



**PENGEMBANGAN MODEL PREDIKSI DAN
DECISION SUPPORT SYSTEM BERBASIS
GENERATIVE AI DAN *MACHINE LEARNING*
UNTUK MANAJEMEN CADANGAN BERAS
PEMERINTAH DI BULOG**

PROPOSAL KUALIFIKASI

Reviana Siti Mardiah

99223133

**PROGRAM DOKTOR TEKNOLOGI INFORMASI
UNIVERSITAS GUNADARMA**

2024

KATA PENGANTAR

Puji dan syukur Penulis panjatkan kehadiran Tuhan Yang Maha Esa yang telah memberikan rahmat dan berkat-Nya, sehingga Penulis dapat menyelesaikan Proposal Disertasi yang berjudul **“Pengembangan Model Prediksi dan *Decision Support System* Berbasis *Generative AI* dan *Machine Learning* untuk Manajemen Cadangan Beras Pemerintah di BULOG”** ini tepat pada waktu yang telah ditentukan. Proposal Disertasi ini disusun untuk memenuhi salah satu syarat memperoleh gelar Doktor Teknologi Informasi (S3) pada Program Doktor, Universitas Gunadarma.

Proses penyusunan Proposal Disertasi ini tidak lepas dari berbagai pihak yang telah memberikan dukungan, bimbingan dan pengarahan yang sangat berharga. Oleh karena itu, Penulis mengucapkan terima kasih atas bantuan dari berbagai pihak dalam penyelesaian Proposal Disertasi ini kepada:

1. Yayasan Pendidikan Gunadarma yang telah memberikan beasiswa kepada Penulis untuk melanjutkan studi Program Doktor Teknologi Informasi di Universitas Gunadarma.
2. Prof. Dr. E. S. Margianti, SE., MM., dan Prof. Suryadi H.S., S.Si., MMSI., selaku Rektor dan Wakil Rektor II Universitas Gunadarma yang telah memberikan kesempatan dan kepercayaan kepada Penulis untuk melanjutkan studi Program Doktor Teknologi Informasi di Universitas Gunadarma.
3. Prof. Dr. Sarifuddin Madenda selaku Ketua Program Doktor Teknologi Informasi Universitas Gunadarma yang telah memberikan motivasi untuk menyelesaikan proposal penelitian ini.
4. Prof. Dr. -ing. Adang Suhendra, S.Si., S.Kom., M.Sc., selaku promotor yang selalu meluangkan waktu, memberikan bimbingan, dan masukan yang sangat bermanfaat bagi Penulis dalam menyelesaikan proposal penelitian ini dengan rasa tanggung jawab.
5. Dr. Fitrianiingsih, S.Kom., M.M.S.I., selaku ko promotor yang selalu meluangkan waktu, memberikan bimbingan dan masukan yang sangat

bermanfaat bagi Penulis dalam menyelesaikan proposal penelitian ini dengan rasa tanggung jawab.

6. Dr. Dian Kemala Putri, S.TP., M.T., selaku ko promotor yang selalu meluangkan waktu, memberikan bimbingan dan masukan yang sangat bermanfaat bagi Penulis dalam menyelesaikan proposal penelitian ini dengan rasa tanggung jawab
7. Orang tua yang selalu memberikan doa, dukungan baik semangat maupun materil, sehingga Penulis mampu menyelesaikan Proposal ini.

Jakarta, Juni 2024

Penulis

DAFTAR ISI

KATA PENGANTAR.....	i
DAFTAR ISI.....	iii
DAFTAR GAMBAR	v
DAFTAR TABEL	vi
DAFTAR RUMUS.....	vii
BAB 1 PENDAHULUAN	1
1.1 Latar Belakang	1
1.2 Rumusan Masalah	6
1.3 Tujuan Penelitian.....	7
1.4 Batasan Penelitian	7
1.5 Kontribusi.....	7
BAB 2 TELAAH PUSTAKA	9
2.1 <i>Artificial Intelligence</i> (AI).....	9
2.2. Konsep Dasar <i>Machine Learning</i> (ML).....	11
2.2.1 Algoritma <i>eXtreme Gradient Boosting</i> (XGBoost)	12
2.2.2 <i>Exploratory Data Analysis</i>	15
2.2.3 <i>Data Preprocessing</i>	17
2.2.4 <i>Features Engineering</i>	20
2.2.5 Augmentasi Data.....	21
2.2.6 <i>Generative Adversarial Network</i> (GAN)	22
2.2.7 <i>Hyperparameter Tuning</i>	26
2.2.8 Teknik Evaluasi.....	27
2.3 Manajemen Cadangan Beras.....	29
2.3.1 Prediksi Produksi dan Permintaan	30
2.3.2 Optimasi.....	31
2.3.3 <i>Decision Support System</i> (DSS)	32
2.4 <i>Black Box Testing</i>	34
2.5. Kajian Penelitian Terdahulu	35

2.5.1 Penelitian Terdahulu Terkait Penggunaan <i>Machine Learning</i> dalam Prediksi Produksi (Hasil Panen).....	35
2.5.2 Penelitian Terdahulu Terkait Penggunaan <i>Machine Learning</i> dalam Prediksi Permintaan.....	42
2.5.3 Penelitian Terdahulu Terkait Integrasi <i>Generative AI</i> dan ML untuk Tujuan Prediksi.....	45
2.5.4 Penelitian Terdahulu Terkait <i>Decision Support System</i> , Prediksi, dan Optimasi	50
2.5.5 <i>Fishbone Diagram</i>	57
BAB 3 METODE PENELITIAN	58
3.1 Tahapan Penelitian	58
3.2 Pemodelan Manajemen Persediaan Beras Perum BULOG	59
3.3. Analisis.....	61
3.3.1 Regulasi Terkait Manajemen Cadangan Beras Pemerintah	68
3.3.2 Parameter yang Dipertimbangkan dalam Menentukan Jumlah CBP71	71
3.4. Pengembangan Model Prediksi Produksi (Hasil Panen) Beras.....	71
3.4.1 Pengumpulan Data.....	73
3.4.2 <i>Exploratory Data Analysis (EDA)</i>	74
3.4.3 <i>Preprocessing Data</i>	79
3.4.4 <i>Features Engineering</i>	80
3.4.5 Pembagian Data (<i>Split Data</i>)	81
3.4.6 Augmentasi Data.....	81
3.4.7 Pemodelan dengan <i>GridSearchCV</i>	81
3.4.8 Evaluasi Model	83
3.5 Pengembangan Model Prediksi Permintaan.....	84
3.6 <i>Prototype Decision Support System (DSS)</i> untuk Pengadaan Cadangan Beras Pemerintah	86
3.7 Uji Coba	90
DAFTAR PUSTAKA.....	91
LAMPIRAN 1 <i>SAMPLE DATASET 22 BARIS PERTAMA</i>	104

DAFTAR GAMBAR

Gambar 2. 1 Sub-bidang dari kecerdasan buatan (AI).....	10
Gambar 2. 2 Teknik <i>Machine Learning</i>	12
Gambar 2. 3 Arsitektur Dasar Algoritma XGBoost	13
Gambar 2. 4 Arsitektur GAN	23
Gambar 2. 5 <i>Fishbone</i> Diagram Usulan Penelitian.....	57
 Gambar 3. 1 Tahapan Penelitian	 59
Gambar 3. 2 Model Manajemen Persediaan Perum BULOG	61
Gambar 3. 3 Perbandingan Harga Gabah.....	63
Gambar 3. 4 Jumlah Impor Beras	64
Gambar 3. 5 Matriks SWOT	68
Gambar 3. 6 Tahapan penelitian prediksi permintaan yang diusulkan	72
Gambar 3. 7 (a) <i>Box Plots</i> , (b) <i>Density Plot</i> , (c) <i>Violin Plots</i> , dan (d) <i>Count Plot</i>	76
Gambar 3. 8 <i>Pair Plot</i>	78
Gambar 3. 9 <i>Heatmap</i> korelasi.....	78
Gambar 3. 10 <i>Flowchart feature engineering</i>	80
Gambar 3. 11 Membagi data menjadi pelatihan dan pengujian	81
Gambar 3. 12 <i>Flowchart</i> pemodelan.....	82
Gambar 3. 13 <i>Flowchart</i> evaluasi model	83
Gambar 3. 14 Tahapan penelitian prediksi permintaan yang diusulkan	85
Gambar 3. 15 Ilustrasi Pemenuhan CBP	87
Gambar 3. 16 Diagram Alir Skenario Dasar Pengambilan Keputusan	88
Gambar 3. 17 <i>Prototype</i> DSS yang Diusulkan.....	90

DAFTAR TABEL

Tabel 2. 1 Ringkasan Penelitian Terdahulu Tentang Prediksi Produksi (Hasil Panen) Padi.....	39
Tabel 2. 2 Ringkasan Penelitian Terdahulu Terkait Prediksi Permintaan	44
Tabel 2. 3 Ringkasan Penelitian Terdahulu Terkait Integrasi Gen AI dan ML untuk Tugas Prediksi	48
Tabel 2. 4 Ringkasan Penelitian Terdahulu tentang DSS, Prediksi, dan Optimasi	53
Tabel 3. 1 Jumlah Impor Beras	63
Tabel 3. 2 Deskripsi Variabel	73
Tabel 3. 3 <i>Missing value</i>	75
Tabel 3. 4 Ringkasan Statistik.....	75
Tabel 3. 5 Parameter Yang Dipilih Dan Nilainya.....	82
Tabel 3. 6 Hasil Evaluasi Model	83

DAFTAR RUMUS

Rumus (2. 1) Fungsi Tujuan GAN	23
Rumus (2. 2) <i>Gradien Fungsi Loss Discriminator</i>	24
Rumus (2. 3) <i>Gradien Fungsi Loss Generator</i>	24
Rumus (2. 4) <i>R-squared</i>	27
Rumus (2. 5) <i>Root Mean Squared Error</i>	28
Rumus (2. 6) <i>Mean Absolute Error</i>	28

BAB 1

PENDAHULUAN

1.1 Latar Belakang

Perkembangan teknologi informasi yang pesat dalam beberapa dekade terakhir telah menyebabkan transformasi besar di berbagai sektor (Anantrasirichai & Bull, 2022). Salah satu pendorong utama transformasi ini adalah kemajuan teknologi kecerdasan buatan (*Artificial Intelligence/AI*) (C. Zhang & Lu, 2021), khususnya di bidang *Machine Learning* (ML) (Alpaydin, 2020). Teknologi ini telah merevolusi berbagai sektor, termasuk kesehatan (Arumugam, Naved, Shinde, Leiva-Chauca, Huaman-Osorio, & Gonzales-Yanac, 2023; Kaur, Singh, & Verma, 2023), energi (C. Zhao, Wan, & Song, 2021), dan manufaktur (Kotsiopoulos, Sarigiannidis, Ioannidis, & Tzovaras, 2021) dengan meningkatkan efisiensi operasional melalui analisis *big data*, pengenalan pola, dan pengambilan keputusan yang lebih baik (Mehmood, Bashir, Rabie, Broderick, & Davies, 2023; Rai, Tiwari, Ivanov, & Dolgui, 2021).

Sektor pertanian dan pangan juga tidak luput dari dampak positif *Machine Learning* (ML). Teknologi ini telah terbukti efektif dalam mengoptimalkan berbagai operasi, termasuk perencanaan, manajemen persediaan, dan distribusi produk, yang berkontribusi pada peningkatan efisiensi operasional di sektor pertanian dan pangan (Kuradusenge, Hitimana, Hanyurwimfura, Rukundo, Mtonga, Mukasine, Uwitonze, Ngabonziza, & Uwamahoro, 2023; Lutoslawski, Hernes, Radomska, Hajdas, Walaszczyk, & Kozina, 2021).

Pangan merupakan kebutuhan dasar manusia yang harus dipenuhi. Hak untuk mendapatkan pangan merupakan salah satu hak asasi manusia, seperti yang dijelaskan dalam Pasal 27 Undang-Undang Dasar 1945 dan Deklarasi Roma 1996. Sebagai kebutuhan dasar dan salah satu hak asasi manusia, pangan memainkan peran penting dalam stabilitas sosial, ekonomi, dan politik suatu negara (BULOG, 2024; Clapp, Moseley, Burlingame, & Termine, 2022).

Negara memiliki kewajiban untuk memenuhi kebutuhan pangan penduduknya sebagaimana pemenuhan hak-hak lainnya. Mengabaikan hak ini

dapat dianggap sebagai pelanggaran hak asasi manusia. Ketika hak atas pangan terus menerus diabaikan, maka hal itu sama saja dengan kejahatan genosida (Maskun, Naswar, Ahmad, Ilmar, Napang, & Nugraha, 2021). Hal ini menyoroti perlunya mengamankan ketahanan pangan nasional.

Menurut UU No. 18 Tahun 2012, ketahanan pangan adalah kondisi terpenuhinya pangan yang cukup dan berkualitas bagi seluruh masyarakat. Gangguan pada ketahanan pangan, seperti meroketnya harga pangan pada saat krisis ekonomi 1997/1998, dapat memicu gejolak sosial yang membahayakan stabilitas ekonomi dan keamanan nasional (BULOG, 2024). Kejadian tersebut menunjukkan pentingnya menjaga ketahanan pangan untuk menjaga stabilitas ekonomi, politik, dan keamanan nasional (BULOG, 2024; Clapp et al., 2022).

Di Indonesia, pangan sering dikaitkan dengan komoditas beras karena beras merupakan makanan pokok dan memegang peran penting dalam ketahanan pangan. Hal ini tercermin dari tingkat konsumsi beras yang berkisar antara 85,88% hingga 91,73% per tahun (Alfa & Subagyo, 2018) dan fakta bahwa masyarakat Indonesia mengalokasikan lebih dari 50% dari total pengeluarannya untuk membeli beras (Hafizah, Hakim, Harianto, & Nurmalina, 2020). Hal ini menunjukkan pentingnya menjaga ketersediaan beras untuk ketahanan pangan (Octania, 2021).

Menurut Permendag Nomor 4 Tahun 2012 dan Instruksi Presiden Nomor 5 Tahun 2015, Cadangan Beras Pemerintah (CBP) yang dikelola oleh Perum BULOG merupakan instrumen kunci dalam menjaga stabilitas harga beras, ketersediaan pangan, dan ketahanan pangan nasional. CBP berfungsi sebagai *buffer stock* untuk mengintervensi pasar ketika terjadi lonjakan harga dan memenuhi kebutuhan pangan masyarakat yang terdampak bencana alam.

Terlepas dari pentingnya CBP bagi ketahanan pangan, manajemen CBP di Indonesia menghadapi beberapa tantangan yang kompleks, yang meliputi pertumbuhan populasi, penurunan produksi, dan perubahan iklim. Pertumbuhan populasi yang pesat menyebabkan peningkatan permintaan beras yang signifikan. Hal ini tidak sejalan dengan penurunan produksi yang terus terjadi akibat dari perubahan iklim, bencana alam, dan alih fungsi lahan pertanian yang mengancam

ketersediaan persediaan beras (Beltran-Peña, Rosa, & D'Odorico, 2020; Lee, Lin, & Petway, 2018). Hal ini seringkali menyebabkan fluktuasi harga yang tidak terkendali, sehingga membebani masyarakat, terutama pada masyarakat dengan ekonomi rendah, dan dapat mengganggu stabilitas ketahanan pangan nasional dan menyebabkan kerawanan pangan (Octania, 2021).

Perubahan iklim menyebabkan kondisi cuaca ekstrem dan perubahan pola hujan, yang secara langsung berdampak pada pola tanam dan produksi pangan, serta meningkatkan risiko gagal panen dan ketidakstabilan persediaan (Amir, Saqib, Khan, Khan, Bokhari, Zaman-ul-Haq, & Majid, 2020; Farooq, Uzair, Raza, Habib, Xu, Yousuf, Yang, & Ramzan Khan, 2022; Hasegawa, Sakurai, Fujimori, Takahashi, Hijioka, & Masui, 2021; Mukherjee, 2021). Faktor-faktor ini secara akumulatif menciptakan ketidakpastian yang tinggi dalam manajemen CBP sehingga menyulitkan pengambilan keputusan yang akurat dan tepat waktu.

Sesuai Pasal 5 Peraturan Badan Pangan Nasional Nomor 12 Tahun 2022, manajemen Cadangan Beras Pemerintah (CBP) melibatkan beberapa tahapan penting, meliputi pengadaan, penyimpanan, dan distribusi. Tahapan pengadaan dimulai dengan mengidentifikasi kebutuhan persediaan berdasarkan prediksi permintaan (kebutuhan) dan produksi, diikuti dengan perencanaan pengadaan yang melibatkan tugas-tugas penting seperti penentuan jumlah cadangan persediaan.

Metode konvensional yang digunakan dalam memprediksi produksi dan permintaan beras saat ini, seperti metode logika fuzzy dan regresi linier berganda seringkali kurang akurat karena memiliki keterbatasan dalam menangani data yang kompleks, tidak pasti, dan tidak mampu menangkap berbagai faktor eksternal seperti pertumbuhan populasi, penurunan produksi, dan perubahan iklim (Sari, Ananda, & Rani, 2020). Keterbatasan ini menyulitkan dalam penentuan prediksi yang akurat, penentuan jumlah cadangan yang optimal, dan pengambilan keputusan pengadaan. Kesalahan dalam memprediksi produksi dan permintaan dapat menyebabkan kelebihan atau kekurangan persediaan, yang berdampak pada naiknya harga beras dan dapat mengganggu stabilitas ekonomi serta sosial

(Octania, 2021). Ketiadaan sistem pendukung keputusan yang komprehensif juga menghambat pengambilan keputusan strategis terkait pengadaan CBP.

Menghadapi tantangan tersebut, diperlukan pendekatan yang lebih inovatif dan berbasis data. Teknologi *Machine Learning* (ML) menawarkan potensi besar dalam meningkatkan akurasi prediksi, optimasi jumlah cadangan, dan pengambilan keputusan yang tepat. Pemanfaatan teknologi ini dapat membantu mengatasi tantangan kerawanan pangan secara efektif, memastikan ketersediaan pangan, dan mendukung stabilitas harga di pasar nasional (Mehmood et al., 2023; Rai et al., 2021).

Penelitian terdahulu telah mengusulkan berbagai model prediksi produksi dan permintaan. Noorunnahar, Chowdhury, and Mila (2023) membandingkan dua model yaitu ARIMA dan XGBoost untuk memprediksi produksi beras di Bangladesh. Hasil penelitian menunjukkan bahwa model XGBoost mengungguli model ARIMA dengan nilai MAE dan RMSE sebesar 2.779.742 dan 3.195.985, namun model ini tidak melibatkan variabel lain seperti suhu dan curah hujan, serta nilai *error* masih tinggi.

Satpathi, Setiya, Das, Nain, Jha, Singh, and Singh (2023) membandingkan lima metode, yaitu SMLR, ANN, LASSO, ELNET, Ridge Regression, dan *ensemble models* untuk memprediksi hasil panen padi di Chhattisgarh. Hasil penelitian menunjukkan bahwa model prediksi dengan ANN mengungguli model lainnya dengan nilai R^2 sebesar 1 dan 0,99 serta nilai RMSE sebesar 0,003 dan 0,004 yang menunjukkan kemampuan model dalam menangani berbagai variabel prediksi. Namun terdapat beberapa model yang menunjukkan kinerja baik selama pelatihan tetapi tidak selama proses pengujian, yang disebabkan oleh *overfitting*.

Nassibi, Fasihuddin, and Hsairi (2023) membandingkan dua metode yaitu Long Short-Term Memory (LSTM) dan Support Vector Machine (SVM) untuk memprediksi permintaan produk makanan. Hasil penelitian ini menunjukkan bahwa model LSTM mengungguli model SVM dengan nilai MAPE dan RMSE antara 3,78-15,3 dan 2,24-4,64 dan berkontribusi pada pengurangan limbah. Namun, kinerja model ini bergantung pada kualitas dan kelengkapan data.

Penelitian terdahulu telah menunjukkan bahwa *machine learning* (ML) dapat meramalkan produksi dan permintaan dengan lebih akurat daripada metode konvensional seperti ARIMA, namun model-model ini masih memiliki keterbatasan, seperti risiko *overfitting* dan tidak mempertimbangkan variabel eksternal, seperti perubahan iklim. Di sisi lain, kinerja model-model ini juga sangat bergantung pada kualitas dan kuantitas data, yang sering kali menghasilkan prediksi yang kurang akurat.

Solusi untuk mengatasi keterbatasan ini adalah dengan mengintegrasikan *Generative AI* (Gen AI) ke dalam model prediksi berbasis ML. Gen AI dapat menghasilkan data sintetis yang berkualitas tinggi untuk melengkapi dan memperbaiki data yang ada, sehingga meningkatkan kualitas *dataset* yang digunakan untuk prediksi (Yu & Guo, 2023). Teknologi ini juga dapat berguna untuk menciptakan skenario pelatihan yang beragam, yang membantu model belajar dari data yang lebih beragam dan mengurangi *overfitting* (Manjunath & Palayyan, 2023).

Penelitian ini bertujuan untuk mengembangkan model prediksi yang lebih akurat dengan memanfaatkan teknologi gen AI dan ML. Model ini akan memanfaatkan GAN, sebuah tipe *generative AI* yang dapat menghasilkan data yang realistis, sehingga dapat mengatasi ketergantungan pada data berkualitas tinggi dalam ML (Harshvardhan, Gourisaria, Pandey, & Rautaray, 2020). Algoritma ML yang digunakan adalah XGBoost yang telah menunjukkan kemampuan untuk menangani distribusi data yang kompleks, efisien secara komputasi, dan dapat mengatasi *overfitting* melalui pengenalan istilah regularisasi dalam *loss function* (C. Qin, Zhang, Bao, Zhang, Liu, & Liu, 2021; Jiayi Wang & Zhou, 2024; J. Wu, Li, & Ma, 2021). XGBoost ini digunakan untuk menganalisis data historis produksi dan konsumsi beras, data cuaca dari BMKG (Badan Meteorologi, Klimatologi, dan Geofisika), data luas tanam dan produktivitas dari Kementerian Pertanian, dan data lainnya. Model ini juga akan mempertimbangkan variabel lain yang relevan, seperti harga beras (Iqbal, 2019; Sossou & Igue, 2019) dan pendapatan per kapita (Bashir & Yuliana, 2019; Yusuf, Yusuf, Adesope, & Adebayo, 2020).

Penelitian ini juga akan menghitung cadangan beras pemerintah yang optimal dengan menggunakan *Deep Reinforcement learning* (DRL) dengan mempertimbangkan produksi nasional, keadaan darurat, stabilisasi harga, perjanjian internasional, dan kecukupan gizi sesuai dengan Pasal 2 dari Peraturan Badan Pangan Nasional Nomor 12 Tahun 2022. Penggunaan DRL memungkinkan penentuan kebijakan yang adaptif dan dinamis untuk mengoptimalkan cadangan beras dalam menghadapi ketidakpastian (Abu Zwaida, Pham, & Beauregard, 2021; Perez, Hubbs, Li, & Grossmann, 2021).

Penelitian ini juga akan membangun sistem pendukung keputusan (DSS) yang interaktif dan berbasis data untuk mendukung pengambilan keputusan strategis. DSS ini akan mengintegrasikan model prediksi, perhitungan cadangan optimal, dan informasi relevan lainnya untuk memberikan rekomendasi yang komprehensif kepada Perum Bulog terkait pengadaan CBP. DSS ini akan dilengkapi dengan antarmuka pengguna yang intuitif dan visualisasi data yang interaktif, seperti grafik, tabel, dan peta, sehingga memudahkan pengguna dalam memahami informasi dan mengambil keputusan. Melalui pengintegrasian teknologi-teknologi ini, diharapkan model dan sistem yang diusulkan dapat meningkatkan efisiensi operasional Perum BULOG dan menguatkan ketahanan pangan.

1.2 Rumusan Masalah

Rumusan masalah pada penelitian ini adalah sebagai berikut:

1. Bagaimana mengembangkan model *generative AI* yang mampu menghasilkan data sintetis untuk meningkatkan akurasi model prediksi permintaan dan produksi beras?
2. Bagaimana mengembangkan model *machine learning* yang efektif untuk memprediksi permintaan dan produksi beras?
3. Bagaimana merancang model optimasi yang efektif untuk menentukan jumlah cadangan beras pemerintah yang optimal?
4. Bagaimana membangun *prototype decision support system* untuk cadangan beras pemerintah?

1.3 Tujuan Penelitian

Berdasarkan pokok permasalahan yang telah dirumuskan dalam rumusan masalah, maka tujuan penelitian ini adalah sebagai berikut:

1. Mengembangkan model *generative AI* untuk menghasilkan data sintetis yang realistis yang dapat digunakan sebagai data pelatihan untuk model prediksi permintaan dan produksi beras.
2. Mengembangkan model *machine learning* yang efektif dalam memprediksi permintaan dan produksi beras.
3. Merancang model optimasi yang efektif untuk menentukan jumlah cadangan beras pemerintah yang optimal.
4. Membangun *prototype decision support system* untuk cadangan beras pemerintah.

1.4 Batasan Penelitian

Peneliti memberikan beberapa batasan agar penelitian yang dilakukan dapat lebih terfokus. Batasan yang diterapkan pada penelitian ini adalah sebagai berikut:

1. Penelitian ini akan menggunakan data komoditas beras *non-premium*.
2. Penelitian ini berfokus pada objek Cadangan Beras Pemerintah (CBP) yang dikelola oleh Perum BULOG.
3. Penelitian ini menggunakan data yang berasal dari Badan Pusat Statistik (BPS), Perum BULOG, Kementerian Pertanian, Badan Meteorologi Klimatologi dan Geofisika (BMKG), dan Badan Nasional Penanggulangan Bencana sebagai sumber informasi utama.
4. Penelitian ini berfokus pada penggunaan GAN, *eXtreme Gradient Boosting* (XGBoost), dan *Deep Reinforcement Learning* (DRL).

1.5 Kontribusi

Penelitian ini memberikan kontribusi baik dari sisi ilmu pengetahuan maupun teknologi. Dari sisi ilmu pengetahuan, kontribusi penelitian ini berupa

pemahaman mengenai pemanfaatan *Generative Adversarial Networks* (GAN) dan *Extreme Gradient Boosting* (XGBoost) untuk memprediksi permintaan dan produksi beras. Melalui integrasi metode-metode tersebut, penelitian ini memberikan pemahaman yang lebih baik mengenai model prediksi. Penelitian ini juga menunjukkan potensi *Deep Reinforcement Learning* (DRL) menangani permasalahan optimasi cadangan beras pemerintah. Melalui pemanfaatan tersebut, penelitian ini tidak hanya memperkaya literatur yang sudah ada, tetapi juga memicu eksplorasi lebih lanjut terhadap pengembangan algoritma prediksi dan optimasi, sehingga memberikan kontribusi yang signifikan terhadap kemajuan teknologi prediksi dan optimasi di bidang ilmu pengetahuan.

Dari sisi teknologi, penelitian ini menghasilkan model prediksi dan sistem pendukung keputusan yang inovatif untuk manajemen persediaan cadangan beras. Sistem ini juga akan memberikan manfaat bagi para pemangku kepentingan dalam pengambilan kebijakan pengadaan beras, yang mengarah pada ketahanan pangan di Indonesia. Hal ini dapat membawa peningkatan signifikan dalam keakuratan dan efisiensi pengambilan kebijakan, serta menjamin ketahanan pangan yang lebih berkelanjutan.

BAB 2

TELAAH PUSTAKA

2.1 *Artificial Intelligence* (AI)

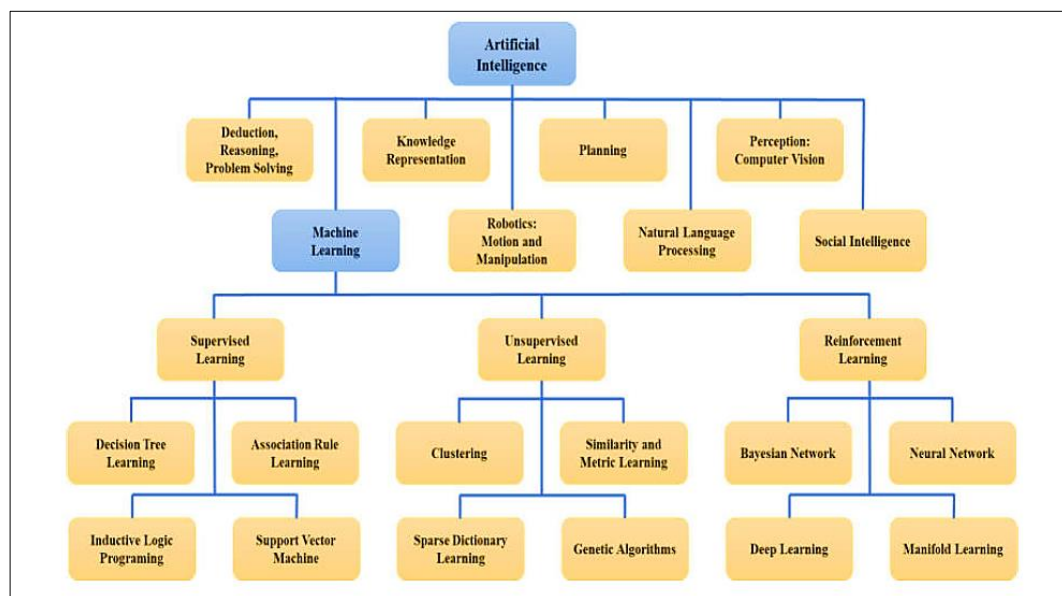
Topik kecerdasan buatan (*Artificial Intelligence* atau AI) telah mendapatkan banyak perhatian dalam selama beberapa tahun terakhir. AI banyak digunakan untuk memecahkan masalah dalam berbagai disiplin ilmu, termasuk teknik, medis, pendidikan, dan pertanian. Luger (2009) mendefinisikan AI sebagai kapasitas mesin atau sistem komputer untuk meniru atau menampilkan kecerdasan manusia. Hal ini mencakup kapasitas sistem untuk mengumpulkan informasi, memahami konteks, melakukan analisis, membuat keputusan, dan belajar dari pengalaman untuk menghadapi tugas-tugas yang kompleks. Tujuan utama AI adalah untuk membangun sistem yang dapat melaksanakan tugas yang biasanya membutuhkan kecerdasan manusia.

Dari sudut pandang yang lain, Webber and Nilsson (2014) menganggap kecerdasan buatan (AI) sebagai sebuah disiplin ilmu komputer yang bertujuan untuk memberikan kemampuan penalaran dan persepsi kepada mesin. Ada pula Shi (2019) yang mendefinisikan kecerdasan buatan sebagai bidang ilmu komputer yang berkonsentrasi mempelajari dan mengembangkan mesin atau sistem yang cerdas. AI melibatkan penggunaan metode dan teknik buatan untuk menciptakan mesin yang dapat berpikir, bernalar, belajar, dan beradaptasi dengan situasi baru seperti yang dilakukan manusia.

Kecerdasan buatan (AI) juga dapat diartikan sebagai percepatan pengembangan paradigma komputasi non-tradisional yang terinspirasi oleh kemampuan dan efisiensi energi otak. Otak manusia unggul terutama dalam tugas-tugas kognitif yang intensif secara komputasi, seperti pengenalan pola dan klasifikasi. Tujuan utama dari AI adalah komputasi neuromorfis yang tidak terpusat, dengan mengandalkan jaringan inti terdistribusi untuk meniru paralelisme masif otak, sehingga secara ketat mengikuti pendekatan yang terinspirasi dari alam untuk pemrosesan informasi. Dengan secara bertahap mengubah blok komputasi yang saling terhubung menjadi jaringan komputasi

berkelanjutan, bentuk materi canggih yang menampilkan fitur dasar kecerdasan dapat dibayangkan, yang mampu belajar dan memproses informasi secara terdesentralisasi. Materi cerdas semacam itu akan berinteraksi dengan lingkungan dengan menerima dan merespons rangsangan eksternal sambil secara internal mengadaptasi strukturnya untuk memfasilitasi distribusi dan penyimpanan informasi sebagai memori.

Pada pengaplikasiannya, kecerdasan buatan (AI) terbagi menjadi beberapa sub-bidang, yaitu pembelajaran mesin (*machine Learning*), *natural language processing* (NLP), sistem pakar (*expert system*), vision, ucapan (*speech*), *planning*, dan *robotic* (Russell & Norvig, 2010). Gambar 2. 1 adalah sub-bidang dari kecerdasan buatan (AI).



Gambar 2. 1 Sub-bidang dari kecerdasan buatan (AI)
(Russell & Norvig, 2010)

Berdasarkan beberapa penjelasan mengenai AI, dapat disimpulkan bahwa AI atau kecerdasan buatan mengacu pada kemampuan mesin atau sistem komputer untuk meniru kecerdasan manusia dengan memproses dan mengumpulkan informasi, memahami konteks, melakukan analisis, membuat penilaian, dan belajar dari pengalaman untuk menangani tugas-tugas yang kompleks. AI adalah bidang multidisiplin yang mencakup berbagai sub-bidang, seperti *machine learning*, *natural language processing*, sistem pakar, *vision*, ucapan, *planning*, dan robotika. Sub-bidang ini berfokus pada pengembangan

algoritma, model, dan sistem yang dapat melakukan tugas-tugas yang kompleks, termasuk rekayasa, dengan akurasi dan efisiensi yang tinggi.

2.2. Konsep Dasar *Machine Learning* (ML)

Machine learning atau pembelajaran mesin adalah sub bidang ilmu komputer yang melibatkan pembuatan algoritma dan model yang dapat meningkatkan kinerjanya berdasarkan pengalaman atau data sampel. Model-model ini memiliki parameter yang perlu dioptimalkan melalui proses yang disebut pembelajaran, dimana program komputer menggunakan data pelatihan atau pengalaman masa lalu untuk menyesuaikan parameter model. Model yang dihasilkan dapat bersifat prediktif, yang berarti dapat membuat prediksi di masa depan, atau deskriptif, yang berarti dapat memperoleh pengetahuan dari data, atau keduanya (Alpaydin, 2020).

Teori statistik adalah dasar dari *machine learning*, karena melibatkan pengambilan kesimpulan dari sampel. Untuk melatih model secara efisien, ilmu komputer memainkan peran penting dalam mengembangkan algoritma yang memecahkan masalah optimasi dan dapat menyimpan dan memproses data dalam jumlah besar. Setelah model dilatih, representasi dan solusi algoritmik untuk inferensi juga harus efisien. Pada beberapa kasus, efisiensi algoritma pembelajaran atau inferensi mungkin sama pentingnya dengan akurasi prediksi (Alpaydin, 2020). Beberapa pendekatan dan teknik telah diusulkan oleh para ahli matematika dan programmer di bidang *machine learning*, yang dapat dilihat pada Gambar 2.3 (Udousoro, 2020).

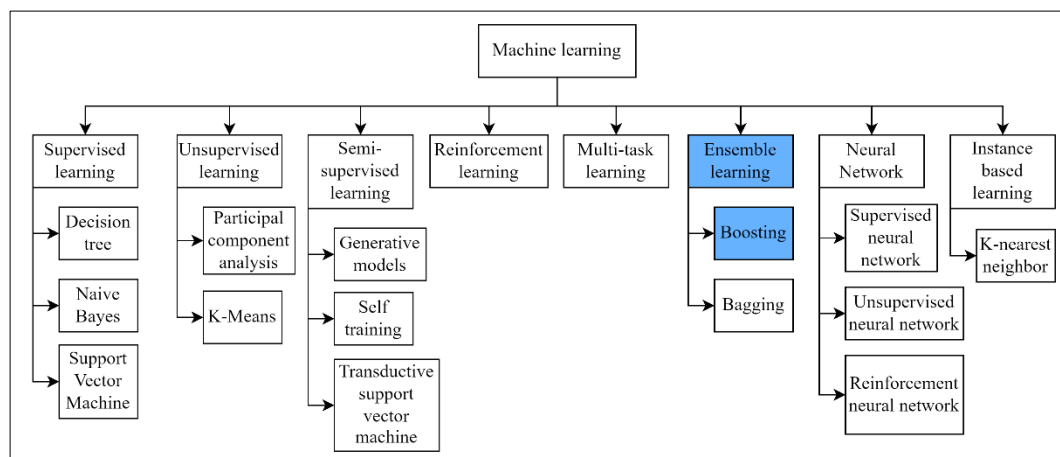
Ensemble learning adalah salah satu teknik *machine learning* yang melibatkan penggabungan beberapa *learner* untuk menciptakan satu *learner* yang lebih kuat. Kombinasi dari beberapa *learner* seperti *decision tree*, *naive Bayes*, dan *neural network* dapat menghasilkan performa yang lebih baik dibandingkan dengan satu *learner* saja (Udousoro, 2020). Terdapat dua teknik utama yang digunakan dalam *ensemble learning*, yaitu *boosting* dan *bagging*.

a. *Boosting*

Teknik *boosting* bekerja dengan mengumpulkan *learner* yang lemah dan mengubahnya menjadi *learner* yang kuat. Dengan mengurangi bias dan varians, teknik *boosting* dapat meningkatkan akurasi model akhir. Salah satu algoritma *boosting* yang paling populer saat ini adalah XGBoost, yang menjadi fokus pada penelitian ini.

b. *Bagging*

Teknik *bagging* (*bootstrap aggregating*) adalah teknik yang digunakan untuk meningkatkan akurasi dan stabilitas algoritma *machine learning*. Teknik ini sangat berguna untuk menangani kumpulan data yang besar.



Gambar 2. 2 Teknik Machine Learning

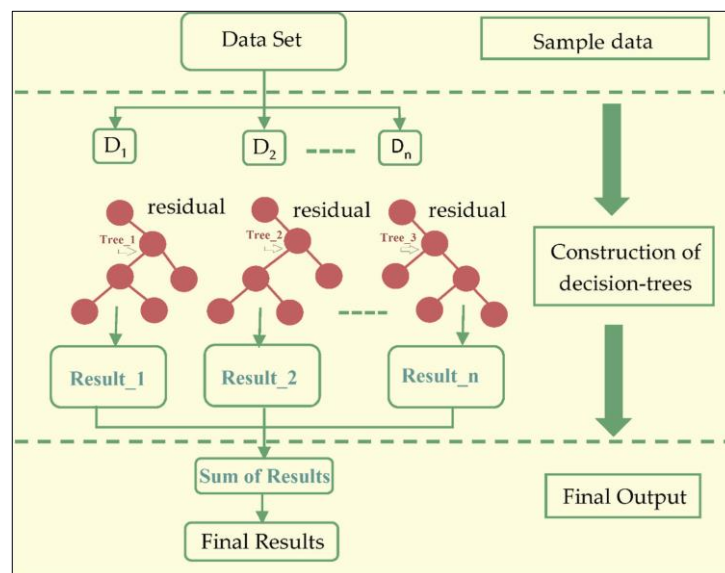
Berdasarkan kajian sebelumnya mengenai ML, disimpulkan bahwa tujuan utama dari ML adalah untuk mengembangkan model dan algoritma yang dapat dilatih dengan menggunakan data sampel atau pengalaman masa lalu. Salah satu teknik ML yang paling efektif adalah *ensemble learning*, yang menggabungkan beberapa *learner* untuk menghasilkan model yang lebih *robust*. Dua teknik utama dalam *ensemble learning* adalah *bagging* dan *boosting*. Teknik *boosting*, seperti algoritma XGBoost, sangat baik dalam menurunkan bias dan varians, yang pada akhirnya meningkatkan akurasi model akhir.

2.2.1 Algoritma *eXtreme Gradient Boosting* (XGBoost)

Extreme Gradient Boosting atau XGBoost adalah algoritma yang sangat terukur dan efisien dalam menangani *dataset* yang besar. Algoritma ini

diperkenalkan pada tahun 2016 dan sejak saat itu telah dikenal dengan kemampuannya yang lebih cepat daripada algoritma populer lainnya dan berskala baik untuk berbagai *dataset*, baik di mesin tunggal maupun di lingkungan terdistribusi (T. Chen & Guestrin, 2016).

Algoritma XGBoost adalah algoritma ML yang menerapkan proses *boosting* untuk menghasilkan model akurat. Model prediktif yang dilatih pada contoh berlabel diterapkan pada contoh baru yang belum terlihat. *Boosting* adalah metode pembelajaran ansambel yang digunakan untuk membangun banyak model secara berurutan, di mana setiap model akan berusaha memperbaiki kekurangan pada model sebelumnya (Ahmed, Raza, Hussain, Aldweesh, Omar, Khan, Eldin, & Nadim, 2023). Gambar 2.3 adalah arsitektur dasar dari algoritma XGBoost.



Gambar 2. 3 Arsitektur Dasar Algoritma XGBoost
(Ahmed et al., 2023)

XGBoost telah berhasil diterapkan dalam berbagai skenario, seperti prediksi persediaan, prediksi keuangan, penilaian kerentanan tanah longsor, dan pemetaan prospek mineral (Kavzoglu & Teke, 2022b; N. Li, Chiang, Down, & Heddle, 2021; W. Wang, Lesner, Ran, Rukonic, Xue, & Shiu, 2020). Di bidang keuangan, XGBoost telah terbukti menjadi alat yang ampuh untuk menilai aplikasi pinjaman dan mengungguli model risiko kredit tradisional (W. Wang et al., 2020).

Pada ilmu kebumian, XGBoost telah digunakan untuk mengevaluasi kerentanan tanah longsor dan memberikan wawasan tentang heterogenitas spasial yang berkontribusi terhadap bencana (Kavzoglu & Teke, 2022b). Selain itu, XGBoost juga telah berhasil memprediksi permintaan RBC dan produk beras di Bangladesh (N. Li et al., 2021; Noorunnahar et al., 2023). XGBoost dilaporkan memiliki kinerja yang lebih baik daripada model *Autoregressive Integrated Moving Average* (ARIMA) dalam memprediksi produksi beras tahunan, dengan ukuran kesalahan yang lebih rendah dalam set pengujian (Noorunnahar et al., 2023).

Keunggulan lain dari XGBoost adalah akurasinya yang tinggi pada data terstruktur, melampaui metode pembelajaran mesin klasik lainnya seperti jaringan saraf. Selain itu, XGBoost telah menunjukkan kemampuan untuk menangani distribusi data yang kompleks dan efisien secara komputasi, sehingga cocok untuk bekerja dengan *dataset* yang besar (J. Wu et al., 2021). Algoritma ini juga dapat mengatasi *overfitting* melalui pengenalan istilah regularisasi dalam *loss function*, yang membantu membatasi kompleksitas model dan meningkatkan kemampuan generalisasinya (C. Qin et al., 2021; Jiayi Wang & Zhou, 2024).

Optimasi *hyperparameter* menggunakan metode seperti *grid search* dan *Particle Swarm Optimization* (PSO) dapat digunakan untuk lebih meningkatkan performa XGBoost dan mengurangi *overfitting* (C. Qin et al., 2021; Soleymani & Mohammadzadeh, 2023; Jiayi Wang & Zhou, 2024). Namun, XGBoost memiliki beberapa keterbatasan, seperti kesulitan menangani data yang tidak terstruktur seperti gambar dan sensitivitas terhadap *outlier*. Penyetelan parameter yang cermat juga diperlukan untuk mencapai kinerja yang optimal, sehingga kurang ramah pengguna dibandingkan dengan algoritma yang lebih sederhana (Zhu & He, 2022).

Berdasarkan kajian sebelumnya mengenai XGBoost, disimpulkan bahwa XGBoost merupakan algoritma yang sangat skalabel dan efisien dalam menangani kumpulan data yang besar. Kecepatan dan kemampuannya untuk menyesuaikan diri dengan baik pada berbagai macam *dataset* membuatnya menjadi pilihan populer di berbagai bidang, termasuk prediksi permintaan. Dibandingkan dengan

metode *machine learning* klasik lainnya, XGBoost dikenal memiliki kinerja yang lebih baik dan mengatasi *overfitting* dengan memperkenalkan istilah regularisasi dalam *loss function*. Namun, XGBoost memiliki beberapa keterbatasan, seperti kesulitan dalam menangani data yang tidak terstruktur seperti gambar dan sensitivitas terhadap *outlier*. Untuk mencapai kinerja yang optimal, diperlukan *tuning* parameter yang tepat.

2.2.2 *Exploratory Data Analysis*

Exploratory Data Analysis (EDA) adalah proses awal yang sangat penting dalam proses pengembangan model *machine learning*. Proses ini melibatkan merangkum, memvisualisasikan, dan menginterpretasikan data untuk mengungkap pola, hubungan, dan informasi yang mendasari sebelum menerapkan teknik pemodelan yang lebih kompleks. EDA sangat penting untuk memahami struktur intrinsik *dataset* besar dengan dimensi tinggi dan potensi masalah seperti *missing* atau *corrupted information*. Melalui ekstraksi pola dan pengujian hipotesis, EDA membantu dalam mengidentifikasi fitur dan anomali yang signifikan (Abukmeil, Ferrari, Genovese, Piuri, & Scotti, 2021).

EDA merupakan proses berulang dan terbuka yang bertujuan untuk menemukan informasi baru. *Machine learning* dapat digunakan untuk menyusun EDA dengan mempelajari urutan langkah analisis, yang dapat meningkatkan reproduktifitas dan mendorong strategi eksplorasi yang lebih baik (Barczewski, Bezerianos, & Boukhelifa, 2020). Pada konteks pengembangan *model machine learning*, EDA sering kali menjadi langkah awal yang menggunakan metode visual untuk merangkum karakteristik data, mengungkapkan informasi seperti non-korelasi antar variabel atau frekuensi kemunculan (Arsyad, Yunita, Krismartopo, Dimar, Dewi, & Madrinovella, 2023).

Masalah keamanan data (privasi) dapat muncul selama proses EDA, karena proses ini dapat menyebabkan hilangnya kerahasiaan data. Hal ini penting untuk diperhitungkan saat menghitung anggaran privasi dalam aplikasi *machine learning* (Nuñez von Voigt, Pauli, Reichert, & Tschorsch, 2020).

EDA diterapkan di berbagai bidang, termasuk pendidikan (Otero-Escobar & Velasco-Ramírez, 2023), energi (Santhoshkumar & Vanila, 2024), dan industri pertanian (Abbate, Centobelli, & Cerchione, 2023). EDA membantu dalam memahami data, mengidentifikasi kesalahan, mendeteksi outlier, dan menemukan hubungan variabel untuk meningkatkan hasil prediksi energi dan, proses pengajaran serta memprediksi kinerja akademik (Otero-Escobar & Velasco-Ramírez, 2023; Santhoshkumar & Vanila, 2024).

Pada *setting* klinis, EDA sangat penting untuk memberikan gambaran lengkap tentang kumpulan data, mengidentifikasi informasi baru, dan meningkatkan kinerja model prediksi *machine learning* (Mrudula & Sowjanya, 2020). Pada bidang kesehatan, EDA dapat menghasilkan keputusan berbasis data, mengungkap pola dan informasi tersembunyi yang sangat penting untuk pengambilan keputusan, seperti memprediksi elemen-elemen yang menyebabkan penyakit jantung (K & Hanumanthappa, 2022).

Metode untuk menerapkan EDA dapat bervariasi tergantung pada data yang dikumpulkan, namun teknik yang umum digunakan adalah rangkuman statistik dan berbagai jenis plot untuk mengidentifikasi pola dan tren (Kadhar & Anand, 2021). Metode EDA kuantitatif dan visual berfokus pada kemampuannya untuk merangkum karakteristik data secara efisien, dengan alat bantu seperti R yang menyediakan kemampuan untuk plotting dan analisis tingkat lanjut (Patel, 2021). Selain R, alat bantu lain yang sering digunakan untuk EDA yaitu Python, Weka, dan KNIME (Mukhiya & Ahmed, 2020).

Python adalah bahasa yang populer untuk EDA, karena relatif mudah dipelajari, keragaman pustaka yang kaya, dan kapasitas penanganan data yang tinggi. Pandas adalah paket yang paling kuat untuk analisis data, yang dapat membersihkan, mengubah, dan menganalisis data. Data dapat disimpan dalam format CSV di komputer. Proses pembersihan, visualisasi, dan penyimpanan data dapat dilakukan dengan mudah. Pandas dibangun di atas paket NumPy, dan memiliki fungsi plotting dari Matplotlib dan algoritma machine learning dari Scikit-learn. Jupyter Notebook menyediakan kemampuan untuk menjalankan kode dalam sel tertentu menggunakan pendekatan berbasis konsol untuk

komputasi dan menyediakan pemrosesan aplikasi berbasis web. Jupyter Notebook mencakup input dan output komputasi serta representasi media yang kaya akan objek (Sahoo, Samal, Pramanik, & Pani, 2019).

Berdasarkan kajian sebelumnya, disimpulkan bahwa EDA adalah proses penting dalam proses pengembangan model *machine learning*, yang berfungsi sebagai alat bagi para peneliti data untuk memahami dan menyiapkan data untuk analisis lebih lanjut. Hal ini ditandai dengan sifatnya yang berulang, penggunaan metode visual dan statistik, dan penerapannya di berbagai bidang. EDA tidak hanya membantu dalam pengembangan model, tetapi juga memastikan integritas data, menginformasikan pertimbangan privasi, dan meningkatkan proses pengambilan keputusan.

2.2.3 Data Preprocessing

Teknik *preprocessing* sangat penting untuk memastikan kualitas data yang dapat dianalisis dan diolah dengan baik (Luengo, García-Gil, Ramírez-Gallego, García, & Herrera, 2020). *Preprocessing* membantu memastikan validitas dan keandalan hasil analisis data dan mengubah data mentah menjadi *smart data* yang dapat dieksploitasi untuk mendapatkan informasi yang berharga (Fan, Chen, Wang, Wang, & Huang, 2021; Luengo et al., 2020).

a. Data cleaning

Data cleaning adalah teknik *preprocessing* penting yang menangani masalah seperti *outlier*, *skewness*, dan *missing value*. *Outlier* dan *skewness* dapat mempengaruhi model *data mining* dan menghasilkan hasil yang bias, sementara *missing value* membuat data tidak dapat dianalisis oleh model (C. Li, 2019).

1. Missing Values Handling

Menangani *missing value* melibatkan penghapusan seluruh sampel atau mengisinya dengan nilai yang sesuai. Mengisinya dengan nilai yang hilang sangat direkomendasikan karena membuang sampel dengan nilai yang hilang akan mengurangi jumlah *dataset* secara dramatis, dan informasi yang terkandung dalam sampel yang dibuang tidak digunakan

(C. Li, 2019). Pada data deret waktu (*time series*), proses pembersihan data yang hilang (*missing value*) dilakukan untuk memastikan kelanjutannya. Hal ini penting untuk prediksi permintaan karena memungkinkan model untuk menangkap pola dan trend dengan akurat, atau dengan kata lain tanpa melakukan pengisian nilai ini, analisis yang dilakukan kemungkinan besar akan menghasilkan hasil yang tidak akurat atau bias (Vallés-Pérez, Soria-Olivas, Martínez-Sober, Serrano-López, Gómez-Sanchís, & Mateo, 2022).

2. *Outlier Detection and Handling*

Deteksi *outlier* adalah teknik *data cleaning* penting lainnya yang membantu mengidentifikasi sampel data yang memiliki jarak yang signifikan dari sebagian besar sampel lainnya. Meskipun beberapa model kuat terhadap *outlier*, deteksi *outlier* masih direkomendasikan dalam proses *preprocessing* data (C. Li, 2019). Salah satu metode yang efektif untuk mendeteksi data yang *outlier* adalah metode deteksi berbasis jarak (*distance*). Pada metode ini, *outliers* dideteksi dengan menghitung jarak antara semua objek data berdasarkan berbagai metrik yang berkaitan dengan jarak. Setelah itu, objek yang tidak memiliki cukup *neighbors* kemungkinan besar merupakan *outlier*. Pendekatan *nearest-neighbor* adalah yang paling umum digunakan dalam metode ini dan juga diterapkan pada penelitian ini (Smiti, 2020). Untuk menangani *outlier* yang telah terdeteksi, salah satu metode yang efektif adalah transformasi logaritmik (*log transformation*) (Choi, Buckley, Kuiper, & Keil, 2022).

3. *Skewness Handling*

Pada *machine learning*, data yang tidak seimbang atau memiliki distribusi miring (*skewed data*) dapat menyebabkan masalah dalam model pembelajaran. Salah satu metode umum untuk menangani *skewness* adalah dengan melakukan transformasi logaritmik (*log transformation*). Data yang sangat miring dapat mengganggu kinerja algoritma *machine learning* yang mengasumsikan distribusi data normal.

Transformasi logaritmik mengubah data yang sangat miring menjadi lebih mendekati distribusi normal (West, 2022).

b. *Data Transformation*

Transformasi data juga penting dalam *preprocessing* karena atribut yang berbeda dapat memiliki representasi data yang berbeda, seperti data kategorikal dan numerik. Tidak semua format data cocok untuk model *data mining*, dan perbedaan antara atribut data dapat mempersulit proses optimasi untuk model-model ini. Transformasi data membantu memodifikasi representasi data untuk memenuhi persyaratan model *data mining* dan mengoptimalkan algoritma yang digunakan oleh model-model ini. Dengan mentransformasi data secara tepat, data tersebut dapat menjadi input yang sesuai untuk model *data mining*, dan proses optimasi dapat menjadi lebih efisien dan efisien (C. Li, 2019).

1. *Numeralization*

Penggunaan data kategorikal umum digunakan, tetapi tidak semua operasi dapat diterapkan terhadap data tersebut. Untuk membuatnya kompatibel dengan model, data ini perlu diubah menjadi data numerik. Ada berbagai teknik untuk numeralisasi, seperti *one-hot encoding*, *sequential encoding*, *customized encoding*, dan *label encoding*.

2. *Discretization*

Diskretisasi data terkadang diperlukan untuk memenuhi persyaratan input model seperti Naive Bayes, yang membutuhkan input yang dapat dihitung. Hal ini juga menghaluskan *noise* dan membuat nilai kontinu dapat dihitung. Metode *unsupervised learning* seperti *binning* dan *clustering*, dan metode *supervised learning* seperti *decision tree*, dapat digunakan untuk diskretisasi data.

3. *Normalization*

Normalisasi diperlukan untuk menangani perbedaan dalam sistem unit, yang dapat membuat beberapa atribut terlihat lebih penting daripada yang lain. Berbagai metode normalisasi seperti normalisasi Min-max,

normalisasi Z-score, normalisasi penskalaan desimal, dan *interquartile range* dapat digunakan untuk mengatasi masalah ini.

4. *Numerical Transformations*

Transformasi numerik dapat membantu mendapatkan fitur tambahan yang mungkin tidak relevan untuk beberapa model data mining seperti neural network, yang memiliki potensi *fitting* yang sangat baik.

2.2.4 *Features Engineering*

Feature engineering (rekayasa fitur) melibatkan manipulasi dan transformasi data dengan cara yang secara optimal merepresentasikan masalah yang mendasari algoritma *machine learning*, sekaligus mengatasi kompleksitas dan bias yang melekat pada data. Ada lima teknik utama untuk *feature engineering*, yaitu *feature improvement*, *feature construction*, *feature selection*, *feature extraction*, dan *feature learning* (Ozdemir, 2022).

a. *Feature Improvement*

Feature Improvement melibatkan transformasi fitur yang ada melalui manipulasi matematis untuk membuatnya lebih dapat digunakan.

b. *Feature Construction*

Feature Construction melibatkan pembuatan fitur baru yang dapat ditafsirkan dari fitur yang sudah ada, sehingga menghasilkan kumpulan fitur yang lebih kaya.

c. *Feature Selection*

Feature Selection melibatkan pemilihan subset fitur terbaik dari kumpulan fitur yang ada untuk meminimalkan redundansi dan memaksimalkan kinerja algoritma pembelajaran (Ozdemir, 2022; Sardar, Hasi Saha, & Rabbi, 2023).

d. *Feature Extraction*

Feature Extraction melibatkan pembuatan fitur baru menggunakan algoritma, berdasarkan asumsi parametrik tentang data.

e. *Feature Learning*

Feature Learning melibatkan pembuatan kumpulan fitur baru, biasanya dengan mengekstraksi struktur dan mempelajari representasi dari data mentah

yang tidak terstruktur, seperti teks, gambar, dan video, yang sering kali menggunakan *deep learning*.

2.2.5 Augmentasi Data

Augmentasi data (*Data Augmentation/ DA*) adalah teknik yang digunakan dalam *machine learning* dan *deep learning* untuk meningkatkan ukuran *dataset* pelatihan dengan membuat versi modifikasi dari data yang ada. Proses ini sangat penting untuk meningkatkan performa model, terutama dalam skenario di mana jumlah data pelatihan yang tersedia terbatas (Bayer, Kaufhold, & Reuter, 2022; Moreno-Barea, Jerez, & Franco, 2020; Onishi & Meguro, 2023). Augmentasi data melibatkan penerapan berbagai transformasi dan modifikasi pada data yang ada untuk menghasilkan data pelatihan baru yang membantu model mempelajari berbagai pola dan variasi yang ada di dalam data (Lin, Kaushik, Dyer, & Muthukumar, 2022; Onishi & Meguro, 2023).

Augmentasi data banyak digunakan di bidang-bidang seperti *deep speaker recognition*, *graph neural networks*, dan tugas-tugas *machine learning* yang melibatkan grafik, gambar, dan teks (P. Li, Li, Hamdulla, & Wang, 2023; G. Liu, Inae, Luo, & Jiang, 2024; J. Zhang, Luo, & Wei, 2023). Dampak dari augmentasi data dapat bervariasi, tergantung pada arsitektur model dan tugas. Sebagai contoh, augmentasi data menginduksi regularisasi spektral implisit melalui kombinasi manipulasi proporsi relatif nilai eigen dari matriks kovarians data dan meningkatkan seluruh spektrum matriks kovarians data secara seragam melalui regresi ridge (Lin et al., 2022).

Generative AI memainkan peran penting dalam augmentasi data, terutama dalam menciptakan variasi yang beragam dan realistis dari data yang ada untuk meningkatkan kinerja model dan generalisasi. Performa teknik augmentasi data berbasis *generative AI* dapat bervariasi, tergantung pada aplikasi spesifik dan kualitas data yang dihasilkan. Teknik *generative AI* dapat menciptakan variasi yang lebih beragam dan realistis dari data yang ada dibandingkan dengan teknik augmentasi data tradisional, seperti *cropping* (Iglesias, Talavera, González-Prieto, Mozo, & Gómez-Canaval, 2023; Missaoui, Gerasimou, & Matragkas, 2023). Dua

teknik *generative* AI yang populer digunakan untuk augmentasi data adalah *Variational Autoencoders* (VAE) dan *Generative Adversarial Networks* (GAN) dan (Iglesias et al., 2023).

Berdasarkan kajian sebelumnya mengenai augmentasi data, disimpulkan bahwa teknik ini merupakan alat yang ampuh dalam *machine learning* yang melibatkan pembuatan versi modifikasi dari data yang sudah ada untuk meningkatkan performa model, ketahanan, dan generalisasi di berbagai aplikasi dan domain. Teknik ini dapat sangat bermanfaat dalam skenario dengan data pelatihan yang terbatas atau batasan keputusan yang ketat, dan pengaruhnya dapat bervariasi tergantung pada arsitektur model dan tugas yang dihadapi. GAN adalah salah satu teknik *generative* AI yang populer untuk augmentasi data.

2.2.6 *Generative Adversarial Network* (GAN)

Generative Adversarial Networks (GAN) adalah model generatif yang memiliki kemampuan untuk menghasilkan data sintetis yang sangat mirip dengan distribusi tertentu. GAN sangat berguna untuk menghasilkan sampel data baru dan menambah kumpulan data yang sudah ada ketika lebih banyak sampel diperlukan untuk berbagai tugas (Bendaoud, Farah, & Ahmed, 2021).

Pada beberapa tahun terakhir, GAN telah mendapatkan banyak perhatian karena kemampuannya yang luar biasa dalam meningkatkan kualitas gambar dan video. Tidak seperti pendekatan statistik tradisional, GAN secara efektif mengelola data multivariat dan nonlinier di berbagai domain, termasuk informasi kategorikal dan konten tekstual. GAN merekonstruksi distribusi probabilitas bersyarat dari data itu sendiri, tidak seperti model *machine learning* diskriminatif yang terutama berfokus pada prediksi probabilitas label berdasarkan data yang diamati (Pérez, Arroba, & Moya, 2023).

Struktur dasar GAN yang diperkenalkan oleh Goodfellow, Pouget-Abadie, Mirza, Xu, Warde-Farley, Ozair, Courville, and Bengio (2014) terdiri dari dua komponen penting yaitu generator (G) dan diskriminator (D). Generator menghasilkan data sintetis yang tidak dapat dibedakan dari data asli, sedangkan diskriminator membedakan antara data asli dan data sintetis. Seperti yang

ditunjukkan pada Gambar 2.4, interaksi ini dimulai dengan generator membuat data sampel dari noise acak dari ruang laten, yang kemudian dievaluasi oleh diskriminator bersama dengan data asli. Hasil evaluasi ini menginformasikan kalkulasi kerugian (*loss calculations*), yang sangat penting untuk pelatihan generator dan diskriminator secara bersamaan.

Pada dinamika operasional GAN, diskriminator bertujuan untuk menyempurnakan pengenalan data asli dengan memaksimalkan probabilitas untuk mengidentifikasi data dengan benar, sekaligus meminimalkan deteksi data sintetis sebagai data asli. Generator bertujuan untuk menghasilkan data sintetis yang semakin mendekati data asli, sehingga meminimalkan kemampuan diskriminator untuk mengidentifikasinya sebagai data sintetis. Hal ini dirumuskan sebagai masalah optimasi *min-max*, yang diwakili oleh fungsi objektif GAN. Fungsi ini mengukur *loss* yang diharapkan untuk data asli dan data sintetis, yang memandu proses pelatihan. Fungsi ini dapat dilihat pada persamaan 2.1.

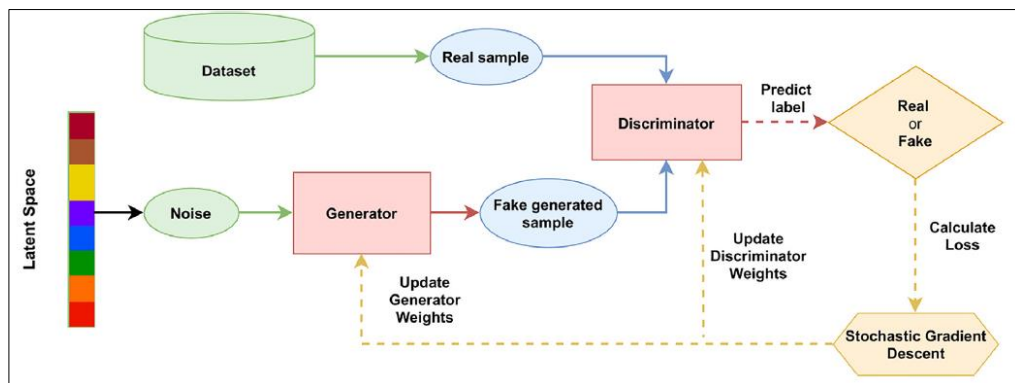
$$\min_G \max_D V(D, G) = \mathbb{E}_{x \sim p_{data}(x)} [\log D(x)] + \mathbb{E}_{z \sim p_z(z)} [\log(1 - D(G(z)))] \quad (2.1)$$

dengan:

$V(D, G)$: fungsi objektif GAN

$\mathbb{E}_{x \sim p_{data}(x)}$: *expected loss* untuk $\log D(x)$

$\mathbb{E}_{z \sim p_z(z)}$: *expected loss* untuk $\log(1 - D(G(z)))$



Gambar 2. 4 Arsitektur GAN
(Bendaoud et al., 2021)

Proses pelatihan melibatkan pembaruan generator dan diskriminator secara iteratif, menggunakan gradien persediaan astik (∇_{θ}) masing-masing. Proses berulang ini, yang diatur oleh persamaan 2.2 dan 2.3 yang menentukan perhitungan kerugian untuk setiap komponen, memastikan bahwa diskriminator dan generator semakin menyempurnakan kinerjanya. Pembaruan diskriminator berfokus pada peningkatan akurasi dalam membedakan data asli dan sintetis, sedangkan pembaruan generator bertujuan untuk menghasilkan *output* sintetis yang lebih realistis. Melalui proses pelatihan ini, GAN telah berevolusi untuk menghasilkan data sintetis berkualitas tinggi, yang menunjukkan dampaknya yang besar pada pembangkitan dan penambahan data.

$$\nabla_{\theta_d} \frac{1}{m} \sum_{i=1}^m [\log D(x^{(i)}) + \log(1 - D(G(z^{(i)})))] \quad (2.2)$$

$$\nabla_{\theta_g} \frac{1}{m} \sum_{i=1}^m \log(1 - D(G(z^{(i)}))) \quad (2.3)$$

Pada bidang data medis, GAN telah digunakan untuk menambah data di mana biaya pengumpulan data sangat mahal dan terdapat masalah privasi. Kerangka kerja *Data Augmentation Optimized for GAN* (DAG) telah diusulkan untuk memastikan bahwa data yang ditambah membantu dalam mempelajari distribusi asli tanpa menyesatkan generator. secara konsisten telah meningkatkan kinerja model GAN yang berbeda dan mencapai skor *Fréchet Inception Distance* (FID) yang tinggi (Tran, Tran, Nguyen, Nguyen, & Cheung, 2021). Selain itu, *Variational Autoencoder* GAN berbasis *Long Short-Term Memory* (VAE-GAN berbasis LSTM) telah dikembangkan untuk deteksi anomali dalam data *time-series*. Metode ini secara efektif mendeteksi anomali dengan memanfaatkan kemampuan pemetaan encoder dan kemampuan diskriminasi diskriminator, yang keduanya merupakan jaringan LSTM (Niu, Yu, & Wu, 2020).

Untuk mengatasi masalah ukuran sampel yang kecil dalam klasifikasi *time-series*, model GAN dengan struktur *Gated Recurrent Unit* (GRU)-

Convolutional Neural Network (CNN) yang dinamakan GC-GAN telah diperkenalkan. Model ini menghasilkan data *time-series* sintetis menggunakan *wasserstein distance* dan *particle swarm optimization*, sehingga meningkatkan akurasi klasifikasi (D. Liu, Wu, Hong, & Wang, 2022). Model TSA-GAN juga telah diusulkan untuk menghasilkan data *time-series* sintetis untuk tugas-tugas klasifikasi, seperti pengenalan aktivitas manusia, dengan mempelajari fitur-fitur yang berguna dari *time-series* yang dikumpulkan. Model ini mencakup strategi pemulihan adaptif mandiri untuk meningkatkan kualitas sampel yang dihasilkan (Z. Li, Ma, Shi, Zhang, Li, & Wu, 2021).

Pada konteks penambahan data biosignal, *Transformer Time-Series Conditional* GAN (TTS-CGAN) telah diperkenalkan. Model ini menghasilkan urutan deret waktu sintetis khusus kelas yang tidak dapat dibedakan dari yang asli, sehingga memperluas ukuran *dataset* untuk penerapan di bidang biomedis (X. Li, Ngu, & Metsis, 2022). Selain itu, *Differentiable Augmentation* (DiffAugment) adalah pendekatan lain yang meningkatkan efisiensi data GAN dengan menerapkan augmentasi yang dapat dibedakan pada data asli dan data sintetis. Pendekatan ini telah terbukti efektif di berbagai arsitektur GAN dan *loss function* (S. Zhao, Liu, Lin, Zhu, & Han, 2020).

Imputation Balanced GAN (IB-GAN) menangani klasifikasi data *time-series* multivariat yang besar dengan ketidakseimbangan kelas yang besar. Metode ini menggunakan teknik imputasi dan *resampling* untuk menghasilkan sampel berkualitas tinggi dan meningkatkan klasifikasi melalui serangkaian data asli dan sintetis yang seimbang antara kelas (Deng, Han, Dreossi, Lee, & Matteson, 2022).

Berdasarkan kajian sebelumnya mengenai GAN, disimpulkan bahwa GAN merupakan salah satu model generatif yang terdiri komponen generator (G) dan diskriminator (D). GAN telah terbukti menjadi alat yang dapat meningkatkan performa model *machine learning* dengan menghasilkan data sintetis yang sangat mirip dengan distribusi data yang sebenarnya. Implementasi GAN untuk augmentasi data *time-series* telah diadopsi secara luas di berbagai bidang, memberikan solusi untuk masalah-masalah seperti ketersediaan data yang terbatas, ketidakseimbangan kelas, dan permintaan data sintetis berkualitas tinggi.

2.2.7 *Hyperparameter Tuning*

Hyperparameter tuning adalah langkah penting dalam *machine learning* untuk memaksimalkan kinerja model (Soleymani & Mohammadzadeh, 2023). Hal ini melibatkan pemilihan *hyperparameter* terbaik yang memungkinkan model untuk berkinerja lebih baik pada tugas target, yang mengarah pada akurasi, presisi, *recall*, dan metrik kinerja lainnya yang lebih tinggi (Machlanski, Samothrakis, & Clarke, 2023; Ottoni, Souza, & Novo, 2023; Vo, Hoang, & Quach, 2023). Hal ini juga memungkinkan model untuk menggeneralisasi lebih baik pada data yang tidak terlihat, sehingga lebih dapat diandalkan dan dapat diterapkan (Machlanski et al., 2023).

Algoritma XGBoost adalah salah satu algoritma *machine learning* yang populer yang memerlukan penyetelan beberapa *hyperparameter* untuk mendukung pemodelan (Dalal, Onyema, & Malik, 2022; Wen, Hu, Zhang, Xiang, & Liao, 2022). *Hyperparameter* yang optimal seperti ``max_depth``, ``gamma``, ``reg_alpha``, ``reg_lambda``, ``colsample_bytree``, ``min_child_weight``, dan ``n_estimator`` memungkinkan model XGBoost bekerja lebih baik dan menggeneralisasi data yang tidak terlihat dengan lebih baik. Parameter-parameter ini mengontrol perilaku model XGBoost dan secara signifikan dapat mempengaruhi akurasi prediksi dan kemampuan generalisasinya (Bloch & Friedrich, 2020; Dalal et al., 2022; Kavzoglu & Teke, 2022a).

- a. ``max_depth``: kedalaman maksimum pohon. Meningkatkan nilai ini akan membuat model menjadi lebih kompleks dan rentan terhadap *overfitting*.
- b. ``gamma``: pengurangan kerugian minimum yang diperlukan untuk membuat partisi lebih lanjut pada simpul daun pohon. Meningkatkan nilai ini membuat model menjadi lebih konservatif.
- c. ``reg_alpha``: istilah regularisasi L1 pada bobot. Meningkatkan nilai ini membuat model menjadi lebih konservatif.
- d. ``reg_lambda``: istilah regularisasi L2 pada bobot. Meningkatkan nilai ini akan membuat model menjadi lebih konservatif.

- e. ``colsample_bytree``: Fraksi kolom yang akan diambil sampelnya secara acak untuk setiap pohon. Mengurangi nilai ini akan mengurangi varians model dan membantu mencegah *overfitting*.
- f. ``min_child_weight``: jumlah minimum bobot contoh yang dibutuhkan dalam sebuah *child*. Menambah nilai ini akan membuat model menjadi lebih konservatif.
- g. ``n_estimator``: Jumlah pohon yang akan dicocokkan.

Hyperparameter yang dipilih dengan tepat menghasilkan model yang lebih tahan terhadap *noise* dan bentuk gangguan data lainnya. Terdapat beberapa tantangan dalam penyetelan *hyperparameter*, seperti peningkatan biaya komputasi, kesulitan dalam menentukan keseimbangan optimal antara *underfitting* dan *overfitting*, dan potensi sensitivitas terhadap inisialisasi (Machlanski et al., 2023). Terlepas dari tantangan-tantangan ini, *tuning hyperparameter* merupakan langkah penting untuk menciptakan model pembelajaran mesin yang kuat dan andal.

Beberapa penelitian telah menunjukkan pentingnya *tuning hyperparameter* di XGBoost untuk memaksimalkan kinerja model XGBoost di berbagai bidang. Memahami dampak dari *hyperparameter* yang berbeda pada model dan menyetelnya dengan tepat sangat penting untuk mencapai kinerja terbaik (Dalal et al., 2022; Wen et al., 2022).

2.2.8 Teknik Evaluasi

Terdapat beberapa teknik evaluasi yang dapat digunakan untuk mengevaluasi model, diantaranya yaitu *R-squared* (R^2), *Root Mean Squared Error* (RMSE) dan *Mean Absolute Error* (MAE) (Chicco, Warrens, & Jurman, 2021).

- a. R^2 mengukur proporsi varians dalam variabel dependen yang dapat diprediksi dari variabel independen. Nilai R^2 berkisar antara 0 – 1, dimana semakin mendekati 1 maka semakin layak suatu model untuk digunakan.

$$R^2 = 1 - \frac{\sum_{i=1}^m (x_i - y_i)^2}{\sum_{i=1}^m (\bar{y} - y_i)^2} \quad (2.4)$$

dengan:

R^2 : *R-squared*

m: jumlah data

x_i : nilai prediksi

y_i : nilai sebenarnya

\bar{y} : rata-rata nilai sebenarnya

- b. RMSE mengukur besarnya tingkat kesalahan prediksi, dimana semakin rendah nilainya (mendekati nol), maka hasil prediksi akan semakin akurat.

$$RSME = \sqrt{\frac{1}{m} \sum_{i=1}^m (x_i - y_i)^2} \quad (2. 5)$$

dengan:

RSME: *Root Mean Squared Error*

m: jumlah data

x_i : nilai prediksi

y_i : nilai sebenarnya

- c. MAE menghitung rata-rata kesalahan kuadrat antara nilai aktual dan nilai yang diramalkan, dengan nilai yang rendah atau mendekati nol menunjukkan kecocokan yang lebih baik antara hasil permintaan dan data aktual.

$$MAE = \frac{1}{m} \sum_{i=1}^m |x_i - y_i| \quad (2. 6)$$

dengan:

MAE: *Mean Absolute Error*

m: jumlah data

x_i : nilai hasil prediksi

y_i : nilai sebenarnya

2.3 Manajemen Cadangan Beras

Pentingnya cadangan beras nasional dalam menjamin ketahanan pangan telah menjadi fokus utama dalam beberapa tahun terakhir, terutama karena meningkatnya permintaan pangan global, perubahan iklim, dan ketidakstabilan ekonomi. Cadangan beras nasional adalah stok strategis yang dikelola oleh pemerintah untuk menstabilkan persediaan dan harga beras, menyediakan bantuan ketika terjadi kekurangan pangan, dan mendukung tujuan ketahanan pangan nasional. Manajemen cadangan beras yang efektif ini sangat penting untuk mengurangi risiko yang terkait dengan gangguan persediaan pangan dan fluktuasi harga (Yulianis & Rachman, 2021).

Fungsi utama cadangan beras nasional adalah sebagai penyangga terhadap gangguan pada persediaan dan fluktuasi harga beras, sehingga penting untuk memastikan persediaan beras yang stabil untuk memenuhi kebutuhan dalam negeri. Stabilitas yang diberikan oleh cadangan ini penting untuk menjaga kepercayaan masyarakat terhadap ketersediaan pangan dan untuk mendukung mata pencaharian petani dan konsumen beras (Kurnia & Iskandar, 2019). Di Indonesia, Cadangan Beras Pemerintah (CBP) telah menjadi kunci dalam menangani fluktuasi ketersediaan dan harga beras akibat berbagai faktor internal dan eksternal (Silalahi, Yudha, Dwiyaniti, Zulvianita, Feranti, & Yustiana, 2019).

Terlepas dari peranan penting cadangan beras dalam menjamin ketahanan pangan nasional, manajemen cadangan tersebut masih menghadapi berbagai tantangan. Beberapa di antaranya meliputi serangan hama dan fasilitas penyimpanan yang tidak memadai (Kandel, Kharel, Njoroge, Smith, Díaz-Valderrama, Timilsina, Paudel, & Baributsa, 2021), dinamika rantai pasok yang kompleks (Guritno, Kristanti, & Tanuputri, 2021; P. Kumar & Yasmeen, 2021), masalah prediksi produksi dan permintaan, serta optimasi cadangan beras (Andriansyah, Rahmi, & Ilyas, 2020; Sasi & Subramanian, 2022). Untuk mengatasi tantangan-tantangan tersebut, sangat penting untuk mengembangkan strategi yang efektif untuk meningkatkan manajemen cadangan beras, terutama di negara dengan manajemen persediaan beras nasional yang kurang efektif. Hal ini meliputi penerapan pendekatan prediksi dan optimasi berbasis teknologi

(Böttcher, Asikis, & Fragkos, 2022; Perez et al., 2021). Penerapan teknologi canggih seperti *Artificial Intelligence* (AI) dapat berperan penting dalam meningkatkan efisiensi manajemen cadangan beras. Teknologi ini memungkinkan respon yang lebih cepat dan akurat terhadap kebutuhan pangan yang mendesak (I. Kumar, Rawat, Mohd, & Husain, 2021).

Berdasarkan kajian sebelumnya, dapat disimpulkan bahwa cadangan beras nasional memainkan peran penting dalam menjamin ketahanan pangan nasional, terutama di masa krisis. Namun, hal ini memiliki beberapa tantangan yang harus dihadapi, terutama dalam memprediksi produksi dan permintaan, serta mengoptimalkan cadangan beras. Penerapan teknologi AI memberikan peluang untuk meningkatkan efektivitas manajemen cadangan beras. Memastikan persediaan beras yang stabil dan memadai melalui manajemen cadangan beras yang efektif sangat penting untuk menjaga ketahanan pangan nasional dan mendukung stabilitas ekonomi masyarakat yang bergantung pada beras seperti Indonesia.

2.3.1 Prediksi Produksi dan Permintaan

Prediksi produksi dan permintaan merupakan komponen penting dalam proses bisnis dan manufaktur, dengan implikasi yang luas bagi kesuksesan perusahaan (Brüggen, Grabner, & Sedatole, 2021). Hal ini melibatkan prediksi produksi dan permintaan di masa depan berdasarkan data historis dan merupakan pekerjaan yang kompleks jika informasi yang tidak lengkap sehingga tidak dapat diprediksi (Yiğit, ESNAF, & KAVUŞ, 2021). Beberapa faktor mempengaruhi keakuratan prediksi produksi dan permintaan, seperti tingkat layanan, pemrosesan sinyal permintaan, dan pengaruh cuaca (Fahimnia, Arvan, Tan, & Siemens, 2023; Giap, Nguyen, Nguyen, & Tran, 2022). Metode prediksi dapat dikategorikan ke dalam metode kualitatif dan kuantitatif. Metode kuantitatif prediksi meliputi metode statistik, *machine learning*, dan *hybrid*, dengan metode *machine learning* dan *hybrid* menawarkan manfaat potensial seperti akurasi yang lebih tinggi dan penurunan kesalahan prediksi (Swaminathan & Venkitasubramony, 2023).

Prediksi produksi dan permintaan sangat penting di berbagai bidang seperti farmasi, otomotif, dan pangan (I.-F. Chen & Lu, 2021; Yamamura, Santana, Masiero, Quintanilha, & Berssaneti, 2022; Yani & Aamer, 2023). Namun, memprediksi produksi dan permintaan ini menimbulkan tantangan karena kompleksitas dan ketidakstasioneritasan faktor-faktor yang mempengaruhi.

Peningkatan akurasi prediksi dapat dilakukan menggunakan teknik-teknik canggih seperti *machine learning*, *deep learning*, dan strategi prediksi *ensemble*. Teknik-teknik ini bertujuan untuk mengatasi ketidakpastian dalam permintaan dan produksi yang membuat prediksi permintaan dan produksi menjadi tugas yang menakutkan (Seyedan & Mafakheri, 2020). Berbagai bidang, termasuk permintaan obat, permintaan mobil, dan permintaan komoditas beras, telah diuntungkan oleh teknik-teknik canggih ini (I.-F. Chen & Lu, 2021; Yamamura et al., 2022; Yani & Aamer, 2023).

Berdasarkan kajian sebelumnya, dapat disimpulkan bahwa prediksi produksi dan permintaan adalah tugas yang memiliki banyak aspek dan tantangan yang membutuhkan integrasi teknik-teknik yang canggih dan pertimbangan berbagai faktor yang mempengaruhi. Keakuratan prediksi ini secara langsung berdampak pada kinerja manajemen cadangan beras, sehingga menekankan perlunya terus meningkatkan model dan metode prediksi.

2.3.2 Optimasi

Berbagai metode optimasi dapat digunakan untuk meningkatkan proses manajemen cadangan beras. Hal ini termasuk pemrograman linear deterministik, pemrograman linear stokastik *multi-stage*, dan *reinforcement learning* (Perez et al., 2021). Pemodelan stokastik menggunakan pemrograman linier stokastik *multi-stage* untuk mengoptimalkan permintaan pengisian ulang persediaan, yang menunjukkan peningkatan keuntungan yang signifikan dibandingkan dengan metode lainnya (Perez et al., 2021).

Pemrograman linier deterministik adalah metode lain yang digunakan untuk menetapkan kebijakan pemesanan ulang yang dinamis, yang menawarkan profitabilitas yang sebanding dengan pembelajaran penguatan (Perez et al., 2021).

Reinforcement learning juga banyak digunakan dalam berbagai bidang, terutama dalam bidang manajemen persediaan, untuk mengotomatiskan keputusan pengisian ulang dan meminimalkan biaya dan tingkat kekurangan, sehingga menciptakan kebijakan persediaan yang lebih seimbang dan tahan terhadap risiko gangguan (Abu Zwaida et al., 2021; Perez et al., 2021). Metode optimasi ini secara kolektif bertujuan untuk meningkatkan akurasi, efisiensi, dan keberlanjutan persediaan di berbagai industri.

2.3.3 Decision Support System (DSS)

Pengambilan keputusan untuk komoditas pangan (terutama beras) menjadi kompleks karena adanya keterlibatan berbagai pemangku kepentingan dengan beragam kepentingan, arus informasi yang cepat, dan waktu pengambilan keputusan yang singkat, sehingga meningkatkan risiko salah mengambil keputusan (Abaku, Edunjobi, & Odimarha, 2024; Rane, Thakker, & Kant, 2021). Informasi asimetris dan informasi yang berlebihan juga dapat mengakibatkan keputusan yang tidak efektif (Tavafoghi, Ouyang, & Teneketzis, 2021; Vosooghizaji, Taghipour, & Canel-Depitre, 2020). Meningkatkan visibilitas informasi sangat penting untuk mengatasi tantangan-tantangan ini. Akses ke informasi yang relevan memungkinkan setiap pemangku kepentingan untuk membuat keputusan dengan informasi yang lebih baik (Tanwar, 2022).

Penerapan teknologi canggih seperti *Decision Support System* (DSS) berbasis AI telah terbukti secara signifikan meningkatkan efisiensi manajemen persediaan. DSS berbasis AI ini memanfaatkan algoritma yang canggih untuk membantu dalam pengambilan keputusan, dengan tujuan meningkatkan efisiensi, akurasi, dan hasil. Adopsi teknologi ini dapat menghasilkan penghematan biaya, manajemen persediaan yang lebih baik, dan proses pengambilan keputusan yang lebih efektif (Mohamed & Saber, 2023; Tanwar, 2022; Teerasoponpong & Sopadang, 2022).

Salah satu aplikasi DSS berbasis AI yang menonjol adalah dalam bidang keuangan, khususnya dalam proses *underwriting* pinjaman. Sistem DSS berbasis AI yang dapat dijelaskan, seperti yang diterapkan dalam automasi *underwriting*

hipotek, menggunakan basis aturan keyakinan (BRB) untuk mengakomodasi pengetahuan manusia dan belajar dari data historis. Sistem ini menunjukkan keseimbangan yang baik antara akurasi dan kemampuan penjelasan, memungkinkan pemahaman yang lebih baik terhadap proses pengambilan keputusan (Sachan, Yang, Xu, Benavides, & Li, 2020).

DSS berbasis AI juga digunakan untuk meningkatkan manajemen proyek melalui model berbasis data pada bidang manajemen proyek konstruksi,. Sistem ini memungkinkan pengambilan keputusan secara *real-time* yang terintegrasi dan dioptimalkan, yang meningkatkan kecepatan dan kualitas keputusan manajemen proyek (Jianqiao Wang, 2023).

Sistem DSS berbasis AI juga telah digunakan untuk mendukung keputusan klinis, terutama dalam manajemen penyakit kardiovaskular. Sistem ini memanfaatkan algoritma AI dan teknik analisis data untuk memberikan rekomendasi yang dipersonalisasi, analisis prediktif, dan pemantauan *real-time*, yang semuanya berkontribusi pada manajemen penyakit yang lebih efektif dan peningkatan hasil kesehatan pasien (Abirami, 2023).

Pada bidang pertanian, DSS berbasis AI juga telah menunjukkan potensinya. Misalnya, teknologi AI telah digunakan untuk meningkatkan keputusan di sektor pertanian dengan memprediksi hasil panen, mengelola nutrisi tanaman, dan mengoptimalkan penggunaan sumber daya. Sebuah penelitian menunjukkan bahwa DSS berbasis AI dapat memberikan rekomendasi yang akurat dan efisien untuk prediksi hasil panen dengan menggunakan algoritma *machine learning* dan *deep learning*, yang meningkatkan kualitas dan kuantitas produksi pertanian (Anbananthen, Subbiah, Chelliah, Sivakumar, Somasundaram, Velshankar, & Khan, 2021).

AI digunakan dalam pengembangan sistem rekomendasi tanaman yang cerdas, yang membantu petani dalam menentukan tanaman yang harus ditanam berdasarkan kondisi lingkungan dan data historis. Sistem ini memanfaatkan data dari berbagai sumber seperti sensor tanah dan cuaca, serta menggunakan ML untuk memberikan rekomendasi yang tepat waktu dan relevan bagi petani (Thilakarathne, Bakar, Abas, & Yassin, 2022).

Berdasarkan kajian sebelumnya, diketahui bahwa meningkatkan visibilitas informasi sangat penting untuk memungkinkan pengambilan keputusan yang lebih efektif. Memanfaatkan teknologi canggih seperti DSS yang dilengkapi dengan algoritma canggih dapat memfasilitasi proses pengambilan keputusan, dengan tujuan meningkatkan efisiensi operasional, ketepatan, dan hasil. Pada akhirnya, pendekatan ini mengarah pada pengurangan biaya dan peningkatan manajemen persediaan.

2.4 *Black Box Testing*

Black box testing adalah metode pengujian perangkat lunak yang fokus pada pengujian fungsionalitas tanpa memerlukan pengetahuan tentang struktur internal atau kode sumber dari sistem yang diuji. Pengujian ini bertujuan untuk memastikan bahwa sistem berfungsi sesuai dengan spesifikasi yang telah ditetapkan dengan menguji berbagai input dan memeriksa output yang dihasilkan (Sasmito & Mutasodirin, 2023). Metode ini memungkinkan penguji untuk mengidentifikasi kesalahan fungsional dengan hanya berfokus pada antarmuka pengguna dan hasil akhir yang diharapkan (TanLi, Jiang, Wang, & Peng, 2020).

Salah satu teknik yang digunakan dalam *black box testing* adalah *Equivalence Partitioning*, yang membagi input data menjadi beberapa kelompok yang dianggap sama untuk tujuan pengujian. Teknik ini digunakan untuk meminimalkan jumlah kasus uji dengan mengasumsikan bahwa semua input dalam satu partisi akan memberikan hasil yang sama (Sasmito & Mutasodirin, 2023). Teknik lain yang sering digunakan adalah *Boundary Value Analysis* (BVA), yang memfokuskan pada pengujian nilai batas dari input untuk menemukan kesalahan yang mungkin terjadi pada batas-batas tersebut (Rahadi & Vikasari, 2020).

Pendekatan lain dalam *black box testing* adalah pengujian berbasis model, seperti penggunaan *Mealy Machine* dalam pengujian sistem reaktif. *Mealy Machine* adalah model yang dapat dipelajari dari sistem yang diuji dan diverifikasi terhadap spesifikasi menggunakan model *checking* (Shijubo, Waga, & Suenaga, 2021).

Berdasarkan kajian terdahulu, disimpulkan bahwa dengan menggunakan teknik *black box testing*, penguji dapat memastikan bahwa perangkat lunak berfungsi sesuai dengan spesifikasi tanpa memerlukan pengetahuan mendalam tentang kode internal, yang membuat metode ini sangat berguna dalam berbagai konteks pengembangan perangkat lunak.

2.5. Kajian Penelitian Terdahulu

Penelitian sebelumnya telah banyak mengeksplorasi penggunaan praktis dari teknologi *Generative AI* (Gen AI) dan *Machine Learning* (ML) untuk mengatasi berbagai masalah pada manajemen persediaan, terutama pada komoditas pangan. Berdasarkan temuan dan keterbatasan yang diidentifikasi dalam penelitian sebelumnya, tahap selanjutnya berfokus pada pengembangan model prediksi, optimasi, dan sistem pendukung keputusan.

2.5.1 Penelitian Terdahulu Terkait Penggunaan *Machine Learning* dalam Prediksi Produksi (Hasil Panen)

Nosratabadi, Ardabili, Lakner, Mako, and Mosavi (2021) mengusulkan model prediksi produksi pangan di tingkat makro suatu negara. Penelitian ini membandingkan dua algoritma ML, meliputi *Multilayer Perceptron* (MLP) dan *Adaptive Network-Based Fuzzy Inference System* (ANFIS), dengan menggunakan data produksi ternak dan pertanian di Iran dari tahun 1961 hingga 2017. Hasil penelitian menunjukkan bahwa model ANFIS dengan fungsi keanggotaan Gbell mengungguli model MLP dan model ANFIS lainnya dalam memprediksi produksi pangan di Iran, dengan RMSE sebesar 33.575.595,74 dan 1.724.426. Namun, model ANFIS hanya menggunakan data historis produksi dan mengabaikan variabel lain seperti perubahan iklim, kebijakan pemerintah, dan kemajuan teknologi, serta nilai *error* masih tinggi.

Noorunnahar et al. (2023) mengusulkan model prediksi produksi padi di Bangladesh. Penelitian ini membandingkan ARIMA dan XGBoost, dengan menggunakan data produksi padi tahunan dari 1961 hingga 2020. Hasil penelitian menunjukkan bahwa model XGBoost mengungguli model ARIMA dalam

memprediksi produksi padi dengan nilai MAE dan RMSE sebesar 2.779.742 dan 3.195.985. Penelitian ini juga menawarkan prediksi jangka panjang yang bermanfaat bagi perencanaan administratif dan ketahanan pangan di Bangladesh. Namun, penelitian ini tidak melibatkan variabel lain seperti suhu dan curah hujan, serta nilai *error* masih tinggi. Untuk itu, para peneliti menyarankan untuk mempertimbangkan variabel lain seperti iklim.

Satpathi et al. (2023) mengusulkan model prediksi hasil panen padi di Chhattisgarh. Penelitian ini membandingkan lima metode, yaitu SMLR, ANN, LASSO, ELNET, Ridge Regression, dan *ensemble models*, dengan menggunakan data hasil panen padi dan data meteorologi (suhu, curah hujan, kelembapan, dan durasi penyinaran matahari) tahun 1998-2018 dari tiga distrik di Chhattisgarh. Hasil penelitian menunjukkan bahwa model prediksi dengan ANN mengungguli model lainnya dengan nilai R^2 sebesar 1 dan 0,99 serta nilai RMSE sebesar 0,003 dan 0,004. Secara keseluruhan, penelitian ini memiliki akurasi yang tinggi, penanganan multikolinearitas yang efektif, dan kemampuan untuk menyertakan banyak variabel prediksi. Namun, beberapa model seperti SMLR dan Ridge Regression menunjukkan kinerja yang tidak selama proses pengujian tetapi baik selama proses pelatihan, yang disebabkan oleh *overfitting*. Untuk itu, para peneliti menyarankan untuk mengeksplorasi model hibrida untuk meningkatkan akurasi prediksi.

Joshua, Priyadharson, and Kannadasan (2021) mengusulkan model prediksi hasil panen padi di Zona Delta Cauvery, bagian timur Tamil Nadu, India Selatan. Penelitian ini membandingkan beberapa algoritma ML, yaitu Support Vector Regression (SVR), General Regression Neural Networks (GRNNs), Radial Basis Function Neural Networks (RBFNNs), dan Back-Propagation Neural Networks (BPNNs), dengan menggunakan data historis hasil panen, curah hujan, suhu, pupuk, nitrogen, fosforus, kalium, dan PH dari Juni 2018-Januari 2019. Hasil penelitian ini menunjukkan bahwa model prediksi dengan GRNN mengungguli model lainnya dengan nilai R^2 , RMSE, MAE, MSE, MAPE, CV, dan NSME sebesar 0,9863; 0,2295; 0,1290; 0,0526; 1,3439; 0,0255; dan 0,0136. Model tersebut menunjukkan tingkat akurasi dan keandalan yang tinggi dalam

memprediksi hasil panen padi, serta efektif dalam mengintegrasikan data historis dan data *real-time* dalam analisisnya. Namun, model yang dibutuhkan membutuhkan data ekstensif untuk prediksi yang akurat.

Guo, Fu, Hao, Zhang, Wu, Jin, Bryant, and Senthilnath (2021) mengusulkan model prediksi hasil panen padi untuk China Selatan. Penelitian ini membandingkan Multiple Linear Regression (MLR) dengan metode ML lainnya seperti Back-Propagation Neural Network (BP), Support Vector Machine (SVM), dan Random Forest (RF), dengan menggunakan data historis hasil panen padi, fenologi padi, iklim, dan data geografis dari tahun 1981-2010. Hasil penelitian ini menunjukkan bahwa model SVM mengungguli model lainnya dengan nilai RMSE dan R^2 sebesar 737 dan 0,33. Model ini mampu mengelola interaksi kompleks antara variabel fenologis dan iklim yang beragam. Namun, model ini tidak mempertimbangkan variasi praktik pertanian, perubahan varietas, atau teknologi pertanian yang digunakan.

Cedric, Adoni, Aworka, Zoueu, Mutombo, Krichen, and Kimpolo (2022) mengusulkan model prediksi hasil panen padi, jagung, ubi kayu, kapas biji, ubi, dan pisang di Afrika Barat. Model ini membandingkan Decision Tree, Regresi Logistik Multivariat, dan k-Nearest Neighbors, dengan menggunakan data pertanian, data curah hujan tahunan, data iklim, data cuaca, dan data kimia dari tahun 1990 hingga 2020. Hasil penelitian ini menunjukkan bahwa k-Nearest Neighbors memberikan kinerja terbaik di antara tiga model *machine learning* dengan nilai sebesar 93,15%. Model yang diusulkan memiliki nilai akurasi prediksi model yang tinggi, integrasi data yang komprehensif, dan potensi untuk membantu dalam pengambilan keputusan untuk perencanaan pertanian. Namun, model tersebut hanya melibatkan variabel iklim dan kimia saja.

Kuradusenge et al. (2023) mengusulkan model prediksi hasil panen untuk komoditi kentang dan jagung. Model ini membandingkan Random Forest (RF), Polynomial Regression (PR), dan Support Vector Regressor (SVR), dengan menggunakan data historis curah hujan, suhu, dan hasil panen dari tahun 2006 hingga 2021. Hasil penelitian menunjukkan bahwa RF mengungguli model lainnya dengan nilai RMSE sebesar 510,8 dan 129,9 untuk komoditas kentang dan

jagung, serta nilai R^2 sebesar 0,875 dan 0,817 untuk komoditas kentang dan jagung. Namun, model yang diusulkan tidak mempertimbangkan variabel lain dan terbatas pada satu wilayah saja sehingga dapat membatasi generalisasi.

Q.-C. Li, Xu, Zhuang, Liu, Yi, and Zhang (2023) mengusulkan model prediksi hasil produksi beras dengan mengintegrasikan berbagai algoritma ML yaitu k-Nearest Neighbor, random forest, dan support vector regression. Model ini dibangun dengan menggunakan data meteorologis harian dan data historis produksi kedelai dari 173 wilayah selama 34 tahun. Hasil penelitian ini menunjukkan bahwa model cukup akurat dengan nilai MAPE berkisar antara 4,43%-5,85%, RMSE 138,47-179,68, MAE 110,08-134,42, dan R^2 0,89-0,94. Secara keseluruhan, model yang diusulkan cukup akurat dalam memprediksi hasil panen melalui integrasi beberapa model ML dan sangat berguna untuk perencanaan di sektor pertanian. Namun, penelitian ini terbatas pada wilayah tertentu di Cina dan mungkin tidak dapat langsung diterapkan pada tanaman atau wilayah lain tanpa penyesuaian terlebih dahulu.

Penelitian-penelitian terdahulu menunjukkan bahwa penggunaan metode *machine learning* (ML) memiliki potensi besar dalam memprediksi produksi atau hasil panen berbagai komoditas pangan. Meskipun model-model ini menunjukkan kinerja yang baik, tetapi masih ada beberapa tantangan yang perlu diatasi seperti tidak melibatkan variabel lain yang relevan terhadap hasil panen, memiliki cakupan geografis yang terbatas, keterbatasan dalam spesifikasi model, dan *overfitting*. Untuk itu, penelitian ini akan berfokus pada pengembangan model prediksi yang lebih *robust* melalui pemilihan dan integrasi variabel-variabel yang dapat mempengaruhi hasil prediksi, menggunakan data dari berbagai wilayah, dan mengimplementasikan ML yang lebih sesuai. Tabel 2.1 adalah ringkasan penelitian terdahulu terkait penerapan ML untuk prediksi produksi (hasil panen) padi.

Tabel 2. 1 Ringkasan Penelitian Terdahulu Tentang Prediksi Produksi (Hasil Panen) Padi

Penulis	Tujuan	Data	Variabel	Metode	Hasil	Kelebihan	Kekurangan
Nosratabadi et al. (2021)	Mengembangkan model untuk memprediksi produksi pangan di tingkat makro suatu negara	Data produksi ternak dan pertanian di Iran tahun 1961- 2017	Data seri waktu produksi tahunan	<i>Multilayer Perceptron</i> (MLP) dan <i>Adaptive Network-Based Fuzzy Inference System</i> (ANFIS)	RMSE MLP: 265.590.099,2 dan 33.575.595,74; RMSE ANFIS: 33.575.595,74 dan 1.724.426	Mampu menangani ketidakpastian dan ketidaktepatan data	Hanya menggunakan data historis produksi dan mengabaikan variabel lain seperti perubahan iklim, kebijakan pemerintah, dan kemajuan teknologi, serta nilai <i>error</i> masih tinggi
Noorunnahar et al. (2023)	Mengembangkan model prediksi produksi padi tahunan di Bangladesh	Data produksi padi tahunan dari 1961 hingga 2020	Data seri waktu produksi tahunan	ARIMA dan XGBoost	MAE dan RMSE ARIMA: 3755137 dan 4093961, MAE dan RMSE XGBoost: 2779742 dan 3195985	Menawarkan prediksi jangka panjang yang bermanfaat bagi perencanaan administratif dan ketahanan pangan	Tidak melibatkan variabel penting lainnya seperti suhu dan curah hujan, serta nilai <i>error</i> masih tinggi
Satpathi et al. (2023)	Menemukan metode terbaik untuk prediksi hasil panen padi di Chhattisgarh	Data produksi padi dan data meteorologi tahun 1998-2018 dari tiga distrik di Chhattisgarh	Data produksi padi, suhu, curah hujan, kelembapan, dan durasi penyinaran matahari	SMLR, ANN, LASSO, ELNET, Ridge Regression, dan Ensemble Models	ANN unggul dengan nilai R^2 sebesar 1,00 dan 0,99, serta nilai RMSE sebesar 0,003 dan 0,004	Memiliki akurasi yang tinggi pada model dan distrik tertentu, penanganan multikolinearitas yang efektif, dan kemampuan untuk menyertakan banyak variabel prediksi	Beberapa model, seperti SMLR dan Ridge Regression, menunjukkan kinerja baik selama pelatihan tetapi tidak selama proses pengujian, yang disebabkan oleh <i>overfitting</i>
Joshua et al. (2021)	Menemukan algoritma ML terbaik untuk	Data historis tentang tanaman	Data produksi padi, curah	SVR, GRNN, RBFNN, dan BPNN	GRNN unggul dengan nilai R^2 , RMSE, MAE,	Memiliki tingkat akurasi dan keandalan yang	Membutuhkan data ekstensif untuk prediksi yang

Penulis	Tujuan	Data	Variabel	Metode	Hasil	Kelebihan	Kekurangan
	prediksi hasil panen padi di India Selatan	padi dari berbagai distrik dalam Zona Delta Cauvery dari Juni 2018-January 2019	hujan, suhu, pupuk, nitrogen, fosforus, kalium, dan pH		MSE, MAPE, CV, dan NSME sebesar 0,9863; 0,2295; 0,1290; 0,0526; 1,3439; 0,0255; dan 0,0136	tinggi dalam memprediksi hasil panen padi, serta efektif dalam mengintegrasikan data historis dan data <i>real-time</i> dalam analisisnya	akurat
Guo et al. (2021)	Mengembangkan model prediksi hasil panen padi menggunakan metode ML berdasarkan data fenologi, iklim, dan geografis	Data produksi padi, data fenologi padi, data iklim, dan data geografis dari tahun 1981-2010	Jumlah hasil panen, tahapan fenologi padi, data iklim (suhu curah hujan, lama penyinaran matahari, dll), dan data geografis (koordinat lokasi dan ketinggian)	Multiple Linear Regression (MLR), Back-Propagation Neural Network (BP), Support Vector Machine (SVM), dan Random Forest (RF).	RMSE dan R^2 untuk BP, SVM, dan RF secara berurutan adalah 800 dan 0,24, 737 dan 0,33, serta 744 dan 0,31	Mampu mengelola interaksi kompleks antara variabel fenologis dan iklim yang beragam	Tidak memperhitungkan variasi praktik pertanian, perubahan varietas, atau teknologi pertanian yang digunakan, yang dapat mempengaruhi akurasi prediksi
Cedric et al. (2022)	Mengembangkan model prediksi hasil panen padi, jagung, ubi kayu, kapas biji, ubi, dan pisang di Afrika Barat	Data pertanian, data curah hujan tahunan, data iklim, data cuaca, dan data kimia dari	Suhu, curah hujan, penggunaan pestisida, dan emisi nitrogen dioksida	Decision Tree, Regresi Logistik Multivariat, dan k-Nearest Neighbors	Koefisien determinasi (R^2) untuk k-Nearest Neighbor adalah 93,15%	Akurasi prediksi model yang tinggi, integrasi data yang komprehensif, dan potensi untuk membantu dalam pengambilan keputusan untuk perencanaan	Hanya melibatkan variabel iklim dan kimia

Penulis	Tujuan	Data	Variabel	Metode	Hasil	Kelebihan	Kekurangan
		tahun 1990-2020				pertanian	
Kuradusenge et al. (2023)	Mengembangkan model prediksi hasil panen untuk komoditi kentang dan jagung	Data historis curah hujan, suhu, dan hasil panen dari tahun 2006-2021	Curah hujan bulanan, suhu harian rata-rata, dan hasil panen musiman (kg/ha)	Random Forest (RF), Polynomial Regression (PR), Support Vector Regressor (SVR)	RF menunjukkan performa terbaik dengan RMSE 510,8 dan 129,9 untuk kentang dan jagung, serta R^2 sebesar 0,875 dan 0,817 untuk kentang dan jagung	Data historis luas	Terbatas pada variabel curah hujan dan suhu, tidak mempertimbangkan variabel lain, terbatas pada satu wilayah penelitian saja
Q.-C. Li et al. (2023)	Mengintegrasikan berbagai algoritma ML untuk meningkatkan akurasi prediksi hasil produksi kedelai	Data meteorologis harian dan data historis produksi kedelai dari 173 wilayah selama 34 tahun	Suhu rata-rata, curah hujan, durasi sinar matahari, dan data historis produksi/ hasil panen	K-nearest neighbor+random forest+support vector regression	Model yang diusulkan unggul dengan dengan MAPE berkisar 4,43%-5,85%, RMSE 138,47-179,68, MAE 110,08-134,42, dan R^2 0,89-0,94	Prediksi akurat, mengintegrasikan beberapa model ML, dan sangat berguna untuk perencanaan di sektor pertanian	Terbatas pada wilayah tertentu di Cina dan mungkin tidak langsung diterapkan pada tanaman atau wilayah lain tanpa penyesuaian

2.5.2 Penelitian Terdahulu Terkait Penggunaan *Machine Learning* dalam Prediksi Permintaan

Lutoslawski et al. (2021) mengusulkan model prediksi untuk permintaan makanan olahan seperti roti dan mentega. Penelitian ini menggunakan metode Nonlinear Autoregressive Exogenous Neural Network (NARXNN) dan data historis permintaan berbagai produk makanan dari April 2017 hingga Juni 2019. Hasil penelitian ini menunjukkan bahwa model yang diusulkan menghasilkan prediksi yang akurat dengan nilai R^2 sebesar 96,2399 hingga 99,6477. Model ini juga adaptif terhadap berbagai jenis produk makanan dan berkontribusi pada pengurangan limbah makanan, namun model ini juga memiliki tingkat kompleksitas yang cukup tinggi dalam *setup* dan *tuning* model, serta tidak mampu untuk menganalisis *dataset* kecil (di bawah 100 baris data).

Nassibi et al. (2023) mengusulkan model prediksi permintaan produk makanan untuk meningkatkan efektivitas manajemen rantai pasok dan mengurangi limbah. Penelitian ini membandingkan dua metode, yaitu Long Short-Term Memory (LSTM) dan Support Vector Machine (SVM), dengan menggunakan data historis penjualan tahun 2020. Hasil penelitian ini menunjukkan bahwa model LSTM mengungguli model SVM dengan nilai MAPE dan RMSE antara 3,78-15,3 dan 2,24-4,64, sedangkan SVM antara 6,8-73,2 dan 7,69-12,35. Secara keseluruhan, model dapat memprediksi permintaan secara akurat, adaptif, dan berkontribusi pada pengurangan limbah. Namun, kinerja model ini bergantung pada kualitas dan kelengkapan data.

Amellal, Amellal, Seghioeur, and Ech-Charrat (2024) mengusulkan pendekatan prediksi permintaan berbasis *Gated Recurrent Unit* (GRU) yang menghasilkan prediksi yang akurat dengan nilai MSE sebesar 0,015. Model ini memiliki keunggulan dalam menangani data berskala besar dan berdimensi tinggi, serta lebih baik dalam mengungkap pola-pola yang kompleks dibandingkan dengan metode statistik tradisional. Namun, model ini dapat mengalami *overfitting* dan membutuhkan teknik regularisasi yang canggih untuk mengatasi masalah ini. Secara keseluruhan, penelitian ini menunjukkan bahwa model yang

diusulkan efektif untuk prediksi permintaan, tetapi penting untuk memperhatikan potensi *overfitting* untuk memastikan kinerja yang optimal.

Rodrigues, Miguéis, Freitas, and Machado (2024) memperkenalkan empat model untuk memprediksi permintaan makanan secara akurat dalam layanan katering makanan untuk mencegah kelebihan dan kekurangan produksi. Model-model tersebut menggunakan *machine learning* yang berbeda, termasuk RF, lightGBM, LSTM, dan transformer. Penelitian ini menemukan bahwa algoritma LSTM menghasilkan perkiraan yang paling dapat diandalkan dengan nilai RMSE berkisar antara 60,9 hingga 173,36, yang dapat mengurangi pemborosan makanan hingga 52%. Namun, model ini hanya efektif dalam memprediksi permintaan untuk satu item menu, bukan keseluruhan menu, yang mengindikasikan perlunya pengembangan lebih lanjut agar model prediksi lebih adaptif untuk keseluruhan menu.

Aci and Yergök (2023) mengusulkan model prediksi berbasis *machine learning* untuk memprediksi lonjakan permintaan makanan selama suatu periode. Penelitian ini membandingkan beberapa algoritma ML, termasuk ANN, Regresi Proses Gaussian, SVR, RT, dan EDT. Model Boosted EDT menunjukkan performa terbaik dengan MSE sebesar 0,51, MAE sebesar 0,5, dan R² sebesar 0,96. Namun, model ini memiliki keterbatasan dalam penerapannya pada konteks yang berbeda. Secara keseluruhan, model yang diusulkan efektif dalam memprediksi permintaan makanan secara akurat.

Penelitian-penelitian terdahulu menunjukkan bahwa penggunaan metode *machine learning* (ML) memiliki potensi besar dalam memprediksi permintaan. Meskipun model-model ini menunjukkan kinerja yang baik, masih terlihat adanya beberapa tantangan yang perlu diatasi seperti kompleksitas tinggi dalam *setup* dan *tuning*, bergantung pada kualitas data, risiko *overfitting*, dan kurang efektif untuk prediksi menyeluruh. Untuk itu, penelitian ini berfokus pada mengembangkan model prediksi yang sederhana, fleksibel, dan memiliki generalisasi yang baik. Tabel 2.2 adalah ringkasan penelitian terdahulu tentang prediksi permintaan.

Tabel 2. 2 Ringkasan Penelitian Terdahulu Terkait Prediksi Permintaan

Penulis	Tujuan	Data	Metode	Hasil	Kelebihan	Kekurangan
Lutoslawski et al. (2021)	Mengembangkan model prediksi permintaan makanan olahan seperti roti dan mentega	Data historis permintaan berbagai produk makanan dari April 2017 - Juni 2019	Nonlinear Autoregressive Exogenous Neural Network (NARXNN)	R^2 : 96,2399 hingga 99,6477	Prediksi akurat, adaptif terhadap berbagai jenis produk makanan, berkontribusi pada pengurangan limbah makanan	Memiliki tingkat kompleksitas yang cukup tinggi dalam <i>setup</i> dan <i>tuning</i> model, serta tidak mampu untuk menganalisis <i>dataset</i> kecil (di bawah 100 baris data)
Nassibi et al. (2023)	Mengembangkan model prediksi permintaan produk makanan untuk manajemen rantai pasok yang lebih efektif	Data historis penjualan tahun 2020	Long Short-Term Memory (LSTM) dan Support Vector Machine (SVM)	MAPE dan RMSE LSTM: antara 3,78-15,3 dan 2,24-4,64; MAPE dan RMSE SVM: antara 6,8-73,2 dan 7,69-12,35	Prediksi akurat, adaptif, dan berkontribusi pada pengurangan limbah	Kinerja model bergantung pada kualitas dan kelengkapan data
Amellal et al. (2024)	Mengembangkan model prediksi permintaan	Data penjualan tahun 2019-2022, total 111.236 baris dengan 10 kolom	Gated Recurrent Unit (GRU)	MSE GRU 0,015 dan R^2 0,982	Prediksi akurat, mampu menangani <i>dataset</i> besar, lebih baik dari model lain, dan dapat memahami ketidakpastian data	Terdapat potensi <i>overfitting</i> pada model yang dapat mengganggu kinerjanya
Rodrigues et al. (2024)	Menemukan model prediksi terbaik untuk katering makanan	Data periode Januari 2017-Desember 2019	RF, lightGBM, LSTM, dan transformer	LSTM RMSE: 60,9-173,36 MAPE: 0,1432-0,3061 R^2 : 0,1397-0,7325	Model dapat menyesuaikan tren data jangka panjang	Model hanya efektif dalam memprediksi permintaan untuk satu item menu daripada untuk keseluruhan menu
Aci and Yergök (2023)	Mengembangkan model prediksi untuk memprediksi lonjakan permintaan makanan selama suatu periode	Data penjualan selama 1 tahun dengan total 2241 catatan penjualan	ANN, Gaussian Process Regression, SVR, RT, dan EDT	Model EDT Boosted menunjukkan performa terbaik dengan MSE 0,51; MAE 0,5, dan R 0,96	Membandingkan beragam teknik <i>machine learning</i> sehingga didapatkan model dengan kinerja terbaik	Keterbatasan dalam penerapan model di konteks yang berbeda

2.5.3 Penelitian Terdahulu Terkait Integrasi *Generative AI* dan ML untuk Tujuan Prediksi

Beberapa kekurangan pada model prediksi produksi dan permintaan, yang diidentifikasi dalam penelitian terdahulu, dapat dikurangi dengan menggunakan teknologi *generative AI* (gen AI). Gen AI dapat menghasilkan data sintetis yang berkualitas tinggi untuk melengkapi dan memperbaiki data yang ada, sehingga meningkatkan kualitas *dataset* yang digunakan untuk prediksi (Yu & Guo, 2023). Teknologi ini juga dapat berguna untuk menciptakan skenario pelatihan yang beragam, yang membantu model belajar dari data yang lebih beragam dan mengurangi *overfitting* (Manjunath & Palayyan, 2023). Gen AI juga dapat meningkatkan prediksi menyeluruh dengan memodelkan interaksi kompleks antar variabel menggunakan model ML yang dilatih dengan data sintetis, sehingga dapat mengatasi berbagai kondisi (Y. Liu, Yang, Yu, Liu, Liu, Lin, Li, Ma, Avdeev, & Shi, 2023). Hal ini menunjukkan bahwa integrasi antara gen AI dan ML memiliki potensi besar dalam mengatasi kekurangan model prediksi yang ada, menyediakan solusi yang lebih komprehensif dan adaptif.

Penerapan *generative AI* dan *Machine Learning* (ML) dalam prediksi produksi (hasil panen) dan permintaan di sektor pertanian dan pangan (terutama dalam konteks data *time series*) juga masih belum banyak diteliti. Di sisi lain, beberapa penelitian juga telah menunjukkan efektivitas pendekatan *hybrid* ini efektif dalam meningkatkan akurasi prediksi di berbagai bidang lain dan memberikan hasil yang menjanjikan.

S. Wang and Yang (2021) mengusulkan model prediksi penjualan untuk *online store*. Penelitian ini menggunakan metode M-GAN-XGBoost, yang menggabungkan LSTM, GAN, dan XGBoost. Penelitian ini menggunakan tiga jenis *Generative Adversarial Networks* (GAN) meliputi WGAN, LSGAN dan BEGAN untuk mencari model yang paling akurat. *Dataset* yang digunakan adalah *dataset* penjualan Jingdong dari Februari 2018 hingga April 2018. Implementasi model ini dilakukan dengan menggunakan *library* TensorFlow dan bahasa pemrograman Python. Hasil penelitian ini menunjukkan bahwa model integrasi LSTM, BEGAN, dan XGBoost mengungguli model lain dengan nilai MSE dan

MAE sekitar 11,9 dan 8,23. Model ini memberikan prediksi yang akurat dan memfasilitasi manajemen sumber daya yang lebih baik, namun model ini juga masih belum optimal dalam memilih dan mengukur variabel langsung dan variabel *lagged* yang berpengaruh dalam mencapai akurasi maksimal.

D. Wu, Hur, and Xiao (2021) mengusulkan model yang akurat dan *robust* untuk memprediksi konsumsi energi pada *smart building*. Model ini menggunakan gabungan beberapa model prediktif meliputi Stochastic Gradient Descent (SGD), Support Vector Classifier (SVC), XGBoost, LightGBM, dan k-Nearest Neighbors (KNN) yang diperkuat dengan GAN (vanilla GAN, InfoGAN, dan cGAN) untuk meningkatkan akurasi prediksi. *Dataset* yang digunakan adalah *dataset* yang dikembangkan oleh Departemen Energi AS bersama dengan tiga laboratorium nasional. Implementasi model ini dilakukan dengan menggunakan *library* keras, bahasa pemrograman Python 3.6.7, dan Adam *optimizer*. Hasil penelitian menunjukkan bahwa model yang ditingkatkan dengan InfoGAN secara signifikan mengurangi kesalahan prediksi dengan nilai MAE sebesar 1.71%, RMSE 1.63% , dan CVRMSE 4.72% dibandingkan dengan model tanpa augmentasi. Hal ini menunjukkan model ini berhasil meningkatkan ketahanan dan menurunkan kesalahan prediksi di berbagai metrik, namun model ini juga hanya menggunakan satu metode *ensemble* dan satu arsitektur GAN pada *dataset* tunggal, serta belum mengoptimasi desain jaringan.

Naveed, Hashmi, Tajved, Sultan, and Imran (2022) mengusulkan model generatif untuk menghasilkan data sintesis berkualitas guna melatih model prediksi beban jaringan dengan menggunakan *supervised machine learning*. Penelitian ini membandingkan dua jenis GAN (TimeGAN dan DoppelGANger) dengan model autoregresif dalam menghasilkan data sintesis dari Telecom Italia *Dataset*. Hasil penelitian ini menunjukkan bahwa model GAN, khususnya DoppelGANger, mengungguli metode autoregresif dalam menghasilkan data sintesis berkualitas untuk pelatihan model prediksi dengan *machine learning*. Meskipun menunjukkan kinerja yang baik dan dapat menghasilkan data sintesis yang beragam, namun penelitian ini memiliki keterbatasan generalisasi model yang kurang dan analisis kuantitatif yang tidak komprehensif.

Chatterjee and Byun (2023) meningkatkan akurasi model prediksi permintaan harian skuter elektrik dengan menggunakan data sintetis. Penelitian ini menggunakan CTGAN untuk menghasilkan data sintetis dan model *ensemble* (CatBoost, RF, XGBoost) untuk memprediksi permintaan. *Dataset* yang digunakan terdiri dari data sewa skuter elektrik di Pulau Jeju, Korea Selatan, serta data cuaca dari Administrasi Meteorologi Korea. Implementasi model dilakukan dengan menggunakan *library NumPy, Pandas, Sklearn, Seaborn, dan PyCaret* dalam bahasa pemrograman Python dengan pengoptimalan melalui *Grid Search*. Hasil penelitian ini menunjukkan bahwa model *ensemble* dengan data sintetis terbukti meningkatkan akurasi prediksi permintaan dan mencapai nilai R^2 tertinggi sebesar 0,79 dan MAPE terendah sebesar 21,22. Meskipun model ini terbukti berhasil meningkatkan akurasi dengan data yang terbatas, namun kinerja model GAN yang digunakan kurang maksimal.

Fekri, Ghosh, and Grolinger (2019) mengusulkan R-GAN untuk menghasilkan data konsumsi energi yang realistis dengan menggunakan GAN, *Recurrent Neural Networks*, dan *ARIMA, Fourier Transform*. Penelitian ini menggunakan dua *dataset*, meliputi *UCI Appliances Energy Prediction Dataset* dan *Building Data Genome Project Dataset*. Hasil penelitian ini menunjukkan bahwa R-GAN efektif untuk menghasilkan data konsumsi energi yang realistis, yang dapat digunakan untuk melatih model ML dengan keandalan tinggi. Model ini mampu menangani dependensi temporal data, meningkatkan kualitas dan realisme data, dan memperkuat stabilitas pelatihan, namun model ini juga memiliki keterbatasan kompleksitas komputasi yang tinggi, potensi masalah dalam generalisasi data sintetis, dan sensitivitas terhadap pemilihan parameter.

Penelitian terdahulu telah menunjukkan bahwa integrasi antara Gen AI dan ML memiliki potensi yang menjanjikan untuk meningkatkan akurasi model prediksi. Kemajuan teknologi ini membuka peluang baru untuk pengembangan di bidang pertanian dan pangan. Tabel 2.3 adalah ringkasan penelitian terdahulu terkait Gen AI dan ML.

Tabel 2. 3 Ringkasan Penelitian Terdahulu Terkait Integrasi Gen AI dan ML untuk Tugas Prediksi

Penulis	Tujuan	Data	Metode	Tipe GAN	Hasil	Kelebihan	Kekurangan
S. Wang and Yang (2021)	Mengembangkan model prediksi penjualan untuk online store	<i>Dataset</i> penjualan Jingdong dari Februari 2018-April 2018	M-GAN-XGBOOST (LSTM-GAN-XGBoost)	WGAN, LSGAN dan BEGAN	Model dengan BEGAN unggul dengan nilai MSE dan MAE sekitar 11,9 dan 8,23	Prediksi akurat dan memfasilitasi manajemen sumber daya yang lebih baik	Model masih belum optimal dalam memilih dan mengukur variabel langsung dan variabel lagged yang berpengaruh dalam mencapai akurasi maksimal
D. Wu et al. (2021)	Mengembangkan model yang akurat dan <i>robust</i> untuk memprediksi konsumsi energi pada <i>smart building</i>	https://www.energy.gov/eere/buildings / commercial-reference-buildings	<i>Ensemble stacking</i> (SGD+SVR+XG Boost+LightGBM+KNN)+GAN	Vanilla GAN, Info GAN, and Conditional GAN (cGAN)	Model yang ditingkatkan dengan InfoGAN secara signifikan mengurangi kesalahan prediksi dengan nilai MAE sebesar 1.71%, RMSE 1.63% , dan CVRMSE 4.72% dibandingkan dengan model tanpa augmentasi	Model menunjukkan peningkatan ketahanan dan mengurangi kesalahan prediksi di berbagai metrik	Hanya menggunakan satu metode <i>ensemble</i> dan satu arsitektur GAN pada <i>dataset</i> tunggal, serta belum mengoptimasi desain jaringan
Naveed et al. (2022)	Mengembangkan model generatif untuk menghasilkan data sintetis yang berkualitas	Telecom Italia <i>Dataset</i>	GANs+model autoregresif untuk menghasilkan data sintetis, serta algoritma <i>supervised machine learning</i> untuk prediksi	TimeGAN dan DoppelGANger	Model GAN, khususnya DoppelGANger, mengungguli metode autoregresif dalam menghasilkan data sintetis yang berkualitas untuk pelatihan model <i>machine learning</i>	Model menunjukkan kinerja yang baik dan dapat menghasilkan data sintetis yang beragam	Evaluasi model yang terbatas, generalisasi model yang kurang, dan analisis kuantitatif yang tidak komprehensif
Chatterjee and Byun (2023)	Meningkatkan akurasi prediksi permintaan	Data penyewaan skuter listrik di Pulau Jeju, Korea	<i>Ensemble model</i> (CatBoost+RF+XGBoost)+GAN	CTGAN	Model <i>ensemble</i> dengan data sintetis meningkatkan akurasi prediksi dengan R^2	Meningkatkan akurasi dengan data yang terbatas	Kinerja model GAN yang digunakan kurang

Penulis	Tujuan	Data	Metode	Tipe GAN	Hasil	Kelebihan	Kekurangan
	harian skuter elektrik dengan menggunakan data sintetis	Selatan, dan data cuaca dari Administrasi Meteorologi Korea			tertinggi 0,79 dan MAPE terendah 21,22		maksimal
Fekri et al. (2019)	Mengembangkan R-GAN untuk menghasilkan data konsumsi energi yang realistis	UCI <i>Appliances Energy Prediction Dataset</i> dan <i>Building Data Genome Project</i>	GAN+ARIMA+ Fourier Transform+Recurrent Neural Networks	R-GAN, WGAN, MH-GAN	R-GAN efektif untuk menghasilkan data konsumsi energi yang realistis, yang dapat digunakan untuk melatih model ML dengan keandalan tinggi	Dapat menangani dependensi temporal data, meningkatkan kualitas dan realisme data, dan memperkuat stabilisasi pelatihan	Kompleksitas komputasi yang tinggi, potensi masalah dalam generalisasi data sintetis, dan sensitivitas terhadap pemilihan parameter

2.5.4 Penelitian Terdahulu Terkait *Decision Support System*, Prediksi, dan Optimasi

Dellino, Laudadio, Mari, Mastronardi, and Meloni (2018) mengembangkan *Decision Support System* (DSS) dengan mengintegrasikan prediksi penjualan dan perencanaan pesanan makanan segar. Penelitian ini menggunakan berbagai model peramalan seperti ARIMA, ARIMAX, dan model fungsi transfer, serta optimasi multi-objektif. Hasil penelitian menunjukkan bahwa sistem mampu menghasilkan prediksi yang akurat dan usulan pesanan yang seimbang, serta dapat beradaptasi dengan kebutuhan dan preferensi pengguna. Sistem ini secara *seamless* mengintegrasikan prediksi dan perencanaan pesanan, serta menawarkan fleksibilitas dalam pemilihan model berdasarkan kriteria yang ditetapkan pengguna. Namun, sistem ini memiliki kompleksitas tinggi dalam implementasi dan penggunaannya, serta skalabilitas yang belum sepenuhnya dieksplorasi. Peneliti merekomendasikan untuk memperluas teknik prediksi seperti *Exponential Smoothing* dan *Neural Networks*, mengembangkan alat canggih untuk penyetelan parameter, mengevaluasi dan meningkatkan DSS dalam menghadapi ketidakpastian permintaan dan adaptasi untuk produk dengan masa simpan acak, serta menambahkan penilaian manajerial dalam prediksi dan mengembangkan DSS.

Namany, Govindan, Alfagih, McKay, and Al-Ansari (2020) mengusulkan sebuah model pengambilan keputusan dinamis untuk sektor pangan di Qatar dengan menggunakan *Agent-Based Modeling* (ABM) dengan fokus pada pangan tomat. Model ini menggabungkan variabel keputusan seperti *grow*, *import*, dan *hybrid*, serta mempertimbangkan faktor-faktor seperti permintaan tomat, harga lokal dan internasional, kebutuhan air tanaman (*Plant Water Requirements/CWR*), kapasitas produksi dan impor, serta kontrak berjangka untuk mempengaruhi keputusan pemilihan strategi. Model yang diimplementasikan menggunakan Python dan *library* MESA ini mensimulasikan profil impor dan produksi tomat selama lima tahun, yang menawarkan berbagai keuntungan termasuk fleksibilitas, kemampuan beradaptasi dengan beragam skenario, pendekatan holistik yang mencakup pertimbangan ekonomi dan lingkungan, dan kemampuan untuk

menangkap perubahan temporal. Namun, efektivitasnya bergantung pada kualitas dan ketersediaan data, kompleksitas penafsiran model ABM, dan potensi penyederhanaan asumsi. Para peneliti merekomendasikan penyempurnaan lebih lanjut, seperti mengintegrasikan tanaman lain, faktor lingkungan yang lebih luas (misalnya, emisi dari sistem energi dan air), dan mengeksplorasi strategi ketahanan pangan tingkat lanjut.

Teerasoponpong and Sopadang (2022) mengusulkan sistem pendukung keputusan (DSS) bagi UKM pangan untuk mengatasi ketidakpastian kondisi bisnis, dengan memanfaatkan simulasi-optimasi yang menggabungkan *Artificial Neural Network* (ANN) dan *Genetic Algorithm* (GA) untuk merepresentasikan masalah dan optimasi solusi pendukung keputusan. DSS yang dihasilkan memfasilitasi pemilihan pemasok yang direkomendasikan dan jumlah pesanan yang optimal, sehingga mengurangi biaya keseluruhan yang terkait dengan pengadaan bahan baku, interval pemesanan, dan persediaan. Namun, kinerja DSS tergantung pada kualitas dan ketersediaan data. Para peneliti menganjurkan penelitian di masa depan untuk mengembangkan sistem berbasis AI, mengeksplorasi pendekatan optimasi dengan mengurangi waktu komputasi, dan meningkatkan visibilitas data dan informasi menggunakan untuk meningkatkan kegunaan DSS.

H. Chen, Chen, Lin, and Zhuang (2021) mengusulkan model *Deep Reinforcement Learning-Supply Chain Management* (DR-SCM) untuk mengoptimalkan keputusan produksi dan persediaan dalam rantai pasok agrikultur, khususnya jagung. Tujuannya adalah untuk meningkatkan keuntungan petani, keterlacakan produk, dan efisiensi pengambilan keputusan. Penelitian ini menggunakan *Deep Reinforcement Learning* (DRL) yang diimplementasikan menggunakan TensorFlow 1.4.0 dan NumPy di Python 3.6. Hasil penelitian menunjukkan bahwa model yang diusulkan mampu mencapai keuntungan yang lebih tinggi dibandingkan metode heuristik dan Q-learning dalam berbagai skenario ASC. Namun, algoritma yang digunakan menerapkan mekanisme *replay memory* yang menunjukkan relevansi kuat, mengurangi efisiensi pelatihan untuk

kinerja optimal. Para peneliti merekomendasikan penerapan algoritma yang lebih canggih seperti *asynchronous advantage actor-critic*.

C. Zhao et al. (2021) mengusulkan pendekatan yang terintegrasi yang menggabungkan probabilitas dan optimasi yang kuat untuk kuantifikasi cadangan energi, yang dihitung berdasarkan interval prediksi daya angin. Metode yang digunakan mencakup Extreme Learning Machine (ELM) dan Mixed Integer Linear Programming (MILP), dengan mempertimbangkan faktor-faktor seperti kapasitas tenaga angin, target tenaga angin, dan biaya denda atas kekurangan cadangan. Pendekatan ini terbukti mengungguli metode tradisional dengan mengurangi biaya operasional secara signifikan sambil tetap mempertahankan keandalan yang tinggi. Pendekatan ini memiliki efisiensi komputasi yang tinggi, keandalan yang lebih baik, dan adaptasi dengan berbagai biaya cadangan, namun kompleksitas dari model MILP dan ketergantungan pada data masukan yang akurat. Untuk itu, para peneliti merekomendasikan untuk menyempurnakan model prediksi dengan algoritma yang lebih canggih.

Penelitian terdahulu menyoroti pentingnya menggunakan strategi pengambilan keputusan yang berbasis data dalam menghadapi ketidakpastian diberbagai sektor, termasuk pertanian. Untuk itu, penelitian ini bertujuan untuk mengembangkan sebuah sistem pendukung keputusan (DSS) yang mengintegrasikan model prediksi dan optimasi. Tujuannya adalah untuk memfasilitasi perumusan kebijakan berbasis data terkait cadangan persediaan beras pemerintah. Tabel 2.4 menyajikan perbandingan penelitian terdahulu terkait DSS, prediksi, dan optimasi.

Tabel 2. 4 Ringkasan Penelitian Terdahulu tentang DSS, Prediksi, dan Optimasi

Penelitian	Tujuan	DSS*	Prediksi	Gen AI**	ML**	Optimasi	Metode	Variabel Keputusan	Parameter	Hasil	Kelebihan	Kekurangan
Dellino, G., Laudadio, T., Mari, R., Mastronardi, N., & Meloni, C. (2018)	Mengembangkan DSS dengan mengintegrasikan prediksi penjualan dan perencanaan pesanan untuk makanan segar	✓	✓	x	x	✓	ARIMA, ARIMAX, model fungsi transfer, dan optimasi multi-objektif	Rencana pengadaan	Permintaan, kesegaran produk, tingkat persediaan, jadwal pengiriman dan <i>lead time</i> , ketidakpastian persediaan, biaya, serta promosi dan acara khusus	Sistem berhasil menghasilkan prediksi yang akurat dan usulan pesanan yang seimbang, yang dapat menyesuaikan berbagai kebutuhan dan preferensi pengguna	Integrasi <i>seamless</i> antara prediksi dan perencanaan pesanan, serta fleksibilitas dalam pemilihan model berdasarkan kriteria yang ditetapkan pengguna	Memiliki kompleksitas tinggi dalam implementasi dan penggunaannya, serta skalabilitas sistem belum sepenuhnya dieksplorasi
Namany, S., Govindan, R., Alfagih, L., McKay, G., & Al-Ansari, T. (2020)	Merancang model pengambilan keputusan dinamis yang inovatif untuk sektor pangan di Qatar	✓	x	x	x	x	Agent-Based Modeling (ABM)	Grow, import, hybrid	<i>Demand, price, CRW</i> , kapasitas produksi dan impor, dan <i>forward contract</i>	Skenario dasar merekomendasikan gabungan produksi lokal dan impor dengan fokus pada impor. Skenario CWR menyarankan impor sepenuhnya untuk menghemat	Model ini dapat disesuaikan dengan berbagai skenario dan kondisi pasar, dapat mempertimbangkan faktor ekonomi dan lingkungan, dapat menangkap perubahan dari waktu ke waktu, dapat membantu pembuat	Akurasi model bergantung pada kualitas dan ketersediaan data, model ABM bisa menjadi kompleks dan sulit untuk diinterpretasikan, dan model ini mungkin membuat asumsi penyederhanaan untuk mengurangi kompleksitas

Penelitian	Tujuan	DSS*	Prediksi	Gen AI**	ML**	Optimasi	Metode	Variabel Keputusan	Parameter	Hasil	Kelebihan	Kekurangan
										air. Skenario harga mendorong produksi lokal karena lebih ekonomis	kebijakan memahami implikasi	
Teerasoponpong, S., & Sopadang, A. (2022)	Mengusulkan DSS untuk pengadaan dan manajemen persediaan untuk UKM	✓	✓	x	✓	✓	Artificial Neural Network (ANN) dan Genetic Algorithm (GA)	Jumlah pesanan bahan baku tertentu dari pemasok tertentu	Jumlah pesanan dan persediaan, biaya, penggunaan dan permintaan, waktu tunggu, tingkat layanan, diskon dan insentif, serta ketersediaan dan kapasitas	DSS dapat memberikan solusi pendukung keputusan yang mengintegrasikan pemilihan pemasok yang direkomendasikan dan jumlah pesanan yang optimal	DSS dapat beradaptasi dengan berbagai parameter dan kondisi ketidakpastian	Kualitas dan ketersediaan data dapat mempengaruhi kinerja DSS

Penelitian	Tujuan	DSS*	Prediksi	Gen AI**	ML**	Optimasi	Metode	Variabel Keputusan	Parameter	Hasil	Kelebihan	Kekurangan
Chen, H., Chen, Z., Lin, F., & Zhuang, P. (2021)	Mengusulkan model untuk mengoptimalkan keputusan produksi dan persediaan untuk mencapai keuntungan yang optimal	✓	x	x	✓	✓	Algoritma DQN	Kuantitas dan waktu pengiriman	Harga pasar, biaya transportasi dan operasional, dan tingkat permintaan	Model terbukti mampu mencapai keuntungan yang lebih tinggi dibandingkan dengan metode heuristik dan Q-learning dalam berbagai skenario ASC	Membantu meningkatkan keuntungan petani melalui pengambilan keputusan yang optimal dan dapat beradaptasi dengan perubahan dalam ASC	Algoritma yang digunakan menerapkan mekanisme <i>replay memory</i> untuk memfasilitasi konvergensi, namun <i>experience data</i> dalam <i>replay memory</i> menunjukkan relevansi yang kuat, yang dapat menyebabkan rendahnya efisiensi pelatihan untuk mencapai kinerja yang optimal
C. Zhao et al. (2021)	Mengusulkan pendekatan yang terintegrasi antara probabilitas dan optimasi yang kuat untuk kuantifikasi cadangan operasi berdasarkan interval	✓	✓	x	✓	✓	Extreme Learning Machine (ELM) dan Mixed Integer Linear Programming (MILP)	Jumlah cadangan energi	Kapasitas tenaga angin, target tenaga angin, dan biaya denda untuk kekurangan cadangan	Pendekatan yang diusulkan mengungguli pendekatan tradisional dengan secara signifikan mengurangi biaya operasional sambil mempertahankan keandalan	Efisiensi komputasi tinggi, keandalan yang lebih baik, dan adaptasi dengan berbagai biaya cadangan	Kompleksitas model MILP dan ketergantungan pada data masukan yang akurat

Penelitian	Tujuan	DSS*	Prediksi	Gen AI**	ML**	Optimasi	Metode	Variabel Keputusan	Parameter	Hasil	Kelebihan	Kekurangan
	prediksi daya angin									yang tinggi		
Penelitian Usulan (2024)	Mengusulkan DSS yang mengintegrasikan model prediksi dan optimasi untuk manajemen cadangan beras pemerintah	✓	✓	✓	✓	✓	GAN dan <i>machine learning</i>	Jumlah CBP	Data total produksi, data permintaan beras, data persediaan aktual, dan lainnya			

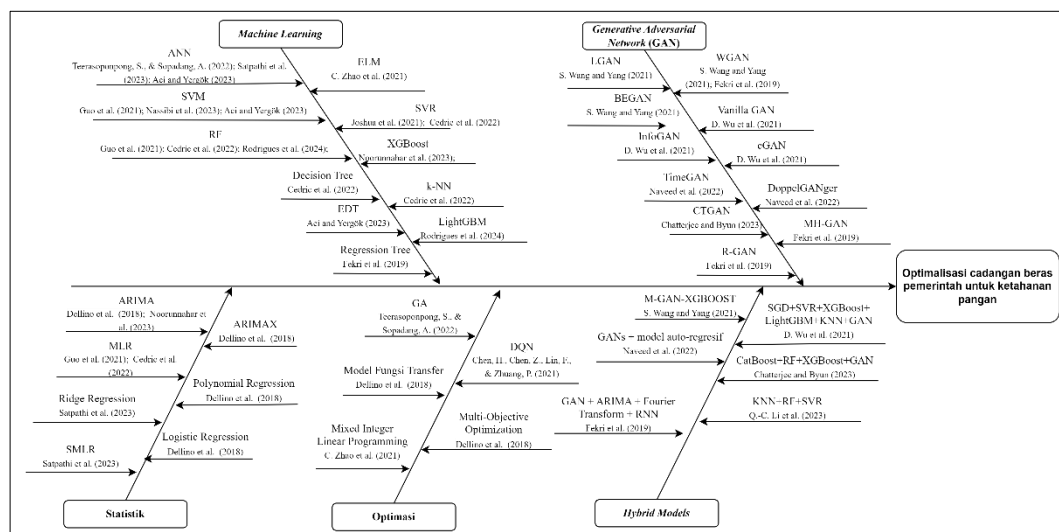
*DSS=*Decision Support System*

**Gen AI=*Generative AI*

***ML=*Machine Learning*

2.5.5 Fishbone Diagram

Penelitian terdahulu telah menunjukkan berbagai pendekatan dan metode yang berbeda untuk memprediksi produksi dan permintaan, optimasi persediaan, serta sistem Pendukung Keputusan (DSS) di berbagai sektor. Masing-masing model tersebut memiliki keterbatasan yang perlu diatasi, seperti bergantung pada kualitas data dan *overfitting*. Di sisi lain, integrasi teknologi Gen AI dan pembelajaran mesin dalam sektor pertanian dan pangan masih tergolong baru dan belum banyak dieksplorasi. Penelitian ini bertujuan untuk mengembangkan model yang lebih adaptif, yang mampu mengatasi fluktuasi dalam produksi dan permintaan, serta memberikan rekomendasi untuk jumlah cadangan beras pemerintah yang optimal. Penelitian ini juga akan membangun DSS yang akan memberikan rekomendasi strategi pengadaan beras yang efektif untuk menjamin ketahanan pangan. Penelitian ini akan menggunakan teknologi Gen AI dan ML dengan mempertimbangkan variabel-variabel yang mempengaruhi manajemen persediaan cadangan beras pemerintah. Gambar 2.5 memperlihatkan *fishbone diagram* yang menggambarkan usulan penelitian ini.



Gambar 2. 5 Fishbone Diagram Usulan Penelitian

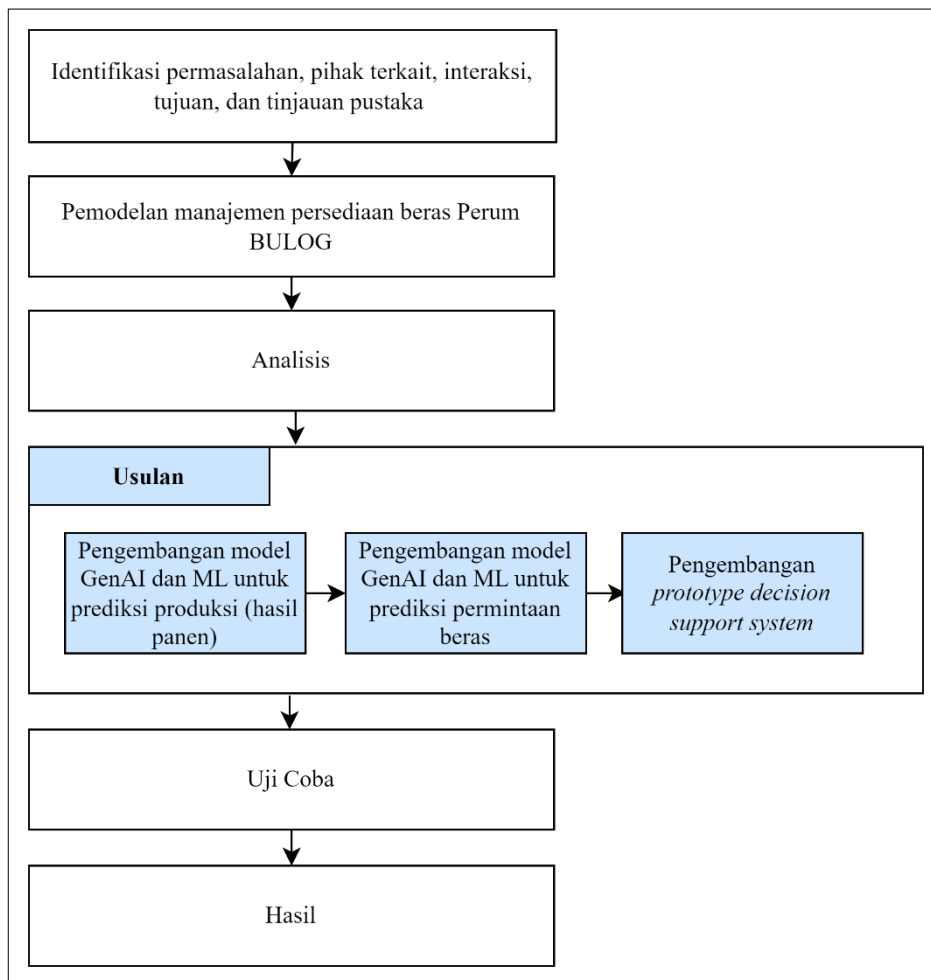
BAB 3

METODE PENELITIAN

3.1 Tahapan Penelitian

Tahapan penelitian merupakan gambaran dari langkah-langkah atau proses yang akan dilakukan dalam suatu penelitian. Penelitian ini terdiri dari lima tahapan. Tahap pertama adalah mengidentifikasi permasalahan, pihak terkait, interaksi, tujuan, dan tinjauan pustaka. Tahapan kedua adalah membuat model manajemen persediaan beras Perum BULOG berdasarkan hasil wawancara awal dengan pihak terkait. Tahap ketiga adalah melakukan analisis terhadap model manajemen persediaan beras Perum BULOG untuk mengidentifikasi area yang perlu ditingkatkan. Hasil analisis ini akan digunakan untuk merumuskan solusi terhadap permasalahan yang ada.

Tahap keempat adalah pengembangan solusi berbasis teknologi yang terdiri dari pengembangan berbagai model dan *prototype* sistem yang akan diuji. Usulan yang pertama adalah model *generative* AI untuk menghasilkan data sintetis yang realistis yang dapat digunakan sebagai data pelatihan untuk model prediksi. Model ini kemudian diintegrasikan ke dalam model ML prediksi produksi (hasil panen). Usulan yang kedua adalah model prediksi permintaan beras yang merupakan model yang mirip dengan model prediksi produksi (hasil panen) beras dengan beberapa penyesuaian agar sesuai dengan karakteristik data untuk prediksi permintaan beras. Usulan yang ketiga adalah pengembangan *prototype decision support system* yang mengintegrasikan model prediksi dan optimasi untuk mendukung kebijakan terkait pengadaan cadangan beras. Tahap kelima adalah uji coba terhadap *prototype decision support system*. Gambar 3.1 adalah tahapan pada penelitian ini.



Gambar 3. 1 Tahapan Penelitian

3.2 Pemodelan Manajemen Persediaan Beras Perum BULOG

Manajemen persediaan cadangan beras nasional telah menjadi perhatian penting dalam beberapa tahun terakhir karena meningkatnya permintaan pangan global, perubahan iklim, dan ketidakstabilan ekonomi. Cadangan ini merupakan stok strategis yang diawasi oleh pemerintah untuk menstabilkan persediaan dan harga beras, memberikan bantuan saat terjadi kekurangan pangan, dan mendukung tujuan ketahanan pangan nasional. Manajemen persediaan cadangan beras yang efektif sangat penting untuk memitigasi risiko yang terkait dengan gangguan persediaan dan fluktuasi harga beras (Yulianis & Rachman, 2021), yang pada akhirnya akan menjamin ketahanan pangan dan stabilitas ekonomi (Octania, 2021; Usdianto & Setiyowati, 2023).

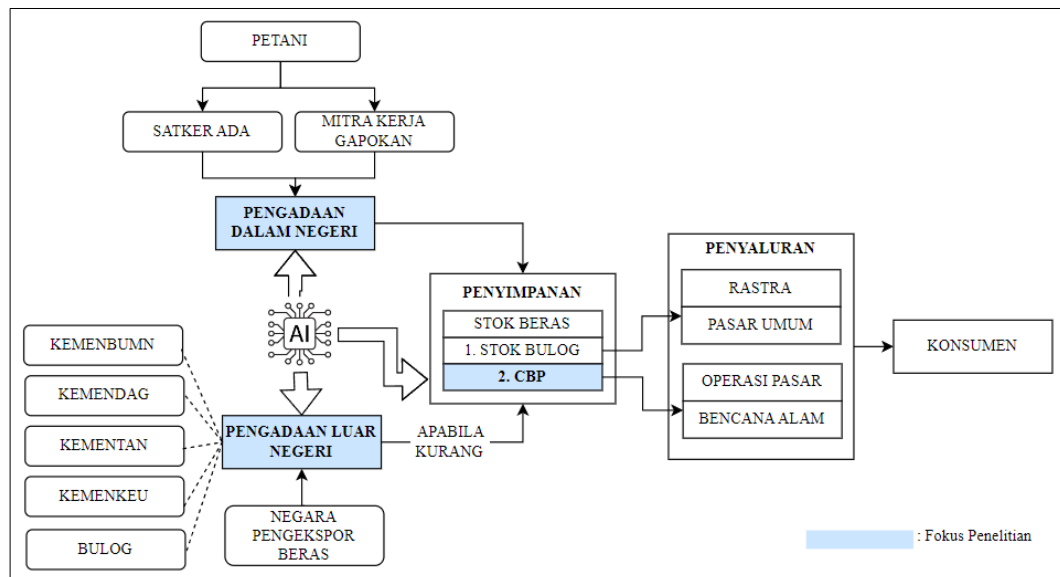
Para pemangku kepentingan yang terlibat dalam manajemen cadangan beras ini termasuk Kementerian Pertanian, Kementerian Perdagangan, Kementerian Keuangan, dan Kementerian Badan Usaha Milik Negara sebagai regulator, serta Perum BULOG yang bertanggung jawab untuk mengelola persediaan beras pemerintah dan stabilisasi harga di tingkat produsen dan konsumen (Octania, 2021). Perum BULOG bertanggung jawab untuk menjaga stabilitas harga dengan membeli gabah dan beras dari petani dengan harga yang ditentukan pemerintah ketika harga beli gabah turun, sehingga melindungi petani dari kerugian, dan menjual beras dengan harga yang lebih rendah daripada harga pasar ketika terjadi kenaikan harga beras, untuk memastikan keterjangkauan harga beras bagi masyarakat (Octania, 2021). Lembaga ini bertanggung jawab atas manajemen salah satu komponen cadangan beras nasional, yaitu Cadangan Beras Pemerintah (CBP), termasuk pada pengadaan (dalam negeri dan impor), penyimpanan, dan penyaluran beras untuk kebutuhan stabilisasi harga, bantuan pangan, dan keadaan darurat (Fang, Chen, Zhang, Pei, Gao, & Wang, 2020; Octania, 2021).

Beberapa penelitian telah menekankan peran penting Perum BULOG dalam manajemen persediaan cadangan beras di Indonesia. Melalui manajemen CBP, Perum BULOG memainkan peran penting dalam menjaga ketahanan pangan nasional, terutama saat terjadi fluktuasi harga atau gangguan persediaan. Keberadaan CBP yang dikelola Perum BULOG tidak hanya menstabilkan harga beras di pasar, tetapi juga menjamin ketersediaan beras bagi masyarakat, sehingga berkontribusi terhadap stabilitas ekonomi nasional (Octania, 2021; Putro, Purwaningsih, Sensuse, & Suryono, 2022; Silalahi et al., 2019). Mengingat peran penting ini, penerapan teknologi AI dapat membantu Perum BULOG dalam mengoptimalkan berbagai aspek dalam manajemen cadangan beras pemerintah, seperti prediksi permintaan dan produksi (hasil panen), optimasi cadangan beras, dan pengambilan keputusan yang lebih baik.

AI dapat digunakan untuk menganalisis data historis dan *real-time* guna menghasilkan prediksi yang akurat mengenai permintaan dan produksi (hasil panen) beras (Mehmood et al., 2023; Rai et al., 2021), sehingga memungkinkan

Perum BULOG untuk mengoptimalkan cadangan beras, menghindari kelebihan atau kekurangan cadangan beras, dan mengambil keputusan yang lebih baik dalam manajemen cadangan beras pemerintah (H. Qin, 2023).

Berdasarkan kajian model manajemen persediaan beras Perum BULOG, maka penelitian ini berfokus pada pemanfaatan teknologi AI untuk efektivitas manajemen cadangan beras pemerintah, terutama pada proses pengadaan. Gambar 3.1 menggambarkan model manajemen persediaan beras Perum BULOG.



Gambar 3. 2 Model Manajemen Persediaan Perum BULOG

3.3. Analisis

Analisis ini bertujuan untuk mengungkap kelemahan dan proses yang kompleks dalam manajemen persediaan cadangan beras pemerintah di Perum BULOG. Penelitian ini menggunakan metode analisis SWOT untuk mengidentifikasi titik-titik lemah yang krusial dalam manajemen persediaan cadangan beras pemerintah dan mengembangkan strategi untuk meningkatkan efisiensi dan efektivitas pengelolaan persediaan cadangan beras di Indonesia

Analisis SWOT bertujuan untuk mengetahui kekuatan, kelemahan, peluang, dan ancaman bisnis. Analisis lingkungan internal difokuskan untuk mengetahui kekuatan dan kelemahan, sedangkan analisis lingkungan eksternal difokuskan untuk mengetahui peluang dan ancaman (Putra, Pujangkoro, & Situmorang, 2022).

1. Analisis SWOT

1. *Strengths* (Kekuatan):

S1: Dukungan Pemerintah

Perum BULOG didukung oleh berbagai kebijakan pemerintah yang bertujuan menjaga stabilitas harga dan ketersediaan beras, seperti yang diatur dalam UU No. 18 Tahun 2012 tentang Pangan dan Perpres No. 48 Tahun 2016 tentang penugasan kepada Perum BULOG. Hal ini memberikan akses terhadap dukungan kebijakan dan finansial yang kuat, termasuk alokasi anggaran khusus untuk pengadaan CBP (Anggraini, Faqih, Sangadji, Kadarisman, & Revany, 2021; Octania, 2021; Utomo, 2020).

S2: Infrastruktur Logistik yang Memadai

Perum BULOG memiliki infrastruktur logistik yang cukup baik, termasuk gudang penyimpanan yang tersebar di berbagai daerah, yang berperan penting dalam menjaga kecukupan persediaan cadangan beras (Octania, 2021; Utomo, 2020)

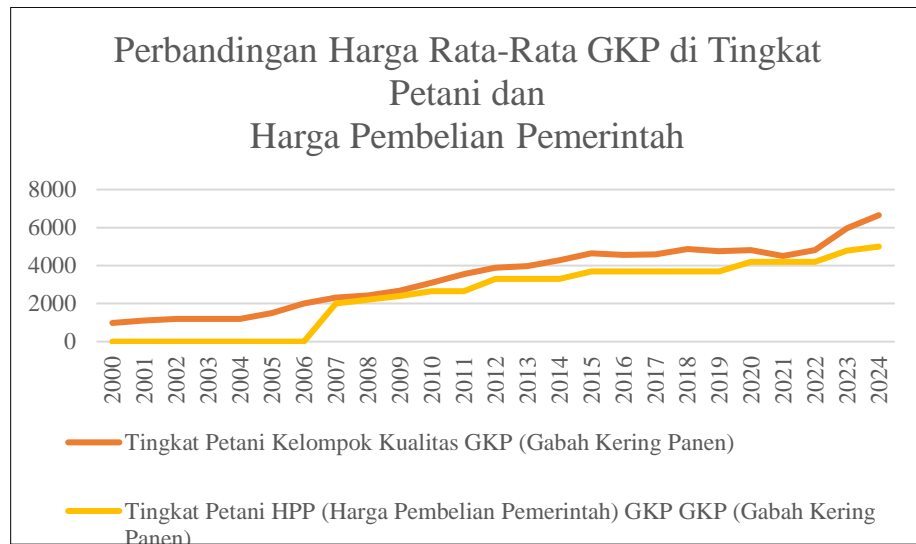
S3: Pengalaman dan Keahlian

Perum BULOG memiliki pengalaman puluhan tahun dalam manajemen persediaan beras, mulai dari pengadaan hingga distribusi, yang menjadi keunggulan dalam menjaga stabilitas harga dan persediaan beras (Anggraini et al., 2021; Utomo, 2020).

2. *Weaknesses* (Kelemahan):

W1: Ketidakmampuan untuk Bersaing dengan Sektor Swasta

Perum BULOG sering menghadapi tantangan dalam bersaing dengan sektor swasta, yang mampu menawarkan harga lebih tinggi kepada petani (Octania, 2021). Gambar 3.3 menggambarkan perbedaan yang signifikan antara harga gabah di tingkat petani dengan harga beli yang ditetapkan pemerintah. Kondisi ini menyebabkan petani lebih memilih menjual hasil panennya ke sektor swasta.



Gambar 3.3 Perbandingan Harga Gabah (BPS, 2023)

W2: Ketergantungan pada Impor

Meskipun Perum BULOG memprioritaskan pengadaan dalam negeri untuk memenuhi CBP, tetapi masih terdapat ketergantungan pada impor beras, terutama selama periode penurunan produksi dalam negeri, yang membuat persediaan CBP rentan terhadap fluktuasi harga dan kebijakan perdagangan internasional (Octania, 2021; Utomo, 2020). Tabel 3.1 dan Gambar 3.4 menunjukkan banyaknya jumlah impor beras yang dilakukan Perum BULOG setiap tahunnya.

Tabel 3.1 Jumlah Impor Beras (BPS, 2024)

Negara Asal	2017	2018	2019	2020	2021	2022	2023
	Berat Bersih: Ton						
India	32209.7	337999	7973.3	10594.4	215386.46	178533.57	69715.7
Thailand	108944.8	795600.1	53278	88593.1	69360.037	80182.506	1381921.2
Vietnam	16599.9	767180.9	33133.1	88716.4	65692.874	81828.039	1147705.3
Pakistan	87500	310990	182564.9	110516.5	52479.011	84407	309309.7
Myanmar	57475	41820	166700.6	57841.4	3790	3830	141204
Jepang	72.1	0.2	90	0.3	230.291	56.087	61.5
Tiongkok	2419	227.7	24.3	23.8	42.601	6	7
Lainnya	54.3	6.5ssss	744.6	0.3	760.146	364.065	12933.3
Total	305274.8	2253824.4	444508.8	356286.2	407741.42	429207.27	3062857.6



**Gambar 3. 4 Jumlah Impor Beras
(BPS, 2024)**

W3: Manajemen Persediaan yang Tidak Efisien

Terdapat ketidakmampuan dalam memprediksi permintaan dan persediaan secara akurat dan *real-time*, yang menyebabkan kelebihan dan kekurangan persediaan (Anggraini et al., 2021; Utomo, 2020).

W4: Keterbatasan Teknologi dan Transparansi

Sistem yang ada saat ini tidak memiliki kapasitas untuk mencatat transaksi secara transparan, yang mengakibatkan meningkatnya risiko kebocoran dan penipuan di sepanjang rantai pasok (Anggraini et al., 2021; Utomo, 2020).

W5: Koordinasi Antar Lembaga

Koordinasi antara Perum BULOG dan lembaga pemerintah lainnya, seperti Kementerian Pertanian dan Kementerian Perdagangan, terkadang menghadapi tantangan karena adanya perbedaan data, sehingga berdampak pada kelancaran pengambilan keputusan kebijakan impor (Siahaan, 2023).

3. *Opportunities* (Peluang):

O1: Memanfaatkan Teknologi Canggih

Integrasi teknologi canggih seperti *blockchain* dan AI berpotensi meningkatkan presisi, transparansi, dan efisiensi dalam mengelola persediaan CBP (Anggraini et al., 2021; Putro et al., 2022; Utomo, 2020).

O2: Pengembangan Sistem yang Terintegrasi

Terdapat peluang untuk mengembangkan sistem yang lebih terintegrasi dan canggih yang dapat memfasilitasi pengelolaan CBP yang lebih baik (Anggraini et al., 2021; Putro et al., 2022; Utomo, 2020).

4. *Threats* (Ancaman):

T1: Ketidakstabilan Harga dan Persediaan Global

Ketidakstabilan harga dan persediaan beras di pasar global dapat mempengaruhi kemampuan Perum BULOG untuk mengimpor beras dalam jumlah yang cukup dan dengan harga yang stabil (Octania, 2021).

T2: Dampak Perubahan Iklim

Perubahan iklim yang ekstrim dapat mengganggu produksi beras dalam negeri, sehingga meningkatkan risiko ketidakcukupan CBP dan fluktuasi harga beras di tingkat konsumen (Octania, 2021; Saud, Wang, Fahad, Alharby, Bamagoos, Mjrashi, Alabdallah, AlZahrani, AbdElgawad, & Adnan, 2022).

2. Strategi

Setelah dilakukan analisis SWOT, maka dapat dirumuskan strategi yang dapat digunakan Perum BULOG untuk meningkatkan kekuatan, mengatasi kelemahan, memanfaatkan peluang, dan meminimalkan potensi ancaman. Berikut adalah rincian strategi-strategi tersebut.

a. Strategi S-O (*Leveraging Strengths to Optimize Opportunities*)

SO1: Pemanfaatan Teknologi *Blockchain* untuk Transparansi dan Efisiensi

Memanfaatkan dukungan pemerintah untuk mengadopsi teknologi *blockchain* dalam manajemen pangan yang dapat meningkatkan transparansi, efisiensi, dan keamanan dalam transaksi dan pencatatan. Teknologi ini membantu dalam pencatatan, mengotomatisasi kontrak, memverifikasi transaksi, dan melacak keaslian produk dari produsen ke konsumen (S1, O1).

SO2: Pengembangan Sistem Terintegrasi

Meningkatkan infrastruktur yang sudah ada dengan mengembangkan sistem manajemen data yang terintegrasi dan canggih untuk memperkuat pengelolaan CBP secara lebih efektif (S2, O2).

SO3: Pengembangan Model Prediksi Permintaan dan Produksi (Hasil Panen)

Memanfaatkan AI untuk mengembangkan model prediksi permintaan dan produksi (hasil panen) beras yang akurat, sehingga memungkinkan pengambilan keputusan pengadaan yang lebih tepat untuk menghindari kelebihan atau kekurangan persediaan (S3, O1, O2).

SO4: Optimalisasi Program Peningkatan Produksi Pangan

Memanfaatkan tenaga ahli dan infrastruktur yang ada untuk mendukung program pemerintah dalam meningkatkan produksi pangan lokal, mengurangi ketergantungan impor, serta memperkuat stabilitas harga dan persediaan beras (S1, S3, O3).

b. Strategi S-T (*Using Strengths to Counter Threats*)

ST1: Pengembangan Alat Pendukung Keputusan untuk Pengadaan Impor
Mengembangkan *Decision Support Tool* (DST) dengan menggunakan input dari model prediksi untuk membantu menentukan kebijakan impor beras. DST akan membantu mengidentifikasi jumlah impor yang optimal berdasarkan analisis kebutuhan dan persediaan beras dalam negeri, sehingga mengurangi risiko ketidakstabilan persediaan dan harga beras (S2, T1).

ST2: Optimalisasi Manajemen Krisis dengan Prediksi Produksi (Hasil Panen)

Menerapkan model prediksi hasil panen untuk mempersiapkan dan merespons secara efektif dampak perubahan iklim pada produksi beras. Model prediksi ini dimanfaatkan untuk meningkatkan ketahanan pangan dan kesiapan dalam menghadapi fluktuasi hasil panen yang tidak terduga (S3, T2).

c. Strategi W-O (*Minimizing Weaknesses by Seizing Opportunities*)

WO1: Mengadopsi Teknologi *Blockchain* untuk Meningkatkan Kepercayaan dan Efisiensi

Mengadopsi teknologi *blockchain* untuk mengatasi keterbatasan teknologi saat ini, seperti sistem yang kurang transparan. *Blockchain* akan meningkatkan kepercayaan dan kredibilitas dalam operasi Perum BULOG, memfasilitasi transaksi yang lebih aman dan audit yang dapat diverifikasi (W1, W4, O1).

WO2: Peningkatan Koordinasi Antar Lembaga Melalui Sistem Terintegrasi

Membangun sistem terintegrasi yang melibatkan semua lembaga terkait untuk mengatasi masalah kurangnya koordinasi dan perbedaan data, serta memfasilitasi pengambilan keputusan yang lebih cepat dan akurat (W5, O2).

d. Strategi W-T (*Minimizing Weaknesses and Avoiding Threats*)

WT1: Mengoptimalkan Kebijakan Impor dengan Model Prediksi

Menggunakan model prediksi untuk mengurangi ketergantungan pada impor dengan mengidentifikasi jumlah produksi (hasil panen) dalam negeri yang dapat memenuhi permintaan, sekaligus mengetahui perlu atau tidaknya dilakukan impor (W2, W3, T1).

WT2: Peningkatan Manajemen Persediaan Melalui Analisis Tingkat Lanjut

Meningkatkan sistem manajemen persediaan dengan model prediksi agar lebih responsif terhadap perubahan permintaan dan kondisi darurat, serta mengurangi risiko kekurangan persediaan dan mengatasi fluktuasi harga (W3, T2).

Setelah melakukan analisis SWOT dan merumuskan strategi, maka hasilnya akan dirangkum dalam bentuk matriks SWOT yang terlampir pada Gambar 3.5.

<div style="text-align: center;"> Faktor Internal Faktor Eksternal </div>	Kekuatan (S) 1. Dukungan Pemerintah (S1) 2. Infrastruktur Logistik yang Memadai (S2) 3. Pengalaman dan Keahlian (S3)	Kelemahan (W) 1. Ketidakmampuan untuk Bersaing dengan Sektor Swasta (W1) 2. Ketergantungan pada Impor (W2) 3. Manajemen Persediaan yang Tidak Efisien (W3) 4. Keterbatasan Teknologi dan Transparansi (W4) 5. Kurangnya Koordinasi Antar Lembaga (W5)
	Strategi S-O 1. Pemanfaatan Teknologi <i>Blockchain</i> untuk Transparansi dan Efisiensi (S1, O1) 2. Pengembangan Sistem Terintegrasi (S2, O2) 3. Pengembangan Model Prediksi Permintaan dan Produksi (Hasil Panen) (S3, O1, O2)	Strategi W-O 1. Mengadopsi Teknologi <i>Blockchain</i> untuk Meningkatkan Kepercayaan dan Efisiensi (W1, W4, O1) 2. Peningkatan Koordinasi Antar Lembaga Melalui Sistem Terintegrasi (W5, O2)
Ancaman (T) 1. Ketidakpastian Harga dan Persediaan Global (T1) 2. Dampak Perubahan Iklim (T2)	Strategi S-T 1. Pengembangan Alat Pendukung Keputusan untuk Pengadaan Impor (S2, T1) 2. Optimalisasi Manajemen Krisis dengan Prediksi Produksi (Hasil Panen) (S3, T2)	Strategi W-T 1. Mengoptimalkan Kebijakan Impor dengan Model Prediksi (W2, W3, T1) 2. Peningkatan Manajemen Persediaan Melalui Analisis Tingkat Lanjut (W3, T2)

Gambar 3. 5 Matriks SWOT

Berdasarkan analisis SWOT yang dilakukan terhadap Perum BULOG, telah diidentifikasi beberapa strategi penting yang menjadi fokus penelitian ini. Strategi tersebut meliputi pengembangan model prediksi permintaan dan produksi, optimasi manajemen krisis dengan prediksi produksi, pembuatan alat pendukung keputusan untuk pengadaan, optimasi kebijakan impor dengan model prediksi, dan peningkatan manajemen persediaan melalui analisis lanjutan. Strategi-strategi ini bertujuan untuk meningkatkan kinerja Perum BULOG dalam manajemen Cadangan Beras Pemerintah (CBP) dan dapat memberikan rekomendasi berbasis data untuk mendukung kebijakan pengadaan cadangan beras.

3.3.1 Regulasi Terkait Manajemen Cadangan Beras Pemerintah

Pemerintah telah menetapkan regulasi dalam manajemen cadangan beras untuk menjamin ketahanan pangan dan menangani potensi keadaan darurat.

Regulasi-regulasi ini bertujuan untuk menjamin ketersediaan beras selama keadaan darurat dan memanfaatkan cadangan beras yang dikelola pemerintah. Berikut adalah regulasi-regulasi utama yang berkaitan dengan cadangan beras pemerintah.

1. Peraturan Presiden Nomor 125 Tahun 2022 tentang Penyelenggaraan Cadangan Pangan Pemerintah

Pasal 2 Ayat (1): Pemerintah menetapkan jenis dan jumlah Cadangan Pangan Pemerintah (CPP) untuk menjamin ketersediaan pangan di seluruh wilayah Indonesia.

Pasal 4 Ayat (1): Penugasan kepada badan usaha milik negara untuk mengelola CPP, termasuk Perum BULOG dan/atau BUMN Pangan.

Pasal 8 Ayat (1): Pendanaan untuk penyelenggaraan CPP bersumber dari APBN dan/atau sumber pendanaan lain yang sah dan tidak mengikat.

2. Peraturan Badan Pangan Nasional Nomor 12 Tahun 2022 tentang Penyelenggaraan Cadangan Beras Pemerintah

Pasal 2: Penetapan jumlah CBP dilakukan dengan mempertimbangkan:

- a. Produksi beras dan/atau gabah secara nasional
- b. Penanggulangan keadaan darurat dan kerawanan pangan
- c. Pengendalian dan stabilisasi harga dan pasokan beras dan/atau gabah pada tingkat produsen dan konsumen
- d. Pelaksanaan perjanjian internasional dan bantuan pangan kerja sama internasional
- e. Angka kecukupan gizi yang dianjurkan

Pasal 3 Ayat (1): Penetapan jumlah CBP sebagaimana dimaksud dalam Pasal 2 ditetapkan oleh Kepala Badan.

Pasal 3 Ayat (4): Penetapan jumlah CBP sebagaimana dimaksud pada ayat (1) dilakukan paling sedikit (satu) kali dalam 1 (satu) tahun.

Pasal 4 Ayat (3): Target pengadaan CBP terdiri atas volume pengadaa dalam negeri dan/atau pengadaan luar negeri.

Pasal 5 Ayat (1): Penyelenggaraan CBP dilakukan melalui pengadaan, pengelolaan, dan penyaluran.

Pasal 5 Ayat (2): Penyelenggaraan CBP sebagaimana dimaksud pada ayat (1) dilaksanakan oleh Badan Pangan Nasional melalui penugasan kepada Perum BULOG.

Pasal 10 Ayat (1): Dalam hal pengadaan CBP melalui produksi dalam negeri tidak mencukupi untuk pemenuhan cadangan, menjaga stabilitas harga dalam negeri, dan/atau memenuhi kebutuhan Pemerintah, dapat dilakukan pengadaan CBP dari luar negeri dengan tetap menjaga kepentingan produsen dan konsumen dalam negeri.

3. Peraturan Menteri Perdagangan Nomor 127 Tahun 2018 tentang Pengelolaan Cadangan Beras Pemerintah

Pasal 2 Ayat (1): Pengelolaan cadangan beras pemerintah untuk menjamin ketersediaan pasokan dan stabilisasi harga.

Pasal 5 Ayat (1): Penugasan kepada Perum BULOG untuk melakukan pengadaan beras dari produksi dalam negeri berdasarkan Harga Pembelian Pemerintah (HPP).

4. Peraturan Menteri Koordinator Bidang Perekonomian Republik Indonesia Nomor 5 Tahun 2018 tentang Koordinasi Pengelolaan Cadangan Beras Pemerintah untuk Stabilisasi Harga

Pasal 1 Ayat (3): Rapat koordinasi adalah rapat yang dipimpin oleh Menteri Koordinator Bidang Perekonomian.

Pasal 2 Ayat (2): Rapat koordinasi sebagaimana dimaksud pada ayat (1) paling sedikit melibatkan Menteri Pertanian, Menteri Perdagangan, Menteri Keuangan, Menteri Badan Usaha Milik Negara, dan Direktur Utama Perum BULOG.

5. Keputusan Kepala Badan Pangan Nasional Nomor 379.1/TS.03.03/K/11/2023:

Menetapkan jumlah persediaan minimal berbagai komoditas pangan yang harus dimiliki pemerintah hingga akhir tahun 2024. Untuk beras, jumlah minimal yang harus dimiliki adalah 2,4 juta ton dengan persediaan akhir tahun minimal 1,2 juta ton.

3.3.2 Parameter yang Dipertimbangkan dalam Menentukan Jumlah CBP

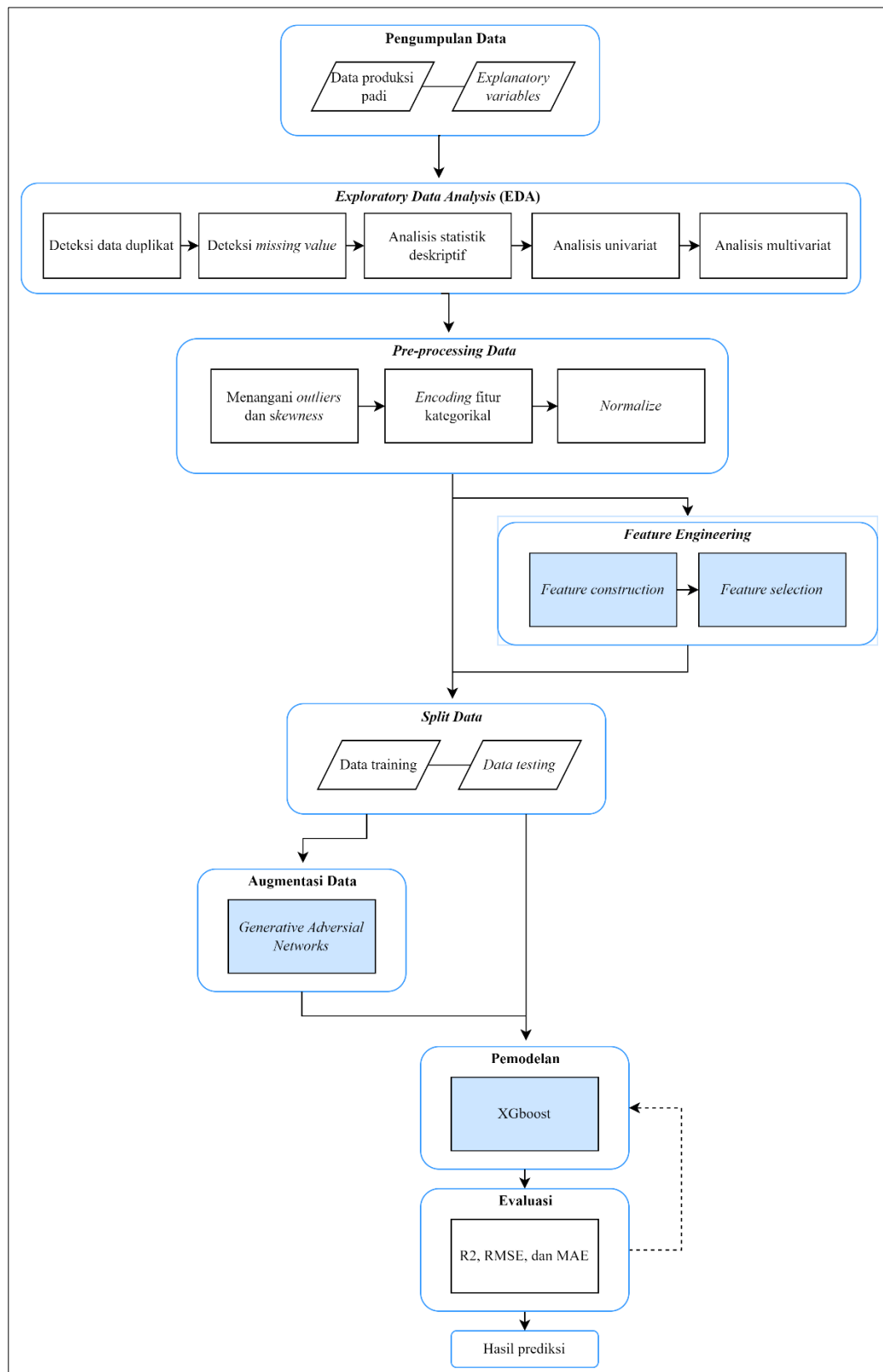
Berdasarkan Pasal 2 dari Peraturan Badan Pangan Nasional Nomor 12 Tahun 2022, terdapat beberapa parameter kunci yang dipertimbangkan dalam penetapan jumlah CBP. Berikut adalah parameter-parameter tersebut:

1. Produksi beras dan/atau gabah secara nasional, merupakan total hasil produksi beras atau gabah di seluruh negeri.
2. Penanggulangan keadaan darurat dan kerawanan pangan, merupakan potensi dan realitas situasi darurat atau krisis pangan yang memerlukan penggunaan cadangan.
3. Pengendalian dan stabilisasi harga dan pasokan beras dan/atau gabah, merupakan kemampuan cadangan untuk mengontrol fluktuasi harga dan pasokan beras di pasar.
4. Pelaksanaan perjanjian internasional dan bantuan pangan kerja sama internasional, merupakan kewajiban dan komitmen internasional yang mempengaruhi jumlah beras yang perlu disimpan.
5. Angka kecukupan gizi yang dianjurkan, merupakan kebutuhan gizi yang harus dipenuhi melalui konsumsi beras oleh populasi.

3.4. Pengembangan Model Prediksi Produksi (Hasil Panen) Beras

Gambar 3.6 menggambarkan delapan tahapan yang dijalani dalam pengembangan model prediksi untuk produksi (hasil panen) beras pada penelitian ini. Setiap tahap dirancang untuk memastikan keakuratan dan efektivitas model dalam memprediksi hasil panen.

Tahap pertama adalah pengumpulan data, yang meliputi data historis produksi padi dan data *explanatory variable*. Tahap kedua adalah eksplorasi data (EDA), yang meliputi penanganan data duplikat, penanganan nilai yang hilang (*missing value*), analisis statistik deskriptif, analisis univariat, dan analisis multivariat.



Gambar 3. 6 Tahapan penelitian prediksi permintaan yang diusulkan

Tahap ketiga adalah *preprocessing data* (pra-pemrosesan), yang meliputi penanganan *outlier* dan *skewness*, pengkodean (*encoding*) fitur kategorikal, serta normalisasi (*normalize*). Tahap keempat adalah *feature engineering* (rekayasa fitur), yang melibatkan konstruksi fitur dan seleksi fitur. Tahap kelima adalah membagi data (*split data*) menjadi data pelatihan (*training*) dan data pengujian (*testing*).

Tahap keenam adalah augmentasi data. Tahap ketujuh adalah pemodelan menggunakan algoritma XGBoost, yang dilatih menggunakan *dataset* pelatihan. Tahap kedelapan adalah evaluasi hasil, yang meliputi pengukuran performa model yang dapat dilihat dari nilai R^2 , *Root Mean Squared Error* (RMSE) dan *Mean Absolute Error* (MAE). Jika performa model perlu ditingkatkan, *hyperparameter* dapat disesuaikan kembali di tahap ketujuh. Setelah performa terbaik tercapai, model dapat diintegrasikan ke dalam model *decision support system*.

3.4.1 Pengumpulan Data

Pengembangan model prediksi produksi (hasil panen) ini akan menggunakan data yang dikumpulkan dari *website* Badan Pusat Statistik (BPS) dan BMKG. *Dataset* ini berisi variabel provinsi, tahun, produksi, luas panen, curah hujan, kelembapan, dan suhu rata-rata untuk periode 1993-2020. Penjelasan mengenai masing-masing variabel dapat dilihat di Tabel 3.2 dan sampel dari *dataset* ini dapat dilihat di Lampiran 1.

Tabel 3. 2 Deskripsi Variabel

Variabel	Deskripsi	Tipe Data
Provinsi	Informasi mengenai nama provinsi yang menghasilkan padi/beras	<i>Categorical</i>
Tahun	Tahun produksi (panen)	<i>Date</i>
Produksi	Jumlah hasil produksi (hasil panen) dalam satu tahun	<i>Numerical</i>
Luas Panen	Ukuran luas lahan yang dipanen di provinsi tersebut	<i>Numerical</i>
Curah Hujan	Curah hujan yang diterima provinsi tersebut dalam satu tahun	<i>Numerical</i>
Kelembapan	Tingkat kelembapan udara rata-rata di provinsi tersebut dalam satu tahun	<i>Numerical</i>
Suhu Rata-rata	Suhu udara rata-rata di provinsi tersebut dalam satu tahun	<i>Numerical</i>

Dataset yang digunakan dalam penelitian ini mencakup berbagai informasi yang berkaitan dengan hasil panen dari berbagai provinsi. Setiap baris mencakup beberapa variabel yang penting untuk analisis prediksi produksi (hasil panen).

Variabel 'Provinsi' menunjukkan nama provinsi di Indonesia di mana produksi padi/beras diukur, sementara variabel 'Tahun' menunjukkan tahun pengumpulan data produksi padi/beras. Variabel 'Produksi' menunjukkan jumlah total produksi (panen) beras dalam ton yang diproduksi di provinsi tersebut pada tahun tertentu. Variabel 'Luas Panen' menunjukkan luas lahan pertanian di provinsi tersebut yang ditanami padi dan dipanen untuk menghasilkan beras dalam satuan hektar.

Variabel 'Curah Hujan' menunjukkan rata-rata curah hujan dalam milimeter yang diterima oleh provinsi tersebut pada tahun tersebut, sedangkan variabel 'Kelembaban' menampilkan tingkat kelembaban udara rata-rata di provinsi tersebut dalam persentase. Terakhir, variabel 'Suhu Rata-rata' menunjukkan suhu udara rata-rata di provinsi tersebut dalam derajat *Celcius*, yang diukur selama satu tahun.

3.4.2 Exploratory Data Analysis (EDA)

Exploratory Data Analysis (EDA) adalah tahapan penting dalam pemodelan prediksi. Hal ini melibatkan analisis data historis secara menyeluruh, menggunakan statistik deskriptif, visualisasi data univariat dan multivariat seperti plot untuk mengidentifikasi pola, tren, dan korelasi. EDA memainkan peran penting dalam mengelola kualitas data. EDA juga memandu seleksi dan rekayasa fitur, yang sangat penting untuk mengembangkan model prediksi yang akurat (Santhoshkumar & Vanila, 2024).

EDA memungkinkan prediksi produksi (hasil panen) yang tepat dengan membangun fondasi yang kuat untuk pemodelan tingkat lanjut. Penelitian ini menggunakan Python untuk EDA.

a. Deteksi data duplikat

Sangat penting untuk mendeteksi dan menghapus data duplikat untuk menjaga keakuratan analisis dan pengembangan model prediktif. Data duplikat dapat menyebabkan hasil yang bias dan tidak dapat diandalkan (Koumarelas, Jiang, & Naumann, 2020). Pada penelitian ini, tidak ditemukan adanya data duplikat dalam *dataset* yang digunakan.

b. Deteksi nilai yang hilang (*missing value*)

Proses deteksi dan penanganan data hilang (*missing value*) dilakukan untuk memastikan keberlanjutan dalam data deret waktu (*time series*). Hal ini penting untuk prediksi produksi (hasil panen) karena memungkinkan model untuk menangkap pola dan *trend* dengan akurat, atau dengan kata lain tanpa melakukan pengisian nilai nol ini, analisis yang dilakukan kemungkinan besar akan menghasilkan hasil yang tidak akurat atau bias (Vallés-Pérez et al., 2022). Pada penelitian ini tidak dilakukan proses *replace missing value* karena *dataset* yang digunakan tidak memuat *missing value* seperti yang terlihat pada Tabel 3.3.

Tabel 3. 3 Missing value

Variabel	Missing value (%)
Provinsi	0
Tahun	0
Produksi	0
Luas Panen	0
Curah Hujan	0
Kelembapan	0
Suhu Rata-rata	0

c. Analisis statistik deskriptif

Analisis statistik deskriptif dilakukan untuk mendapatkan wawasan awal mengenai properti statistik dari *dataset*. Hal ini termasuk penghitungan rata-rata, median, standar deviasi, dan lainnya yang memberikan gambaran umum tentang bentuk dan distribusi data. Tabel 3.4 adalah ringkasan analisis statistik deskriptif *dataset* hasil panen.

Tabel 3. 4 Ringkasan Statistik

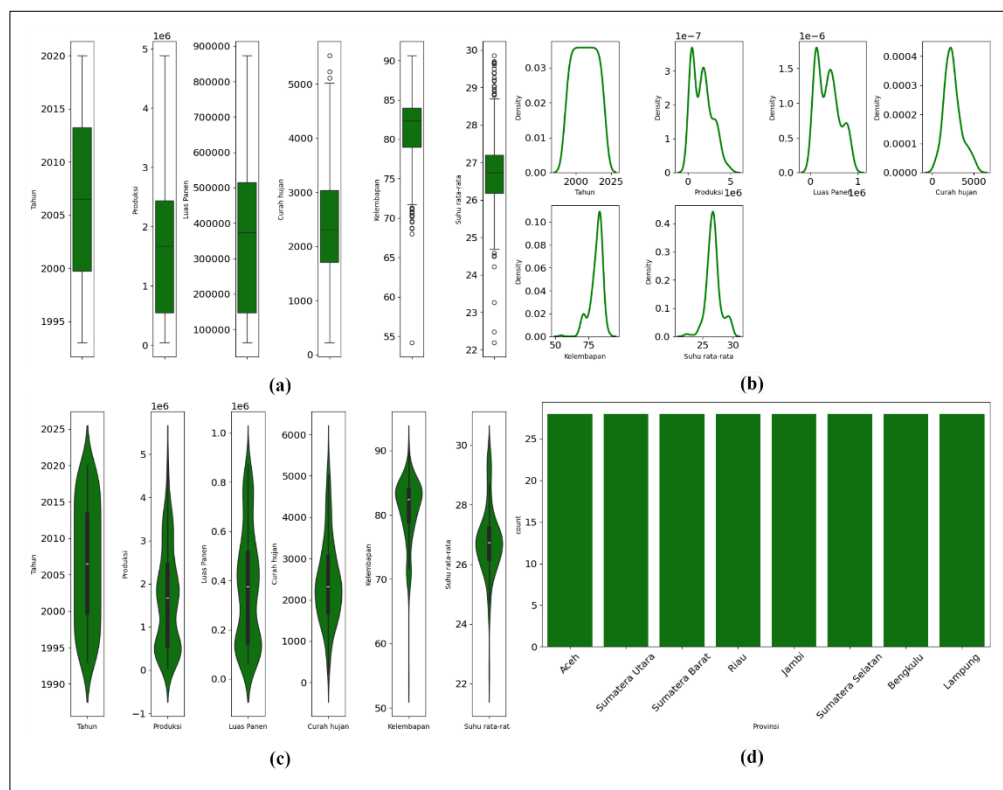
	Tahun	Produksi	Luas Panen	Curah hujan	Kelembapan	Suhu rata-rata
<i>Count</i>	224	224	224	224	224	224
<i>Mean</i>	2006,5	1,679701e+06	374349,96692	2452.490759	80,948705	26,801964
<i>Std</i>	8,095838	1,161387e+06	232751,161987	1031,972625	4,87868	1,197041
<i>Min</i>	1993	4,293800e+04	63142,04	222,5	54,2	22,19
25%	1999,75	5,488570e+05	146919,5	1703,525	78,975	26,1775
50%	2006,5	1,667773e+06	373551,5	2315,7	82,375	26,73
75%	2013,25	2,436851e+06	514570,25	3039,7	84	27,2
<i>Max</i>	2020	4,881089e+06	872737	5522	90,6	29,85

Berdasarkan Tabel 3.4 diketahui bahwa distribusi data cenderung normal karena perbedaan antara nilai *mean* dan *median* tidak terlalu besar. Produksi dan luas panen menunjukkan variasi yang besar dengan adanya *outlier*, yang

ditunjukkan oleh perbedaan antara *mean* dan *median* yang cukup besar. Curah hujan menunjukkan rentang nilai yang luas, namun perbedaan antara *mean* dan *median* relatif kecil. Kelembaban menunjukkan variasi yang cukup stabil dengan perbedaan antara *mean* dan *median* yang tidak besar. Suhu rata-rata menunjukkan data yang konsisten dengan perbedaan minimal antara *mean* dan *median*.

d. Analisis univariat

Analisis univariat dilakukan untuk setiap variabel dalam *dataset*, menggunakan teknik *box plots*, *density plot*, *violin plots*, dan *count plot*. Hal ini membantu dalam memvisualisasikan distribusi, mendeteksi *outliers*, dan memahami variasi dalam data (Shabdin, Ya'acob, & Sjarif, 2020).



Gambar 3. 7 (a) Box Plots, (b) Density Plot, (c) Violin Plots, dan (d) Count Plot

Box Plot memberikan visualisasi distribusi berbasis waktu dari berbagai variabel seperti produksi, area panen, curah hujan, kelembaban, dan suhu rata-rata. Produksi dan luas panen menunjukkan tren yang meningkat dari waktu ke waktu, sementara curah hujan, kelembaban, dan suhu rata-rata

relatif stabil, dengan *outlier* yang terlihat pada variabel curah hujan, kelembaban, dan suhu rata-rata.

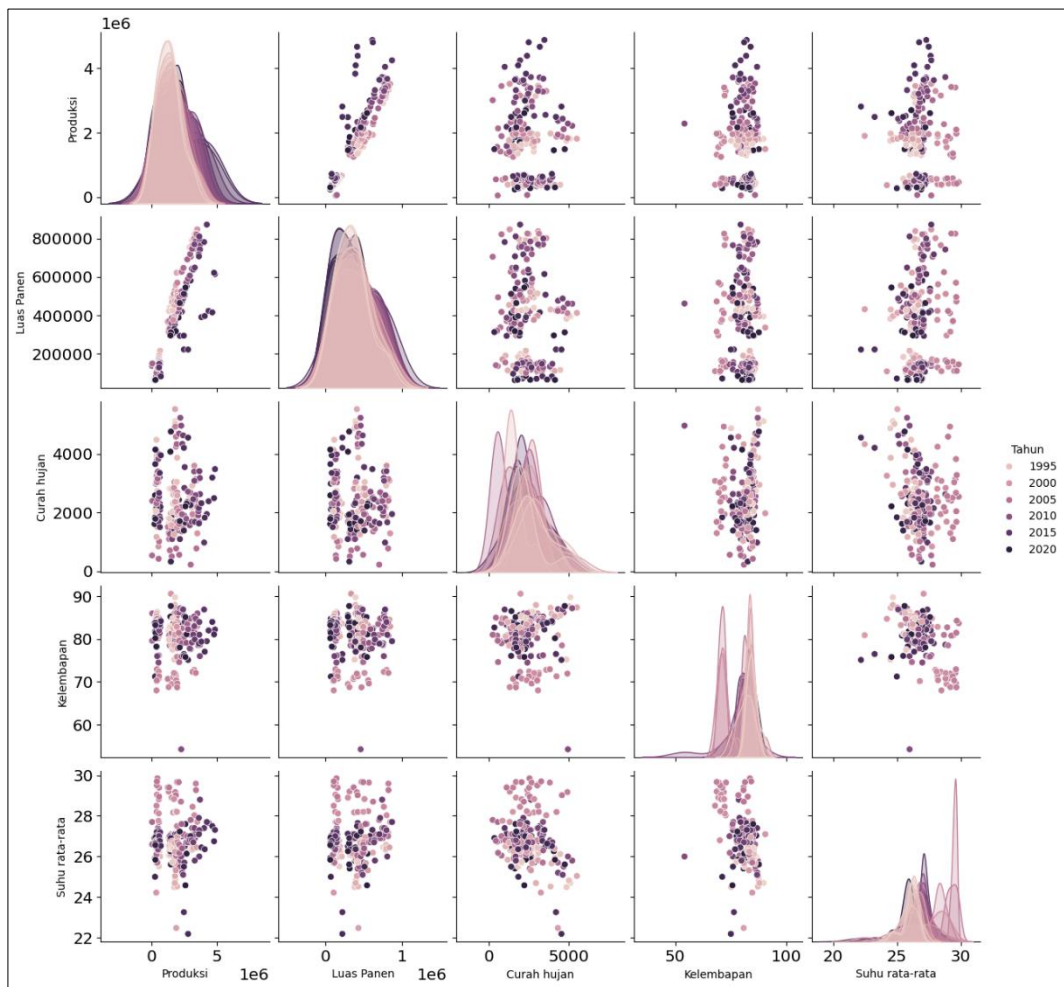
Density Plot menggambarkan distribusi probabilitas dari setiap variabel. Distribusi produksi dan luas panen menunjukkan pola miring ke kanan (*right-skewed*), yang mengindikasikan adanya konsentrasi nilai di batas bawah kisaran dan beberapa nilai yang sangat tinggi. Distribusi curah hujan dan suhu terlihat lebih simetris, sementara distribusi kelembaban menunjukkan bukti bimodalitas dengan dua puncak yang berbeda.

Hal serupa juga terjadi pada *violin plot* yang memberikan wawasan tentang distribusi probabilitas data. Variabel produksi dan luas panen menunjukkan distribusi yang condong ke kanan. Sebaliknya, variabel curah hujan dan suhu menunjukkan distribusi yang lebih simetris, sedangkan variabel kelembaban menunjukkan pola bimodal. Terakhir, *count plot* menunjukkan jumlah pengamatan, terlihat bahwa semua provinsi memiliki jumlah observasi yang sama.

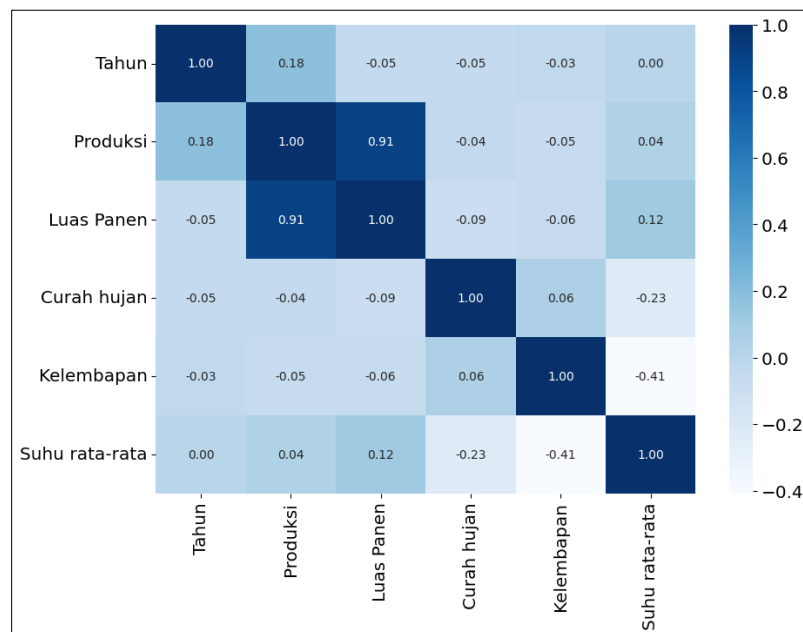
e. Analisis multivariat

Analisis multivariat melibatkan visualisasi data multivariat menggunakan *pair plot* dan *heatmap* korelasi untuk mengeksplorasi hubungan antar variabel. Visualisasi ini membantu mengidentifikasi pola dan hubungan yang mungkin tidak terlihat ketika menganalisis variabel secara terpisah (Ratilal & Reddy, 2023). Gambar 3.8 menunjukkan *pair plot* dan Gambar 3.9 menunjukkan *heatmap* korelasi.

Pair plot menggambarkan distribusi dan hubungan antara variabel, dengan fokus pada korelasi yang kuat antara produksi dan luas panen, yang mengindikasikan bahwa luas panen secara signifikan mempengaruhi tingkat produksi. Hubungan ini diperkuat oleh *heatmap* korelasi, yang menunjukkan koefisien korelasi sebesar 0,91, yang menandakan adanya korelasi yang kuat antara kedua variabel tersebut.



Gambar 3. 8 Pair Plot



Gambar 3. 9 Heatmap korelasi

3.4.3 *Preprocessing Data*

Tahap *preprocessing* dilakukan untuk mempersiapkan *dataset* untuk keperluan analisis *machine learning*. Berdasarkan hasil EDA yang telah dilakukan, maka tahapan *preprocessing* yang dilakukan meliputi penanganan *outlier* dan *skewness*, pengkodean (*encoding*) fitur kategorikal, serta normalisasi (*normalize*). Tahap ini sangat penting untuk menghasilkan model prediksi berbasis *machine learning* yang akurat dan efektif.

a. Penanganan *outliers* dan *skewness*

Berdasarkan analisis *box plots* dan *violin plots*, terlihat bahwa *dataset* yang digunakan mengandung banyak *outlier*, sehingga perlu menerapkan metode untuk mengatasinya. Di sisi lain, analisis *density plot* menunjukkan bahwa terdapat data yang *skew* atau miring pada *dataset* yang dapat mempengaruhi kinerja model. Untuk menangani keduanya, *transformasi log* akan digunakan dalam penelitian ini (Choi et al., 2022).

b. Pengkodean (*encoding*) fitur kategorikal

Untuk memproses variabel input kategorikal dengan cara yang dapat dipahami dan digunakan untuk analisis, maka digunakan representasi *label encoding*. Representasi ini mengubah variabel kategorikal menjadi representasi numerik berdasarkan penomoran kategori secara berurutan, di mana setiap kategori diberi nilai numerik yang unik (Dahouda & Joe, 2021). Pada penelitian ini *label encoding* diterapkan pada variabel 'Provinsi'.

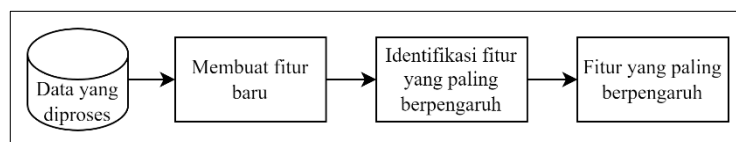
c. Normalisasi (*normalize*)

Pada penelitian ini, proses normalisasi data dilakukan dengan menggunakan metode *Interquartile Range* (IQR) atau *robust normalization*, yang melibatkan penghitungan rentang interkuartil data dan menggunakannya untuk menormalkan data. IQR adalah rentang antara kuartil pertama (Q1) dan kuartil ketiga (Q3), yang mewakili 50% bagian tengah data. Dengan menormalkan data, metode IQR dapat membantu meningkatkan akurasi prediksi (Vaitheeshwari & SathieshKumar, 2019).

3.4.4 *Features Engineering*

Feature engineering dilakukan untuk meningkatkan interpretabilitas model dengan membuat fitur-fitur yang lebih mudah dimengerti. Fitur-fitur yang terstruktur dengan baik dapat membantu dalam menjelaskan alasan dibalik prediksi yang dibuat oleh model. Proses *feature engineering* terdiri dari lima teknik utama, yaitu *feature improvement*, *feature construction*, *feature selection*, *feature extraction*, dan *feature learning* (Ozdemir, 2022).

Pada penelitian ini dilakukan dua perlakuan, dimana model dibangun dengan dan tanpa proses *feature engineering*. Hal ini dilakukan untuk membuktikan hasil penelitian Swaminathan and Venkitasubramony (2023) yang menyatakan bahwa penggunaan *feature engineering* sangat diperlukan untuk pemodelan prediksi untuk menghasilkan prediksi yang lebih akurat dan dapat diandalkan. Selain itu, penggunaan *features engineering*, keterlibatan *explanatory variabel* dan data yang lebih banyak akan menjadi salah satu fokus utama pada penelitian ini. Penelitian ini melibatkan *feature construction* dan *feature selection* seperti yang terlihat pada Gambar 3.10.



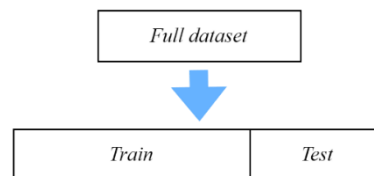
Gambar 3. 10 Flowchart *feature engineering*

Feature construction melibatkan proses pembuatan fitur baru dari data yang sudah ada. Hal ini melibatkan penggabungan data untuk membuat fitur yang lebih informatif dan relevan dengan masalah yang sedang dihadapi (Ozdemir, 2022). Pada penelitian ini dibuat fitur ‘Efisiensi Produksi’ dari ‘Produksi’ dan ‘Luas Panen’.

Seleksi fitur (*feature selection*) berperan penting dalam meningkatkan akurasi dalam model prediksi. Teknik ini melibatkan identifikasi fitur yang paling berpengaruh terhadap hasil panen. Pemilihan fitur yang tepat dapat memfasilitasi prediksi hasil panen padi/beras masa depan. Proses seleksi ini esensial untuk menyesuaikan model prediksi dengan dinamika hasil panen yang berubah-ubah (Bedi, 2023).

3.4.5 Pembagian Data (*Split Data*)

Pada proses pembuatan model *machine Learning*, data dapat dibagi menjadi dua set, yaitu data pelatihan (*training*) dan data pengujian (*testing*). Mesin dilatih dengan data pelatihan dan diuji dengan data pengujian untuk mengevaluasi keandalan dan performanya dalam mencapai tujuan penelitian. Untuk menentukan rasio pembagian data yang optimal, penelitian ini menggunakan tiga rasio yang umum digunakan yaitu 70:30, 80:20, dan 90:10 untuk data pelatihan dan pengujian (Grigorev, 2021; Nguyen, Ly, Ho, Al-Ansari, Le, Tran, Prakash, & Pham, 2021).



Gambar 3. 11 Membagi data menjadi pelatihan dan pengujian (Grigorev, 2021)

3.4.6 Augmentasi Data

Augmentasi data (*Data Augmentation/ DA*) adalah teknik yang digunakan dalam *machine learning* dan *deep learning* untuk meningkatkan ukuran *dataset* pelatihan dengan membuat versi modifikasi dari data yang ada. Proses ini sangat penting untuk meningkatkan performa model, terutama dalam skenario di mana jumlah data pelatihan yang tersedia terbatas (Bayer et al., 2022; Moreno-Barea et al., 2020; Onishi & Meguro, 2023). *Generative AI* (GenAI) akan dimanfaatkan untuk menambah data *training* dalam penelitian ini, khususnya dengan menggunakan *Generative Adversarial Networks* (GAN).

3.4.7 Pemodelan dengan *GridSearchCV*

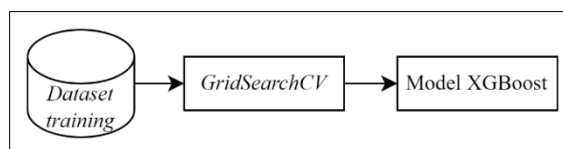
XGBoost merupakan model *machine learning* canggih yang membangun ansambel pohon keputusan secara berurutan, mengoptimalkan *loss function* dan menangani ketidakseimbangan data, *missing value*, serta *overfitting*. Prediksi produksi beras menggunakan algoritma XGBoost membutuhkan penyetingan

beberapa *hyperparameter* sebagai penunjang pemodelan. Proses untuk memilih *hyperparameter* terbaik dalam model *machine learning* untuk meningkatkan kinerja dikenal sebagai penyetelan *hyperparameter* (*tuning*) (Soleymani & Mohammadzadeh, 2023). Model prediksi pada penelitian ini menggunakan parameter *n_estimator*, *max_depth*, dan *learning_rate* yang diadopsi dari penelitian Ørebæk and Geitle (2021). *Range* parameter yang dipertimbangkan pada penelitian ini dapat dilihat di Tabel 3.5.

Tabel 3. 5 Parameter Yang Dipilih Dan Nilainya (Ørebæk & Geitle, 2021)

<i>Hyperparameter</i>	<i>Value Range</i>
<i>Learning_rate</i>	0,1-1
<i>n_estimator</i>	100-250
<i>max_depth</i>	1- 14

Uji coba satu persatu kombinasi yang telah diatur dan ditambah dengan *GridSearchCV* akan mampu menentukan kombinasi mana yang menghasilkan performa terbaiknya. Pada konteks *machine learning*, *10-fold cross-validation* seringkali lebih disukai karena memberikan estimasi yang lebih dapat diandalkan untuk kinerja model, terutama ketika ukuran *dataset* terbatas (Marcot & Hanea, 2021). Untuk itu, pada penelitian ini nilai *CV* disetting 10 yang mana setiap kombinasi model dan parameter divalidasi sebanyak 10 kali dan data akan dibagi menjadi 10 bagian sama besar secara acak, yang mana 9 bagian akan digunakan untuk *training* dan 1 bagian untuk *validasi*. Dari hasil uji coba secara acak, didapatkan kombinasi terbaik yaitu *n_estimator* 115, *learning_rate* 0,28, serta *max_depth* 1. Kombinasi ini yang dimasukkan ke dalam algoritma XGBoost seperti yang terlihat di Gambar 3.12.



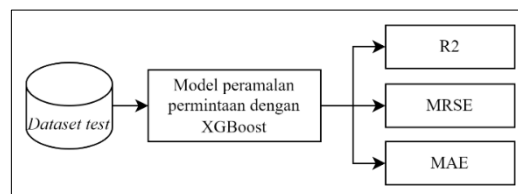
Gambar 3. 12 Flowchart pemodelan

Jika ingin meningkatkan performa model, maka dapat dilakukan penyetelan ulang atau *tuning* model. Proses ini melibatkan penyesuaian *hyperparameter* dan konfigurasi model untuk mencapai hasil yang lebih baik. Dengan penyetelan ulang, dapat mengeksplorasi kombinasi *hyperparameter* yang berbeda atau menggunakan teknik optimasi lanjut, seperti *GridserchCV*

(Liyanage, Basnayake, Gamage, Prabhashi, Kasthuriarachchi, & Abeywardhana, 2023). Meskipun memerlukan upaya yang cukup besar, penyetelan ulang dapat menghasilkan peningkatan signifikan dalam performa prediksi. Untuk itu, dalam konteks penelitian ini, akan terus dilakukan juga evaluasi dan penelitian pada tahap pemodelan akan terus dilakukan untuk memastikan bahwa model sesuai dengan tujuan penelitian.

3.4.8 Evaluasi Model

Setelah model dilatih, model tersebut menjalani tahap evaluasi di mana tingkat kesalahan diuji untuk memastikan model berfungsi dengan baik. Penelitian ini menggunakan *R-squared* (R^2), *Root Mean Squared Error* (RMSE) dan *Mean Absolute Error* (MAE) untuk menguji tingkat kesalahan. R^2 mengukur proporsi varians dalam variabel dan berkisar antara 0 – 1, dimana semakin mendekati 1 maka semakin layak suatu model untuk digunakan. RMSE mengukur besarnya tingkat kesalahan prediksi, dimana semakin rendah nilainya (mendekati nol), maka hasil prediksi akan semakin akurat. Di sisi lain, MAE menghitung rata-rata kesalahan kuadrat antara nilai aktual dan nilai yang diramalkan, dengan nilai yang rendah atau mendekati nol menunjukkan kecocokan yang lebih baik antara hasil permintaan dan data aktual (Chicco et al., 2021). Gambar 3.13 adalah skema evaluasi model pada penelitian ini.



Gambar 3. 13 Flowchart evaluasi model

Pada tahap evaluasi, model juga diuji dengan dua rancangan yang berbeda, yaitu model tanpa *features engineering*, serta pada model dengan *features engineering*. Tabel 3.6 adalah hasil evaluasi model pada penelitian ini.

Tabel 3. 6 Hasil Evaluasi Model

Dataset	Kondisi tanpa <i>feature engineering</i>			Kondisi dengan <i>feature engineering</i>		
	R^2	RMSE	MAE	R^2	RMSE	MAE
70:30	0,773	0,441	0,176	0,933	0,209	0,115
80:20	0,795	0,418	0,155	0,956	0,163	0,099
90:10	0,976	0,111	0,079	0,976	0,109	0,076

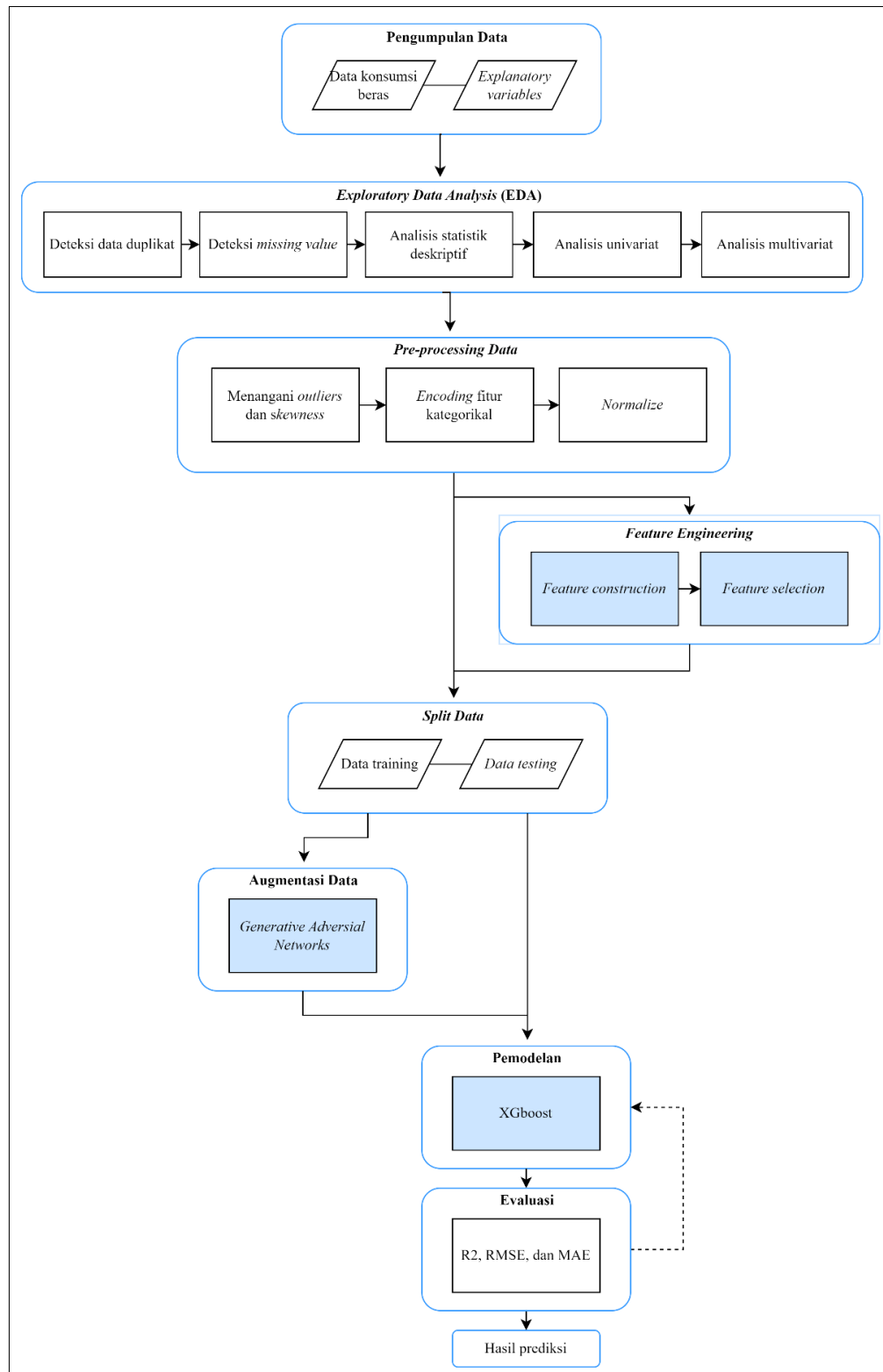
Hasil evaluasi menunjukkan pengaruh positif dari penggunaan *Feature Engineering* (FE) terhadap kinerja model *machine learning* dalam skema pembagian *dataset* yang berbeda (70:30, 80:20, dan 90:10). Khususnya, penggunaan *feature engineering* membantu meningkatkan nilai R^2 , yang menandakan peningkatan kemampuan model dalam menjelaskan variabilitas data yang diamati. Selanjutnya, nilai RMSE dan MAE menurun dengan implementasi FE, mengindikasikan bahwa kesalahan pada model prediksi menjadi lebih kecil dan prediksi menjadi lebih akurat secara umum. Analisis lebih lanjut pada perbedaan skema pembagian data menunjukkan bahwa model dengan proporsi data pelatihan yang lebih besar (90:10) menunjukkan hasil yang paling stabil dan akurat. Hal ini menyoroti pentingnya FE dalam meningkatkan efektivitas model dan menunjukkan keuntungan dari alokasi yang lebih besar pada data pelatihan dalam pengembangan model prediksi berbasis *machine learning*.

Nilai RMSE terendah yang diperoleh adalah 0,109, yang menunjukkan potensi model untuk menghasilkan prediksi produksi yang sangat akurat jika terus dikembangkan. Nosratabadi et al. (2021) mendapatkan nilai RMSE 33.575.595,74 untuk model prediksinya, hal ini menunjukkan bahwa model prediksi yang diusulkan memiliki potensi untuk menyaingi model yang ada dalam hal keakuratan hasil prediksi. Meskipun hal ini tidak mutlak karena model ini dibangun dengan algoritma dan data yang berbeda. Namun, potensi model ini menghasilkan prediksi yang sangat akurat masih terlihat. Berdasarkan hasil ini, penelitian ini akan fokus pada penambahan data pelatihan menggunakan *Generative Adversarial Network* (GAN) dan pengembangan lebih lanjut pada teknik *feature engineering* untuk meningkatkan kinerja prediksi.

3.5 Pengembangan Model Prediksi Permintaan

Model prediksi yang digunakan untuk prediksi permintaan adalah model yang sama dengan yang digunakan untuk memprediksi produksi atau hasil panen. Namun, beberapa penyesuaian dilakukan untuk memastikan bahwa model sesuai dengan karakteristik data permintaan beras, terutama pada tahapan *features engineering*, augmentasi data, dan pemodelan dengan *GridSearchCV*. Gambar

3.14 adalah tahapan pengembangan model prediksi permintaan pada penelitian ini.



Gambar 3. 14 Tahapan penelitian prediksi permintaan yang diusulkan

Pengembangan model prediksi permintaan beras ini terdiri dari delapan tahap. Tahap pertama adalah pengumpulan data, yang meliputi data historis konsumsi beras (Bashir & Yuliana, 2019) dan data *explanatory variable*, seperti harga beras (Iqbal, 2019; Sossou & Igue, 2019), pendapatan per kapita (Bashir & Yuliana, 2019; Yusuf et al., 2020), populasi (Bashir & Yuliana, 2019), rasio kemiskinan (Qian, Ito, & Zhao, 2020), dan peristiwa khusus seperti bulan ramadhan dan hari raya (Hasanah, 2020). Tahap kedua adalah eksplorasi data (EDA), yang meliputi penanganan data duplikat, penanganan nilai yang hilang (*missing value*), analisis statistik deskriptif, analisis univariat, dan analisis multivariat.

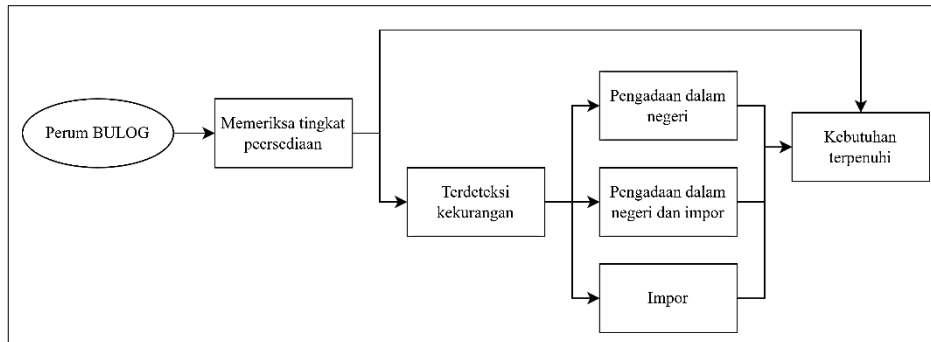
Tahap ketiga adalah *preprocessing data* (pra-pemrosesan), yang meliputi penanganan *outlier* dan *skewness*, pengkodean (*encoding*) fitur kategorikal, dan normalisasi (*normalize*). Tahap keempat adalah *feature engineering* (rekayasa fitur), yang melibatkan konstruksi fitur dan seleksi fitur. Tahap kelima adalah membagi data (*split data*) menjadi data pelatihan (*training*) dan data pengujian (*testing*).

Tahap keenam adalah augmentasi data. Tahap ketujuh adalah pemodelan menggunakan algoritma XGBoost, yang dilatih menggunakan *dataset* pelatihan. Tahap kedelapan adalah evaluasi hasil, yang meliputi pengukuran performa model yang dapat dilihat dari nilai R^2 , *Root Mean Squared Error* (RMSE), dan *Mean Absolute Error* (MAE). Jika performa model perlu ditingkatkan, *hyperparameter* dapat disesuaikan kembali di tahap ketujuh. Setelah performa terbaik tercapai, model dapat diintegrasikan ke dalam model *decision support system*.

3.6 *Prototype Decision Support System (DSS) untuk Pengadaan Cadangan Beras Pemerintah*

Decision Support System (DSS) pada penelitian ini akan digunakan untuk mendukung proses pengambilan keputusan internal terkait kebijakan pengadaan cadangan beras pemerintah. DSS ini akan mengintegrasikan data hasil prediksi produksi (hasil panen), permintaan, dan persediaan aktual untuk menentukan variabel keputusan pengadaan yang optimal. Alat ini memungkinkan Perum

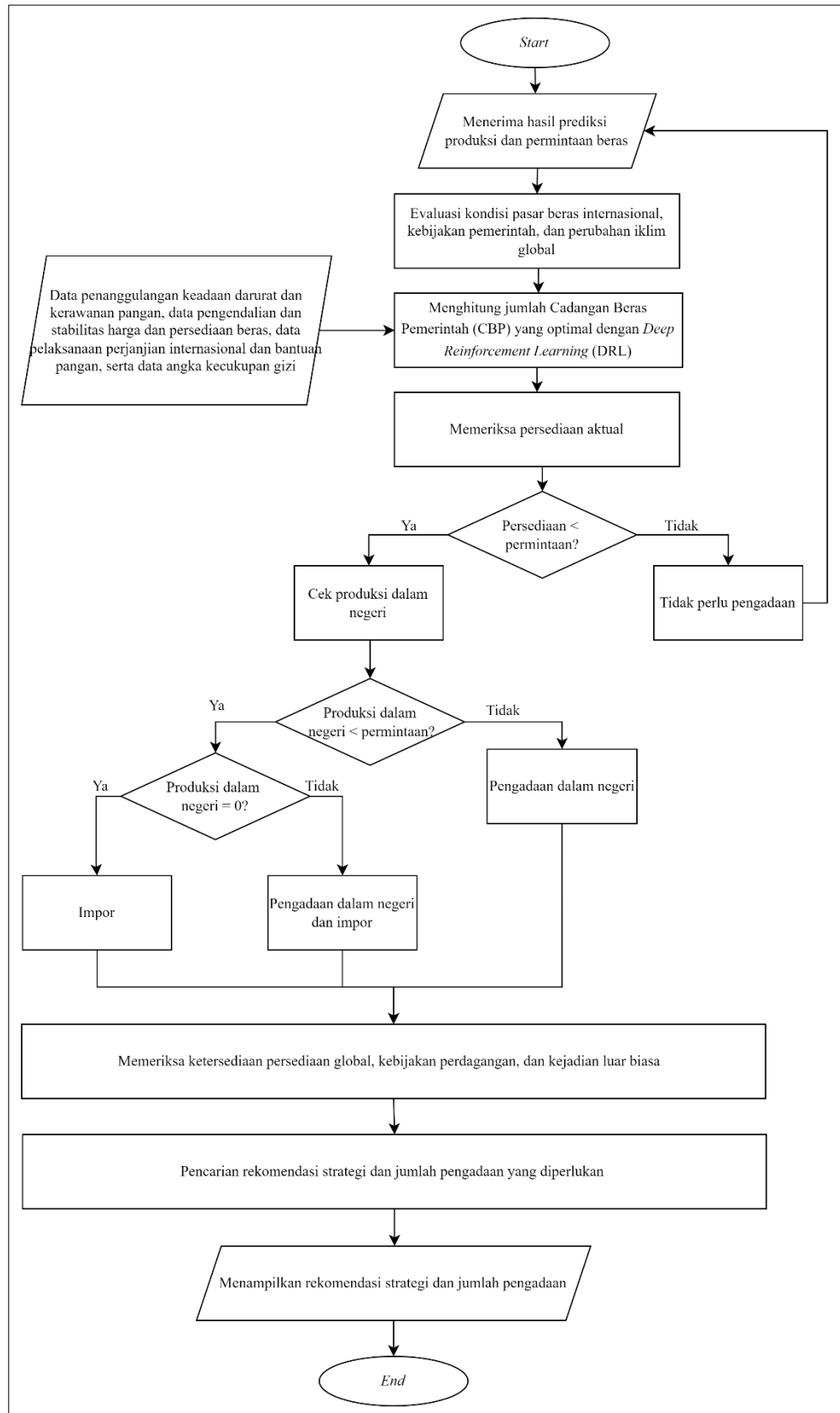
BULOG untuk merencanakan dan mengelola Cadangan Beras Pemerintah (CBP) secara efisien, sesuai dengan kebutuhan masyarakat. Gambar 3.15 adalah ilustrasi dari proses pemenuhan CBP.



Gambar 3. 15 Ilustrasi Pemenuhan CBP

Gambar 3.15 menggambarkan alur proses pemenuhan CBP yang dilakukan oleh Perum BULOG dan lembaga terkait. Proses ini dimulai dengan Perum BULOG memeriksa tingkat persediaan CBP. Jika ditemukan adanya kekurangan persediaan, maka langkah selanjutnya adalah menyusun strategi untuk mengatasi kekurangan tersebut. Terdapat beberapa opsi untuk mengatasi kekurangan tersebut, meliputi pengadaan dalam negeri, kombinasi antara pengadaan dalam negeri dan impor, atau impor penuh jika pengadaan dalam negeri tidak mencukupi.

Gambar 3.16 menguraikan proses pengambilan keputusan untuk mengelola persediaan CBP. Proses ini dimulai dengan penerimaan data tentang prediksi produksi (hasil panen) dalam negeri dan kebutuhan beras mendatang. Selanjutnya, evaluasi kondisi pasar beras internasional, kebijakan pemerintah, dan perubahan iklim global. Lalu dilakukan perhitungan terhadap jumlah Cadangan Beras Pemerintah (CBP) yang optimal, yang menentukan jumlah beras yang harus disimpan untuk kebutuhan darurat. Untuk mencapai optimasi ini, penelitian ini akan menggunakan metode *Deep Reinforcement Learning* (DRL). Berbeda dengan metode heuristik, DRL lebih kuat dengan hasil konvergensi yang stabil dan lebih cocok untuk masalah pengambilan keputusan (Z. Zhang, Zhang, & Qiu, 2019). Tahap berikutnya adalah memeriksa jumlah persediaan aktual.



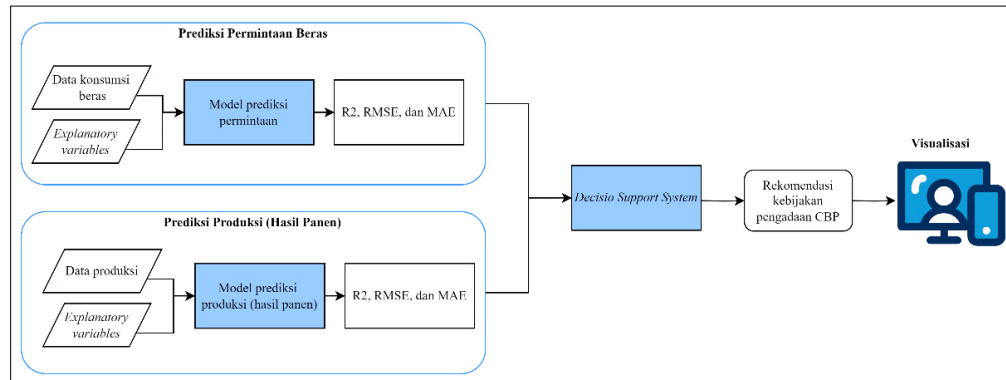
Gambar 3. 16 Diagram Alir Skenario Dasar Pengambilan Keputusan

Berdasarkan data persediaan aktual, dilakukan evaluasi apakah jumlah tersebut sudah mencukupi untuk memenuhi prediksi permintaan atau kebutuhan beras mendatang. Jika persediaan cukup, maka tidak diperlukan tindakan pengadaan tambahan. Sebaliknya, jika persediaan dinilai belum mencukupi kebutuhan, maka akan dilakukan penyerapan atau pengadaan dari produksi (hasil panen) dalam negeri. Pada kondisi di mana produksi (hasil panen) dalam negeri yang ada pada petani atau mitra kerja Perum BULOG sangat minim, maka akan dilakukan pengadaan dari sumber dalam negeri dan impor untuk mengisi kekurangan tersebut. Jika kondisi di mana produksi (hasil panen) dalam negeri kosong.

Penentuan jumlah pengadaan didasarkan pada penghitungan jumlah CBP yang diperlukan untuk memenuhi kekurangan antara kebutuhan dan persediaan yang tersedia dengan turut mempertimbangkan ketersediaan persediaan global, kebijakan perdagangan, dan kejadian luar biasa, seperti bencana alam atau gangguan politik yang bisa mempengaruhi pengadaan.

Output dari DSS ini adalah sebuah rekomendasi untuk strategi dan jumlah pengadaan beras. Rekomendasi ini dapat digunakan oleh Perum BULOG untuk membuat kebijakan pengadaan beras, khususnya dalam menentukan kebutuhan akan impor beras. Keseluruhan proses ini penting untuk memastikan ketersediaan CBP dan menghindari potensi krisis pangan.

DSS yang diusulkan adalah sebuah *platform* yang mengintegrasikan dua model prediksi dan proses optimasi untuk cadangan persediaan beras pemerintah. Sistem ini membantu Perum BULOG dan lembaga terkait dalam mengambil kebijakan strategis terkait pengadaan cadangan beras pemerintah. Hal ini untuk memastikan ketersediaan Cadangan Beras Pemerintah (CBP), sehingga mendukung stabilitas harga dan ketahanan pangan nasional. Gambar 3.17 adalah *prototype* DSS yang diusulkan.



Gambar 3. 17 Prototype DSS yang Diusulkan

3.7 Uji Coba

Uji coba sistem bertujuan untuk memvalidasi apakah sistem yang dikembangkan sesuai dengan tujuan awal dan layak untuk digunakan. Pada tahap ini, *prototype* sistem akan menjalani evaluasi secara menyeluruh untuk memastikan bahwa fungsionalitasnya telah memenuhi standar yang diharapkan dan untuk mengidentifikasi potensi *error* (Sutiah & Supriyono, 2021). Tahap uji coba pada penelitian ini akan menggunakan metode *black box testing*, yang dapat menemukan *error* secara cepat dan efisien di berbagai kategori. Hal ini termasuk fungsi yang salah atau hilang, kesalahan *interface*, masalah dengan struktur data atau akses *database* eksternal, kesalahan kinerja, serta kesalahan yang terkait dengan operasi dan *shutdown* sistem (Corso, Moss, Koren, Lee, & Kochenderfer, 2021).

DAFTAR PUSTAKA

- Abaku, E. A., Edunjobi, T. E., & Odimarha, A. C. (2024). Theoretical approaches to AI in supply chain optimization: Pathways to efficiency and resilience. *J International Journal of Science Technology Research Archive*, Vol. 6 (1), 092-107.
- Abbate, S., Centobelli, P., & Cerchione, R. (2023). The digital and sustainable transition of the agri-food sector. *J Technological Forecasting Social Change*, Vol. 187, 122222.
- Abirami, M. (2023). *AI Clinical Decision Support System (AI-CDSS) for Cardiovascular Diseases*. Paper presented at the 2023 International Conference on Computer Science and Emerging Technologies (CSET).
- Abu Zwaيدا, T., Pham, C., & Beauregard, Y. (2021). Optimization of inventory management to prevent drug shortages in the hospital supply chain. *J Applied Sciences*, Vol. 11 (6), 2726.
- Abukmeil, M., Ferrari, S., Genovese, A., Piuri, V., & Scotti, F. (2021). A survey of unsupervised generative models for exploratory data analysis and representation learning. *J ACM Computing Surveys*, Vol. 54 (5), 1-40.
- Aci, M., & Yergök, D. (2023). Demand Forecasting for Food Production Using Machine Learning Algorithms: A Case Study of University Refectory. *J Tehnički vjesnik*, Vol. 30 (6), 1683-1691.
- Ahmed, S., Raza, B., Hussain, L., Aldweesh, A., Omar, A., Khan, M. S., Eldin, E. T., & Nadim, M. A. (2023). The deep learning resnet101 and ensemble xgboost algorithm with hyperparameters optimization accurately predict the lung cancer. *J Applied Artificial Intelligence*, Vol. 37 (1), 2166222.
- Alfa, B., & Subagyo, S. (2018). Analysis of rice supply chain mechanism with simulation approach. *J IPTEK Journal of Proceedings Series*, (3), 37-44.
- Alpaydin, E. (2020). *Introduction to machine learning*: MIT press.
- Amellal, I., Amellal, A., Seghioeur, H., & Ech-Charrat, M. (2024). An integrated approach for modern supply chain management: Utilizing advanced machine learning models for sentiment analysis, demand forecasting, and probabilistic price prediction. *J Decision Science Letters*, Vol. 13 (1), 237-248.
- Amir, S., Saqib, Z., Khan, M. I., Khan, M. A., Bokhari, S. A., Zaman-ul-Haq, M., & Majid, A. (2020). FARMERS' PERCEPTIONS AND ADAPTATION PRACTICES TO CLIMATE CHANGE IN RAIN-FED AREA: A CASE STUDY FROM DISTRICT CHAKWAL, PAKISTAN. *J Pakistan Journal of Agricultural Sciences*, Vol. 57 (2).
- Anantrasirichai, N., & Bull, D. (2022). Artificial intelligence in the creative industries: a review. *J Artificial intelligence review*, 1-68. doi:<https://doi.org/10.1007/s10462-021-10039-7>
- Anbananthen, K. S. M., Subbiah, S., Chelliah, D., Sivakumar, P., Somasundaram, V., Velshankar, K. H., & Khan, M. A. (2021). An intelligent decision support system for crop yield prediction using hybrid machine learning algorithms. *J FResearch*, Vol. 10.

- Andriansyah, A., Rahmi, A., & Ilyas, I. (2020). *Using System Dynamic Model for Predicting Inventory of Rice Necessity*. Paper presented at the Journal of Physics: Conference Series.
- Anggraini, E., Faqih, A., Sangadji, M. A., Kadarisman, M. I., & Revany, R. (2021). Mewujudkan Keberlanjutan Pasokan Pangan dalam Periode Pandemi Covid-19. *J Policy Brief Pertanian, Kelautan, dan Biosains Tropika*, Vol. 3 (1), 71-77.
- Arsyad, K. M., Yunita, A., Krismartopo, H. M. u., Dimar, A. S., Dewi, K., & Madrinovella, I. (2023). Revealing Insights Through Exploratory Data Analysis on Earthquake Dataset. *J Journal of Science Informatics for Society*, Vol. 1 (1), 1-6.
- Arumugam, K., Naved, M., Shinde, P. P., Leiva-Chauca, O., Huaman-Osorio, A., & Gonzales-Yanac, T. (2023). Multiple disease prediction using Machine learning algorithms. *J Materials Today: Proceedings*, Vol. 80, 3682-3685.
- Barczewski, A., Bezerianos, A., & Boukhelifa, N. (2020). *How domain experts structure their exploratory data analysis: Towards a machine-learned storyline*. Paper presented at the Extended Abstracts of the 2020 CHI Conference on Human Factors in Computing Systems.
- Bashir, A., & Yuliana, S. (2019). Identifying factors influencing rice production and consumption in Indonesia. *J Jurnal Ekonomi Pembangunan: Kajian Masalah Ekonomi dan Pembangunan*, Vol. 19 (2), 172-185.
- Bayer, M., Kaufhold, M.-A., & Reuter, C. (2022). A survey on data augmentation for text classification. *J ACM Computing Surveys*, Vol. 55 (7), 1-39.
- Bedi, T. (2023). A Framework for Predictive Futuristic Trend and Demand Analysis for Reengineering the Fashion Industry.
- Beltran-Peña, A., Rosa, L., & D'Odorico, P. (2020). Global food self-sufficiency in the 21st century under sustainable intensification of agriculture. *J Environmental Research Letters*, Vol. 15 (9), 095004.
- Bendaoud, N. M. M., Farah, N., & Ahmed, S. B. (2021). Comparing generative adversarial networks architectures for electricity demand forecasting. *J Energy Buildings*, Vol. 247, 111152.
- Bloch, L., & Friedrich, C. M. (2020). *Using bayesian optimization to effectively tune random forest and XGBoost hyperparameters for early Alzheimer's disease diagnosis*. Paper presented at the International Conference on Wireless Mobile Communication and Healthcare.
- Böttcher, L., Asikis, T., & Fragkos, I. (2022). Solving inventory management problems with inventory-dynamics-informed neural networks. *J arXiv e-prints*, arXiv: 2201.06126.
- BPS. (2023). *Rata-Rata Harga Gabah Menurut Kelompok Kualitas dan HPP di Tingkat Petani dan Tingkat Penggilingan (Rupiah/kg), 2000-2022*. Retrieved from: <https://www.bps.go.id/id/statistics-table/1/MTU5NiMx/rata-rata-harga-gabah-menurut-kelompok-kualitas-dan-hpp-di-tingkat-petani-dan-tingkat-penggilingan--rupiah-kg---2000-2022.html>

- BPS. (2024). *Impor Beras Menurut Negara Asal Utama, 2017-2023*. Retrieved from: <https://www.bps.go.id/id/statistics-table/1/MTA0MyMx/impor-beras-menurut-negara-asal-utama--2017-2023.html>
- Brüggen, A., Grabner, I., & Sedatole, K. L. (2021). The folly of forecasting: The effects of a disaggregated demand forecasting system on forecast error, forecast positive bias, and inventory levels. *J The Accounting Review*, Vol. 96 (2), 127-152.
- BULOG. (2024). Ketahanan Pangan. <https://www.bulog.co.id/beraspangan/ketahanan-pangan/>
- Cedric, L. S., Adoni, W. Y. H., Aworka, R., Zoueu, J. T., Mutombo, F. K., Krichen, M., & Kimpolo, C. L. M. (2022). Crops yield prediction based on machine learning models: Case of West African countries. *J Smart Agricultural Technology*, Vol. 2, 100049.
- Chatterjee, S., & Byun, Y.-C. (2023). A synthetic data generation technique for enhancement of prediction accuracy of electric vehicles demand. *J Sensors*, Vol. 23 (2), 594.
- Chen, H., Chen, Z., Lin, F., & Zhuang, P. (2021). Effective management for blockchain-based agri-food supply chains using deep reinforcement learning. *J IEEE Access*, Vol. 9, 36008-36018.
- Chen, I.-F., & Lu, C.-J. (2021). Demand forecasting for multichannel fashion retailers by integrating clustering and machine learning algorithms. *J Processes*, Vol. 9 (9), 1578.
- Chen, T., & Guestrin, C. (2016). *Xgboost: A scalable tree boosting system*. Paper presented at the Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining.
- Chicco, D., Warrens, M. J., & Jurman, G. (2021). The coefficient of determination R-squared is more informative than SMAPE, MAE, MAPE, MSE and RMSE in regression analysis evaluation. *J PeerJ Computer Science*, Vol. 7, e623. doi:<https://doi.org/10.7717/peerj-cs.623>
- Choi, G., Buckley, J. P., Kuiper, J. R., & Keil, A. P. (2022). Log-transformation of independent variables: must we? *J Epidemiology*, Vol. 33 (6), 843-853.
- Clapp, J., Moseley, W. G., Burlingame, B., & Termine, P. (2022). The case for a six-dimensional food security framework. *J Food Policy*, Vol. 106, 102164.
- Corso, A., Moss, R., Koren, M., Lee, R., & Kochenderfer, M. (2021). A survey of algorithms for black-box safety validation of cyber-physical systems. *J Journal of Artificial Intelligence Research*, Vol. 72, 377–428-377–428.
- Dahouda, M. K., & Joe, I. (2021). A deep-learned embedding technique for categorical features encoding. *J IEEE Access*, Vol. 9, 114381-114391. doi:<https://doi.org/10.1109/ACCESS.2021.3104357>
- Dalal, S., Onyema, E. M., & Malik, A. (2022). Hybrid XGBoost model with hyperparameter tuning for prediction of liver disease with better accuracy. *J World Journal of Gastroenterology*, Vol. 28 (46), 6551.
- Dellino, G., Laudadio, T., Mari, R., Mastronardi, N., & Meloni, C. J. I. J. o. P. R. (2018). A reliable decision support system for fresh food supply chain

- management. *International Journal of Production Research*, Vol. 56 (4), 1458-1485.
- Deng, G., Han, C., Dreossi, T., Lee, C., & Matteson, D. S. (2022). *Ib-gan: A unified approach for multivariate time series classification under class imbalance*. Paper presented at the Proceedings of the 2022 SIAM International Conference on Data Mining (SDM).
- Fahimnia, B., Arvan, M., Tan, T., & Siemsen, E. (2023). A hidden anchor: The influence of service levels on demand forecasts. *J Journal of Operations Management*, Vol. 69 (5), 856-871.
- Fan, C., Chen, M., Wang, X., Wang, J., & Huang, B. (2021). A review on data preprocessing techniques toward efficient and reliable knowledge discovery from building operational data. *J Frontiers in Energy Research*, Vol. 9, 652801.
- Fang, W., Chen, W., Zhang, W., Pei, J., Gao, W., & Wang, G. (2020). Digital signature scheme for information non-repudiation in blockchain: a state of the art review. *J EURASIP Journal on Wireless Communications Networking*, Vol. 2020, 1-15.
- Farooq, M. S., Uzair, M., Raza, A., Habib, M., Xu, Y., Yousuf, M., Yang, S. H., & Ramzan Khan, M. (2022). Uncovering the research gaps to alleviate the negative impacts of climate change on food security: a review. *J Frontiers in plant science*, Vol. 13, 927535.
- Fekri, M. N., Ghosh, A. M., & Grolinger, K. (2019). Generating energy data for machine learning with recurrent generative adversarial networks. *J Energies*, Vol. 13 (1), 130.
- Giap, Q. H., Nguyen, D. L., Nguyen, T. T. Q., & Tran, T. M. D. (2022). Applying Neural Network And Levenberg-Marquardt Algorithm for Load Forecasting in IA-Grai District, Gia Lai Province. *J Tạp chí Khoa học và Công nghệ-Đại học Đà Nẵng*, 13-18.
- Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A., & Bengio, Y. (2014). Generative adversarial nets. *J Advances in neural information processing systems*, Vol. 27.
- Grigorev, A. (2021). *Machine Learning Bookcamp: Build a Portfolio of Real-life Projects*. Shelter Island: Manning Publications co.
- Guo, Y., Fu, Y., Hao, F., Zhang, X., Wu, W., Jin, X., Bryant, C. R., & Senthilnath. (2021). Integrated phenology and climate in rice yields prediction using machine learning methods. *J Ecological Indicators*, Vol. 120, 106935.
- Guritno, A. D., Kristanti, N. E., & Tanuputri, M. R. (2021). *Collaborative Strategy for the Supply Chain of Rice: A Case Study on Demak and Sukoharjo Regency, Central Java, Indonesia*.
- Hafizah, D., Hakim, D. B., Harianto, H., & Nurmalina, R. (2020). Analysing food consumption in Indonesia. *J Int J Progress Sci Tech*, Vol. 20 (2), 340-347.
- Harshvardhan, G., Gourisaria, M. K., Pandey, M., & Rautaray, S. S. (2020). A comprehensive survey and analysis of generative models in machine learning. *J Computer Science Review*, Vol. 38, 100285.

- Hasanah, A. F. (2020). Analisis perilaku konsumen masyarakat Ponorogo sesaat dan sesudah datangnya bulan Ramadhan. *J IJoIS: Indonesian Journal of Islamic Studies*, Vol. 1 (2), 95-106.
- Hasegawa, T., Sakurai, G., Fujimori, S., Takahashi, K., Hijioka, Y., & Masui, T. (2021). Extreme climate events increase risk of global food insecurity and adaptation needs. *J Nature Food*, Vol. 2 (8), 587-595.
- Iglesias, G., Talavera, E., González-Prieto, Á., Mozo, A., & Gómez-Canaval, S. (2023). Data augmentation techniques in time series domain: a survey and taxonomy. *J Neural Computing Applications*, Vol. 35 (14), 10123-10145.
- Iqbal, M. Z. (2019). The Effects of a Large Rice Price Increase on Household Welfare and Poverty in Rural Bangladesh. *Economics Bulletin*, Vol. 39 (1), 295-309.
- Joshua, V., Priyadharson, S. M., & Kannadasan, R. (2021). Exploration of machine learning approaches for paddy yield prediction in eastern part of Tamilnadu. *J Agronomy*, Vol. 11 (10), 2068.
- K, U., & Hanumanthappa, M. (2022). HEART DISEASE DATA ANALYSIS USING EXPLORATORY DATA ANALYSIS METHOD. *J International Journal of Engineering Applied Sciences Technology*, Vol. 7 (4).
- Kadhar, K. M. A., & Anand, G. (2021). Analyzing the Data. In *J Data Science with Raspberry Pi*. Berkeley: Apress.
- Kandel, P., Kharel, K., Njoroge, A., Smith, B. W., Díaz-Valderrama, J. R., Timilsina, R. H., Paudel, G. P., & Baributsa, D. (2021). On-farm grain storage and challenges in Bagmati Province, Nepal. *J Sustainability*, Vol. 13 (14), 7959.
- Kaur, H., Singh, H., & Verma, K. (2023). *Using Machine Learning to Predict Patient Health Outcomes*. Paper presented at the 2023 14th International Conference on Computing Communication and Networking Technologies (ICCCNT).
- Kavzoglu, T., & Teke, A. (2022a). Advanced hyperparameter optimization for improved spatial prediction of shallow landslides using extreme gradient boosting (XGBoost). *J Bulletin of Engineering Geology the Environment*, Vol. 81 (5), 201.
- Kavzoglu, T., & Teke, A. (2022b). Predictive Performances of ensemble machine learning algorithms in landslide susceptibility mapping using random forest, extreme gradient boosting (XGBoost) and natural gradient boosting (NGBoost). *Arabian Journal for Science Engineering*, Vol. 47 (6), 7367-7385.
- Kotsiopoulos, T., Sarigiannidis, P., Ioannidis, D., & Tzovaras, D. (2021). Machine learning and deep learning in smart manufacturing: The smart grid paradigm. *J Computer Science Review*, Vol. 40, 100341.
- Koumarelas, I., Jiang, L., & Naumann, F. (2020). Data Preparation for Duplicate Detection. *J Journal of Data Information Quality*, Vol. 12 (3), 1-24.
- Kumar, I., Rawat, J., Mohd, N., & Husain, S. (2021). Opportunities of artificial intelligence and machine learning in the food industry. *J Journal of Food Quality*, Vol. 2021 (1), 4535567.

- Kumar, P., & Yasmeen. (2021). Supply Chain Management of Food Grains in India. *J India's Economy Society: Lateral Explorations*, 53-74.
- Kuradusenge, M., Hitimana, E., Hanyurwimfura, D., Rukundo, P., Mtonga, K., Mukasine, A., Uwitonze, C., Ngabonziza, J., & Uwamahoro, A. (2023). Crop yield prediction using machine learning models: Case of Irish potato and maize. *J Agriculture*, Vol. 13 (1), 225.
- Kurnia, L. A., & Iskandar, D. D. (2019). Food independence determinant (Rice) In Supporting The Availability Of National Rice. *J Jurnal Ekonomi Pembangunan: Kajian Masalah Ekonomi dan Pembangunan*, Vol. 17 (1), 55-72.
- Lee, H.-L., Lin, Y.-P., & Petway, J. R. (2018). Global agricultural trade pattern in a warming world: regional realities. *J Sustainability*, Vol. 10 (8), 2763.
- Li, C. (2019). Preprocessing methods and pipelines of data mining: An overview. *J arXiv preprint arXiv:1906.08510*. doi:<https://doi.org/10.48550/arXiv.1906.08510>
- Li, N., Chiang, F., Down, D. G., & Heddle, N. M. (2021). A decision integration strategy for short-term demand forecasting and ordering for red blood cell components. Vol. 29, 100290.
- Li, P., Li, L., Hamdulla, A., & Wang, D. (2023). Visualizing data augmentation in deep speaker recognition. *J arXiv preprint arXiv:16070*.
- Li, Q.-C., Xu, S.-W., Zhuang, J.-Y., Liu, J.-J., Yi, Z., & Zhang, Z.-X. (2023). Ensemble learning prediction of soybean yields in China based on meteorological data. *J Journal of Integrative Agriculture*, Vol. 22 (6), 1909-1927.
- Li, X., Ngu, A. H. H., & Metsis, V. (2022). Tts-cgan: A transformer time-series conditional gan for biosignal data augmentation. *J arXiv preprint arXiv:13676*.
- Li, Z., Ma, C., Shi, X., Zhang, D., Li, W., & Wu, L. (2021). *Tsa-gan: A robust generative adversarial networks for time series augmentation*. Paper presented at the 2021 International Joint Conference on Neural Networks (IJCNN).
- Lin, C.-H., Kaushik, C., Dyer, E. L., & Muthukumar, V. (2022). The good, the bad and the ugly sides of data augmentation: An implicit spectral regularization perspective. *J arXiv preprint arXiv:05021*.
- Liu, D., Wu, Y., Hong, D., & Wang, S. (2022). Time series data augmentation method of small sample based on optimized generative adversarial network. *J Concurrency Computation: Practice Experience*, Vol. 34 (27), e7331.
- Liu, G., Inae, E., Luo, T., & Jiang, M. (2024). Rationalizing Graph Neural Networks with Data Augmentation. *J ACM Transactions on Knowledge Discovery from Data*, Vol. 18 (4), 1-23.
- Liu, Y., Yang, Z., Yu, Z., Liu, Z., Liu, D., Lin, H., Li, M., Ma, S., Avdeev, M., & Shi, S. (2023). Generative artificial intelligence and its applications in materials science: Current situation and future perspectives. *J Journal of Materiomics*, Vol. 9 (4), 98-816.

- Liyanage, S., Basnayake, N., Gamage, M., Prabhashi, P., Kasthuriarachchi, S., & Abeywardhana, L. (2023). *Sustainable Growth Through Automation: Machine Learning and Computer Vision Advancements in Sri Lankan Floriculture*. Paper presented at the 2023 5th International Conference on Advancements in Computing (ICAC).
- Luengo, J., García-Gil, D., Ramírez-Gallego, S., García, S., & Herrera, F. (2020). Big data preprocessing. *J Cham: Springer*.
- Luger, G. F. (2009). *Artificial intelligence: structures and strategies for complex problem solving*: Pearson education.
- Lutoslawski, K., Hernes, M., Radomska, J., Hajdas, M., Walaszczyk, E., & Kozina, A. (2021). Food demand prediction using the nonlinear autoregressive exogenous neural network. *J IEEE Access*, Vol. 9, 146123-146136.
- Machlanski, D., Samothrakis, S., & Clarke, P. (2023). Hyperparameter tuning and model evaluation in causal effect estimation. *J arXiv preprint arXiv:01412*. doi:<https://doi.org/10.48550/arXiv.2303.01412>
- Manjunath, M. C., & Palayyan, B. P. (2023). An Efficient Crop Yield Prediction Framework Using Hybrid Machine Learning Model. *J Revue d'Intelligence Artificielle*, Vol. 37 (4).
- Marcot, B. G., & Hanea, A. M. (2021). What is an optimal value of k in k-fold cross-validation in discrete Bayesian network analysis? *J Computational Statistics*, Vol. 36 (3), 2009-2031.
- Maskun, Naswar, Ahmad, Ilmar, A., Napang, M., & Nugraha, R. (2021). The state responsibility on food availability on the pandemic. *J Gaceta Sanitaria*, Vol. 35, S345-S347.
- Mehmood, U., Bashir, A. K., Rabie, K., Broderick, J., & Davies, S. (2023). *Vending Machine Product Demand Prediction Using Machine Learning Algorithms*. Paper presented at the 2023 International Symposium on Networks, Computers and Communications (ISNCC).
- Missaoui, S., Gerasimou, S., & Matragkas, N. (2023). *Semantic Data Augmentation for Deep Learning Testing using Generative AI*. Paper presented at the 2023 38th IEEE/ACM International Conference on Automated Software Engineering (ASE).
- Mohamed, M., & Saber, N. E. (2023). Intelligent Decision Support Machine Learning Based Optimizing Inventory Management. *J American Journal of Business Operations Research*, Vol. 9 (2), 41-49.
- Moreno-Barea, F. J., Jerez, J. M., & Franco, L. (2020). Improving classification accuracy using data augmentation on small data sets. *J Expert Systems with Applications*, Vol. 161, 113696.
- Mrudula, O., & Sowjanya, A. M. (2020). Understanding Clinical Data using Exploratory Analysis. *J International Journal of Recent Technology and Engineering*, Vol. 8 (5).
- Mukherjee, D. (2021). Food Security Under The Era Of Climate Change Threat. *J Journal of Advanced Agriculture Horticulture Research*, Vol. 1 (1), 1-4.

- Mukhiya, S. K., & Ahmed, U. (2020). *Hands-On Exploratory Data Analysis with Python: Perform EDA techniques to understand, summarize, and investigate your data*: Packt Publishing Ltd.
- Namany, S., Govindan, R., Alfagih, L., McKay, G., & Al-Ansari, T. (2020). Sustainable food security decision-making: an agent-based modelling approach. *J Journal of Cleaner Production*, Vol. 255, 120296.
- Nassibi, N., Fasihuddin, H., & Hsairi, L. (2023). Demand Forecasting Models for Food Industry by Utilizing Machine Learning Approaches. *J International Journal of Advanced Computer Science Applications*, Vol. 14 (3).
- Naveed, M. H., Hashmi, U. S., Tajved, N., Sultan, N., & Imran, A. J. I. A. (2022). Assessing deep generative models on time series network data. Vol. 10, 64601-64617.
- Nguyen, Q. H., Ly, H.-B., Ho, L. S., Al-Ansari, N., Le, H. V., Tran, V. Q., Prakash, I., & Pham, B. T. (2021). Influence of data splitting on performance of machine learning models in prediction of shear strength of soil. *J Mathematical Problems in Engineering*, Vol. 2021, 1-15. doi:<https://doi.org/10.1155/2021/4832864>
- Niu, Z., Yu, K., & Wu, X. (2020). LSTM-based VAE-GAN for time-series anomaly detection. *J Sensors*, Vol. 20 (13), 3738.
- Noorunnahar, M., Chowdhury, A. H., & Mila, F. A. (2023). A tree based eXtreme Gradient Boosting (XGBoost) machine learning model to forecast the annual rice production in Bangladesh. *J PloS one*, Vol. 18 (3), e0283452.
- Nosratabadi, S., Ardabili, S., Lakner, Z., Mako, C., & Mosavi, A. (2021). Prediction of food production using machine learning algorithms of multilayer perceptron and ANFIS. *J Agriculture*, Vol. 11 (5), 408.
- Núñez von Voigt, S., Pauli, M., Reichert, J., & Tschorsch, F. (2020). *Every query counts: Analyzing the privacy loss of exploratory data analyses*. Paper presented at the Data Privacy Management, Cryptocurrencies and Blockchain Technology: ESORICS 2020 International Workshops, DPM 2020 and CBT 2020, Guildford, UK, September 17–18, 2020, Revised Selected Papers 15.
- Octania, G. (2021). *Peran Pemerintah dalam Rantai Pasok Beras Indonesia*. Retrieved from
- Onishi, S., & Meguro, S. (2023). Rethinking data augmentation for tabular data in deep learning. *J arXiv preprint arXiv:10308*.
- Ørebæk, O.-E., & Geitle, M. (2021). *Exploring the Hyperparameters of XGBoost Through 3D Visualizations*. Paper presented at the AAAI Spring Symposium: Combining Machine Learning with Knowledge Engineering.
- Otero-Escobar, A. D., & Velasco-Ramírez, M. L. (2023). *Study on Exploratory Data Analysis Applied to Education*. Paper presented at the 2023 IEEE International Conference on Engineering Veracruz (ICEV).
- Otoni, A. L. C., Souza, A. M., & Novo, M. S. (2023). Automated hyperparameter tuning for crack image classification with deep learning. *J Soft Computing*, Vol. 27 (23), 18383-18402.
- Ozdemir, S. (2022). *Feature Engineering Bookcamp*: Simon and Schuster.

- Patel, D. T. (2021). Quantitative and Visual Exploratory Data Analysis for Machine Intelligence. In R. R. A. D. Samanta, S. Pramanik, & S. Dutta (Ed.), *Source Title: Methodologies and Applications of Computational Statistics for Machine Intelligence* (pp. 97-117): IGI Global.
- Perez, H. D., Hubbs, C. D., Li, C., & Grossmann, I. E. (2021). Algorithmic approaches to inventory management optimization. *J Processes*, Vol. 9 (1), 102.
- Pérez, J., Arroba, P., & Moya, J. M. (2023). Data augmentation through multivariate scenario forecasting in Data Centers using Generative Adversarial Networks. *J Applied Intelligence*, Vol. 53 (2), 1469-1486.
- Putra, M. R. J., Pujangkoro, S. A., & Situmorang, S. H. (2022). *Marketing Mix Strategy and SWOT Analysis on Beraskita Products Perum Bulog Regional Division North Sumatra*. Paper presented at the 19th International Symposium on Management (INSYMA 2022).
- Putro, P. A. W., Purwaningsih, E. K., Sensuse, D. I., & Suryono, R. R. (2022). Model and implementation of rice supply chain management: A literature review. *J Procedia Computer Science*, Vol. 197, 453-460.
- Qian, J., Ito, S., & Zhao, Z. (2020). The effect of price support policies on food security and farmers' income in China. *J Australian Journal of Agricultural Resource Economics*, Vol. 64 (4), 1328-1349.
- Qin, C., Zhang, Y., Bao, F., Zhang, C., Liu, P., & Liu, P. (2021). XGBoost optimized by adaptive particle swarm optimization for credit scoring. *J Mathematical Problems in Engineering*, Vol. 2021, 1-18. doi:<https://doi.org/10.1155/2021/6655510>
- Qin, H. (2023). Research on Machine Learning and Intelligent Decision Support System Based on Risk Prediction. *J Advances in Engineering Technology Research*, Vol. 4 (1), 582-582.
- Rahadi, N. W., & Vikasari, C. (2020). Pengujian Software Aplikasi Perawatan Barang Miliki Negara Menggunakan Metode Black Box Testing Equivalence Partitions. *J Jurnal Infotekmesin*, Vol. 11 (01), 57-61.
- Rai, R., Tiwari, M. K., Ivanov, D., & Dolgui, A. (2021). Machine learning in manufacturing and industry 4.0 applications. *J International Journal of Production Research*, Vol. 59 (16), 4773-4778. doi:<https://doi.org/10.1080/00207543.2021.1956675>
- Rane, S. B., Thakker, S. V., & Kant, R. (2021). Stakeholders' involvement in green supply chain: a perspective of blockchain IoT-integrated architecture. *J Management of Environmental Quality: An International Journal*, Vol. 32 (6), 1166-1191.
- Ratilal, P. D., & Reddy, P. (2023). The Importance of Data Visualization in Exploratory Data Analysis. *J Journal of Advanced Zoology*, Vol. 44.
- Rodrigues, M., Miguéis, V., Freitas, S., & Machado, T. (2024). Machine learning models for short-term demand forecasting in food catering services: A solution to reduce food waste. *J Journal of Cleaner Production*, Vol. 435, 140265.
- Russell, S. J., & Norvig, P. (2010). *Artificial intelligence: a modern approach* (Third Edition ed.). New Jersey: Pearson Education.

- Sachan, S., Yang, J.-B., Xu, D.-L., Benavides, D. E., & Li, Y. (2020). An explainable AI decision-support-system to automate loan underwriting. *J Expert Systems with Applications*, Vol. 144, 113100.
- Sahoo, K., Samal, A. K., Pramanik, J., & Pani, S. K. (2019). Exploratory data analysis using Python. *J International Journal of Innovative Technology Exploring Engineering*, Vol. 8 (12), 4727-4735.
- Santhoshkumar, T., & Vanila, S. (2024). Exploratory Data Analysis and Energy Predictions With Advanced AI and ML Techniques. In *AI Approaches to Smart and Sustainable Power Systems* (pp. 336-370): IGI Global.
- Sardar, M. A. S., Hasi Saha, M. N. S., & Rabbi, M. F. (2023). Intrusion Detection in Electric Vehicles using Machine Learning with Model Explainability. *Journal of Information Hiding and Multimedia Signal Processing*.
- Sari, Y. R., Ananda, L. R., & Rani, M. (2020). *The Analysis of Fuzzy Logic Method and Multiple Linear Regression in Determining National Rice Production to Improve Food Self-Sufficiency in Indonesia*. Paper presented at the IOP Conference Series: Materials Science and Engineering.
- Sasi, A., & Subramanian, T. (2022). Comparative analysis of ARIMA and double exponential smoothing for forecasting rice sales in fair price shop. *J Journal of Statistics Management Systems*, Vol. 25 (7), 1601-1619.
- Sasmito, G. W., & Mutasodirin, M. A. (2023). *Black Box Testing with Equivalence Partitions Techniques in Transcrop Applications*. Paper presented at the 2023 6th International Conference of Computer and Informatics Engineering (IC2IE).
- Satpathi, A., Setiya, P., Das, B., Nain, A. S., Jha, P. K., Singh, S., & Singh, S. (2023). Comparative analysis of statistical and machine learning techniques for rice yield forecasting for Chhattisgarh, India. *J Sustainability*, Vol. 15 (3), 2786.
- Saud, S., Wang, D., Fahad, S., Alharby, H. F., Bamagoos, A. A., Mjrashi, A., Alabdallah, N. M., AlZahrani, S. S., AbdElgawad, H., & Adnan, M. (2022). Comprehensive impacts of climate change on rice production and adaptive strategies in China. *J Frontiers in microbiology*, Vol. 13, 926059.
- Seyedan, M., & Mafakheri, F. (2020). Predictive big data analytics for supply chain demand forecasting: methods, applications, and research opportunities. *J Journal of Big Data*, Vol. 7 (1), 1-22.
- Shabdin, N. I., Ya'acob, S., & Sjarif, N. N. A. (2020). *Relationship types in visual analytics*. Paper presented at the Proceedings of the 2020 6th International Conference on Computer and Technology Applications.
- Shi, Z. (2019). *Advanced artificial intelligence* (Vol. 4): World Scientific.
- Shijubo, J., Waga, M., & Suenaga, K. (2021). *Efficient black-box checking via model checking with strengthened specifications*. Paper presented at the International Conference on Runtime Verification.
- Siahaan, H. M. (2023, 31 Januari 2023). Pentingnya Pembenahan Data Beras Nasional. *Kompas.com*. Retrieved from <https://money.kompas.com/read/2023/01/31/112238126/pentingnya-pembenahan-data-beras-nasional?page=all>

- Silalahi, N. H., Yudha, R. O., Dwiyaniti, E. I., Zulvianita, D., Feranti, S. N., & Yustiana, Y. (2019). Government policy statements related to rice problems in Indonesia. *J Journal of Biological Science, Technology, Management*, Vol. 1 (1), 35-41.
- Smiti, A. (2020). A critical overview of outlier detection methods. *J Computer Science Review*, Vol. 38, 100306. doi:<https://doi.org/10.1016/j.cosrev.2020.100306>
- Soleymani, S., & Mohammadzadeh, S. (2023). Comparative Analysis of Machine Learning Algorithms for Solar Irradiance Forecasting in Smart Grids. *J arXiv preprint arXiv:2307.13791*.
- Sossou, S., & Igue, C. B. (2019). Impact of Rising International Market Prices of Rice on Welfare and Poverty in Senegal. *Asian Journal of Agricultural Extension*, Vol. 29 (3), 1-12.
- Sutiah, S., & Supriyono, S. (2021). *Software testing on e-learning Madrasahs using Blackbox testing*. Paper presented at the IOP Conference Series: Materials Science and Engineering.
- Swaminathan, K., & Venkitasubramony, R. (2023). Demand forecasting for fashion products: A systematic review. *J International Journal of Forecasting*. doi:<https://doi.org/10.1016/j.ijforecast.2023.02.005>
- TanLi, M., Jiang, Y., Wang, X., & Peng, R. (2020). *Black-box approach for software testing based on fat-property*. Paper presented at the MATEC Web of Conferences.
- Tanwar, S. (2022). Blockchain for Supply Chain Management. In *Blockchain Technology: From Theory to Practice* (pp. 321-353): Springer.
- Tavafoghi, H., Ouyang, Y., & Teneketzis, D. (2021). A unified approach to dynamic decision problems with asymmetric information: Nonstrategic agents. *J IEEE Transactions on Automatic Control*, Vol. 67 (3), 1105-1119.
- Teerasoponpong, S., & Sopadang, A. (2022). Decision support system for adaptive sourcing and inventory management in small-and medium-sized enterprises. *J Robotics computer-integrated manufacturing*, Vol. 73, 102226.
- Thilakarathne, N. N., Bakar, M. S. A., Abas, P. E., & Yassin, H. (2022). A cloud enabled crop recommendation platform for machine learning-driven precision farming. *J Sensors*, Vol. 22 (16), 6299.
- Tran, N.-T., Tran, V.-H., Nguyen, N.-B., Nguyen, T.-K., & Cheung, N.-M. (2021). On data augmentation for gan training. *J IEEE Transactions on Image Processing*, Vol. 30, 1882-1897.
- Udousoro, I. C. (2020). Machine learning: A review. *Semiconductor Science Information Devices*, Vol. 2 (2), 5-14.
- Usdianto, D. A., & Setiyowati, R. (2023). MATHEMATICAL MODEL OF RICE COMMODITY SUPPLY CHAIN IN INDONESIA. *J BAREKENG: Jurnal Ilmu Matematika dan Terapan*, Vol. 17 (2), 1139-1148.
- Utomo, B. (2020). Tantangan dan peran bulog di era industri 4.0. *J Jurnal Pangan*, Vol. 29 (1), 71-86.
- Vaitheeshwari, R., & SathieshKumar, V. (2019). *Performance analysis of epileptic seizure detection system using neural network approach*. Paper presented

- at the 2019 International Conference on Computational Intelligence in Data Science (ICCIDS).
- Vallés-Pérez, I., Soria-Olivas, E., Martínez-Sober, M., Serrano-López, A. J., Gómez-Sanchís, J., & Mateo, F. (2022). Approaching sales forecasting using recurrent neural networks and transformers. *J Expert Systems with Applications*, Vol. 201, 116993. doi:<https://doi.org/10.1016/j.eswa.2022.116993>
- Vo, H.-T., Hoang, T. N., & Quach, L.-D. (2023). An Approach to Hyperparameter Tuning in Transfer Learning for Driver Drowsiness Detection Based on Bayesian Optimization and Random Search. *J International Journal of Advanced Computer Science Applications*, Vol. 14 (4).
- Vosooghizadeh, M., Taghipour, A., & Canel-Depitre, B. (2020). Supply chain coordination under information asymmetry: a review. *J International Journal of Production Research*, Vol. 58 (6), 1805-1834.
- Wang, J. (2023). *Intelligent Decision Support System for Building Project Management Based on Artificial Intelligence*. Paper presented at the Journal of Physics: Conference Series.
- Wang, J., & Zhou, S. (2024). Particle swarm optimization-XGBoost-based modeling of radio-frequency power amplifier under different temperatures. *J International Journal of Numerical Modelling: Electronic Networks, Devices Fields*, e3168. doi:<https://doi.org/10.1002/jnm.3168>
- Wang, S., & Yang, Y. (2021). M-GAN-XGBOOST model for sales prediction and precision marketing strategy making of each product in online stores. *J Data Technologies Applications*, Vol. 55 (5), 749-770.
- Wang, W., Lesner, C., Ran, A., Rukonic, M., Xue, J., & Shiu, E. (2020). *Using small business banking data for explainable credit risk scoring*. Paper presented at the Proceedings of the AAAI Conference on Artificial Intelligence.
- Webber, B. L., & Nilsson, N. J. (2014). *Readings in artificial intelligence*: Morgan Kaufmann.
- Wen, H., Hu, J., Zhang, J., Xiang, X., & Liao, M. (2022). Rockfall susceptibility mapping using XGBoost model by hybrid optimized factor screening and hyperparameter. *J Geocarto International*, Vol. 37 (27), 16872-16899.
- West, R. M. (2022). Best practice in statistics: The use of log transformation. *J Annals of Clinical Biochemistry*, Vol. 59 (3), 162-165.
- Wu, D., Hur, K., & Xiao, Z. (2021). A GAN-enhanced ensemble model for energy consumption forecasting in large commercial buildings. *J IEEE Access*, Vol. 9, 158820-158830.
- Wu, J., Li, Y., & Ma, Y. (2021). *Comparison of XGBoost and the neural network model on the class-balanced datasets*. Paper presented at the 2021 IEEE 3rd international conference on frontiers technology of information and computer (ICFTIC).
- Yamamura, C. L. K., Santana, J. C. C., Masiero, B. S., Quintanilha, J. A., & Berssaneti, F. T. (2022). Forecasting New Product Demand Using Domain Knowledge and Machine Learning: A proposed method uses machine learning and an expert's domain knowledge to enhance the accuracy of

- new product predictions. *J Research-Technology Management*, Vol. 65 (4), 27-36.
- Yani, L. P. E., & Aamer, A. (2023). Demand forecasting accuracy in the pharmaceutical supply chain: a machine learning approach. *J International Journal of Pharmaceutical Healthcare Marketing*, Vol. 17 (1), 1-23. doi:<http://dx.doi.org/10.1108/IJPHM-05-2021-0056>
- Yiğit, F., ESNAF, Ş., & KAVUŞ, B. Y. (2021). A Poisson-Regression, Support Vector Machine and Grey Prediction Based Combined Forecasting Model Proposal: A Case Study in Distribution Business. *J Turkish Journal of Forecasting*, Vol. 5 (2), 23-35.
- Yu, H., & Guo, Y. (2023). *Generative artificial intelligence empowers educational reform: current status, issues, and prospects*. Paper presented at the Frontiers in Education.
- Yulianis, N., & Rachman, B. (2021). *The implementation and synergy of Indonesian national food reserves*. Paper presented at the IOP Conference Series: Earth and Environmental Science.
- Yusuf, W., Yusuf, S., Adesope, A., & Adebayo, O. (2020). Determinants of rice import demand in Nigeria. *J Journal of Applied Sciences Environmental Management*, Vol. 24 (5), 923-931.
- Zhang, C., & Lu, Y. (2021). Study on artificial intelligence: The state of the art and future prospects. *Journal of Industrial Information Integration*, Vol. 23, 100224.
- Zhang, J., Luo, D., & Wei, H. (2023). *Mixupexplainer: Generalizing explanations for graph neural networks with data augmentation*. Paper presented at the Proceedings of the 29th ACM SIGKDD Conference on Knowledge Discovery and Data Mining.
- Zhang, Z., Zhang, D., & Qiu, R. C. (2019). Deep reinforcement learning for power system applications: An overview. *J CSEE Journal of Power Energy Systems*, Vol. 6 (1), 213-225.
- Zhao, C., Wan, C., & Song, Y. (2021). Operating reserve quantification using prediction intervals of wind power: An integrated probabilistic forecasting and decision methodology. *J IEEE Transactions on Power Systems*, Vol. 36 (4), 3701-3714.
- Zhao, S., Liu, Z., Lin, J., Zhu, J.-Y., & Han, S. (2020). Differentiable augmentation for data-efficient gan training. *J Advances in neural information processing systems*, Vol. 33, 7559-7570.
- Zhu, Z., & He, K. (2022). Prediction of Amazon's Stock Price Based on ARIMA, XGBoost, and LSTM Models. *J Proceedings of Business Economic Studies*, Vol. 5 (5), 127-136.

LAMPIRAN 1 *SAMPLE DATASET* 22 BARIS PERTAMA

Provinsi	Tahun	Produksi	Luas Panen	Curah hujan	Kelembapan	Suhu rata-rata
Aceh	1993	1329536	323589	1627	82	26.06
Aceh	1994	1299699	329041	1521	82.12	26.92
Aceh	1995	1382905	339253	1476	82.72	26.27
Aceh	1996	1419128	348223	1557	83	26.08
Aceh	1997	1368074	337561	1339	82.46	26.31
Aceh	1998	1404580	365892	1465	82.6	26.84
Aceh	1999	1478712	359817	1778	82.79	26.14
Aceh	2000	1486909	336765	1974.7	90.6	27.1
Aceh	2001	1547499	295212	1688.7	69.48	28.9
Aceh	2002	1314165	315131	1296.8	68.75	29.2
Aceh	2003	1246614	367636	1507.2	70.66	29.4
Aceh	2004	1350748	370966	1097	80.84	29.4
Aceh	2005	1411650	337893	710.5	79.5	26.8
Aceh	2006	1552078	320789	506.5	80.8	26.73
Aceh	2007	1556858	360717	1414	81.5	26.38
Aceh	2008	1402287	329109	1270.4	78.5	27
Aceh	2009	1533369	359375	1577	78.7	26.9
Aceh	2010	1788738	352281	1986	81.4	27.1
Aceh	2011	1772962	380686	1268	79.4	27.1
Aceh	2012	1582393	387803	1098	79.6	26.9
Aceh	2013	2331046	419183	1623.6	80.7	27