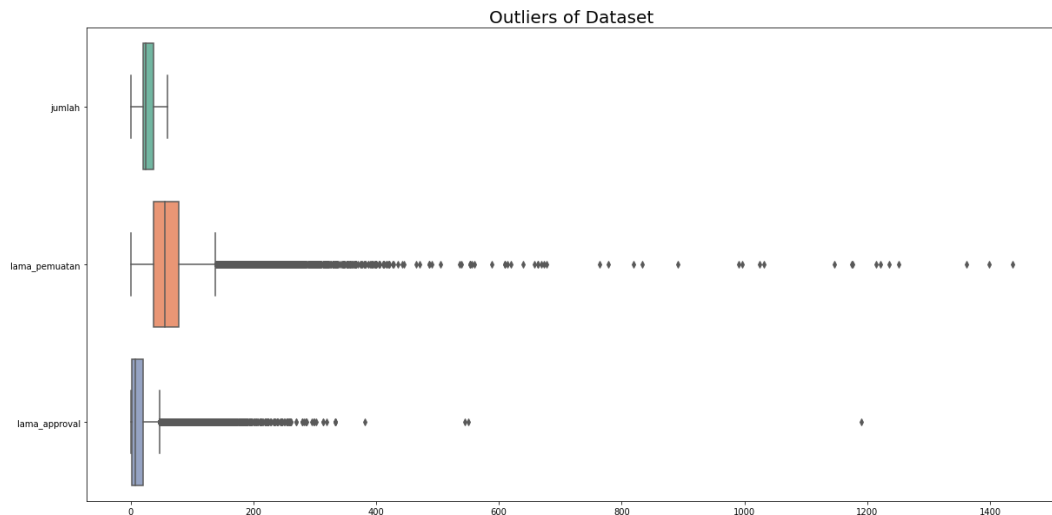
pengolahan data untuk mengatasi masalah data yang tidak valid, inkonsisten, atau hilang.

4.2.2.1 Penghapusan Nilai yang Hilang

Data pemuatan yang telah diimpor kedalam *dataframe* dan telah diseleksi harus dibersihkan terlebih dahulu supaya terhindar dari inkonsistensi seperti nilai yang hilang atau kosong (*missing value*). Dengan menggunakan fungsi *dropna* pada pandas maka semua *cell* yang *value*-nya kosong akan otomatis terhapus.

4.2.2.2 Penanganan Outlier

Outlier merupakan nilai yang menyimpang secara signifikan dari kisaran perilaku normal dalam data. Keberadaan outlier dapat memengaruhi hasil analisis secara negatif, sehingga penting untuk mengenali dan menanganinya dengan tepat. Berikut adalah visualisasi outlier pada data numerik pada dataset pemuatan pupuk menggunakan *boxplot*.



Gambar IV.6 Outlier pada dataset pemuatan pupuk

Terlihat bahwa outlier hanya terdapat pada atribut lama pemuatan dan lama approval, untuk menangani outlier dapat menggunakan metode IQR (Interquartile Range).

```
1 # Menghapus outlier pada kolom 'lama_pemuatan'
2 Q1 = data['lama_pemuatan'].quantile(0.25)
3 Q3 = data['lama_pemuatan'].quantile(0.75)
4 IQR = Q3 - Q1
5 lower_limit = Q1 - 1.5 * IQR
6 upper_limit = Q3 + 1.5 * IQR
7
8 # Menghapus data outlier
9 data= data[(data['lama_pemuatan'] >= lower_limit) & (data['lama_pemuatan'] <= upper_limit)]
10
```

Gambar IV.7 Menghapus outlier pada atribut lama pemuatan

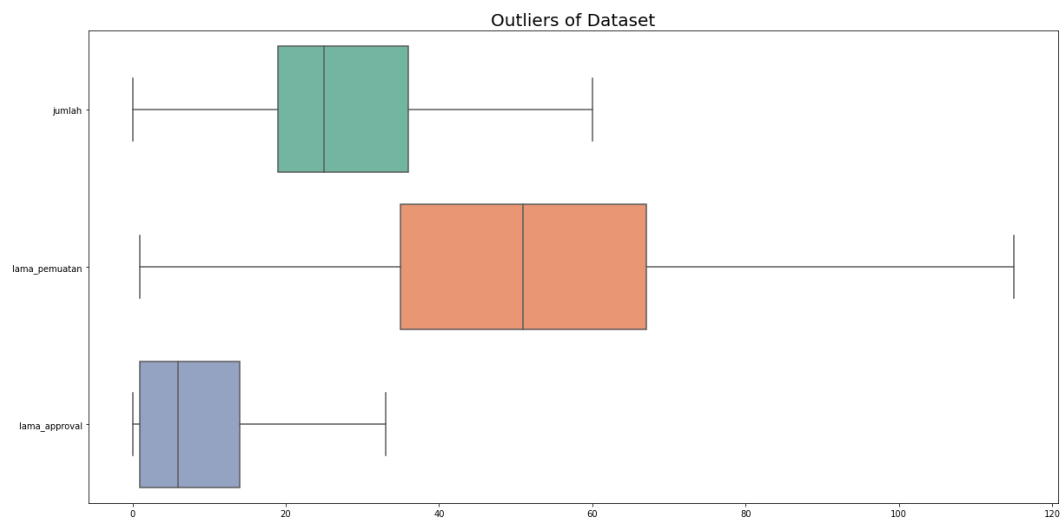
```

1 # Menghapus outlier pada kolom 'lama_approval'
2 Q1 = data['lama_approval'].quantile(0.25)
3 Q3 = data['lama_approval'].quantile(0.75)
4 IQR = Q3 - Q1
5 lower_limit = Q1 - 1.5 * IQR
6 upper_limit = Q3 + 1.5 * IQR
7
8 # Menghapus data outlier
9 data = data[(data['lama_approval'] >= lower_limit) & (data['lama_approval'] <= upper_limit)]
10

```

Gambar IV.8 Menghapus outlier pada atribut lama *approval*

Berikut merupakan hasil dari pembersihan data dari outlier:



Gambar IV.9 Dataset pemuatan pupuk yang sudah bersih dari *outlier*

Dapat terlihat dari *boxplot* bahwa sudah tidak terdapat outlier pada dataset sehingga bisa dilanjutkan untuk proses selanjutnya.

4.2.3 Transformasi dan Pemrosesan Data

Transformasi dan pemrosesan data merupakan tahap penting dalam pengolahan data untuk mengatasi masalah data yang tidak valid, inkonsisten, atau hilang sebelum dilakukan analisis lebih lanjut. Tujuan dari transformasi dan pemrosesan data adalah untuk mengubah dan mempersiapkan data agar sesuai dengan kebutuhan analisis.

4.2.3.1 Pengkodean Ulang Variabel

Pengkodean ulang variabel bertujuan untuk mengatasi masalah data yang tidak valid, inkonsisten, atau hilang sebelum melakukan analisis lebih lanjut. Dengan melakukan pengkodean ulang variabel, data akan diubah dan dipersiapkan agar sesuai dengan kebutuhan analisis yang akan dilakukan. Teknik yang digunakan untuk mengkodekan ulang variabel adalah *One-Hot encoding*.

nama_karu	nama_checker	jumlah	waktu_pembuatan	waktu_pengajuan	waktu_approval	lama_pemuatan	lama_approval
Agung Mahardhika Putra	Muhammad Rizal Azizi	10.0	2022-07-20 11:00:00	2022-07-20 11:09:00	2022-07-20 11:13:00	13	4.0
Arli PF1	Faruq Fakhruddin	40.0	2022-02-09 20:52:00	2022-02-09 20:54:00	2022-02-09 20:55:00	3	0.0
Arli PF1	Moh. Zakiyudin	30.0	2023-02-02 00:43:00	2023-02-02 00:46:00	2023-02-02 00:46:00	3	0.0
Arli PF1	Mahrus Ali	0.3	2022-06-29 10:26:00	2022-06-29 10:29:00	2022-06-29 10:30:00	4	1.0
Arli PF1	Mahrus Ali	10.0	2022-06-28 11:03:00	2022-06-28 11:08:00	2022-06-28 11:09:00	5	0.0

Gambar IV.10 Pengkodean ulang variabel dengan teknik *One-Hot Encoding*

nama_produk_UREA PRL NS @5KG	nama_produk_Urea PRL NS @50 KG	nama_produk_Urea SUB @50 KG	nama_produk_ZA NS @50 KG	nama_produk_ZA PLUS @50 KG	nama_produk_ZK NS @1KG
0	0	0	0	0	0
0	0	0	0	0	0
0	0	0	0	0	0
0	0	0	0	1	0
0	0	0	1	0	0

Gambar IV.11 Pengkodean ulang variabel dengan teknik *One-Hot Encoding* (lanjutan)

4.2.3.2 Penambahan atribut

Tahap ini memecah atribut dari dataset menjadi atribut baru atau penambahan atribut baru dari pemahaman bisnis yang bertujuan untuk memberikan informasi tambahan pada dataset untuk keperluan analisis dan pembentukan model prediksi.

Pada dataset ini ditambahkan atribut *rate* yang merupakan pembagian dari atribut jumlah dengan lama_pemuatan dan atribut ini lah yang akan diprediksi menggunakan model machine learning. Selain itu, ditambahkan juga atribut *shift* kerja yang merupakan waktu *shift* kerja ketika pemuatan pupuk dilakukan oleh tenaga kerja bongkar muat.

nama_produk_Urea PRL NS @50 KG	nama_produk_Urea SUB @50 KG	nama_produk_ZA NS @50 KG	nama_produk_ZA PLUS @50 KG	nama_produk_ZK NS @1KG	nama_produk_ZK NS @50 KG	rate
0	0	0	0	0	0	1.401822
0	0	0	0	0	0	0.784586
0	0	0	0	0	0	0.410540

Gambar IV.12 Penambahan atribut *rate*

4.2.4 Penggabungan Data dari Sumber Berbeda

Dalam penelitian ini, dataset pemuatan pupuk akan digabungkan dengan dataset cuaca untuk memperoleh informasi yang lebih lengkap dan terintegrasi. Penggabungan data dari sumber berbeda memungkinkan kita untuk melihat hubungan antara variabel pemuatan pupuk dengan variabel cuaca, dan mengidentifikasi pola atau tren yang dapat berpengaruh terhadap produktivitas pemuatan pupuk. Dengan melakukan penggabungan data yang komprehensif, diharapkan penelitian ini dapat memberikan wawasan yang lebih mendalam dan akurat terkait faktor-faktor yang mempengaruhi produktivitas pemuatan pupuk. Proses penggabungan dengan dataset cuaca menggunakan fungsi *merge*, *dataframe data3* dan cuaca akan digabungkan berdasarkan kolom "waktu_muat_baru" dari dataframe *data3* dan kolom "datetime" dari dataframe cuaca. Penggabungan dilakukan dengan menggunakan metode *outer*, yang akan menggabungkan semua baris dari kedua dataframe, bahkan jika ada data yang tidak ada di salah satu dataframe. Dalam hal ini, data yang tidak ada di salah satu dataframe akan diisi dengan nilai *null*.

Dataset cuaca memiliki banyak atribut yang tidak semuanya relevan dalam mendukung model prediksi. Oleh karena itu, hanya beberapa atribut yang dipilih dari dataset cuaca, yaitu 'temp', 'humidity', 'windspeed', dan 'conditions'. Pemilihan atribut tersebut didasarkan pada pertimbangan bahwa atribut-atribut tersebut memiliki potensi untuk mempengaruhi atau memberikan informasi yang penting terkait dengan prediksi rate pemuatan pupuk.

nama_karu	nama_checker	waktu_pembuatan	waktu_pengajuan	waktu_approval	lama_approval	shift_kerja	jumlah	lama_pemua
Iksan ZA Urea	Alfin Adam A	2022-01-07 00:07:00	2022-01-07 00:32:00	2022-01-07 00:34:00	1.0	3.0	29.00	54
Derry Phonska 3	Subiantoro	2022-01-07 00:13:00	2022-01-07 00:16:00	2022-01-07 00:47:00	30.0	3.0	26.22	49
Derry Phonska 3	Subiantoro	2022-01-07 00:17:00	2022-01-07 00:37:00	2022-01-07 00:47:00	9.0	3.0	26.22	49
Iksan Urea IB	Gabril Kurniawan	2022-01-07 00:21:00	2022-01-07 00:30:00	2022-01-07 00:35:00	4.0	3.0	39.00	54

Gambar IV.13 Penggabungan dataset pemuatan dengan dataset cuaca

rate	jam_cuaca	tanggal_pembuatan	waktu_muat_baru	datetime	temp	precip	humidity	windspeed	conditions
0.537037	00:00:00	2022-01-07	2022-01-07 00:00:00	2022-01-07 00:00:00	25.0	0.0	94.19	9.4	0.0
0.531522	00:00:00	2022-01-07	2022-01-07 00:00:00	2022-01-07 00:00:00	25.0	0.0	94.19	9.4	0.0
0.531522	00:00:00	2022-01-07	2022-01-07 00:00:00	2022-01-07 00:00:00	25.0	0.0	94.19	9.4	0.0
0.713371	00:00:00	2022-01-07	2022-01-07 00:00:00	2022-01-07 00:00:00	25.0	0.0	94.19	9.4	0.0

Gambar IV.14 Penggabungan dataset pemuatan dengan dataset cuaca (lanjutan)

4.3 Pengujian Hubungan dan Signifikansi

Sebelum melakukan *modelling* perlu dilakukan analisis korelasi dan signifikansi antara variabel independen dan variabel dependen sehingga bisa dijadikan pertimbangan apakah suatu variabel independent dapat dimasukkan kedalam model. Pada pengujian ini uji korelasi menggunakan metode Pearson Correlation Coefficient menggunakan modul `scipy.stats`. berikut adalah hasil dari koefisien korelasi pearson dan signifikansi p-value variabel X terhadap Y.

```
Koefisien korelasi Pearson: -0.489
Nilai p-value: 0.0
```

Gambar IV.15 Hasil uji korelasi dan signifikansi variabel lama pemuatan terhadap rate

Pada pengujian korelasi antara variabel lama pemuatan terhadap rate didapatkan hasil koefisien korelasi pearson sebesar -0.489 dan nilai p-value sebesar 0.0. berikut adalah hasil uji korelasi antara variabel jumlah terhadap rate.

```
Koefisien korelasi Pearson: 0.096
Nilai p-value: 1.9244793612280302e-97
```

Gambar IV.16 Hasil uji korelasi dan signifikansi variabel jumlah terhadap rate

Pada pengujian korelasi antara variabel jumlah terhadap rate didapatkan hasil koefisien korelasi pearson sebesar 0.096 dan nilai p-value sebesar 1.9244793612280302e-97. berikut adalah hasil uji korelasi antara variabel *shift* kerja terhadap rate.

```
Koefisien korelasi Pearson: -0.039
Nilai p-value: 3.937658169852203e-17
```

Gambar IV.17 Hasil uji korelasi dan signifikansi variabel *shift* kerja terhadap rate

Pada pengujian korelasi antara variabel *shift* kerja terhadap rate didapatkan hasil koefisien korelasi pearson sebesar -0.039 dan nilai p-value sebesar 3.937658169852203e-17. berikut adalah hasil uji korelasi antara variabel jam terhadap rate.

Koefisien korelasi Pearson: 0.026
Nilai p-value: 2.769946322379121e-08

Gambar IV.18 Hasil uji korelasi dan signifikansi variabel jam terhadap rate

Pada pengujian korelasi antara variabel jam terhadap rate didapatkan hasil koefisien korelasi pearson sebesar 0.026 dan nilai p-value sebesar 2.769946322379121e-08. berikut adalah hasil uji korelasi antara variabel *temperature* terhadap rate.

Koefisien korelasi Pearson: 0.044
Nilai p-value: 2.3744097591455336e-21

Gambar IV.19 Hasil uji korelasi dan signifikansi variabel *temperature* terhadap rate

Pada pengujian korelasi antara variabel *temperature* terhadap rate didapatkan hasil koefisien korelasi pearson sebesar 0.044 dan nilai p-value sebesar 2.3744097591455336e-21. berikut adalah hasil uji korelasi antara variabel precipitation terhadap rate.

Koefisien korelasi Pearson: -0.000
Nilai p-value: 0.9542631172113246

Gambar IV.20 Hasil uji korelasi dan signifikansi variabel precipitation terhadap rate

Pada pengujian korelasi antara variabel precipitation terhadap rate didapatkan hasil koefisien korelasi pearson sebesar -0.000 dan nilai p-value sebesar 0.9542631172113246. berikut adalah hasil uji korelasi antara variabel *humidity* terhadap rate.

Koefisien korelasi Pearson: -0.039
Nilai p-value: 1.5137562511140666e-17

Gambar IV.21 Hasil uji korelasi dan signifikansi variabel *humidity* terhadap rate

Pada pengujian korelasi antara variabel *humidity* terhadap rate didapatkan hasil koefisien korelasi pearson sebesar -0.039 dan nilai p-value sebesar 1.5137562511140666e-17. berikut adalah hasil uji korelasi antara variabel *windspeed* terhadap rate.

Koefisien korelasi Pearson: 0.009
Nilai p-value: 0.04423494986092527

Gambar IV.22 Hasil uji korelasi dan signifikansi variabel *windspeed* terhadap rate

Pada pengujian korelasi antara variabel *windspeed* terhadap rate didapatkan hasil koefisien korelasi pearson sebesar 0.009 dan nilai p-value sebesar

0.04423494986092527. berikut adalah hasil uji korelasi antara variabel *conditions* terhadap rate.

Koefisien korelasi Pearson: -0.006

Nilai p-value: 0.21178727427075414

Pada pengujian korelasi antara variabel *conditions* terhadap rate didapatkan hasil koefisien korelasi pearson sebesar -0.006 dan nilai p-value sebesar 0.21178727427075414.

4.4 Pengembangan Model Machine Learning

Pada penelitian ini, digunakan model *Artificial Neural network* (ANN) untuk menganalisis dan memprediksi proses pemuatan pupuk. Model ANN merupakan model yang terinspirasi oleh jaringan saraf manusia dan mampu mempelajari pola-pola kompleks dari data.

Pada awalnya, data dibagi menjadi data *training* dan data *testing* dengan perbandingan 80:20 menggunakan metode `train_test_split` dari library `sklearn`. Data *training* digunakan untuk melatih model, sementara data *testing* digunakan untuk menguji kinerja model yang telah dilatih.

Selanjutnya, dilakukan proses *scaling* menggunakan `StandardScaler` dari `sklearn.preprocessing`. Proses *scaling* ini bertujuan untuk menormalkan atau menskalakan fitur-fitur (variabel input) dalam rentang yang seragam. Dengan menggunakan `StandardScaler`, data *training* dan data *testing* diubah sehingga memiliki skala yang seragam, memudahkan proses interpretasi dan perbandingan antar variabel.

Selanjutnya, dilakukan pengaturan *callback* menggunakan `ModelCheckpoint` dari `tensorflow.keras.callbacks`. *Callback* ini digunakan untuk menyimpan model dengan performa terbaik berdasarkan nilai `val_loss` terkecil selama proses pelatihan. Model tersebut disimpan dalam file "`model_checkpoint.h5`" yang dapat digunakan kembali di masa depan.

Arsitektur model ANN yang digunakan terdiri dari tiga layer *Dense*. Terdapat tiga layer *Dense* dengan masing-masing memiliki 128, 64, dan 32 neuron pada layer pertama, kedua, dan ketiga, serta layer terakhir yang memiliki 1 neuron. Fungsi aktivasi *ReLU* (*Rectified Linear Unit*) digunakan pada layer-layer tersebut, kecuali pada layer terakhir yang tidak menggunakan fungsi aktivasi.

Setelah itu, model dikompilasi menggunakan optimizer "`adam`" dengan *learning rate* yang ditentukan sebesar 0.0001 dan *loss function* "`mean_squared_error`". Optimizer `adam` merupakan metode optimasi yang efisien dan sering digunakan dalam pelatihan model *deep learning*. *Loss function* "`mean_squared_error`" digunakan untuk mengukur kesalahan prediksi antara nilai prediksi dan nilai sebenarnya.

Selama proses pelatihan, dilakukan pengamatan terhadap nilai loss dan metrik "mean_absolute_error" pada data *training* dan data *testing*. Metrik "mean_absolute_error" digunakan untuk mengukur besarnya kesalahan rata-rata antara prediksi dan nilai sebenarnya.

Proses pelatihan model dilakukan menggunakan metode fit dari model. Model dilatih dengan 100 epochs dan menggunakan batch_size sebesar 32. Data validasi (X_test, y_test) digunakan selama pelatihan untuk memonitor performa model menggunakan callback ModelCheckpoint yang telah diatur sebelumnya.

Setelah proses pelatihan selesai, model digunakan untuk melakukan prediksi terhadap data *testing* (X_test). Hasil prediksi tersebut akan digunakan dalam tahap analisis dan interpretasi hasil pada bagian selanjutnya. Berikut adalah *model summary* dari pembuatan model machine learning menggunakan ANN.

Model: "sequential_4"

Layer (type)	Output Shape	Param #
dense_14 (Dense)	(None, 128)	5248
dropout_12 (Dropout)	(None, 128)	0
dense_15 (Dense)	(None, 64)	8256
dropout_13 (Dropout)	(None, 64)	0
dense_16 (Dense)	(None, 32)	2080
dropout_14 (Dropout)	(None, 32)	0
dense_17 (Dense)	(None, 1)	33

=====

Total params: 15,617
Trainable params: 15,617
Non-trainable params: 0

Gambar IV.23 Model *summary* prediksi rate pemuatan pupuk

Model yang digunakan dalam penelitian ini adalah model sequential dengan nama "sequential_4". Model ini memiliki empat layer yang berbeda dan menghasilkan output dengan bentuk yang berbeda pula.

Layer pertama adalah dense_14, yang memiliki output shape (None, 128) dan menggunakan 5,248 parameter. Layer ini dapat dianggap sebagai layer input yang menerima input dari data yang diberikan.

Selanjutnya, terdapat dropout_12 yang tidak mengubah bentuk output, namun membantu dalam mencegah overfitting dengan mengabaikan sebagian node pada layer sebelumnya. Dropout_12 tidak memiliki parameter yang dapat diubah.

Layer ketiga adalah dense_15, yang memiliki output shape (None, 64) dan menggunakan 8,256 parameter. Layer ini mengambil input dari dropout_12 dan melakukan komputasi pada node-node yang terhubung penuh atau fully connected.

Kemudian, dropout_13 diterapkan setelah dense_15 dengan tujuan yang sama seperti dropout_12, yaitu mencegah overfitting. Dropout_13 tidak memiliki parameter yang dapat diubah.

Layer selanjutnya adalah dense_16 dengan output shape (None, 32) dan menggunakan 2,080 parameter. Layer ini mengambil input dari dropout_13 dan melakukan komputasi pada node-node yang terhubung penuh.

Terakhir, dropout_14 diterapkan setelah dense_16 untuk menghindari overfitting. Dropout_14 tidak memiliki parameter yang dapat diubah.

Layer terakhir adalah dense_17, yang memiliki output shape (None, 1) dan menggunakan 33 parameter. Layer ini merupakan layer output yang menghasilkan prediksi akhir dari model.

Secara keseluruhan, model ini memiliki total 15,617 parameter yang dapat diubah atau dipelajari selama proses pelatihan. Selama pelatihan, model ini diharapkan dapat menyesuaikan parameter-parameter ini agar mampu menghasilkan prediksi yang akurat berdasarkan fitur input yang diberikan.

4.5 Evaluasi Model Machine Learning

Pada penelitian ini, model *Artificial Neural network* (ANN) telah dikembangkan untuk menganalisis dan memprediksi rate pemuatan pupuk. Evaluasi model sangat penting untuk mengukur sejauh mana model dapat melakukan prediksi yang akurat terhadap data yang diberikan. Pada penelitian ini *metrics* yang digunakan untuk mengukur kinerja model ANN adalah *mean squared error*, *R-squared*, *mean absolute error*, *root mean squared error*, dan *adjusted R-squared*. Berikut adalah hasil *metrics evaluation* dari model ANN.

Table IV.3 *Metrics evaluation* model

Metrics	Skor
Mean Squared Error	0,655854706016367
R-Squared	0,32013653994695335
Mean Absolute Error	0,2690485579808164
Root Mean Squared Error	0,8098485698057181
Adjusted R-Squared	0.31955927151837027

Pada model yang telah dikembangkan mendapatkan skor *mean squared error* (MSE) sebesar 0,66. MSE Menyediakan ukuran rata-rata dari kuadrat perbedaan antara nilai prediksi dengan nilai sebenarnya. MSE mencerminkan keakuratan keseluruhan model dengan semakin rendah nilai MSE menunjukkan prediksi yang lebih baik. Berikut adalah contoh perhitungan MSE.

Diketahui :

$$N = 47150$$

$$y_i = 0.39328732$$

$$\hat{y}_i = 0.55952644$$

Ditanya : MSE

Dijawab :

$$\begin{aligned} MSE &= \frac{1}{N} \sum_{i=1}^N (y_i - \hat{y}_i)^2 \\ MSE &= \frac{1}{47150} \sum_{i=1}^N (0.39328732 - 0.55952644)^2 \\ &= \frac{1}{47150} \sum_{i=1}^N (-0.16623912)^2 \\ &= -0.166239 \end{aligned}$$

Pada model yang telah dikembangkan mendapatkan skor *R-Squared* (R^2) sebesar 0,32. *R-Squared* Memberikan indikasi tentang sejauh mana variasi dalam data target dapat dijelaskan oleh model. *R-squared* berkisar antara 0 hingga 1, dengan semakin tinggi nilai *R-squared* menunjukkan model yang lebih baik dalam menjelaskan variasi data target. Berikut adalah contoh perhitungan *R-Squared*.

Diketahui :

$$\bar{y}_i = 0.704121864955362$$

$$y_i = 0.39328732$$

$$\hat{y}_i = 0.55952644$$

Ditanya : R^2

Dijawab :

$$\begin{aligned} R^2 &= 1 - \frac{\text{sum of squared prediction}}{\text{sum of squared total}} = \frac{\sum_i (y_i - \hat{y}_i)^2}{\sum_i (y_i - \bar{y}_i)^2} \\ &= \frac{\sum_i (0.39328732 - 0.55952644)^2}{\sum_i (0.39328732 - 0.704121864955362)^2} \\ &= -0.08890 \end{aligned}$$

Selain menggunakan R^2 sebagai metrics pengukuran, model yang dibangun juga menggunakan *Adjusted R^2* untuk mengukur kinerja model. *Adjusted R^2* Menyesuaikan R^2 dengan jumlah variabel prediktor dalam model. *Adjusted R^2* memberikan penilaian yang lebih konservatif terhadap kualitas model dengan

mempertimbangkan kompleksitas model. Berikut adalah contoh perhitungan *Adjusted R²*.

Diketahui :

$$R^2 = 0.32013653994695335$$

$$n = 47150$$

$$k = 32$$

Ditanya : R^2

Dijawab :

$$\begin{aligned} Adjusted R^2 &= \left\{ 1 - \left[\frac{(1 - R^2)(n - 1)}{n - k - 1} \right] \right\} \\ &= \left\{ 1 - \left[\frac{(1 - 0.32013653994695335)(47150 - 1)}{47150 - 32 - 1} \right] \right\} \\ &= 0.31955927151837027 \end{aligned}$$

Pada model yang telah dikembangkan mendapatkan skor *mean absolute error* (MAE) sebesar 0,269. MAE Menyediakan ukuran rata-rata dari perbedaan absolut antara nilai prediksi dengan nilai sebenarnya. MAE memberikan gambaran tentang sejauh mana prediksi model secara keseluruhan mendekati nilai sebenarnya. Berikut adalah contoh perhitungan MAE.

Diketahui :

$$N = 47150$$

$$y_i = 0.39328732$$

$$\hat{y}_i = 0.55952644$$

Ditanya : MAE

Dijawab :

$$\begin{aligned} MAE &= \frac{1}{N} \sum_{i=1}^N |y_i - \hat{y}_i| \\ MAE &= \frac{1}{47150} \sum_{i=1}^N |0.39328732 - 0.55952644| \\ &= 0.16623912 \end{aligned}$$

Pada model yang telah dikembangkan mendapatkan skor *root mean squared error* (RMSE) sebesar 0,809. RMSE menghitung akar kuadrat dari MSE untuk memberikan ukuran rata-rata dari perbedaan antara nilai prediksi dan nilai sebenarnya. RMSE digunakan untuk memberikan gambaran tentang kesalahan

prediksi dalam satuan yang sama dengan variabel target. Berikut adalah contoh perhitungan RMSE.

Diketahui :

$$MSE = 0.655854706016367$$

Ditanya : RMSE

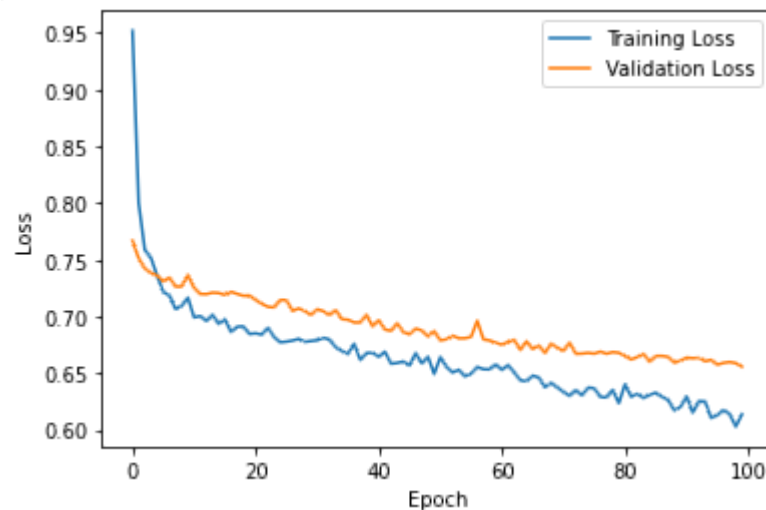
Dijawab :

$$RMSE = \sqrt{MSE}$$

$$RMSE = \sqrt{0.655854706016367}$$

$$= 0.8098485698057181$$

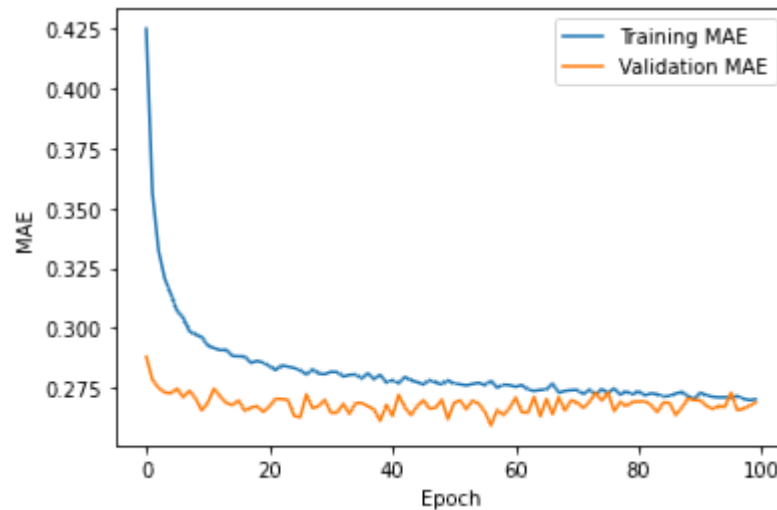
Selain *metrics evaluation* yang disebutkan di atas, evaluasi model juga melibatkan analisis *learning curve loss function*. Berikut adalah grafik loss function pada model ANN.



Gambar IV.24 Grafik *learning curve loss*

Grafik *learning curve loss function* pada laporan ini menggambarkan perubahan nilai loss function (mean squared error) pada sumbu y seiring dengan peningkatan jumlah epoch pada sumbu x. Grafik ini memberikan gambaran tentang bagaimana performa model ANN berkembang seiring dengan proses pelatihan.

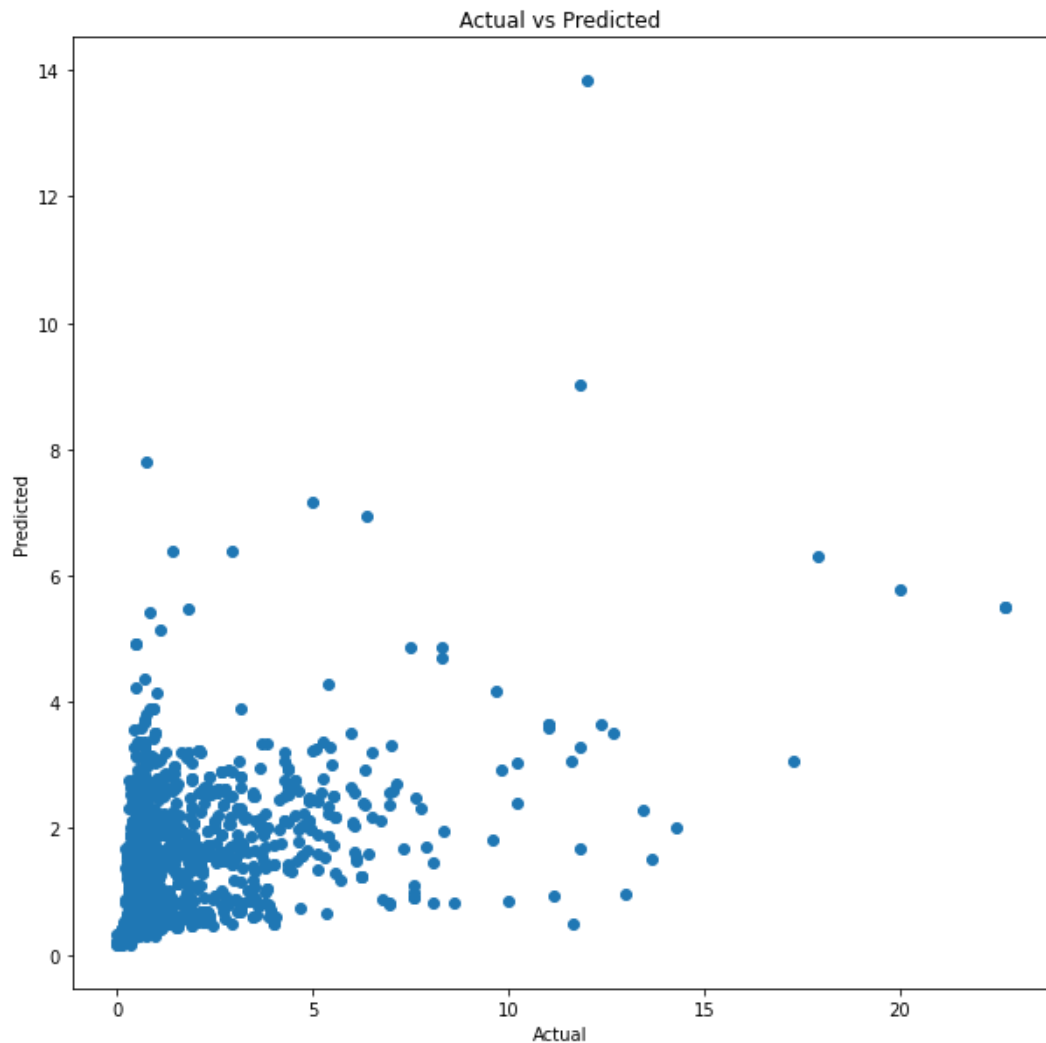
Pada awal pelatihan, *training loss* memiliki skor yang cukup tinggi, yaitu sekitar 0.95, sedangkan *validation loss* memiliki skor sekitar 0.78. Namun, seiring dengan peningkatan jumlah epoch, keduanya mengalami penurunan secara signifikan. *Training loss* berhenti di angka 0.61, sementara *validation loss* berhenti di angka 0.67. Berikut adalah grafik *learning curve MAE*



Gambar IV.25 Grafik *learning curve* MAE

Grafik Plot MAE menunjukkan perubahan tingkat kesalahan (Mean Absolute Error) pada model. Pada bagian *Training* MAE, terlihat bahwa pada awal pelatihan, model memiliki tingkat kesalahan yang cukup tinggi dengan nilai sekitar 0.425. Namun, seiring dengan pertambahan epoch, terjadi penurunan yang signifikan pada tingkat kesalahan model. Hal ini menandakan bahwa model secara bertahap dapat mempelajari pola-pola yang ada dalam data pelatihan, sehingga kesalahan prediksi dapat dikurangi. Pada akhirnya, model berhasil mencapai tingkat kesalahan yang lebih rendah dengan nilai MAE sekitar 0.275.

Di sisi lain, pada bagian *Validation* MAE, terlihat bahwa tingkat kesalahan cenderung konstan dan mengalami sedikit kenaikan. Awalnya, *Validation* MAE memiliki nilai sekitar 0.280, kemudian mengalami penurunan sedikit sebelum mengalami kenaikan lagi. Namun, perbedaan kesalahan tersebut tidak signifikan secara keseluruhan. Pada akhirnya, *Validation* MAE mencapai tingkat kesalahan yang serupa dengan *Training* MAE, yaitu sekitar 0.275. berikut adalah visualisasi *scatter plot* untuk melihat perbandingan hasil prediksi dengan data aktual



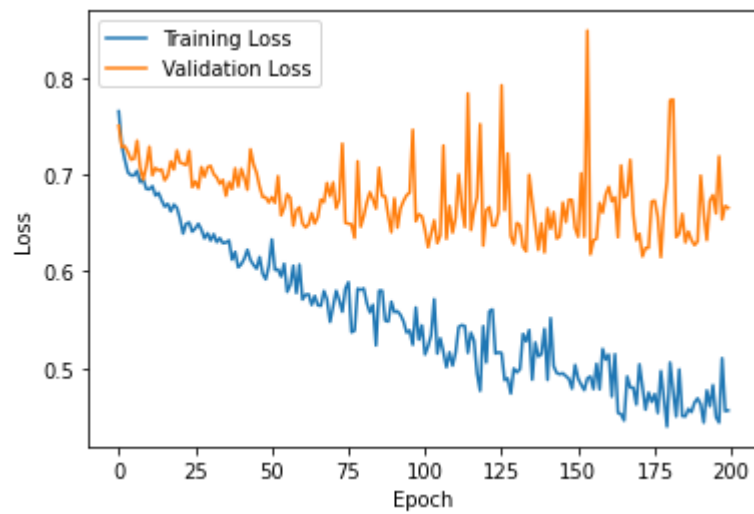
Gambar IV.26 Visualisasi *scatter plot* hasil prediksi dengan data aktual

4.6 *Hyperparameter Tuning*

Model yang telah dibangun perlu dilakukan penyesuaian ulang terhadap parameter model untuk meningkatkan kinerja dan keakuratan prediksi. Hyperparameter tuning perlu dilakukan karena setiap model machine learning memiliki parameter yang harus ditentukan sebelum pelatihan model, dan nilai-nilai default dari hyperparameter tersebut mungkin tidak optimal untuk *dataset* yang digunakan. *Hyperparameter tuning* pada penelitian ini dilakukan dengan berbagai cara yaitu metode *trial and error*, penggunaan *framerowk* AutoKeras, dan penyesuaian data ulang.

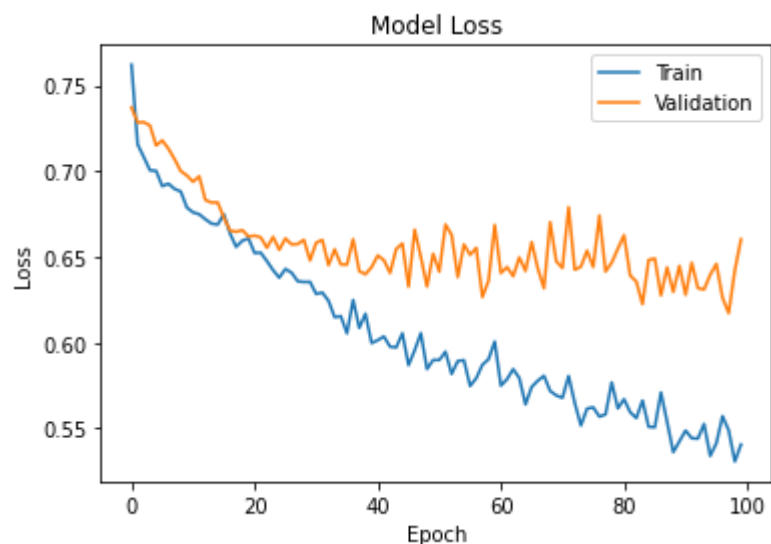
Pada metode *trial and error* model yang telah dibangun disesuaikan ulang parameternya seperti penyesuaian jumlah layer yang semula menggunakan 4 layer *Dense* diubah menjadi 5 layer *Dense* dengan aktivasi *ReLu* (*rectified linear unit*) dan kombinasi *neuron* 256, 128, 64, 32, 1. Selain itu, dilakukan juga perubahan pada jumlah *epoch* menjadi 200, *learning rate* yang semula 0.0001 menjadi 0.001, dan pengubahan teknik *scaler* yang semula menggunakan *StandardScaler* menjadi

MinMaxScaler. Berikut adalah hasil model yang telah dilakukan *Hyperparameter* dengan metode *trial and error*.



Gambar IV.27 Grafik *learning curve* *Hyperparameter* dengan *trial and error*

Pada metode *Hyperparameter* selanjutnya menggunakan *framework* AutoKeras. AutoKeras adalah sebuah *framework machine learning* otomatis yang memungkinkan menemukan kombinasi *hyperparameter* yang optimal untuk model yang dibangun. Dalam proses ini, AutoKeras akan melakukan pencarian otomatis dan eksperimen dengan berbagai kombinasi *hyperparameter* untuk menemukan model yang terbaik. AutoKeras menggunakan teknik pencarian heuristik yang canggih dan algoritma optimasi untuk menavigasi ruang *hyperparameter* dengan cepat dan efisien, model yang terbaik akan disimpan dengan format *.h5*. berikut adalah hasil terbaik dari *running* model dengan AutoKeras.



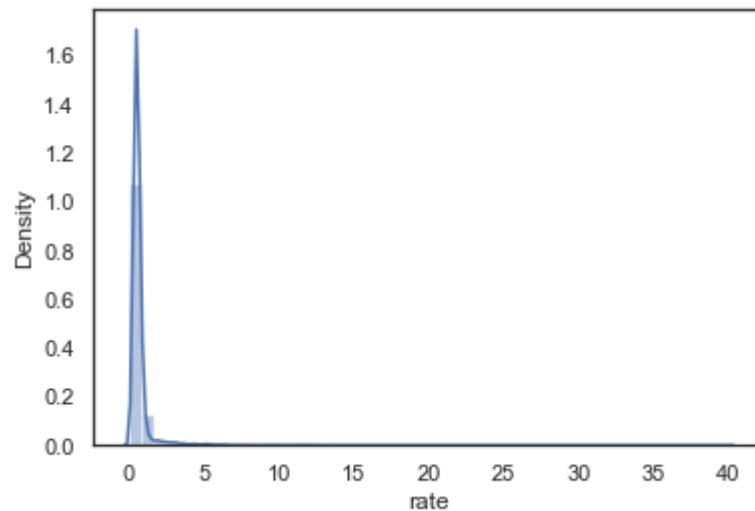
Gambar IV.28 Grafik *learning curve* *Hyperparameter* dengan AutoKeras

Berikut adalah *metrics evaluation* model dengan AutoKeras

Table IV.4 *Metrics evaluation* model dengan AutoKeras

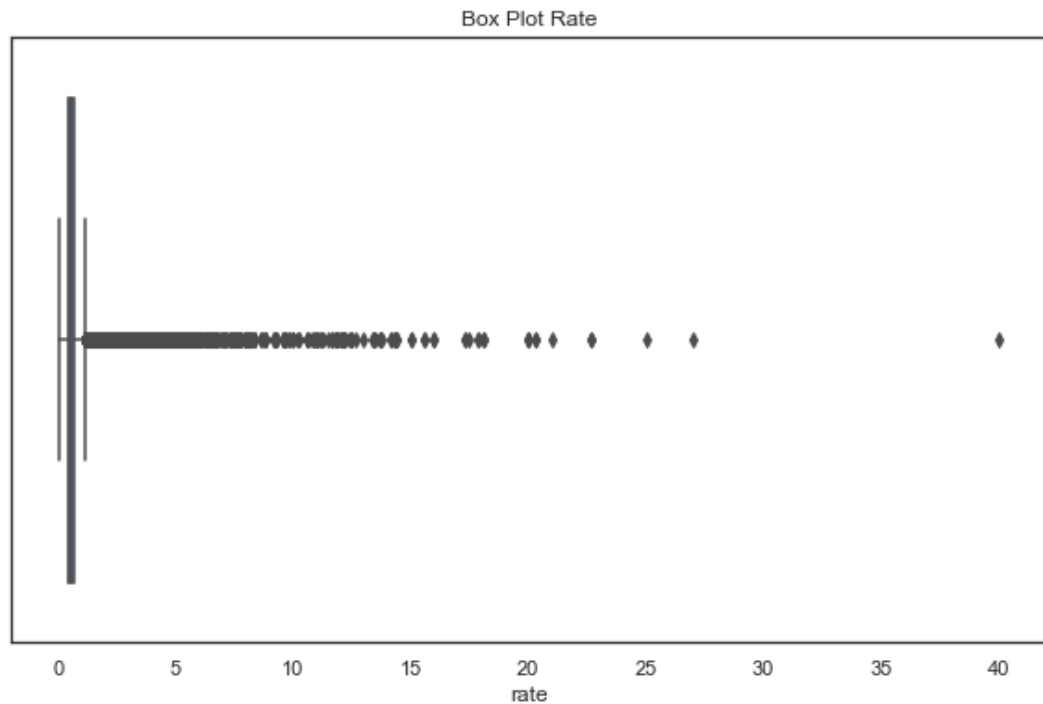
Metrics	Skor
Mean Squared Error	0,6603621643290111
R-Squared	0,31546407792698483
Mean Absolute Error	0,2591609954804092

Metode terakhir yang digunakan untuk meningkatkan kinerja model melalui *Hyperparameter tuning* adalah dengan melakukan penyesuaian ulang data. Berdasarkan hasil eksplorasi data ke-2 ditemukan bahwa data pada atribut tidak berdistribusi normal dan terindikasi terdapat *outlier*. Berikut adalah histogram distribusi atribut rate pada *dataset*.

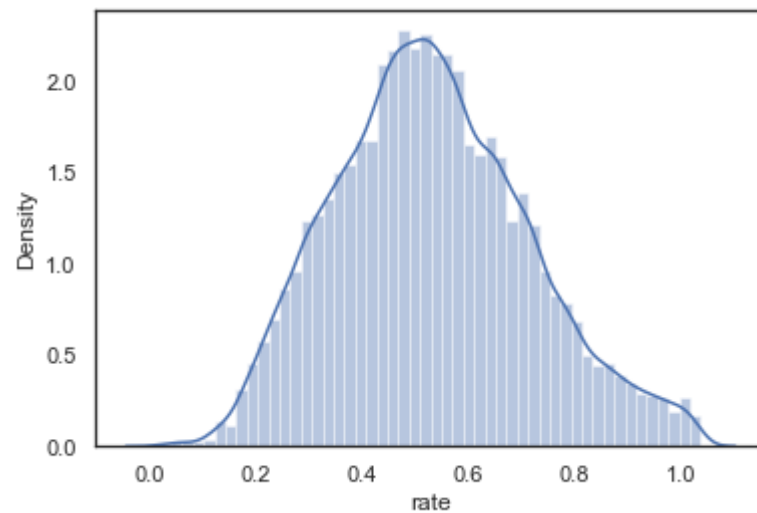


Gambar IV.29 Histogram distribusi atribut rate

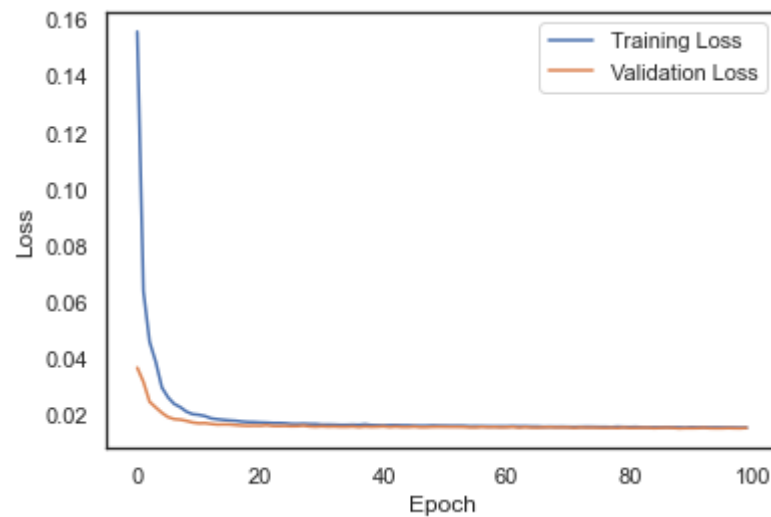
Berikut adalah *boxplot* yang memvisualisasikan *outlier* pada atribut rate



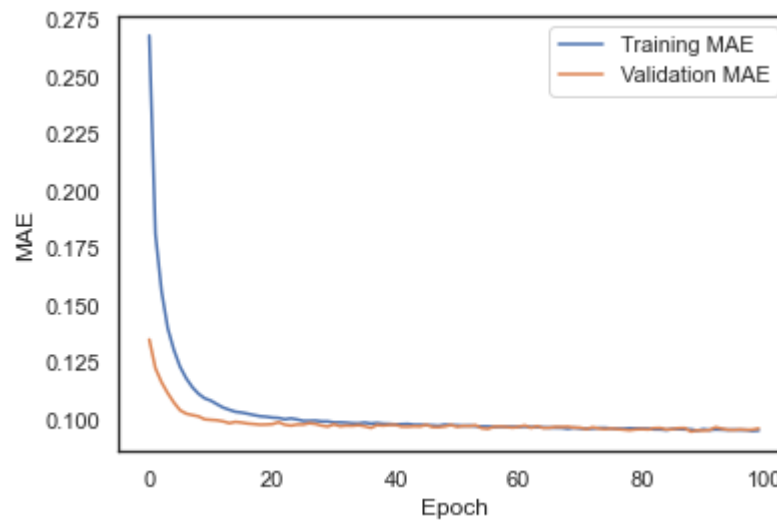
Setelah dilakukan pembersihan *outlier* pada atribut rate menggunakan metode IQR (Interquartile Range) berikut adalah histogram distribusi atribut rate terbaru.



Setelah dilakukan penyesuaian ulang terhadap data, model ANN yang sebelumnya digunakan dijalankan ulang tanpa mengubah parameter apapun. Berikut adalah grafik *learning curve loss function* dan MAE.



Gambar IV.30 Grafik *learning curve loss function Hyperparameter* dengan penghapusan *outlier*

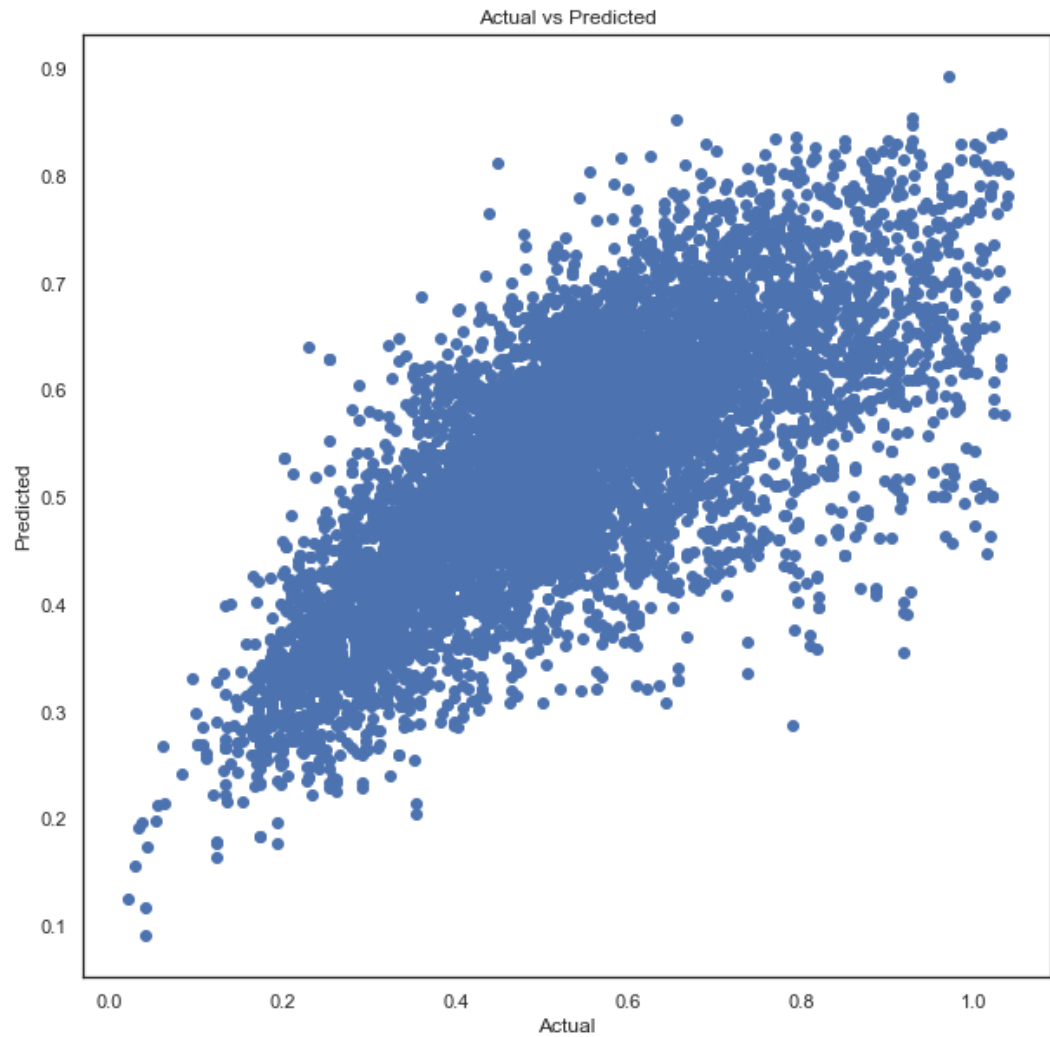


Gambar IV.31 Grafik *learning curve MAE Hyperparameter* dengan penghapusan *outlier*

Berikut adalah skor model setelah dilakukan *Hyperparameter* dengan penghapusan *outlier* dan visualisasi *scatter plot* perbandingan antara hasil prediksi dengan data actual.

Table IV.5 *Metrics evaluation model dengan penghapusan outlier*

Metrics	Skor
Mean Squared Error	0,655854706016367
R-Squared	0,32013653994695335
Mean Absolute Error	0,2690485579808164
Root Mean Squared Error	0,8098485698057181
Adjusted R-Squared	0.31955927151837027



Gambar IV.32 Visualisasi *scatter plot* hasil prediksi dengan data actual Setelah *Hyperparameter* dengan metode penghapusan *outlier*

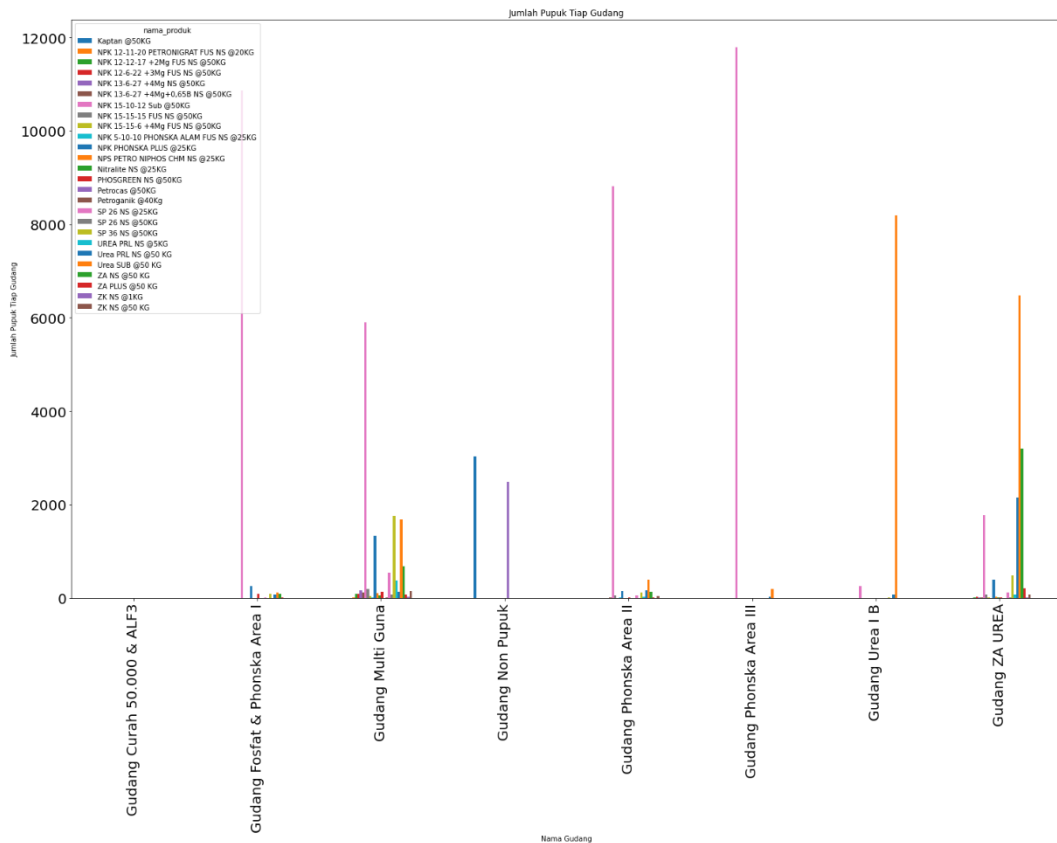
BAB V

ANALISIS DAN INTERPRETASI HASIL

Bab ini menjelaskan mengenai analisis dan interpretasi hasil pada hasil pengolahan data. Adapun analisis yang dilakukan pada penelitian ini adalah analisis deskriptif, analisis *feature selection*, analisis evaluasi model, dan analisis *hyperparameter tuning*.

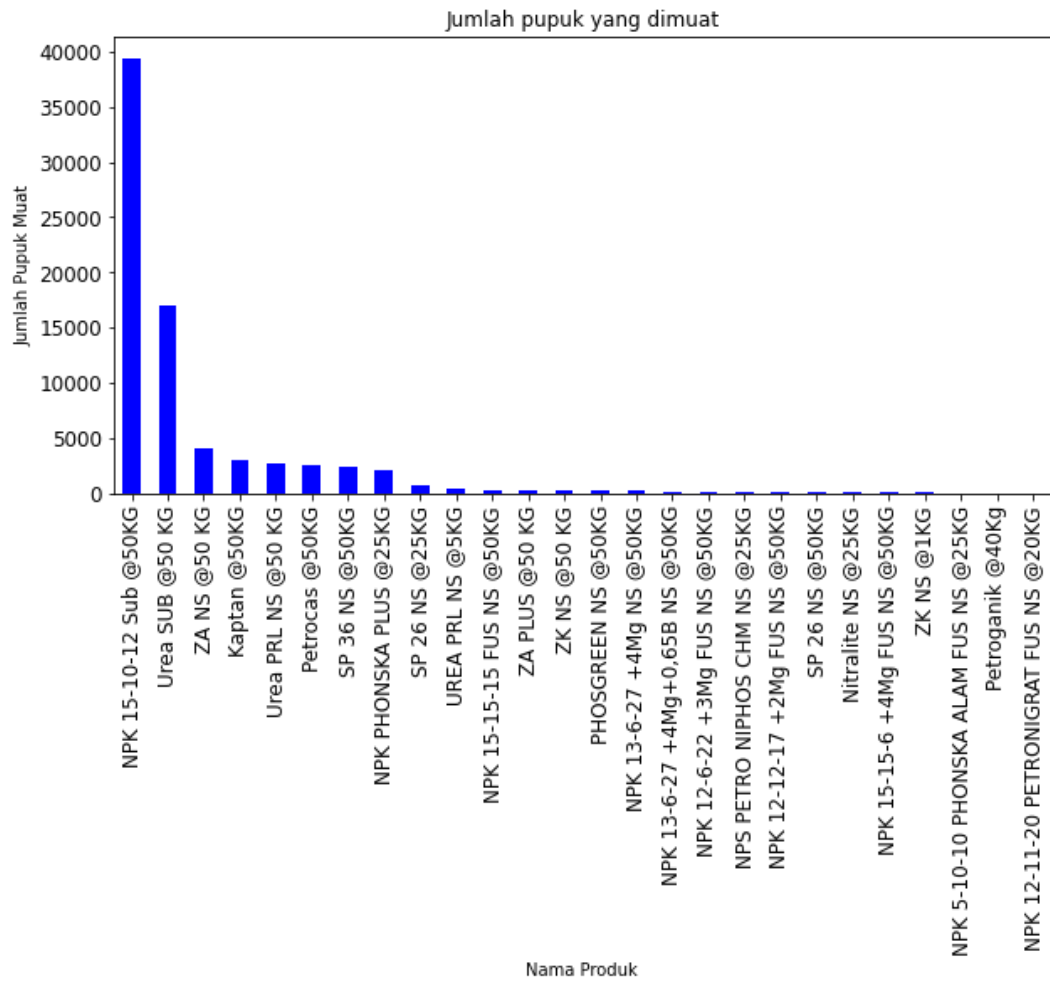
5.1 Analisis Deskriptif

Tahap awal dari penerapan *framework* CRISP-DM adalah pemahaman terhadap data. Eksplorasi data melalui analisis deskriptif penting untuk dilakukan untuk mendapatkan pemahaman yang lebih mendalam tentang karakteristik dataset yang digunakan. Selain itu, tahap ini penting untuk menemukan wawasan baru atau *insight* dari data sehingga bisa dilakukan analisis lanjutan. Berikut adalah visualisasi jumlah pupuk yang dimuat untuk tiap gudang.



Gambar V.1 Jumlah pemuatan pupuk tiap Gudang

Dari gambar dapat dilihat bahwa pupuk NPK 50 Kg merupakan jenis pupuk yang paling banyak dimuat dalam dataset ini terutama pada Gudang Fosfat & Phonska Area I, Gudang Multi Guna, Gudang Phonska Area II, dan Gudang Phonska Area III. Berikut adalah visualisasi jumlah pemuatan secara total.



Gambar V.2 Jumlah pupuk yang dimuat

Dari visualisasi diatas dapat dilihat bahwa pupuk NPK 50 Kg merupakan pupuk yang paling banyak dimuat dan NPK 20 Kg merupakan pupuk yang paling sedikit dimuat selama satu periode dari dataset yang digunakan. Berikut adalah visualisasi jumlah pemutatan pupuk sepanjang tahun 2022



Gambar V.3 Jumlah pemuatan pupuk selama tahun 2022

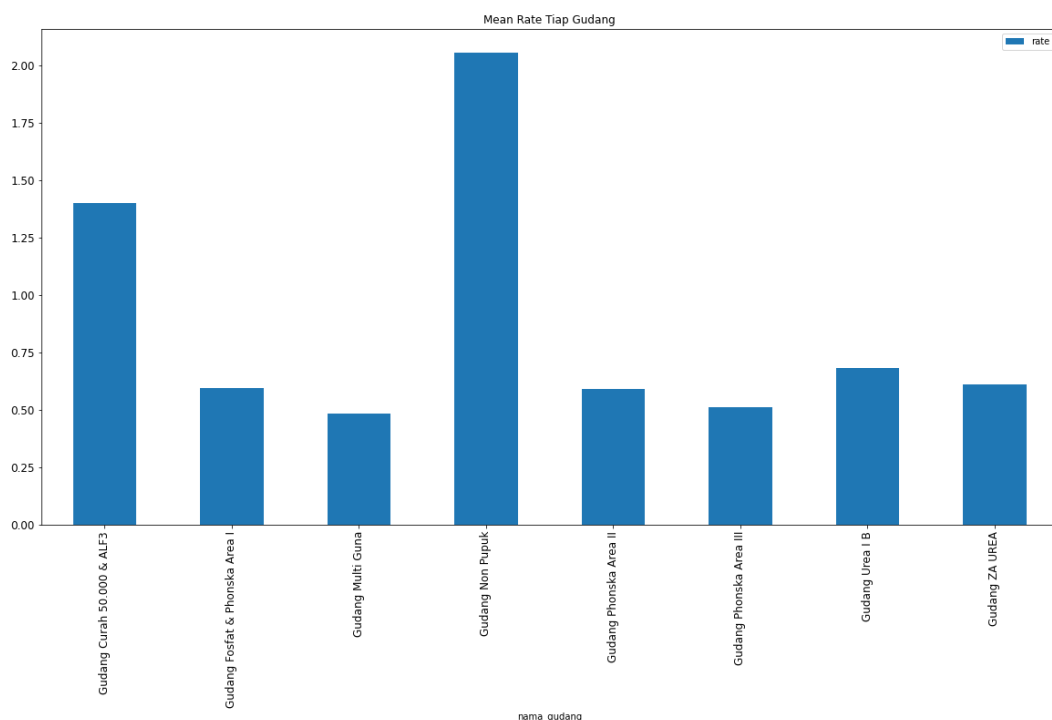
Dari gambar dapat disimpulkan bahwa jumlah pemuatan tertinggi pada bulan Desember 2022 dengan jumlah tonase mencapai 10000 Ton dan pemuatan terkecil pada bulan Mei 2022 dengan pemuatan hanya sekitar 2300 Ton. Berikut adalah nilai minimum, maksimum, dan rata-rata rate pemuatan untuk tiap jenis pupuk.

Table V.1 Data maksimum, minimum, dan *mean* rate

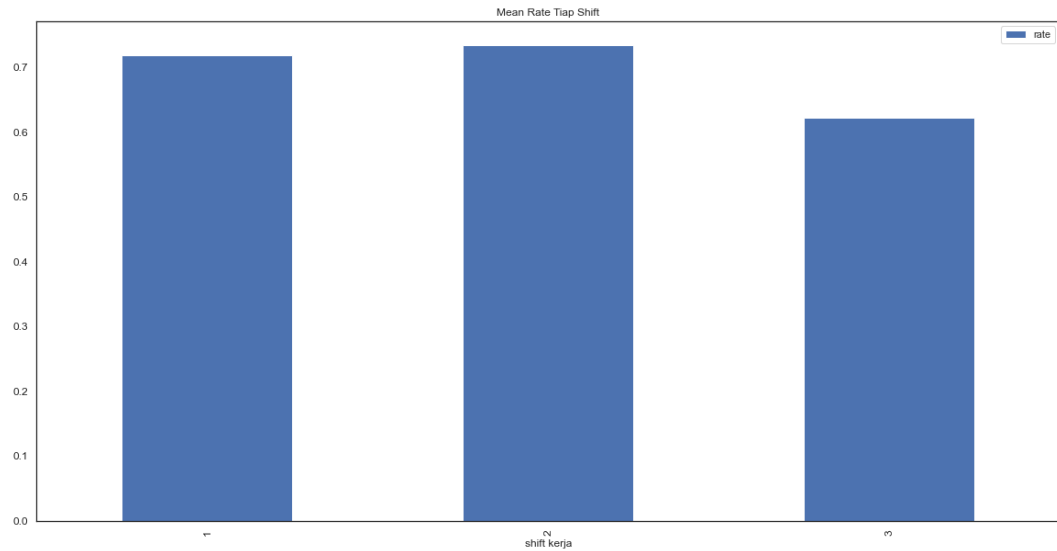
nama_produk	max	min	mean
Kaptan @50KG	40	0,025714286	2,402901945
NPK 12-11-20 PETRONIGRAT FUS NS @20KG	0,345815832	0,196768061	0,249785407
NPK 12-12-17 +2Mg FUS NS @50KG	0,862068966	0,13559322	0,36138076
NPK 12-6-22 +3Mg FUS NS @50KG	2,681818182	0,119047619	0,585936759
NPK 13-6-27 +4Mg NS @50KG	1,907769231	0,177758887	0,348757718
NPK 13-6-27 +4Mg+0,65B NS @50KG	2,380952381	0,037037037	0,360782547
NPK 15-10-12 Sub @50KG	17,85857143	0,033898305	0,567601027
NPK 15-15-15 FUS NS @50KG	1,890243902	0,014925373	0,402119559
NPK 15-15-6 +4Mg FUS NS @50KG	1,402985075	0,142394822	0,410671023
NPK 5-10-10 PHONSKA ALAM FUS NS @25KG	0,644210526	0,105271739	0,281501449
NPK PHONSKA PLUS @25KG	11,84210526	0,014074074	0,465515299
NPS PETRO NIPHOS CHM NS @25KG	12	0,119047619	0,917222422
Nitralite NS @25KG	1,25	0,095238095	0,391340974
PHOSGREEN NS @50KG	1,166666667	0,040909091	0,401237892
Petrocas @50KG	22,66666667	0,181818182	1,632837929
Petroganik @40Kg	0,541666667	0,179981343	0,356883172

SP 26 NS @25KG	11,66666667	0,102236842	0,455597113
SP 26 NS @50KG	1,512820513	0,010714286	0,448590947
SP 36 NS @50KG	11,665	0,017241379	0,448897391
UREA PRL NS @5KG	1,25	0,0534	0,34203631
Urea PRL NS @50 KG	25	0,014074074	0,6764156
Urea SUB @50 KG	10	0,044871795	0,637016125
ZA NS @50 KG	7,785714286	0,03125	0,52433246
ZA PLUS @50 KG	2,5	0,021428571	0,481395823
ZK NS @1KG	0,557471264	0,036507937	0,27195776
ZK NS @50 KG	2,244444444	0,029411765	0,341962485

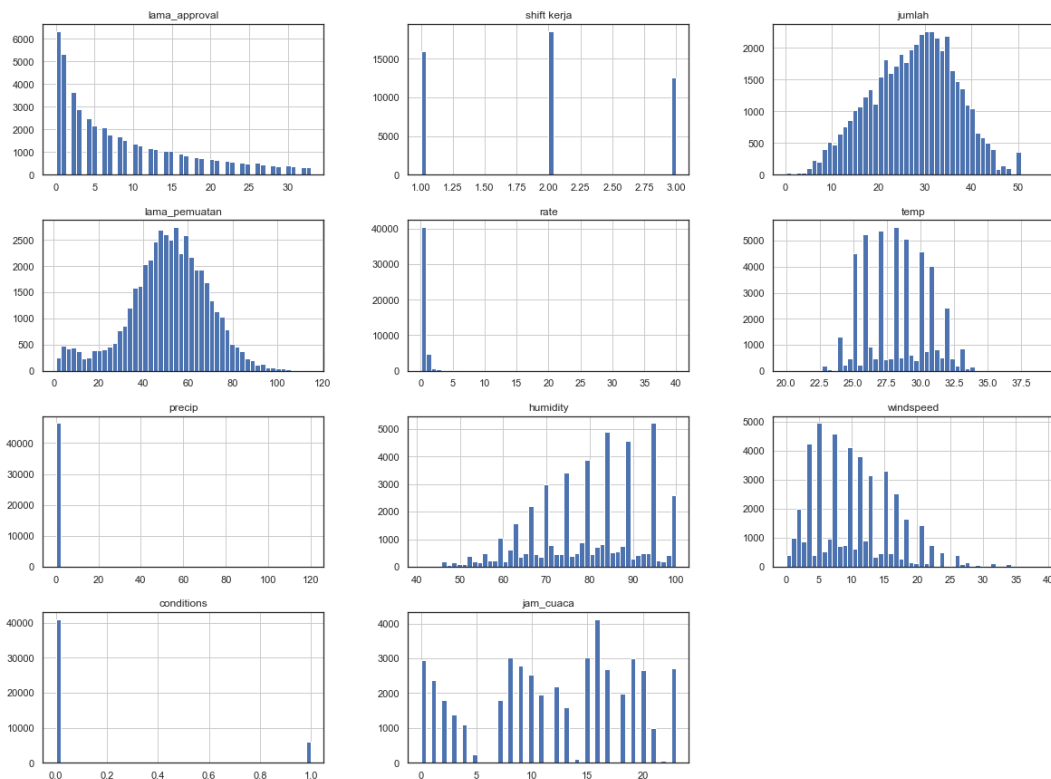
Tabel diatas memperlihatkan nilai maksimum, minimum, dan *mean* dari rate pemuatan pupuk. Dari pengamatan sekilas tampak ada data yang sangat jauh dari rata-rata dan dapat diindikasikan sebagai *outlier*. Untuk pembuktian apakah data tersebut tergolong outlier atau bukan akan dilakukan di analisis selanjutnya. Berikut adalah visualisasi *mean* rate untuk tiap gudang pada dataset.



Dari visualisasi tersebut dapat terlihat bahwa pada Gudang Non Pupuk memiliki tingkat produktivitas tertinggi karena memiliki rata-rata rate tertinggi daripada pemuatan di gudang lain. Sementara itu Gudang Multi Guna memiliki rate paling rendah dibandingkan dengan gudang lain. Namun informasi ini tidak bisa langsung dinyatakan valid karena mempertimbangkan informasi sebelumnya bahwa ditemukan indikasi outlier pada data rate yang dapat menimbulkan bias. Berikut adalah visualisasi *mean* rate untuk tiap *shift* kerja.



Dari visualisasi tersebut dapat diketahui bahwa *shift* kerja ke-2 yang berlangsung dari jam 15:00 hingga jam 23:00 memiliki rate tertinggi dibandingkan dengan *shift* lain, sementara itu *shift* kerja ke-3 yang berlangsung dari jam 23:00 hingga jam 07:00 memiliki skor rate paling rendah dibandingkan dengan *shift* kerja lainnya. Berdasarkan informasi yang didapat ada kemungkinan model ANN dapat memahami pola dari data dan menghasilkan prediksi yang akurat. Berikut adalah histogram distribusi variabel numerik pada *dataset* pemuatan pupuk.



Gambar V.4 Histogram distribusi variabel numerik

Dari histogram tersebut berdasarkan bentuk kurva nya dapat dilihat bahwa atribut rate kemungkinan tidak berdistribusi normal karena bentuk kurva yang tidak simetris. Selain itu, nilai puncak dari atribut rate juga tidak berada di tengah kurva. Berdasarkan informasi distribusi ini ada kemungkinan akan berpengaruh signifikan terhadap performa model ANN.

5.2 Analisis Hubungan dan Pengujian Hipotesis.

Sebelum membangun model prediksi dilakukan analisis korelasi dan signifikansi terutama untuk variabel X terhadap Y sehingga bisa menjadi pertimbangan konstruksi model. Pada penelitian ini uji korelasi menggunakan metode *Pearson Correlation Coefficient* melalui modul *scipy.stats*. berikut adalah hasil dari koefisien korelasi pearson dan signifikansi *p-value* variabel X terhadap Y.

Table V.2 Koefisien korelasi dan *p-value*

X	Y	Koefisien korelasi Pearson (r)	P-Value
Jumlah	rate	0.096	1.9244793612280302e-97
Lama Pemuatan	rate	-0.489	0.0
Shift Kerja	rate	-0.039	3.937658169852203e-17
Jam	rate	0.026	2.769946322379121e-08
Temperature	rate	0.044	2.3744097591455336e-21
Precipitation	rate	-0.000	0.9542631172113246
Humidity	rate	-0.039	1.5137562511140666e-17
Windspeed	rate	0.009	0.04423494986092527
Conditions	rate	-0.006	0.21178727427075414

Korelasi Pearson mengukur kekuatan dan arah hubungan linear antara dua variabel. Rentang nilai korelasi Pearson berkisar antara -1 hingga 1. Semakin dekat nilai korelasi dengan 1 atau -1, semakin kuat hubungan linier antara kedua variabel tersebut. Selain itu, dalam pengujian ini batas signifikansi (α) yang ditetapkan untuk menentukan apakah dapat menolak atau menerima hipotesis nol berdasarkan perbandingan dengan nilai p-value adalah sebesar 0.05, Hipotesis nol (H_0) menyatakan bahwa bahwa tidak ada hubungan (korelasi = 0) antara dua variabel, sementara hipotesis alternatif (H_1) menyatakan bahwa ada hubungan (korelasi \neq 0). jika p-value lebih kecil dari α maka tolak H_0 dan terima H_1 yang artinya terdapat bukti yang cukup untuk menyatakan bahwa ada hubungan yang signifikan antara variabel x dan y, sedangkan jika p-value lebih besar daripada α maka terima H_0 dan tolak H_1 yang artinya tidak cukup bukti untuk menyatakan bahwa ada hubungan yang signifikan antara variabel x dan y.

Hubungan antara variabel jumlah terhadap rate pemuatan pupuk menghasilkan skor koefisien korelasi Pearson sebesar 0.096 sehingga dapat disimpulkan bahwa hubungan antara variabel jumlah terhadap rate pemuatan pupuk adalah positif lemah, tetapi berdasarkan nilai p-value sebesar 1.9244793612280302e-97 yang mana kurang dari batas signifikansi α maka tolak

H0 dan terima H1 yang artinya terdapat bukti statistik yang kuat untuk menyatakan bahwa ada hubungan yang signifikan antara variabel jumlah dengan rate pemuatan pupuk.

Hubungan antara variabel lama pemuatan terhadap rate pemuatan pupuk menghasilkan skor Koefisien korelasi Pearson sebesar -0.489, menunjukkan hubungan negatif yang cukup kuat antara keduanya. Dengan nilai p-value sebesar 0.0 yang jauh lebih kecil dari tingkat signifikansi $\alpha=0.05$, terdapat bukti statistik yang kuat untuk menolak hipotesis nol (H0) dan menerima hipotesis alternatif (H1). Dengan demikian, dapat disimpulkan bahwa terdapat hubungan yang signifikan antara variabel lama pemuatan dengan rate pemuatan pupuk. Hubungan tersebut menunjukkan adanya korelasi negatif yang cukup kuat antara kedua variabel tersebut.

Hubungan antara variabel *shift* kerja terhadap rate pemuatan pupuk menghasilkan skor Koefisien korelasi Pearson sebesar -0.039, menunjukkan hubungan negatif yang sangat lemah antara keduanya. Namun, nilai p-value sebesar $3.937658169852203e-17$ yang jauh lebih kecil dari tingkat signifikansi $\alpha=0.05$ memberikan bukti statistik yang kuat untuk menolak hipotesis nol (H0) dan menerima hipotesis alternatif (H1). Dengan demikian, terdapat bukti statistik yang kuat untuk menyatakan bahwa ada hubungan yang signifikan antara variabel *shift* kerja dengan rate pemuatan pupuk, meskipun hubungannya sangat lemah.

Hubungan antara variabel jam terhadap rate pemuatan pupuk menghasilkan skor koefisien korelasi Pearson sebesar 0.026. Skor tersebut menunjukkan adanya hubungan positif yang sangat lemah antara variabel jam dan rate pemuatan pupuk. Namun, berdasarkan nilai p-value sebesar $2.769946322379121e-08$ yang jauh lebih kecil dari tingkat signifikansi α , kita dapat menolak hipotesis nol (H0) dan menerima hipotesis alternatif (H1). Artinya, terdapat bukti statistik yang kuat untuk menyatakan bahwa ada hubungan yang signifikan antara variabel jam dengan rate pemuatan pupuk.

Hubungan antara variabel *temperature* terhadap rate pemuatan pupuk menghasilkan skor koefisien korelasi Pearson sebesar 0.044. Skor ini menunjukkan adanya hubungan positif yang sangat lemah antara variabel *temperature* dan rate pemuatan pupuk. Namun, dengan nilai p-value sebesar $2.3744097591455336e-21$ yang jauh lebih kecil dari tingkat signifikansi α , kita dapat menolak hipotesis nol (H0) dan menerima hipotesis alternatif (H1). Dengan demikian, terdapat bukti statistik yang kuat untuk menyatakan bahwa ada hubungan yang signifikan antara variabel *temperature* dengan rate pemuatan pupuk.

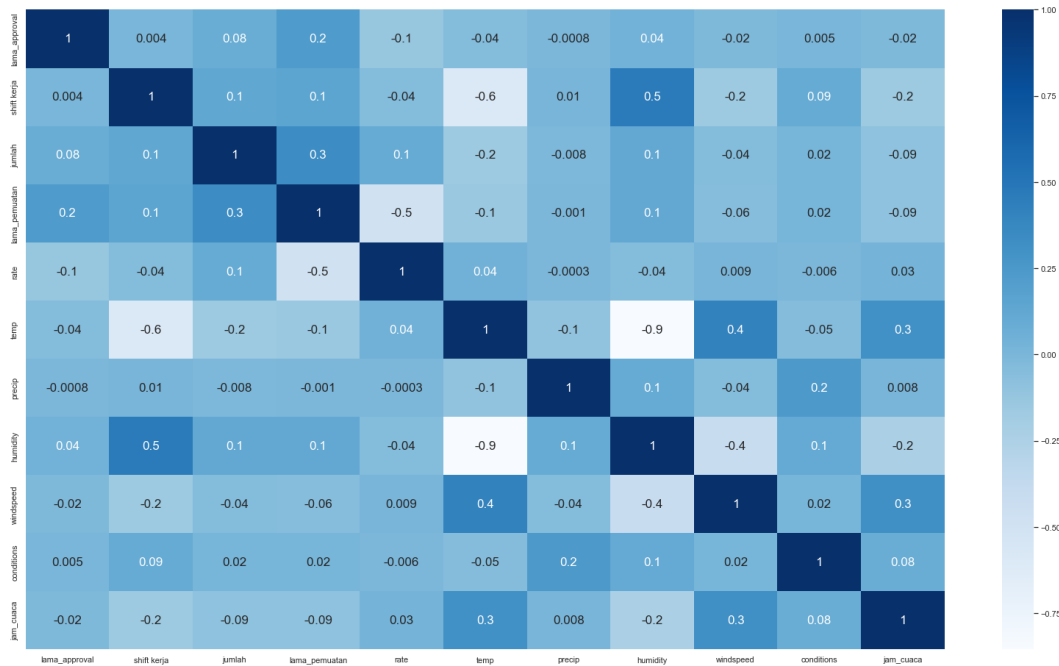
Hubungan antara variabel *precipitation* terhadap rate pemuatan pupuk menghasilkan skor Koefisien korelasi Pearson sebesar -0.000. Skor tersebut menunjukkan bahwa tidak ada hubungan linier yang signifikan antara variabel *precipitation* dan rate pemuatan pupuk. Hal ini didukung oleh nilai p-value sebesar 0.9542631172113246 yang jauh lebih besar dari tingkat signifikansi α . Oleh karena itu, tidak terdapat bukti statistik yang cukup untuk menolak hipotesis nol (H0) dan

menyimpulkan bahwa tidak ada hubungan yang signifikan antara variabel precipitation dengan rate pemuatan pupuk.

Hubungan antara variabel *humidity* terhadap rate pemuatan pupuk menghasilkan skor Koefisien korelasi Pearson sebesar -0.039. Skor tersebut menunjukkan adanya hubungan negatif yang sangat lemah antara variabel *humidity* dan rate pemuatan pupuk. Namun, dengan nilai p-value sebesar $1.5137562511140666e-17$ yang jauh lebih kecil dari tingkat signifikansi α , kita dapat menolak hipotesis nol (H_0) dan menerima hipotesis alternatif (H_1). Artinya, terdapat bukti statistik yang kuat untuk menyatakan bahwa ada hubungan yang signifikan antara variabel *humidity* dengan rate pemuatan pupuk.

Hubungan antara variabel *windspeed* terhadap rate pemuatan pupuk menghasilkan skor Koefisien korelasi Pearson sebesar 0.009. Skor tersebut menunjukkan adanya hubungan positif yang sangat lemah antara variabel *windspeed* dan rate pemuatan pupuk. Dengan nilai p-value sebesar 0.04423494986092527 yang lebih kecil dari tingkat signifikansi $\alpha=0.05$, terdapat bukti statistik yang cukup untuk menolak hipotesis nol (H_0) dan menerima hipotesis alternatif (H_1). Oleh karena itu, dapat disimpulkan bahwa ada hubungan yang signifikan antara variabel *windspeed* dengan rate pemuatan pupuk.

Hubungan antara variabel *conditions* (hujan atau tidak hujan) terhadap rate pemuatan pupuk menghasilkan skor Koefisien korelasi Pearson sebesar -0.006. Skor tersebut menunjukkan adanya hubungan negatif yang sangat lemah antara variabel *conditions* dan rate pemuatan pupuk. Namun, dengan nilai p-value sebesar 0.21178727427075414 yang lebih besar dari tingkat signifikansi $\alpha=0.05$, tidak terdapat bukti statistik yang cukup untuk menolak hipotesis nol (H_0). Oleh karena itu, tidak dapat disimpulkan bahwa terdapat hubungan yang signifikan antara variabel *conditions* dengan rate pemuatan pupuk. Berikut adalah visualisasi korelasi antar variabel pada dataset menggunakan heatmap.



Gambar V.5 Heatmap untuk visualisasi korelasi antar variabel

5.3 Evaluasi Model

Pada penelitian ini, model *Artificial Neural network* (ANN) telah dikembangkan untuk memprediksi rate pemuatan pupuk. Evaluasi model sangat penting untuk mengukur sejauh mana model dapat melakukan prediksi yang akurat terhadap data yang diberikan.

Hasil evaluasi model *machine learning* menunjukkan bahwa model yang telah dikembangkan memiliki tingkat kesalahan yang relatif tinggi dalam memprediksi rate pemuatan pupuk dalam kurun waktu tertentu. Skor *Mean Squared Error* (MSE) sebesar 0.6558 mengindikasikan adanya kesalahan yang signifikan antara nilai prediksi dan nilai sebenarnya. MSE yang tinggi menunjukkan bahwa model memiliki tingkat ketepatan yang rendah dalam memprediksi rate pemuatan pupuk, di mana perbedaan antara nilai prediksi dan nilai sebenarnya cenderung besar.

Skor *R-Squared* (R²) sebesar 0.320 menggambarkan kemampuan model dalam menjelaskan variasi dalam data target. Nilai R² yang relatif rendah menandakan bahwa model hanya mampu menjelaskan sekitar 32% variasi yang terdapat dalam data. Dalam konteks ini, R² yang rendah menunjukkan bahwa model tidak efektif dalam menggambarkan hubungan antara fitur-fitur yang digunakan dan rate pemuatan pupuk yang diobservasi.

Mean Absolute Error (MAE) sebesar 0.269 mengindikasikan tingkat kesalahan rata-rata antara prediksi model dan nilai sebenarnya. MAE yang lebih rendah menunjukkan bahwa model memiliki tingkat ketepatan yang lebih tinggi dalam memprediksi jumlah pemuatan pupuk, di mana perbedaan antara nilai prediksi dan nilai sebenarnya cenderung lebih kecil.

Root Mean Squared Error (RMSE) sebesar 0.8098 menunjukkan nilai rata-rata kesalahan prediksi model dengan memperhitungkan perbedaan kuadrat antara nilai prediksi dan nilai sebenarnya. RMSE yang lebih rendah mengindikasikan bahwa model memiliki tingkat akurasi yang lebih tinggi dalam memprediksi rate pemuatan pupuk, di mana perbedaan antara nilai prediksi dan nilai sebenarnya cenderung lebih kecil.

Untuk mengukur performa model selain menggunakan skor *metrics evaluation* juga bisa dilihat dari grafik *learning curve* untuk *loss function* dan *validation loss* MAE. Dari grafik *learning curve loss function*, terlihat bahwa *training loss* (loss pada data pelatihan) secara konsisten mengalami penurunan seiring dengan peningkatan jumlah *epoch*. Penurunan yang stabil ini menunjukkan bahwa model secara bertahap mempelajari pola-pola yang ada dalam data pelatihan dan mampu mengurangi kesalahan prediksi seiring berjalannya waktu.

Selain itu, *validation loss* (loss pada data validasi) juga mengalami penurunan yang sejalan dengan *training loss*. Hal ini menunjukkan bahwa model mampu melakukan generalisasi dengan baik pada data baru yang belum pernah dilihat sebelumnya. Perbedaan antara *training loss* dan *validation loss* juga terlihat relatif kecil, menunjukkan bahwa model tidak mengalami *overfitting* yang berarti.

Grafik *learning curve loss function* ini memberikan bukti bahwa model ANN yang dikembangkan mampu belajar dari data pelatihan dan menghasilkan prediksi yang semakin baik seiring dengan peningkatan jumlah *epoch*. Meskipun terdapat beberapa fluktuasi pada awal pelatihan, model berhasil menemukan pola-pola yang relevan dalam data sehingga *loss function* dapat dikurangi secara signifikan.

Grafik *learning curve* MAE menunjukkan bahwa model ANN yang dikembangkan memiliki kemampuan yang baik dalam mengurangi kesalahan prediksi selama pelatihan. Penurunan yang signifikan pada *training* MAE menunjukkan bahwa model dapat belajar dan memperbaiki prediksi seiring dengan peningkatan *epoch*. Meskipun *validation* MAE cenderung konstan dan mengalami fluktuasi kecil, tingkat kesalahan yang relatif rendah dan serupa dengan *training* MAE menunjukkan bahwa model mampu melakukan generalisasi dengan baik pada data validasi.

Grafik Plot MAE ini memberikan indikasi positif bahwa model ANN yang dikembangkan mampu menghasilkan prediksi dengan tingkat kesalahan yang rendah dan konsisten pada data pelatihan maupun data validasi. grafik ini menunjukkan bahwa model pada awal pelatihan memiliki tingkat kesalahan yang relatif tinggi, namun dengan peningkatan jumlah *epoch*, model berhasil memperbaiki prediksi dan mengurangi tingkat kesalahan. Tingkat kesalahan yang relatif rendah dan serupa antara *Training* MAE dan *Validation* MAE menunjukkan bahwa model dapat melakukan generalisasi dengan baik pada data yang tidak digunakan saat pelatihan. Meskipun *Validation* MAE mengalami sedikit fluktuasi, tingkat kesalahan tetap berada dalam kisaran yang stabil.

Berdasarkan hasil evaluasi tersebut, dapat disimpulkan bahwa model yang dikembangkan termasuk kedalam kategori *good fit* atau *robust* karena tidak mengalami *overfitting* dan *underfitting*. Selain itu, berdasarkan grafik *learning curve* model ANN yang dibangun mampu belajar dari data pelatihan dan menghasilkan prediksi dan generalisasi yang semakin baik seiring dengan peningkatan jumlah *epoch*. Namun berdasarkan skor evaluasi model belum mampu memberikan prediksi yang akurat dalam memprediksi rate pemuatan pupuk. Skor evaluasi yang tinggi pada MSE dan R2 serta rendahnya skor pada MAE dan RMSE menunjukkan bahwa model perlu diperbaiki dan ditingkatkan untuk meningkatkan akurasi prediksinya.

5.4 Analisis Hyperparameter Tuning

Setelah membangun model dan mengevaluasi kinerja model dilakukan teknik *Hyperparameter* untuk meningkatkan skor dan akurasi prediksi model. Berdasarkan percobaan *Hyperparameter* yang dilakukan maka dipilih *Hyperparameter* dengan metode penyesuaian data ulang karena terbukti meningkatkan kinerja model secara signifikan.

Berdasarkan *Hyperparameter* yang dilakukan didapatkan hasil *Mean Squared Error* sebesar 0,655854706016367. Semakin rendah nilai MSE, semakin baik model dalam melakukan prediksi yang akurat. Dalam kasus ini, nilai MSE yang rendah menunjukkan bahwa model memiliki tingkat kesalahan yang relatif rendah dalam memprediksi nilai target.

R-squared memiliki nilai sebesar 0,5449207091676362. Nilai *R-squared* berkisar antara 0 hingga 1, dan semakin mendekati 1, semakin baik model dalam menjelaskan variasi data. Dalam kasus ini, nilai *R-squared* yang sekitar 0.545 menunjukkan bahwa sekitar 54.5% variasi data target dapat dijelaskan oleh model.

MAE memiliki nilai sebesar 0.09657313230269346. Semakin rendah nilai MAE, semakin baik model dalam melakukan prediksi yang akurat. Dalam kasus ini, nilai MAE yang rendah menunjukkan bahwa model memiliki tingkat kesalahan yang relatif rendah dalam memprediksi nilai target.

RMSE memiliki nilai sebesar 0,12424787965055246. RMSE merupakan akar kuadrat dari MSE dan digunakan untuk memberikan interpretasi dalam satuan yang sama dengan data asli. Semakin rendah nilai RMSE, semakin baik model dalam melakukan prediksi yang akurat. Dalam kasus ini, nilai RMSE yang rendah menunjukkan bahwa model memiliki tingkat kesalahan yang relatif rendah dalam memprediksi nilai target.

Adjusted R-squared memiliki nilai sebesar 0.5445147508350472. *Adjusted R-squared* merupakan variasi dari *R-squared* yang memperhitungkan jumlah fitur atau variabel yang digunakan dalam model. Nilai *Adjusted R-squared* yang sekitar 0.545 menunjukkan bahwa sekitar 54.5% variasi data target dapat dijelaskan oleh model dengan mempertimbangkan jumlah fitur yang digunakan.

Secara keseluruhan, hasil evaluasi model menunjukkan bahwa model yang telah dibangun ulang memiliki tingkat kesalahan yang relatif rendah dan mampu menjelaskan sebagian variasi data target dengan cukup baik.

Berdasarkan grafik *learning curve* Pada awal pelatihan, *training loss* memiliki nilai yang sedikit tinggi sekitar 0.16, sementara *validation loss* sudah memiliki nilai yang relatif rendah yaitu 0.0367. Hal ini menunjukkan bahwa model pada awalnya mungkin memiliki tingkat kesalahan yang cukup tinggi dalam memprediksi data pelatihan, tetapi secara relatif baik dalam memprediksi data validasi. Namun, seiring dengan peningkatan jumlah *epoch*, *training loss* secara stabil turun hingga mencapai nilai sekitar 0.0155. Penurunan yang stabil ini menunjukkan bahwa model secara bertahap mempelajari pola-pola yang ada dalam data pelatihan dan mampu mengurangi kesalahan prediksi seiring dengan bertambahnya waktu.

Selain itu, *validation loss* juga mengalami penurunan yang stabil seiring dengan bertambahnya *epoch* dan berhenti di nilai yang hampir sama dengan *training loss*, yaitu sekitar 0.0154. Hal ini menunjukkan bahwa model mampu melakukan generalisasi dengan baik pada data validasi yang belum pernah dilihat sebelumnya. Perbedaan yang sangat kecil antara *training loss* dan *validation loss* menunjukkan bahwa model tidak mengalami *overfitting* ataupun *underfitting* yang berarti.

BAB VI

KESIMPULAN DAN SARAN

Bab ini menjelaskan mengenai kesimpulan berdasarkan pengolahan data, analisis dan interpretasi hasil yang telah dilakukan. Selain itu juga diberikan saran bagi perusahaan agar dapat dilakukan perbaikan

6.1 Kesimpulan

Subbab ini menjelaskan mengenai kesimpulan, analisis dan interpretasi hasil yang telah dilakukan. Kesimpulan dari penelitian ini adalah sebagai berikut.

1. Berdasarkan analisis deskriptif dari eksplorasi data, ditemukan beberapa pola pada variabel independen yang berpotensi mempengaruhi tingkat pemuatan pupuk. Misalnya, terdapat pola pada jenis pupuk tertentu atau pola pada *shift* kerja yang berkorelasi dengan tingkat pemuatan pupuk. Hal ini menunjukkan adanya faktor-faktor yang berperan dalam menentukan rate pemuatan pupuk di PT Petrokimia Gresik.
2. Berdasarkan hasil analisis korelasi yang telah dilakukan, ditemukan beberapa temuan penting terkait hubungan antara variabel independen dengan rate pemuatan pupuk. Analisis menunjukkan bahwa beberapa variabel seperti jumlah, lama pemuatan, *shift* kerja, jam, *temperature*, *humidity*, *windspeed*, dan *conditions* memiliki korelasi yang rendah dengan rate pemuatan pupuk. Meskipun demikian, variabel-variabel ini tetap memiliki signifikansi yang tinggi terhadap tingkat pemuatan pupuk. Hal ini menunjukkan bahwa meskipun korelasinya rendah, variabel-variabel tersebut masih memiliki pengaruh yang signifikan terhadap rate pemuatan pupuk. Kesimpulan ini mengindikasikan adanya faktor-faktor lain yang mungkin mempengaruhi hubungan antara variabel independen dengan rate pemuatan pupuk, yang tidak terlihat dalam analisis korelasi. Oleh karena itu, meskipun korelasi antara variabel-variabel tersebut rendah, tetap perlu mempertimbangkan variabel-variabel tersebut dalam model prediksi untuk mendapatkan gambaran yang lebih komprehensif mengenai faktor-faktor yang mempengaruhi rate pemuatan pupuk.
3. Setelah melakukan pemodelan, evaluasi, dan *tuning hyperparameter*, model prediksi yang dikembangkan dapat dipertimbangkan untuk dilakukan

deployment karena memiliki skor yang cukup bagus dan akurat dalam memprediksi rate pemuatan pupuk. Skor evaluasi yang diperoleh adalah sebagai berikut: Mean Absolute Error (MAE): 0.09657, Mean Squared Error (MSE): 0.01544, Root Mean Squared Error (RMSE): 0.12425, R-squared (R²): 0.54492, dan Adjusted R-squared: 0.54451. nilai MAE yang rendah menunjukkan bahwa model memiliki tingkat kesalahan yang relatif rendah dalam memprediksi nilai target. Skor MSE dan RMSE yang rendah juga menunjukkan tingkat kesalahan prediksi model yang relatif kecil. Skor R-squared yang sebesar 0.54492 mengindikasikan bahwa sekitar 54.49% variabilitas dari variabel dependen dapat dijelaskan oleh variabel independen yang digunakan dalam model. Hal ini menunjukkan adanya hubungan yang signifikan antara variabel independen dengan variabel dependen. Skor Adjusted R-squared yang hampir sama dengan R-squared menandakan bahwa penambahan variabel independen dalam model tidak memberikan peningkatan yang signifikan dalam menjelaskan variabilitas data. Berdasarkan skor evaluasi tersebut, model prediksi yang dikembangkan menunjukkan kinerja yang memadai dalam memprediksi rate pemuatan pupuk.

6.2 Saran

Subbab ini menjelaskan mengenai saran berdasarkan hasil pengolahan data dan analisis yang telah dilakukan. Saran dari penelitian ini antara lain:

1. Bagi Perusahaan
 - a. PT Petrokimia Gresik sebaiknya menerapkan standar operasional atau prosedur yang jelas terkait pengambilan data pemuatan pupuk oleh operator di lapangan. Hal ini akan membantu memastikan validitas dan konsistensi data yang dikumpulkan, sehingga memudahkan analisis lebih lanjut. Beberapa langkah yang dapat dilakukan seperti membuat panduan dan instruksi yang rinci mengenai metode pengambilan data pemuatan pupuk dan melakukan pelatihan dan pemantauan terhadap operator di lapangan.

- b. Melakukan validasi data secara berkala. Perusahaan perlu melakukan validasi data secara rutin untuk memastikan keakuratan dan integritas data yang dikumpulkan. Hal ini dapat dilakukan dengan membandingkan data yang dihasilkan dengan standar atau referensi yang ada.

2. Bagi Peneliti

- a. Penelitian ini sebaiknya perlu mempertimbangkan penggunaan variabel tambahan yang memiliki potensi mempengaruhi rate pemuatan pupuk. Dengan memperluas variabel penelitian, diharapkan dapat memperoleh wawasan yang lebih komprehensif tentang faktor-faktor yang memengaruhi tingkat pemuatan pupuk dan menghasilkan model yang lebih akurat.
- b. Penelitian ini sebaiknya perlu mempertimbangkan analisis regresi lanjutan untuk menguji pengaruh variabel independen secara lebih rinci terhadap variabel dependen. Dalam analisis regresi dapat dipertimbangkan penggunaan metode seperti regresi linear berganda atau regresi logistik. Analisis regresi yang lebih mendalam akan memberikan pemahaman yang lebih baik tentang sejauh mana variabel independen mempengaruhi rate pemuatan pupuk.
- c. Model ANN yang telah dibangun sebaiknya dilakukan evaluasi dan penyetelan *hyperparameter* yang lebih mendalam dengan cara mengeksplorasi berbagai kombinasi *hyperparameter* dan metode penyetelan yang lebih cermat untuk memaksimalkan performa model. Ini termasuk mencoba berbagai metode optimasi, mengatur tingkat pembelajaran (*learning rate*), atau mengubah arsitektur model untuk mencapai hasil yang lebih baik.

DAFTAR PUSTAKA

- Abasova, J., Tanuska, P., & Rydzi, S. (2021). Big data—knowledge discovery in production industry data storages—implementation of best practices. *Applied Sciences (Switzerland)*, 11(16). <https://doi.org/10.3390/app11167648>
- De La Bourdonnaye, F., & Daniel, F. (2021). *Evaluating categorical encoding methods on a real credit card fraud detection database*. <http://www.lusisai.com>
- Fatma Adelia, P., Koesriharti, & Sunaryo. (2013). Pengaruh Penambahan Unsur Hara Mikro (Fe Dan Cu) Dalam Media Paitan Cair Dan Kotoran Sapi Cair Terhadap Pertumbuhan Dan Hasil Bayam Merah (*Amaranthus Tricolor L.*) Dengan Sistem Hidroponik Rakit Apung. *Jurnal Produksi Tanaman*, 1(3), 126195. <https://doi.org/10.21176/PROTAN.V1I3.30>
- M, S., V.J, R., & . A. (2018). Overview of Data Mining. *IJARCCCE*, 7(11), 109–113. <https://doi.org/10.17148/ijarcce.2018.71123>
- Mansyur, N. I., Pudjiwati, E. H., & Murti Laksono, A. (2021). *Pupuk dan Pemupukan* (Z. Hanum, Ed.; 1st ed., Vol. 1). Syiah Kuala University Press. https://books.google.co.id/books?hl=en&lr=&id=eiwyEAAAQBAJ&oi=fnd&pg=PP1&dq=pupuk+merupakan&ots=TCub-B9fso&sig=1-hXc7yKnIjDhGQAcF1D737R4Z8&redir_esc=y#v=onepage&q=pupuk%20merupakan&f=false
- Sakellaropoulos, T., Vougas, K., Narang, S., Koinis, F., Kotsinas, A., Polyzos, A., Moss, T. J., Piha-Paul, S., Zhou, H., Kardala, E., Damianidou, E., Alexopoulos, L. G., Aifantis, I., Townsend, P. A., Panayiotidis, M. I., Sfikakis, P., Bartek, J., Fitzgerald, R. C., Thanos, D., ... Gorgoulis, V. G. (2019). A Deep Learning Framework for Predicting Response to Therapy in Cancer. *Cell Reports*, 29(11), 3367-3373.e4. <https://doi.org/10.1016/j.celrep.2019.11.017>
- Vinutha, H. P., Poornima, B., & Sagar, B. M. (2018). Detection of outliers using interquartile range technique from intrusion dataset. *Advances in Intelligent Systems and Computing*, 701, 511–518. https://doi.org/10.1007/978-981-10-7563-6_53
- Tahwia, A., Heniegal, A., Elgamal, M., & Tayeh, B. (2021). The prediction of compressive strength and non-destructive tests of sustainable concrete by using artificial *neural networks*. *Computers and Concrete*.