

**PREDIKSI OMSET BISNIS RESTORAN SOTO-KWALI PAK WASIS
MENGUNAKAN METODE RANDOM FOREST
DAN LOGISTIC REGRESSION**

THESIS

**Oleh :
ERFAN AINUL YAKIN
NIM. 220605210002**



**PROGRAM STUDI MAGISTER INFORMATIKA
FAKULTAS SAINS DAN TEKNOLOGI
UNIVERSITAS ISLAM NEGERI MAULANA MALIK IBRAHIM
MALANG
2023**

**PREDIKSI OMSET BISNIS RESTORAN SOTO-KWALI PAK WASIS
MENGUNAKAN METODE RANDOM FOREST
DAN LOGISTIC REGRESSION**

THESIS

**Diajukan kepada:
Universitas Islam Negeri Maulana Malik Ibrahim Malang
Untuk memenuhi Salah Satu Persyaratan dalam
Memperoleh Gelar Magister Komputer (M.Kom)**

**Oleh :
ERFAN AINUL YAKIN
NIM. 220605210002**

**PROGRAM STUDI MAGISTER INFORMATIKA
FAKULTAS SAINS DAN TEKNOLOGI
UNIVERSITAS ISLAM NEGERI MAULANA MALIK IBRAHIM
MALANG
2023**

**PREDIKSI OMSET BISNIS RESTORAN SOTO-KWALI PAK WASIS
MENGUNAKAN METODE RANDOM FOREST
DAN LOGISTIC REGRESSION**

THESIS

**Diajukan kepada:
Fakultas Sains dan Teknologi
Universitas Islam Negeri Maulana Malik Ibrahim Malang
Untuk memenuhi Salah Satu Persyaratan dalam
Memperoleh Gelar Magister Komputer (M.Kom)**

**Oleh :
ERFAN AINUL YAKIN
NIM. 220605210002**

**PROGRAM STUDI MAGISTER INFORMATIKA
FAKULTAS SAINS DAN TEKNOLOGI
UNIVERSITAS ISLAM NEGERI MAULANA MALIK IBRAHIM
MALANG
2023**

**PREDIKSI OMSET BISNIS RESTORAN SOTO-KWALI PAK WASIS
MENGUNAKAN METODE RANDOM FOREST
DAN LOGISTIC REGRESSION**

THESIS

**Oleh :
ERFAN AINUL YAKIN
NIM. 220605210002**

Telah diperiksa dan disetujui untuk di uji:
Tanggal 12 Oktober 2023

Pembimbing I,



Dr. Ririen Kusumawati, S.Si., M.Kom
NIP 19720309 200501 2 002

Pembimbing II,



Dr. Usman Pagalay, M.Si
NIP. 19650414 200312 1 001

Mengetahui,
Ketua Program Studi Magister Informatika
Fakultas Sains dan Teknologi
Universitas Islam Negeri Maulana Malik Ibrahim Malang



Dr. Cuyyo Crysdian
NIP. 19740424 200901 1 008

**PREDIKSI OMSET BISNIS RESTORAN SOTO-KWALI PAK WASIS
MENGUNAKAN METODE RANDOM FOREST
DAN LOGISTIC REGRESSION**

THESIS

Oleh :
ERFAN AINUL YAKIN
NIM. 220605210002

**Telah Dipertahankan di Depan Dewan Penguji Thesis
dan Dinyatakan Diterima Sebagai Salah Satu Persyaratan
Untuk Memperoleh Gelar Magister Komputer (M.Kom)
Tanggal 12 Oktober 2023**

Susunan Dewan Penguji

Penguji Utama : Prof. Dr. Sri Harini, M.Si
NIP 19731014 200112 2 002

Ketua Penguji : Dr. M. Imamuddin, LC., MA
NIP 19740602 200901 1 010

Sekretaris Penguji : Dr. Ririen Kusumawati, S.Si., M.Kom
NIP 19720309 200501 2 002

Anggota Penguji : Dr. Usman Pagalay, M.Si
NIP. 19650414 200312 1 001

Tanda Tangan

()

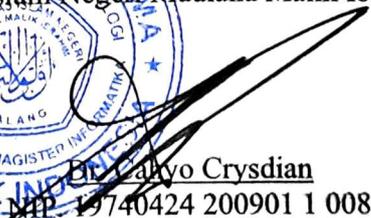
()

()

()

Mengetahui dan Mengesahkan,
Ketua Program Studi Magister Informatika
Fakultas Sains dan Teknologi
Universitas Islam Negeri Maulana Malik Ibrahim Malang




Dr. Cahyo Crysdian
NIP. 19740424 200901 1 008

PERNYATAAN KEASLIAN TULISAN

Saya yang bertanda tangan dibawah ini:

Nama : Erfan Ainul Yakin

NIM : 220605210002

Program Studi : Magister Informatika

Fakultas : Sains dan Teknologi

Menyatakan dengan sebenarnya bahwa Thesis yang saya tulis ini benar-benar merupakan hasil karya saya sendiri, bukan merupakan pengambilalihan data, tulisan atau pikiran orang lain yang saya akui sebagai tulisan atau pikiran saya sendiri, kecuali dengan mencantumkan sumber cuplikan pada daftar pustaka.

Apabila dikemudian hari terbukti atau dapat dibuktikan Thesis ini hasil jiplakan, maka saya bersedia menerima sanksi atas perbuatan tersebut.

Malang, 12 Oktober 2023

Yang Membuat Pernyataan



Erfan Ainul Yakin
NIM. 220605210002

PERSEMBAHAN

Atas berkah dan rahmat Allah SWT, Thesis ini bisa saya selesaikan dengan lengkap dan baik sesuai dengan arahan bapak ibu pembimbing, oleh karena itu hasil kerja keras atas keberhasilan dari penulisan Thesis ini merupakan dukungan dari beberapa pihak yaitu :

1. Keluarga kecil tercinta antara lain istri (Lely Wijayanti, S.Kom., M.M.), putra pertama (Harun Arayyan Kahfi) dan putra kedua (Maulana Malik Ahsan).
2. Keluarga besar, yaitu ibu, ayah, ayah mertua, kakek, nenek, paman, kakak sepupu dan saudara-saudara yang lain.
3. Team software house CV. Do Event Teknologi atas dukungan data pada studi kasus.
4. Divisi EDP di PT. Kotakom atas dukungan fasilitas hardware maupun software dalam pengerjaan penelitian ini.
5. Bapak ibu dosen Magister Informatika Fakultas Sains dan Teknologi Universitas Islam Negeri Maulana Malik Ibrahim Malang.
6. Teman-teman angkatan 6 Magister Informatika Fakultas Sains dan Teknologi Universitas Islam Negeri Maulana Malik Ibrahim Malang.
7. Teman-teman mahasiswa Magister Informatika Fakultas Sains dan Teknologi Universitas Islam Negeri Maulana Malik Ibrahim Malang.
8. Serta rekan-rekan yang tidak mungkin disebutkan satu persatu atas dukungan terselesaikannya Thesis ini.

KATA PENGANTAR

Assalamu'alaikum Wr. Wb.

Awal kata, saya ucapkan puja dan puji syukur atas kehadiran Allah SWT karena telah dilimpahkannya kesehatan jasmani dan rohani serta rizki yang berkah sehingga Thesis yang berjudul “Prediksi Omset Bisnis Restoran Soto-Kwali Pak Wasis Menggunakan Metode Random Forest dan Logistic Regression” dapat diselesaikan dengan baik.

Sholawat serta salam tetap saya curahkan kepada junjungan nabi besar Muhammad SAW atas jasa beliau, Thesis ini dapat memberikan nilai dakwah dalam agama Islam sehingga menjadi berkah baik di dunia maupun di akhirat, amin.

Penulis ucapan terima kasih kepada pihak yang telah menemani serta mendukung dengan kesabaran dan keikhlasan dalam proses penulisan Thesis ini. Ucapan ini lebih khusus saya sampaikan kepada :

1. Ibu bapak pembimbing yaitu Dr. Ririen Kusumawati, S.Si.,M.Kom dan Dr. Usman Pagalay, M.Si yang sudah memberikan nasihat, arahan serta pengalaman berharga dalam proses pengerjaan Thesis ini.
2. Ibu bapak dosen Magister Informatika Fakultas Sains dan Teknologi Universitas Islam Negeri Maulana Malik Ibrahim Malang atas ilmu dalam proses belajar selama menempuh S2.
3. Keluarga kecil dan keluarga besar tercinta atas dukungan moral dan do'a pada setiap kegiatan formal maupun informal.
4. Teman-teman mahasiswa, rekan kerja dan semua pihak yang mengenal pribadi saya atas jalinan silaturahmi yang baik.

Sebelumnya saya mohon maaf bila dalam penulisan Thesis ini masih banyak kekurangan dan kesalahan. Saya berharap Thesis ini dapat memberi manfaat bagi kemajuan ilmu teknologi khususnya bagi saya sendiri, amin.

Wassalamu'alaikum Wr. Wb

Malang, 12 Oktober 2023

Penulis

DAFTAR ISI

HALAMAN JUDUL	i
HALAMAN PENGAJUAN	ii
HALAMAN PERSETUJUAN	iii
HALAMAN PENGESAHAN	iv
PERNYATAAN KEASLIAN TULISAN	v
PERSEMBAHAN	vi
KATA PENGANTAR	vii
DAFTAR ISI	viii
DAFTAR GAMBAR	x
DAFTAR TABEL	xii
ABSTRAK	xiv
ABSTRACT	xv
المخلص	xvi
BAB I. PENDAHULUAN	1
1.1. Latar Belakang	1
1.2. Pernyataan Masalah	3
1.3. Tujuan Penelitian.....	3
1.4. Batasan Masalah.....	4
1.5. Manfaat Penelitian	4
BAB II. TINJAUAN PUSTAKA	5
2.1. Prediksi dengan Machine Learning.....	5
2.2. Kerangka Teori Prediksi pada Machine Learning	7
2.3. Prediksi Algoritma Random Forest.....	9
2.4. Prediksi Algoritma Logistic Regression	11
BAB III. DESAIN PENELITIAN	14
3.1. Riset Prosedur	14
3.2. Data Collection.....	15
3.3. Data Engineering.....	16
3.4. Desain Sistem.....	18
3.5. Implementasi Sistem	19

3.6. Experiment	20
3.7. Evaluation Method	21
3.7.1. MAE (Mean Absolute Error)	21
3.7.2. MAPE (Mean Absolute Percentage Error)	22
BAB IV. METODE RANDOM FOREST	23
4.1 Desain	23
4.2 Uji Coba	26
4.2.1 Skenario 1	28
4.2.2 Skenario 2	32
4.2.3 Skenario 3	36
4.2.4 Skenario 4	40
4.3 Evaluasi	45
BAB V. METODE LOGISTIC REGRESSION	47
5.1 Desain	47
5.2 Uji Coba	49
5.2.1 Skenario 1	51
5.2.2 Skenario 2	55
5.2.3 Skenario 3	59
5.2.4 Skenario 4	62
5.3 Evaluasi	67
BAB VI. PEMBAHASAN	69
6.1. Perbandingan Performance Random Forest dan Logistic Regression	69
6.1.1. Perbandingan Metode Menggunakan <i>Accuracy</i>	69
6.1.2. Perbandingan Metode Menggunakan <i>Duration</i>	72
6.1.3. Perbandingan Metode Menggunakan <i>MAE</i> dan <i>MAPE</i>	73
6.2. Implementasi Metode Terbaik	75
6.3. Prediksi Omset Bisnis Restoran dalam Pandangan Islam	76
BAB VII. KESIMPULAN	80
7.1. Kesimpulan	80
7.2. Saran	82
DAFTAR PUSTAKA	83
LAMPIRAN	87

DAFTAR GAMBAR

Gambar 2.1 : Kerangka Teori Prediksi dalam Machine Learning	8
Gambar 2.2 : Skema Model Random Forest Prediksi Kinerja Mahasiswa.....	10
Gambar 2.3 : Skema Model Random Forest Prediksi Kematian COVID-19	11
Gambar 2.4 : Skema Model Logistic Regression Prediksi Resiko Drop-out.....	12
Gambar 2.5 : Skema Model Logistic Regression Prediksi momentum air ir.....	13
Gambar 3.1 : Diagram Riset Prosedur	14
Gambar 3.2 : Cara pengumpulan data.....	15
Gambar 3.3 : Diagram data prosesing.....	16
Gambar 3.4 : Desain Sistem.....	19
Gambar 3.5 : Ilustrasi Confusion Matrix.	20
Gambar 4.1 : Flowchart Prediksi Model Metode Random Forest	23
Gambar 4.2 : Diagram Topologi Metode Random Forest	25
Gambar 4.3 : Interface Prediksi Metode Random Forest.....	26
Gambar 4.4 : Diagram plot prediksi Random Forest skenario 1	30
Gambar 4.5 : Diagram line prediksi Random Forest skenario 1	31
Gambar 4.6 : Diagram plot prediksi Random Forest skenario 2	34
Gambar 4.7 : Diagram line prediksi Random Forest skenario 2.....	35
Gambar 4.8 : Diagram plot prediksi Random Forest skenario 3	39
Gambar 4.9 : Diagram line prediksi Random Forest skenario 3.....	39
Gambar 4.10 : Diagram plot prediksi Random Forest skenario 4.....	43
Gambar 4.11 : Diagram line prediksi Random Forest skenario 4.....	44
Gambar 4.12 : Diagram Evaluasi Performance Metode Random Forest.....	45
Gambar 5.1 : Curva Logistic Regression	47
Gambar 5.2 : Flowchart Prediksi Model Metode <i>Logistic Regression</i>	48
Gambar 5.3 : Interface Prediksi Metode Logistic Regression	50
Gambar 5.4 : Curva prediksi Logistic Regression skenario 1.....	54
Gambar 5.5 : Diagram line prediksi Logistic Regression skenario 1	54
Gambar 5.6 : Curva prediksi Logistic Regression skenario 2.....	57
Gambar 5.7 : Diagram line prediksi Logistic Regression skenario 2	58
Gambar 5.8 : Curva prediksi Logistic Regression skenario 3.....	61

Gambar 5.9 : Diagram line prediksi Logistic Regression skenario 3	62
Gambar 5.10 : Curva prediksi Logistic Regression skenario 4.....	65
Gambar 5.11 : Diagram line prediksi Logistic Regression skenario 4	66
Gambar 5.12 : Diagram Evaluasi Performance Metode Logistic Regression	67
Gambar 6.1 : Perbandingan <i>Accuracy</i> RF dan LR.....	71
Gambar 6.2 : Perbandingan nilai <i>duration</i> RF dan LR.....	73
Gambar 6.3 : Perbandingan nilai <i>MAE</i> RF dan LR.....	75
Gambar 6.4 : Perbandingan nilai <i>MAPE</i> RF dan LR	75

DAFTAR TABEL

Tabel 2.1 Ringkasan Jurnal	7
Tabel 2.2 Peringkat kinerja dengan ukuran error	8
Tabel 2.3 Peringkat kinerja dengan ukuran akurasi	9
Tabel 3.1 Detail dataset	17
Tabel 3.2 Penggalan Data Final	18
Tabel 3.3 Perbandingan Data Training dan Data Testing	19
Tabel 3.4 Skenario experiment	20
Tabel 4.1 Hasil Prediksi Random Forest $n_estimators = 50$ skenario 1	28
Tabel 4.2 Hasil Prediksi Random Forest $n_estimators = 100$ skenario 1	28
Tabel 4.3 Hasil Prediksi Random Forest $n_estimators = 500$ skenario 1	29
Tabel 4.4 Hasil Prediksi Random Forest $n_estimators = 1000$ skenario 1	29
Tabel 4.5 Rank <i>Variable Important</i> Random Forest skenario 1	30
Tabel 4.6 Performance metode Random Forest Skenario 1	31
Tabel 4.7 Hasil Prediksi Random Forest $n_estimators = 50$ skenario 2	32
Tabel 4.8 Hasil Prediksi Random Forest $n_estimators = 100$ skenario 2	32
Tabel 4.9 Hasil Prediksi Random Forest $n_estimators = 500$ skenario 2	33
Tabel 4.10 Hasil Prediksi Random Forest $n_estimators = 1000$ skenario 2	33
Tabel 4.11 Rank <i>Variable Important</i> Random Forest skenario 2	34
Tabel 4.12 Performance metode Random Forest Skenario 2	35
Tabel 4.13 Hasil Prediksi Random Forest $n_estimators = 50$ skenario 3	36
Tabel 4.14 Hasil Prediksi Random Forest $n_estimators = 100$ skenario 3	36
Tabel 4.15 Hasil Prediksi Random Forest $n_estimators = 500$ skenario 3	37
Tabel 4.16 Hasil Prediksi Random Forest $n_estimators = 1000$ skenario 3	38
Tabel 4.17 Rank <i>Variable Important</i> Random Forest skenario 3	38
Tabel 4.18 Performance metode Random Forest Skenario 3	40
Tabel 4.19 Hasil Prediksi Random Forest $n_estimators = 50$ skenario 4	40
Tabel 4.20 Hasil Prediksi Random Forest $n_estimators = 100$ skenario 4	41
Tabel 4.21 Hasil Prediksi Random Forest $n_estimators = 500$ skenario 4	42
Tabel 4.22 Hasil Prediksi Random Forest $n_estimators = 1000$ skenario 4	42
Tabel 4.23 Rank <i>Variable Important</i> Random Forest skenario 4	43

Tabel 4.24 Performance metode Random Forest Skenario 4.....	44
Tabel 4.25 Evaluasi Performance Metode Random Forest.....	45
Tabel 5.1 Hasil Prediksi Logistic Regression <i>max_iter</i> = 100 skenario 1	52
Tabel 5.2 Hasil Prediksi Logistic Regression <i>max_iter</i> = 500 skenario 1	52
Tabel 5.3 Hasil Prediksi Logistic Regression <i>max_iter</i> = 1000 skenario 1	52
Tabel 5.4 Rank <i>Coefficient</i> Logistic Regression skenario 1	53
Tabel 5.5 Performance metode Logistic Regression Skenario 1	55
Tabel 5.6 Hasil Prediksi Logistic Regression <i>max_iter</i> = 100 skenario 2	55
Tabel 5.7 Hasil Prediksi Logistic Regression <i>max_iter</i> = 500 skenario 2	56
Tabel 5.8 Hasil Prediksi Logistic Regression <i>max_iter</i> = 1000 skenario 2	56
Tabel 5.9 Rank <i>Coefficient</i> Logistic Regression skenario 2	57
Tabel 5.10 Performance metode Logistic Regression Skenario 2	58
Tabel 5.11 Hasil Prediksi Logistic Regression <i>max_iter</i> = 100 skenario 3	59
Tabel 5.12 Hasil Prediksi Logistic Regression <i>max_iter</i> = 500 skenario 3	59
Tabel 5.13 Hasil Prediksi Logistic Regression <i>max_iter</i> = 1000 skenario 3	60
Tabel 5.14 Rank <i>Coefficient</i> Logistic Regression skenario 3	60
Tabel 5.15 Performance metode Logistic Regression Skenario 3	62
Tabel 5.16 Hasil Prediksi Logistic Regression <i>max_iter</i> = 100 skenario 4	63
Tabel 5.17 Hasil Prediksi Logistic Regression <i>max_iter</i> = 500 skenario 4	63
Tabel 5.18 Hasil Prediksi Logistic Regression <i>max_iter</i> = 1000 skenario 4	64
Tabel 5.19 Rank <i>Coefficient</i> Logistic Regression skenario 4	64
Tabel 5.20 Performance metode Logistic Regression Skenario 4	66
Tabel 5.21 Evaluasi Performance Metode Logistic Regression	67
Tabel 6.1 Perbandingan <i>accuracy</i> Metode RF dan LR.....	70
Tabel 6.2 Perbandingan <i>duration</i> Metode RF dan LR.....	72
Tabel 6.3 Perbandingan <i>MAE</i> dan <i>MAPE</i> Metode RF dan LR.....	74
Tabel 6.4 Performance Rata-rata, Min, Max dan Std berdasarkan omset.....	76

ABSTRAK

Yakin, Erfan Ainul, 2023, Prediksi Omset Bisnis Restoran Soto-Kwali Pak Wasis Menggunakan Metode Random Forest Dan Logistic Regression, Program Magister Informatika Universitas Islam Negeri Maulana Malik Ibrahim, Pembimbing: (1) Dr. Ririen Kusumawati, S.Si., M.Kom (2) Dr. Usman Pagalay, M.Si

Kata Kunci : Prediksi, Omset, Random Forest, Logistic Regression

Perusahaan kecil yang berhubungan dengan manajemen memerlukan pendekatan yang berbeda terutama untuk kualitas mutu. Kualitas mutu sangat berhubungan dengan administrasi sistem yang dipakai oleh perusahaan kecil tersebut. Penelitian ini dilakukan untuk memprediksikan tingkat pendapatan dari bisnis restoran Soto Kwali Pak Wasis dengan menggunakan Machine Learning yang mana bisnis restoran Soto Kwali Pak Wasis sekarang berskala kecil. Metode *Random Forest* (RF) dipilih karena mampu melakukan prediksi dengan hasil yang optimal dengan akurasi yang tinggi. Sedangkan metode alternatif *Logistic Regression* (LR) dipilih karena kemampuan yang dapat menghasilkan prediksi dengan dengan durasi yang cepat. Hasil dari uji coba prediksi antara kedua metode menghasilkan metode terbaik diraih oleh metode *Logistic Regression* (LR) dengan perolehan nilai *accuracy* 97% dengan *duration* 10s dan level error yang paling rendah mendapat nilai *MAE* 0,22273, *MAPE* 22,2424. Dengan demikian metode *Logistic Regression* (LR) merupakan metode yang paling cocok dibandingkan dengan metode *Random Forest* (RF) karena metode *Logistic Regression* (LR) sangat efisien dengan nilai akurasi yang baik dan sangat cocok untuk melakukan prediksi omset bisnis restoran Soto Kwali Pak Wasis dengan kecepatan waktu yang cepat dan optimasi metode sangat sederhana sehingga tidak memerlukan beban biaya pengadaan yang tinggi.

ABSTRACT

Yakin, Erfan Ainul, 2023, Prediction of Revenue Restaurant's Business Soto-Kwali Pak Wasis Using Random Forest and Logistic Regression, Master Study in Computer Science, State Islamic University Maulana Malik Ibrahim, Advisors: (1) Dr. Ririen Kusumawati, S.Si.,M.Kom (2) Dr. Usman Pagalay, M.Si

Keywords : Prediction, Revenue, Random Forest, Logistic Regression

Small companies dealing with management require a different approach, especially to quality quality. Quality quality is closely related to the administration of the system used by the small company. This research was conducted to predict the level of income from the Soto Kwali Pak Wasis restaurant business using Machine Learning where the Soto Kwali Pak Wasis restaurant business is now small-scale. The Random Forest (RF) method was chosen because it is able to predict optimal results with high accuracy. The alternative method of Logistic Regression (LR) was chosen because of the ability to produce predictions with a fast duration. The results of the prediction trial between the two methods resulted in the best method achieved by the Logistic Regression (LR) method with an accuracy value of 97% with a duration of 10s and the lowest error level received an MAE value of 0.22273, MAPE 22.2424. Thus, the Logistic Regression (LR) method is the most suitable method compared to the Random Forest (RF) method because the Logistic Regression (LR) method is very efficient with good accuracy values and is very suitable for predicting the turnover of the Soto Kwali Pak Wasis restaurant business with fast time speed and method optimization is very simple so that it does not require a high procurement cost burden.

الملخص

اليقين، عرفان عين، ٢٠٢٣، التنبؤ بدخل عمل تجاري لمطعم سوتو كوالي السيد وسيس باستخدام طريقة الغابة العشوائية والانحدار اللوجستي، قسم الهندسة المعلوماتية، كلية الدراسات العليا بجامعة مولانا مالك إبراهيم الإسلامية الحكومية مالانج، المشرف الأول: د. ريرين كوسوماواتي، الماجستير. المشرف الثاني: د. عثمان باغالالي، الماجستير.

الكلمات الرئيسية: التنبؤ، الدخل، الغابة العشوائية، الانحدار اللوجستي.

تتطلب الشركة الصغيرة التي تتعامل مع الإدارة مدخلا مختلفا خاصة لترقية جودتها. ترتبط ترقية الجودة ارتباطا وثيقا بإدارة النظام الذي تستخدمه الشركة الصغيرة. تم إجراء هذا البحث للتنبؤ بمستوى الدخل من عمل تجاري لمطعم سوتو كوالي السيد وسيس باستخدام الآلي التعليمي حيث كان المطعم صغير الحجم. تم اختيار طريقة الغابة العشوائية (RF) لأنها قادرة على التنبؤ بأفضل النتائج بدقة عالية. بينما تم اختيار الطريقة البديلة للانحدار اللوجستي (LR) بسبب القدرة على إنتاج تنبؤات بمدى سرعة. أسفرت نتائج تجربة التنبؤ بين الطريقتين عن أفضل طريقة حققتها طريقة الانحدار اللوجستي (LR) بقيمة دقة ٩٧% مع مدة ١٠ ثوان وحصل أدنى مستوى خطأ على قيمة MAE ٠.٢٢٢٧٣ و $MAPE$ ٢٢.٢٤٢٤. وبالتالي، فإن طريقة الانحدار اللوجستي (LR) هي الطريقة الأنسب مقارنة بطريقة الغابة العشوائية (RF) لأن طريقة الانحدار اللوجستي (LR) فعالة للغاية مع قيمة دقة جيدة ومناسبة جدا للتنبؤ بدخل عمل تجاري لمطعم سوتو كوالي السيد وسيس بسرعة زمنية عالية وعمليتها بسيطة للغاية بحيث لا تتطلب تكاليف شراء عالية.

BAB I

PENDAHULUAN

1.1. Latar Belakang

Azimovna *et al.* (2022) menyatakan bisnis pada perusahaan kecil yang berhubungan dengan manajemen memerlukan pendekatan yang berbeda terutama untuk kualitas mutu. Kualitas mutu sangat berhubungan dengan administrasi sistem yang dipakai oleh perusahaan kecil tersebut. Pengusaha pada perusahaan kecil selalu terfokus pada hal eksternal dan memenuhi permintaan pasar sehingga pengetahuan manajemen dan administrasi sering terabaikan.

Restoran soto kwali Pak Wasis merupakan salah satu dari bisnis yang dimulai dari skala yang kecil dan sekarang sudah berkembang sampai mempunyai 5 cabang. Restoran ini berlokasi di Surabaya dan Sidoarjo dengan skala usaha menengah masing-masing restoran. Struktur bisnis yang dipakai restoran ini masih manual dan sulit untuk melakukan analisa data. Jumlah penjualan pada restoran ini sangat tinggi dan restoran buka atau melayani pelanggan setiap hari.

Pemilik dari restoran soto kwali Pak Wasis telah menjalin kerja sama dengan salah satu pihak pengembang aplikasi administrasi sederhana untuk mencatat jumlah item makanan yang terjual. Dalam aplikasi administrasi tersebut fitur untuk laporan hanya tersedia dalam bentuk rekapitulasi sehingga untuk evaluasi omset yang telah didapat sulit untuk dilakukan.

Data yang diambil dari aplikasi administrasi soto kwali Pak Wasis merupakan data mentah berbentuk jumlah masing-masing item makanan yang terjual. Zahoor *et al.* (2020) menjelaskan teknik dasar untuk mengolah data yang dapat dianalisa menjadi prediksi dapat menggunakan Machine Learning. Prediksi menggunakan data kategori level omset untuk analisa pendapatan dari kinerja bisnis tersebut.

Prediksi pada masa sekarang sangat diperlukan untuk berbagai bidang terutama bidang bisnis. Sirisha *et al.* (2022) menjelaskan Analisa dari pengukuran nilai pendapatan suatu bisnis merupakan tolak ukur yang penting dalam menilai perkembangan suatu bisnis agar mampu bertahan dan dapat berkembang. Kumar *et al.* (2021) dalam penelitiannya prediksi omset menggunakan Machine Learning

sangat penting untuk pertumbuhan dan omset yang lebih baik. Teknologi Machine learning dan Data Science berpengaruh terhadap penentuan algoritma atau metode yang digunakan untuk prediksi dapat ditentukan lebih cepat.

Salah satu algoritma prediksi dalam Machine Learning adalah Random Forest dengan menggunakan parameter *Gini* dan *Entropy*. Proses pada algoritma Random Forest menghasilkan tingkat akurasi tinggi pada dataset dasar tanpa mengubah dimensi atau attribute dari dataset tersebut, tetapi algoritma Random Forest membutuhkan spesifikasi komputer yang tinggi karena kalkulasi yang diproses melalui banyak tahap sesuai yang dijelaskan Phase (2019).

Algoritma alternatif dengan kalkulasi yang lebih ringan sehingga lebih lebih cepat dalam menunjukkan hasil yaitu Logistic Regression seperti disinggung Kuhle *et al.*(2018). Dalam penelitian Kuhle *et al.* (2018) tentang prediksi menggunakan algoritma Logistic Regression menghasilkan akurasi yang baik, tetapi algoritma ini memerlukan korelasi antar variabel independen supaya mendapatkan hasil yang presisi.

Umam *et al.*(2019) dalam penelitiannya tentang prediksi yang sesuai dengan anjuran agama Islam sehingga dapat berinvestasi berdasarkan dengan data pengetahuan yang telah ada, sesuai dengan penggalan ayat Al Qur'an surat Al Hasyr 18 sebagai berikut :

يَا أَيُّهَا الَّذِينَ ءَامَنُوا اتَّقُوا اللَّهَ وَارْتَقُوا إِلَيْهِ فَلَسْ مَا قَدَّمْتُمْ لِعَدِيبِهِ وَاتَّقُوا اللَّهَ إِنَّ اللَّهَ خَبِيرٌ بِمَا تَعْمَلُونَ

Wahai orang-orang yang beriman! Bertakwalah kepada Allah dan hendaklah setiap orang memperhatikan apa yang telah diperbuatnya untuk hari esok (akhirat), dan bertakwalah kepada Allah. Sungguh, Allah Mahateliti terhadap apa yang kamu kerjakan.

Tafsir dari surat Al Hasyr sesuai dengan artikel Irfan *et al.*(2022) dengan rujukan buku tafsir Turjuman al-Mustafid menjelaskan orang yang beriman dan bertakwa kepada Allah dan hidup dimasa kini harusnya mempelajari dan memperhatikan masa kini yang dapat dijadikan acuan untuk mempersiapkan hari esok. Merujuk pada tafsir ayat surat Al Hasyr, prediksi model yang dilakukan pada penelitian ini sangat berkesinambungan dengan tuntunan Allah kepada umat

manusia yang beriman dan bertakwa supaya dapat mempelajari dan memperhatikan dari kondisi data yang telah didapat atau masa kini untuk menjadikan dasar acuan penentuan rencana hari yang akan datang supaya mendapatkan hasil yang lebih baik.

Dalam tuntunan agama Islam serta dengan pesatnya berkembangnya teknologi maka perlunya dalam bisnis supaya tetap berkembang dari masa ke masa maka penulis melakukan penelitian tentang prediksi omset bisnis restoran soto kwali pak wasis menggunakan metode Random Forest dan Logistic Regression diharapkan mampu memberikan analisa dengan optimasi terbaik sesuai dengan kelebihan dan kekurangan masing-masing metode.

1.2. Pernyataan Masalah

Pernyataan masalah dalam penelitian ini yang akan dibahas meliputi :

1. Bagaimana mendapatkan hasil prediksi omset dari bisnis restoran soto kwali Pak Wasis dengan proses yang paling efisien?.
2. Bagaimana evaluasi hasil akurasi dan durasi proses metode Random Forest dan Logistic Regression dalam memprediksikan omset bisnis restoran soto kwali Pak Wasis?.

1.3. Tujuan Penelitian

Sesuai dengan pernyataan masalah dalam penelitian ini, maka tujuan penelitian untuk menjawab dari pernyataan masalah yaitu :

1. Menganalisa hasil prediksi dengan mengolah data penjualan makanan pada bisnis restoran soto kwali Pak Wasis untuk mendapatkan hasil yang optimal dengan durasi proses yang paling efisien.
2. Mendapatkan hasil akurasi dan durasi proses metode algoritma prediksi yang paling optimal sehingga dapat dijadikan penetapan metode baru.

1.4. Batasan Masalah

Penelitian ini menggunakan Batasan masalah yang terdiri atas :

1. Input variabel data yang digunakan dalam penelitian ini terbatas pada bisnis restoran soto kwali Pak Wasis yang mencakup nilai omset harian.
2. Output dalam penelitian ini dibatasi oleh 3 level yaitu tinggi, standart dan rendah
3. Data yang digunakan dalam penelitian ini terbatas pada kondisi 6 bulan kebelakang dari maret 2023.

1.5. Manfaat Penelitian

Manfaat dari penelitian ini sebagai berikut :

1. Untuk mengetahui hasil prediksi omset dari bisnis restoran soto kwali pak wasis.
2. Ditemukannya metode yang terbaik dan sesuai dengan kondisi data yang disediakan oleh bisnis restoran soto kwali pak wasis.
3. Menjadi bahan kajian untuk menganalisa kinerja dari pendapatan omset bisnis restoran soto kwali pak wasis.

BAB II

TINJAUAN PUSTAKA

2.1. Prediksi dengan Machine Learning

Alam *et al.* (2019) memiliki algoritma dalam prediksi model menggunakan Machine Learning, terdiri atas diantaranya Linear Regression, REPTree, Locally Weighted Regression, Piecewise Regression, Logistic Regression. Algoritma ini merupakan jenis-jenis algoritma yang bisa diterapkan dalam prediksi data. Algoritma tersebut merupakan partisi dari algoritma dan balanced data dengan metode rekursif ganda dengan Teknik ensemble. Hasil dari pengujian algoritma Logistic Regression dalam imbalanced data menghasilkan Teknik rekursif yang efektif dan meluas ke Analisa regresi dengan nilai error 0.01043 pada pengujian dataset compactive, dibandingkan dengan nilai Liner Regression error 0.21557, REPTree error 0.01328, LWR error 0.0344, Piecewise Regression error 0.0943.

Ghorbani *et al.* (2020) memiliki Random Forest yang merupakan salah satu algoritma dalam Machine Learning untuk memprediksikan dataset. Pengujian menggunakan metode SVM-SMOTE dalam pengambilan sampel, menjadikan Random Forest menjadi algoritma yang efisien dengan performa yang bagus dalam memproses prediksi yang berupa dataset minoritas dan mayoritas dengan nilai akurasi mencapai 73% dibandingkan dengan hasil KNN 68.70%, ANN 67.84%, XG-boost 72%, SVM 67.56%, LR 59.35%, Decision Tree 67.56%, naïve bayes 54.49%. Tetapi diperlukan penelitian lebih lanjut menggunakan peningkatan hybrid algoritma untuk menghasilkan prediksi yang lebih bagus.

Bajaj *et al.* (2020) menerangkan prediksi dalam bisnis penjualan retail menggunakan machine learning sangat efektif. Algoritma yang dipakai yaitu Logistic Regression, KNN dan Random Forest. Akurasi tertinggi diperoleh oleh algoritma random forest mencapai 93.5%.

Chen *et al.* (2021) menjelaskan teknologi machine learning dalam prediksi dataset yang didistribusikan dengan peer-to-peer dapat dianalisa. Algoritma Random Forest menjadi algoritma dengan nilai akurasi tinggi mencapai 65% dengan Teknik pengambilan sampel *RUS* atau *Random UnderSampling*

dibandingkan dengan algoritma Neural Network 58%, dan Logical Regression 60%. Analisa hanya sebatas menggunakan dataset yang diambil dengan Teknik *RUS* atau *Random UnderSampling*.

Bujang *et al.* (2021) menerangkan penerapan prediksi menggunakan Machine Learning sangat populer menggunakan Algoritma Random Forest. Random Forest menghasilkan nilai akurasi 99.5%. Hasil tersebut sangat baik bila dibandingkan dengan algoritma lain diantaranya : Decision Tree (J48) 99.3%, Support Vector Machine (SVM) 98.9%, Naïve Bayes (NB) 97.5%, K-Nearest Neighbor (kNN) 99.3%, Regresi Logistik (LR) 98.6%. Dengan begitu algoritma Random Forest menunjukkan hasil yang sebanding dan menjanjikan dalam prediksi data.

Emmanuel *et al.* (2021) memiliki Machine Learning dengan Algoritma Random Forest merupakan algoritma yang populer. Dalam Analisa studi kasus yang lain algoritma tersebut digunakan untuk memproses prediksi data dengan metode pengambilan sampel RF imputation RSME. Selain algoritma Random Forest terdapat algoritma KNN dalam perbandingan hasil uji data. Secara umum memang algoritma Random Forest sangat baik dalam menangani missing data tersebut dengan nilai performance 0.0549, dan KNN 0.2099 dengan dataset 5% tetapi bila data set diatas 10% hasilnya KNN 0.2382 dan RF 0.2860. Sehingga kesimpulan dari penelitian dalam jurnal ini disebutkan presisi dan akurasi algoritma tergantung pada jenis data yang dianalisa dan ada tidaknya metode lain yang mendukung algoritma tersebut.

Bagui (2021) memiliki analisa presisi algoritma tergantung pada jenis data dan metode lain yang mendukung. Algoritma ANN atau artificial Neural Network juga bisa efektif dalam menangani prediksi data, tentunya dengan metode pengambilan sampel yang mendukung algoritma tersebut. ANN bisa mendapatkan hasil yang efektif dengan metode pengambilan sampel menggunakan RURO dan RU-SMOTE. ANN dengan metode sampel RURO dengan nilai presisi 89% dan RU-SMOTE dengan nilai presisi 90%.

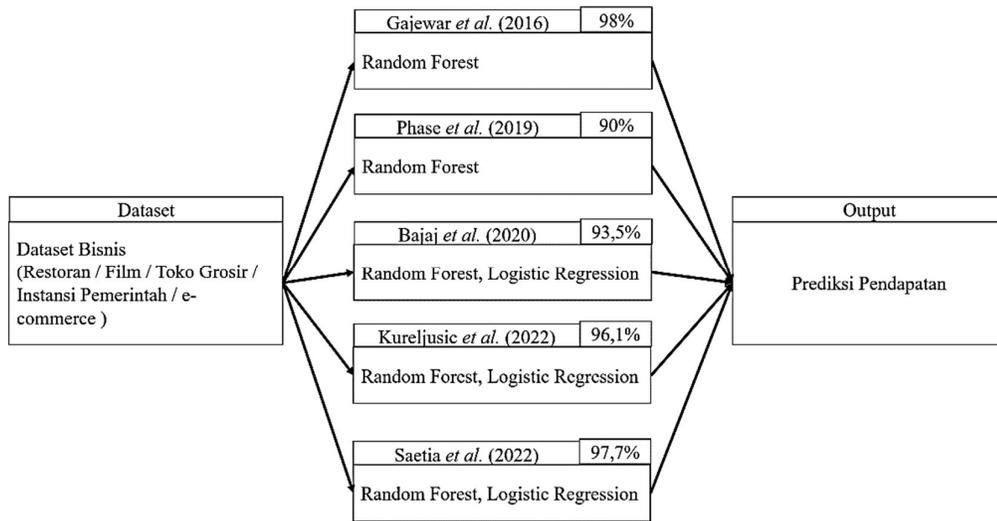
2.2. Kerangka Teori Prediksi pada Machine Learning

Penelitian dengan topik prediksi menggunakan machine learning dirangkum dalam tabel 2.1 yang menjadi acuan penulis dalam menentukan posisi penelitian. Tabel tersebut menjelaskan topik, dataset yang digunakan, metode dan performance dari masing-masing penelitian.

Tabel 2.1 Ringkasan Jurnal

Peneliti	Topik	Dataset	Metode	Performance
Gajewar <i>et al.</i> (2016)	Prediksi pendapatan produksi film	Data Lokasi Produser	Random Forest	98%
Raul <i>et al.</i> (2016)	Prediksi pendapatan restoran	Data Restoran	Random Forest	90%
Phase <i>et al.</i> (2019)	Prediksi pendapatan pegawai	pegawai pemerintah	Random Forest	90%
Singh <i>et al.</i> (2020)	Prediksi pendapatan e-commerce	Data website e-commerce	Random Forest	90.99%
Bajaj <i>et al.</i> (2020)	Prediksi pendapatan toko retail	Supermarket dan toko grosir	Random Forest Logistic Regression	93.5%
Siddamsetty <i>et al.</i> (2021)	Prediksi pendapatan restoran	Data restoran	Random Forest	91.50%
Kureljusic <i>et al.</i> (2022)	Prediksi pendapatan film	Produce Film	Random Forest Logistic Regression	96.15%
Lin <i>et al.</i> (2022)	Prediksi pendapatan hydropower	Data Hydropower	Logistic Regression	94%
Saetia <i>et al.</i> (2022)	Prediksi pendapatan restoran	Data jumlah stok makanan	Random Forest Logistic Regression	97.70%
Sahinbas <i>et al.</i> (2022)	Prediksi pendapatan restoran	Data restoran	Random Forest	98.09%

Kerangka teori menunjukkan gambaran tentang proses yang berkaitan dalam prediksi. Ada 3 parameter yang dipakai dalam theoretical framework yaitu dataset, proses dan output. theoretical framework ditunjukkan dalam gambar 2.1.



Gambar 2.1 : Kerangka Teori Prediksi dalam Machine Learning

Pengurutan hasil kinerja dengan ukuran error ditunjukkan pada Tabel 2.2. Error yang paling kecil merupakan ranking paling tinggi yang diperoleh dalam literatur Alam *et al* (2019) menggunakan metode Logistic Regression dengan nilai error 0.01043. Metode tersebut merupakan salah satu metode yang akan dianalisa dan dibandingkan dalam penelitian ini.

Tabel 2.2 Peringkat kinerja dengan ukuran error

Rank	Metode	Error
1	Alam <i>et al.</i> (2019) - Logistic Regression	0.01043
2	Emmanuel <i>et al.</i> (2021) - Random Forest	0.0549
3	Emmanuel <i>et al.</i> (2021) - K Nearest Neighbor (KNN)	0.2099
4	Alam <i>et al.</i> (2019) - Linear Regression	0.21577

Pengurutan hasil kinerja dengan ukuran akurasi ditunjukkan pada Tabel 2.3. Nilai akurasi yang mendekati angka 100 merupakan ranking akurasi paling tinggi. Bujang *et al* (2021) mendapatkan skor akurasi 99.50% dengan metode Random Forest menjadikan metode selanjutnya yang akan dibahas dalam penelitian ini.

Tabel 2.3 Peringkat kinerja dengan ukuran akurasi

Rank	Metode	Accuracy
1	Bujang <i>et al.</i> (2021) - Random Forest	99.50%
2	Bujang <i>et al.</i> (2021) - Decision Tree	99.30%
3	Bujang <i>et al.</i> (2021) - K Nearest Neighbor (KNN)	99.30%
4	Bujang <i>et al.</i> (2021) - Logistic Regresion	98.60%
5	Bajaj <i>et al.</i> (2020) - Random Forest	93.50%
6	Ghorbani <i>et al.</i> (2020) - Random Forest	73.00 %
7	Ghorbani <i>et al.</i> (2020) - K Nearest Neighbor (KNN)	68.70%
8	Ghorbani <i>et al.</i> (2020) - Artificial Neural Network (ANN)	67.84%
9	Ghorbani <i>et al.</i> (2020) - Support Vector Machine (SVM)	67.56%
10	Ghorbani <i>et al.</i> (2020) - Decision Tree	65.56%
11	Chen <i>et al.</i> (2020) - Random Forest	65.00%
12	Chen <i>et al.</i> (2020) - Logistic Regresion	60.00%
13	Ghorbani <i>et al.</i> (2020) - Logistic Regresion	59.35%

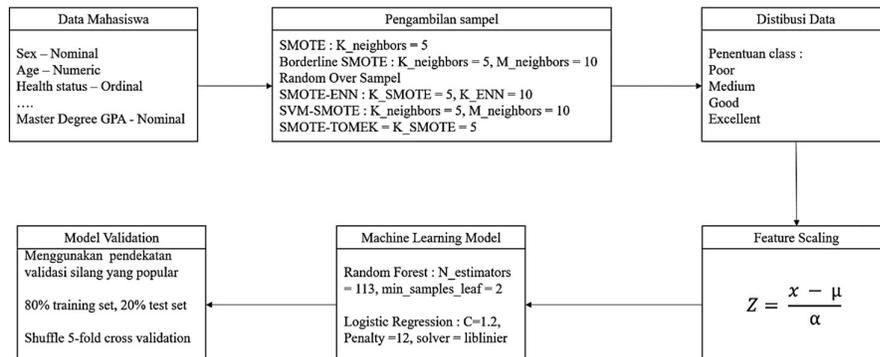
Diagram teoritical framework memuat nilai performance masing-masing metode. Performance diukur menggunakan nilai error dan akurasi, sehingga mengacu dalam Tabel 2.2 dan 2.3 dapat disimpulkan metode yang mendapat ranking 1 akan menjadi metode dalam prediksi. Oleh karena itu dalam penelitian ini akan ditindaklanjuti dengan membandingkan dua metode yaitu Logistic Regression dengan Error 0.01043 dan Random Forest dengan akurasi 99.50%.

2.3. Prediksi Algoritma Random Forest

Penerapan prediksi dengan metode Random Forest diterapkan untuk mengetahui performance dari hasil uji data dengan indikator akurasi, presisi, recall dan F1 score. Indikator tersebut menjadi dasar dalam analisa prediksi yang dapat digunakan dalam menyimpulkan keberhasilan metode melakukan training dataset.

Ghorbani *et al.* (2020) memiliki penelitian yang membahas prediksi dengan salah satunya menggunakan metode Random Forest dengan dataset yang diambil dari mahasiswa perguruan tinggi untuk menentukan kinerja mahasiswa dalam proses belajar. Tipe dari dataset tersebut berupa numerik dan ordinary atau text. Proses yang dilakukan dalam prediksi dimulai dari pengambilan sampel dan dilakukan analisa distribusi data dengan menetapkan class dalam dataset,

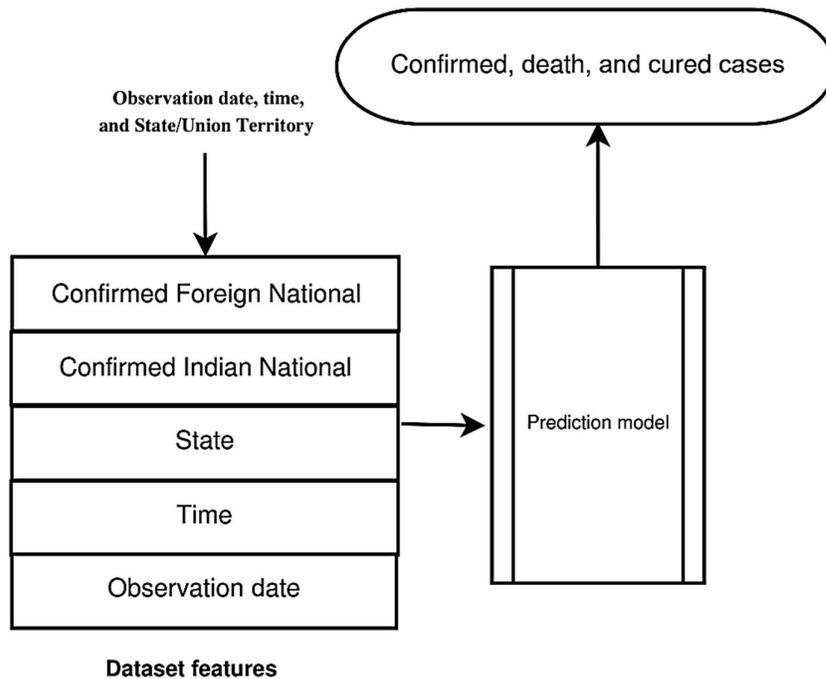
selanjutnya dilakukan feature scaling, penerapan algoritma model dan melakukan evaluasi model atau model validation, sehingga hasil dari proses tersebut adalah prediksi kinerja mahasiswa. Skema alur ditunjukkan pada Gambar 2.2.



Gambar 2.2 : Skema Model Random Forest Prediksi Kinerja Mahasiswa
(Sumber : diolah dari Bujang *et al.* 2021)

Bujang *et al.* (2021) menerangkan prediksi yang digunakan untuk menentukan tingkatan nilai mahasiswa menggunakan input yang berupa data mahasiswa yang terdiri atas angka dan text. Dalam proses model juga menggunakan indikator akurasi, presisi, recall dan F1 score. Dan hasil atau output dari proses tersebut berupa prediksi dari nilai mahasiswa.

Gupta *et al.* (2021) memiliki prediksi dengan algoritma Random Forest untuk menentukan kematian pada kasus covid-19. Dataset yang digunakan dalam penelitian ini merupakan dataset dari kasus covid-19 yang berupa numerik. Proses yang dilakukan dengan melakukan pengambilan sampel dan penerapan model dan evaluasi model menggunakan k-fold cross validation dengan inisialisasi k=7, indikator yang digunakan terdiri atas : akurasi dan error. Hasil atau outputnya merupakan prediksi dari tingkat kematian dari kasus covid-19. Skema alur kinerja algoritma ditunjukkan pada Gambar 2.3.



Gambar 2.3 : Skema Model Random Forest Prediksi Kematian COVID-19
(Sumber : diolah dari Gupta *et al.*2021)

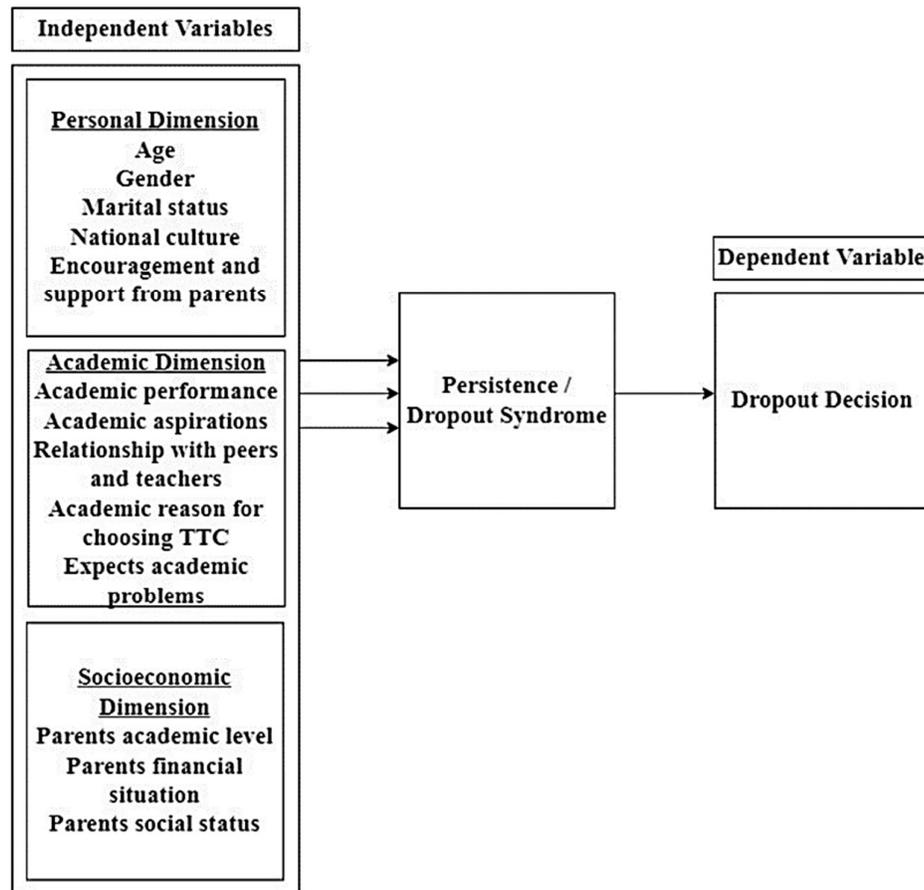
Penerapan algoritma Random Forest dalam prediksi sangat efektif untuk dilakukan analisa karena algoritma ini masih populer hingga sekarang dalam melakukan prediksi.

2.4. Prediksi Algoritma Logistic Regression

Prediksi menggunakan algoritma Logistic Regression mengacu pada perhitungan statistika. Logistic Regression membutuhkan variabel independent dan variabel dependent. Variabel independent dalam penerapan Logistic Regression perlu diubah menjadi numerik atau angka bila data berbentuk kategorikal, dan variabel dependent juga akan berbentuk nilai angka.

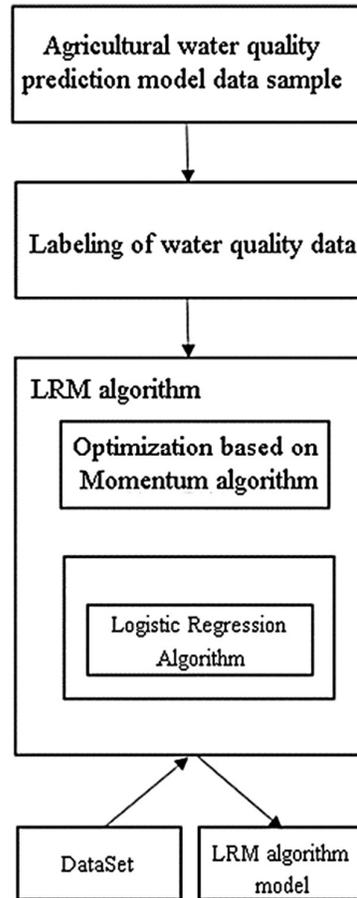
Singh *et al.* (2021) memiliki prediksi tentang drop-out siswa dengan menggunakan Logistic Regression korelasi prediksi yang baik terhadap variabel independent. Variabel independent yang dipakai diantaranya : umur, jenis kelamin, status kawin, kondisi orang tua, pendapatan orang tua dan lain sebagainya. Pada proses prediksi uji data yang menghasilkan nilai R (korelasi antar independent

variabel) dibandingkan dengan nilai B (koefisien Logistic Regresion) sehingga dapat menghasilkan prediksi dropout siswa. Skema model ditunjukkan pada gambar 2.4.



Gambar 2.4 : Skema Model Logistic Regression Prediksi Resiko Drop-out siswa.
(Sumber : diolah dari Singh *et al.*2021)

Gai *et al.* (2023) menerangkan prediksi dalam analisa momentum air untuk irigasi menggunakan Logistic Regression mendapatkan hasil yang maksimal. Variabel yang saling terkait diantaranya : kadar oksigen, temperature, residu clorine dan PH air. Korelasi dari uji data menghasilkan nilai positif atau negatif yang digunakan dalam melakukan prediksi menentukan kualitas air. Hasil prediksi mendapatkan tingkat error yang sangat kecil. Skema model ditunjukkan pada gambar 2.5.



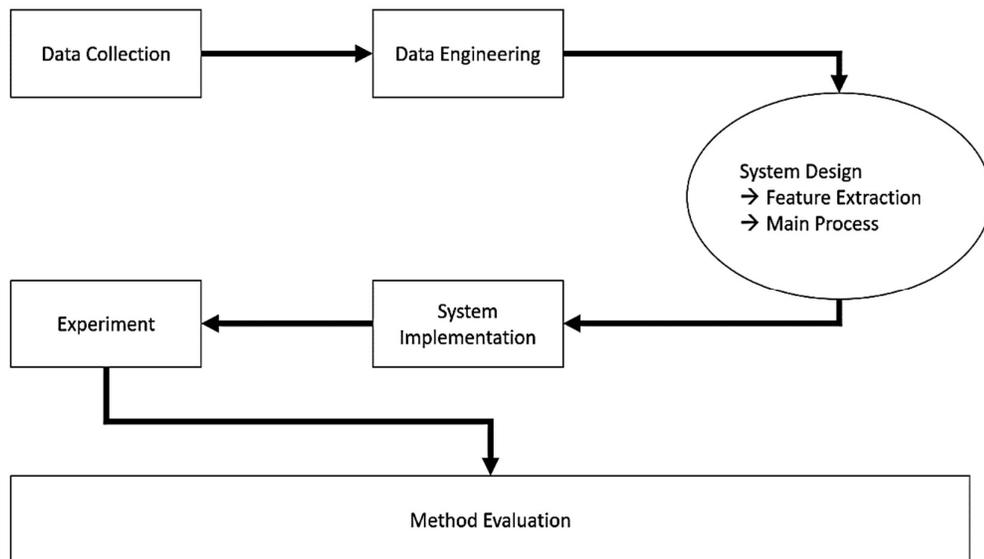
Gambar 2.5 : Skema Model Logistic Regression Prediksi momentum air irigasi.
(Sumber : diolah dari Gai *et al.*2023)

Metode dengan menggunakan algoritma Logistic Regression sangat efektif, cepat dan ringan sehingga metode ini sangat sering digunakan untuk melakukan data sains yang berguna dalam menunjang melakukan prediksi.

BAB III DESAIN PENELITIAN

3.1. Riset Prosedur

Riset Prosedur dalam penelitian ini melakukan beberapa tahapan, setiap tahapan masih berhubungan dengan yang lain dan meski dalam tahapan awal belum tercapai secara maksimal maka tetap bisa dilakukan tahapan selanjutnya



Gambar 3.1 : Diagram Riset Prosedur

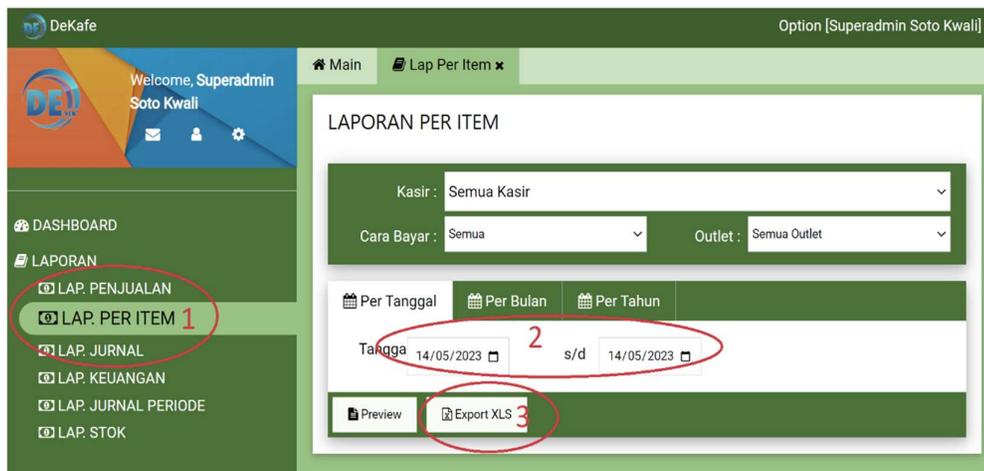
Tahapan riset prosedur sebagai berikut:

1. Menentukan dan mengumpulkan data yang akan digunakan dalam penelitian ini. Data yang akan dipakai dengan jumlah lebih dari 50 data, dan data tersebut akan dilakukan pre processing terlebih dahulu.
2. Menyiapkan data dan preprocessing data supaya data tersebut cocok dan efektif dalam proses dengan algoritma yang telah ada. Dalam tahap ini data akan memasuki proses filtering dan membuang data yang tidak diperlukan.
3. Merancang system merupakan kegiatan dalam memilah feature dari data yang akan dipakai, kegiatan selanjutnya adalah implementasi algoritma yang sudah ditentukan peneliti dengan data yang sudah siap.

4. Pengujian rancangan sistem dalam dunia nyata menggunakan perangkat dan aplikasi yang telah ditentukan dengan batasan-batasan sesuai dengan fungsi dari penelitian ini.
5. Dalam experiment menjelaskan perbandingan hasil dari sistem implementation dengan kondisi nyata dalam kehidupan dan juga melakukan perhitungan manual dengan data yang lebih sedikit untuk memastikan hasil sistem implementation sudah sesuai.
6. Dari proses 2 algoritma yang telah diuji, maka akan dilakukan tahap perbandingan antara 2 algoritma tersebut. Hasilnya akan dievaluasi dan dapat ditarik kesimpulan untuk penelitian yang akan datang.

3.2. Data Collection

Data dalam penelitian ini menggunakan data omset harian di bisnis restoran soto kwali Pak Wasis. Data tersebut adalah data primer yang diambil pada bulan maret 2023. Data diambil dengan akses dari sistem laporan dari instansi terkait dan diexport pertanggal dari bulan maret 2023 mundur 6 bulan ke belakang sampai oktober 2022.



Gambar 3.2 : Cara pengumpulan data

Pada gambar 3.2 terdapat alur dalam pengambilan data setelah berhasil login pada sistem di instansi terkait. Alur pengambilan data pada sistem terkait sebagai berikut :

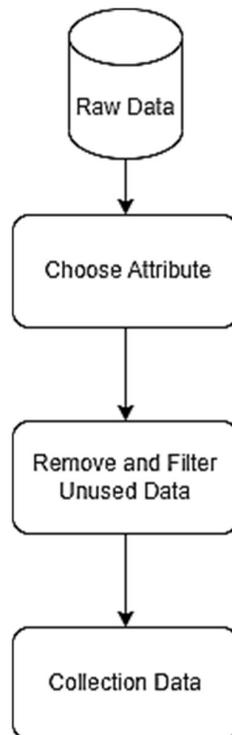
1. Pilih menu laporan di masing-masing aplikasi
2. Tentukan tanggal mulai dan tanggal pasangannya
3. Klik export xls dan akan mengunduh file yang berbentuk xls.

Data diambil secara online yang berjumlah kurang lebih 180 data.

Komponen input dari data tersebut merupakan nilai omset yang berbentuk numerik dengan variabel masing-masing tanggal dan jumlah item terjual. Sedangkan untuk output merupakan level dari omset tersebut dibagi menjadi tiga kategori yaitu : tinggi, standart dan rendah.

3.3. Data Engineering

Data Engineering merupakan tahapan yang harus dilakukan oleh peneliti karena data yang dikumpulkan merupakan data primary yang masih mentah. Tahapan data engineering dijelaskan dalam gambar 3.3.



Gambar 3.3 : Diagram data proseding

Tahap pertama yang dilakukan yaitu memilih attribute yang akan digunakan untuk dataset tersebut. Tahap selanjutnya memilih dari masing-masing data mentah

dengan cara dibuang data yang tidak digunakan dan disaring data yang penting untuk penelitian ini. Penjelasan tahapan data engineering sebagai berikut:

1. Raw Data merupakan data mentah yang diambil dari proses collection data.
2. Pemilihan attribute dilakukan dengan tujuan data yang dipakai sesuai dengan main proses sehingga dapat dilakukan test dan training data. Pada tahap ini pemilihan attribute menggunakan tanggal dan item makanan yang terjual bentuk numeric dan memberi label level sesuai dengan penetapan pemilik restoran. Detail dataset ditunjukkan pada tabel 3.1 :

Tabel 3.1 Detail dataset

No	Attribut	Type	Variabel	Ket
1	Tanggal	Date	independen	Tanggal Pembukuan
2	Air Mineral	Numeric	independen	Jumlah Item terjual
3	Black Garlic 30 K	Numeric	independen	Jumlah Item terjual
4	Dawet	Numeric	independen	Jumlah Item terjual
5	Es Batu	Numeric	independen	Jumlah Item terjual
6	Galantin	Numeric	independen	Jumlah Item terjual
7	Jeruk Hangat / Dingin	Numeric	independen	Jumlah Item terjual
8	Kacang Menté	Numeric	independen	Jumlah Item terjual
9	Kerupuk	Numeric	independen	Jumlah Item terjual
10	Kripik Kentang	Numeric	independen	Jumlah Item terjual
11	Nasi Putih	Numeric	independen	Jumlah Item terjual
12	Perkedel	Numeric	independen	Jumlah Item terjual
13	Sate Babat	Numeric	independen	Jumlah Item terjual
14	Sate Paru	Numeric	independen	Jumlah Item terjual
15	Sate Puyuh	Numeric	independen	Jumlah Item terjual
16	Sate Tetelan	Numeric	independen	Jumlah Item terjual
17	Sate Usus	Numeric	independen	Jumlah Item terjual
18	Soto Ayam	Numeric	independen	Jumlah Item terjual
19	Soto Ayam Besar	Numeric	independen	Jumlah Item terjual
20	Soto Ayam bungkus 15 K	Numeric	independen	Jumlah Item terjual
21	Soto Ayam bungkus 20 K	Numeric	independen	Jumlah Item terjual
22	Soto Ayam bungkus 20 KB	Numeric	independen	Jumlah Item terjual
23	Soto ayam kecil	Numeric	independen	Jumlah Item terjual
24	Soto bungkus 12 K	Numeric	independen	Jumlah Item terjual
25	Soto Daging Besar	Numeric	independen	Jumlah Item terjual
26	Soto Daging kecil	Numeric	independen	Jumlah Item terjual
27	Soto Daging Sapi	Numeric	independen	Jumlah Item terjual

Lanjutan Tabel Detail Dataset

No	Attribut	Type	Variabel	Ket
28	Tahu Bakso	Numeric	independen	Jumlah Item terjual
29	Teh Hangat / Dingin	Numeric	independen	Jumlah Item terjual
30	Teh Tawar	Numeric	independen	Jumlah Item terjual
31	Tempe Mendoan	Numeric	independen	Jumlah Item terjual
32	level	Text	dependen	tinggi, standart & rendah

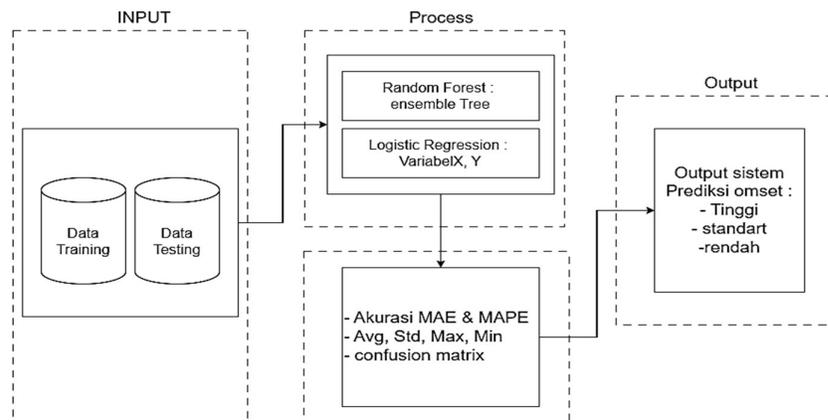
3. Proses filterisasi data merupakan tahap merangkai data menjadi data yang siap untuk dilakukan testing atau filtering. Bentuk data akan ditunjukkan pada tabel 3.2

Tabel 3.2 Penggalan Data Final

Tanggal	Air Mineral	Black Garlic 30 K	Da-wet	Galan-tin	...	The Tawar	Tempe Mendoan	level
2022-10-01	85	2	32	123	...	41	781	Standart
2022-10-02	105	1	36	223	...	62	1238	Tinggi
2022-10-03	24	0	0	54	...	21	337	Rendah
2022-10-04	51	0	19	81	...	27	744	Rendah
2022-10-05	47	0	10	107	...	49	645	Rendah
2022-10-06	44	1	15	89	...	49	585	Rendah
2022-10-07	51	32	7	66	...	41	632	Rendah
2022-10-08	90	1	37	123	...	49	922	Tinggi
2022-10-09	97	1	25	164	...	56	1159	Tinggi
2022-10-10	39	3	21	48	...	11	336	Rendah
2022-10-11	59	0	12	72	...	54	629	Rendah
2022-10-12	56	2	22	71	...	35	657	Rendah

3.4. Desain Sistem

Desain sistem merupakan tahap sistem melakukan prediksi terhadap omset dengan menggunakan algoritma Random Forest (RF) dan algoritma Logistic Regression (LR). Dari hasil proses kedua algoritma akan dibandingkan dan dilakukan analisa untuk menentukan algoritma yang paling baik. Alur dalam penelitian ini ditunjukkan pada gambar 3.4 :



Gambar 3.4 : Desain Sistem

Pada gambar 3.4 tahap awal sistem memakai data training untuk memproses algoritma, selanjutnya system akan menggunakan data testing untuk melakukan akurasi dan perbandingan. Dari akurasi yang dihasilkan sistem akan menentukan prediksi dari dataset tersebut.

3.5. Implementasi Sistem

Sistem akan diimplementasikan dengan menggunakan spesifikasi pada mesin sebagai berikut :

1. CPU memakai 4 core dengan kecepatan 2.5Ghz
2. Ram 16GB
3. Hardisk memakai SATA SSD
4. OS Win10
5. Language Programing menggunakan python.

Data dalam tahap implementasi memakai perbandingan data training dan data testing sesuai pada Tabel 3.3.

Tabel 3.3 Perbandingan Data Training dan Data Testing

No	Percobaan	Perbandingan Data Training dan Testing
1	P1	90:10
2	P2	80:20
3	P3	70:30
4	P4	60:40

Implementasi akan dilakukan beberapa tahapan dan masing-masing tahapan yang lebih detail akan dibahas dalam bab tersendiri.

3.6. Experiment

Tahap experiment dengan menggunakan acuan confusion matrix yang digunakan untuk membandingkan hasil dari prediksi data. Confusion matrix yang dipakai sesuai dengan yang tertera pada gambar 3.5 di bawah ini :

		True Class	
		Positive	Negative
Predicted Class	Positive	True Positive Count (TP)	False Positive Count (FP)
	Negative	False Negative Count (FN)	True Negative Count (TN)

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \quad (3.1)$$

Gambar 3.5 : Ilustrasi Confusion Matrix.

Predicted class akan diisi oleh hasil dari perhitungan data yang telah dilakukan training sedangkan *true class* merupakan kondisi real dari kondisi sesungguhnya.

Experiment memakai beberapa skenario untuk menguji tingkat efektifitas dari metode dengan menggunakan acuan akurasi dan kecepatan metode dalam menghasilkan prediksi. Skenario tersebut dipaparkan dalam tabel 3.4 sebagai berikut :

Tabel 3.4 Skenario experiment

Metode	Skenario	Training : Testing	Modifikasi Metode
Random Forest	S1	90:10	Tree states : 50,100,500,1000
	S2	80:20	
	S3	70:30	
	S4	60:40	
Logistic Regression	S1	90:10	Iterasi : 100,500,1000
	S2	80:20	
	S3	70:30	
	S4	60:40	

Experiment yang akan dilakukan terdiri dari 4 skenario dengan perbedaan jumlah dari training dan testing masing-masing metode. Skenario 1 menggunakan perbandingan training dan testing 90:10, skenario 2 menggunakan perbandingan 80:20, skenario 3 menggunakan perbandingan 70:30 dan skenario 4 menggunakan 60:40. Untuk mengetahui performace dari masing-masing metode maka modifikasi dari Random Forest memakai acuan Tree state yang dimulai dari 50,100,500 dan 1000, sedangkan untuk Logistic Regression memakai acuan Iterasi mulai dari 100,500 dan 1000.

3.7. Evaluation Method

Hasil dari akurasi dan kecepatan proses akan menunjukkan perbandingan dari 2 algoritma yang telah dilakukan training. Pada tahap evaluation method akan dilakukan rekapitulasi hasil dari uji coba dan akan dilakukan Analisa dengan membuat grafik dan tabel. Analisa ini digunakan untuk menunjukkan keberhasilan dari uji coba yang dilakukan oleh 2 algoritma tersebut sehingga dapat digunakan untuk penelitian lebih lanjut.

3.7.1. MAE (Mean Absolute Error)

Nilai akurasi yang telah didapat dari proses metode akan dilakukan evaluasi dengan menggunakan *MAE* supaya rata-rata mutlak dari nilai actual atau nyata dapat dibandingkan dengan nilai prediksi. Menurut Emmanuel *et al.* (2021) rumus *MAE* didefinisikan pada persamaan (3.2) :

$$MAE = \frac{1}{m} \sum_{i=1}^m |y_i - \hat{y}_i| \quad (3.2)$$

Di mana :

m = ukuran sampel

y_i = nilai aktual ke-i

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

3.7.2. MAPE (Mean Absolute Percentage Error)

MAPE mempunyai rumus hampir sama dengan *MAE*, hanya saja *MAPE* menyediakan hasil dalam bentuk persen. Rumus *MAPE* ditambahkan dengan perkalian 100.

$$\text{MAPE} = \frac{1}{m} \sum_{i=1}^m |y_i - \hat{y}_i| \times 100 \quad (3.3)$$

BAB IV METODE RANDOM FOREST

4.1 Desain

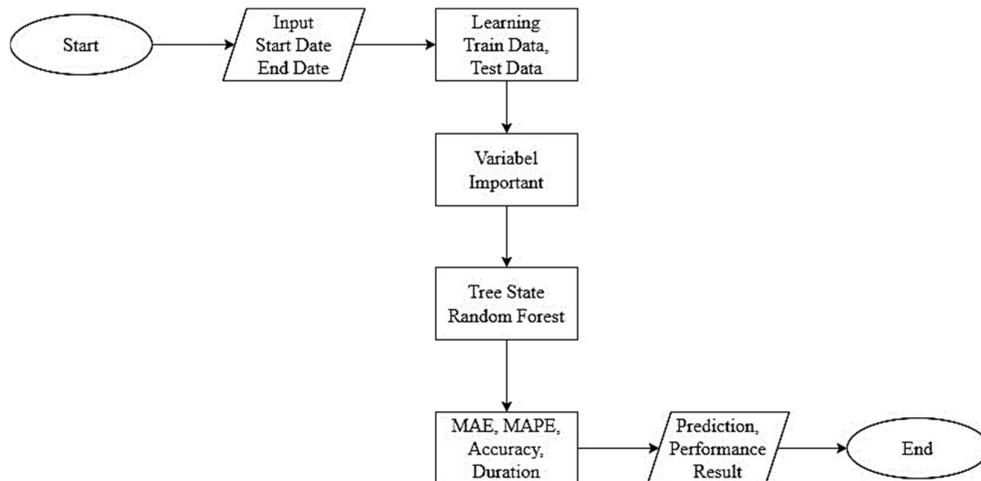
Metode Random Forest dibangun berdasarkan pada perhitungan algoritma *CART* (*Classification and Regression Trees*) yang dilakukan berulang kali dan ditentukan hasil akhir menggunakan voting, Chen *et al.* (2020). Metode *CART* menggunakan ukuran dari persamaan *gini index* yang dijabarkan pada persamaan (4.1) :

$$Gini(S) = \sum_{j=1}^n P_j^2 \tag{4.1}$$

$$Gini_A(S) = \frac{|S_1|}{|S|} Gini(S_1) + \frac{|S_2|}{|S|} Gini(S_2)$$

Di mana P_j adalah peluang data dari variabel j .

Kumpulan dari nilai bobot *gini index* divoting untuk mendapatkan bobot akurasi sehingga hasil prediksi dapat ditemukan. Diagram alur aplikasi prediksi dengan menggunakan metode Random Forest ditunjukkan pada flowchart gambar 4.1 berikut:



Gambar 4.1 : Flowchart Prediksi Model Metode Random Forest
(Sumber : diolah dari Russel. 2018)

Pada gambar 4.1 sistem dimulai dengan memasukkan tanggal awal dan tanggal akhir untuk pengujian prediksi. Parameter tanggal akan disesuaikan dengan data training untuk dilakukan proses *Supervised Learning* sehingga menghasilkan data yang siap untuk dilakukan perhitungan metode Random Forest tersebut. Pada tahap ini bila data training masih belum berbentuk numeric, sistem akan mengubah menjadi numeric.

Tahap Flowchart pada Gambar 4.1 selanjutnya adalah menentukan *Variable Important*. Random Forest yang dibangun menggunakan dasar perhitungan *CART* menggunakan acuan *Mean Decrease in Impurity (MDI)* sebagai dasar dalam melakukan proses perhitungan *Variable Important*. Untuk mendapatkan nilai *MDI* dengan metode menjumlahkan penurunan gangguan dari variabel independen pada proses perhitungan *gini index* yang terkait pada partisi variabel independen tersebut Chaibi *et al.* (2022). Penurunan gangguan dihitung menggunakan persamaan (4.2) :

$$\Delta V(t) = V(t) - p_L V(t_L) - p_R V(t_R) \quad (4.2)$$

Di mana, t_R dan t_L mewakili node anak dari partisi variabel independen dengan peluang p_R dan p_L dengan proposional dari node anak masing-masing.

Menurut Xiao *et al.* (2019) proses dari *MDI* berdasarkan dari nilai partisi variabel independen dilakukan perhitungan rata-rata skor dari seluruh hasil q dari seluruh *tree process*. Rumus perhitungan *MDI* ditunjukkan pada persamaan (4.3) :

$$MDI = \frac{1}{q} \sum_q \sum_{t \in q} p(t) \Delta V(t) \quad (4.3)$$

Di mana, $p(t) = N_t/N$ menunjukkan proporsi dari hasil node t .

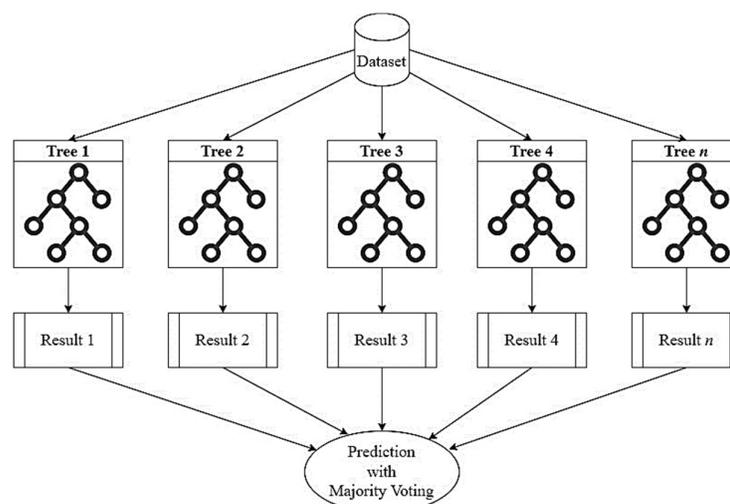
Tahap *Tree State Random Forest* pada flowchart Gambar 4.1 merupakan proses sistem dalam melakukan *customize* metode supaya mendapatkan nilai error dan hasil prediksi yang baik. Menurut Hongyang *et al.* (2022), evaluasi metode Random Forest untuk mendapatkan nilai error dan prediksi yang stabil dengan menggunakan *n_estimators* atau *tree state* pada 50, 100, 500 dan 1000. Pada tahap

ini sistem akan melakukan proses perhitungan *Out-of-bag(OOB)* error sehingga dapat dianalisa perangnya pada *Variable Important*. Persamaan *OOB* error sebagai berikut :

$$J = \frac{1}{n - m} \sum_{i=1}^m (x_1 - x_2)^2 \quad (4.4)$$

di mana, J merupakan *Variable Important*, n merupakan nilai *tree state* dan x_1 dan x_2 merupakan nilai masing *OOB* error dari *tree*.

Proses dari flowchart Random Forest saling terkait untuk mendapatkan hasil prediksi yang maksimal sehingga untuk pada tahap proses *MAE, MAPE, accuracy* dan *duration* dapat diproses dengan baik. Dalam kutipan buku Farnham *et al.*(2019) menjelaskan proses metode Random Forest melakukan perhitungan *gini index* kemudian dilanjutkan dengan menentukan *Feature important* dan melakukan proses perhitungan *OOB* sehingga dapat dihitung *MAE, MAPE* dan *accuracy*. Metode Random Forest merupakan metode boosting yang berdasarkan dari metode *tree* yang dihitung secara berulang dengan pengambilan dataset dan feature secara random. Hasil dari masing-masing *tree* akan dilakukan *Majority Voting* untuk mendapatkan hasil yang paling baik. Diagram topologi metode Random Forest ditunjukkan pada gambar 4.2.



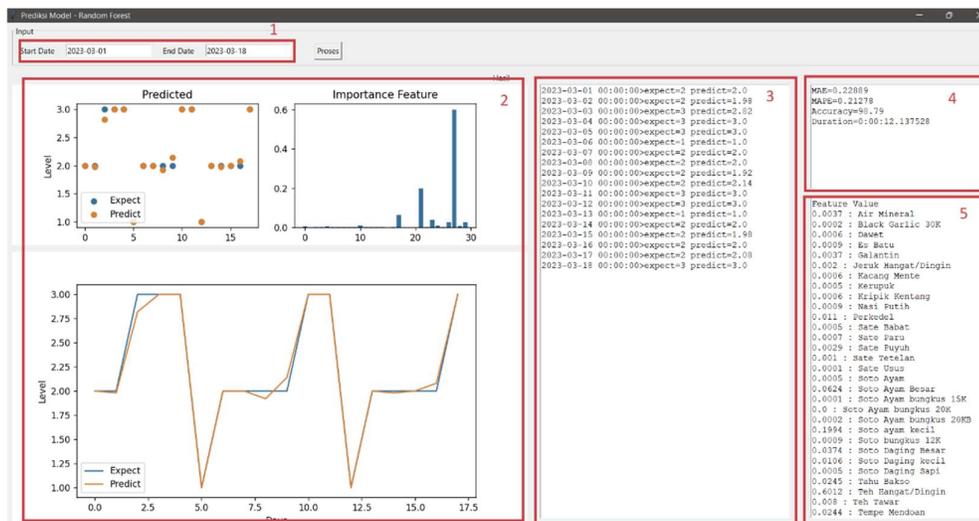
Gambar 4.2 : Diagram Topologi Metode Random Forest
(Sumber : diolah dari Farnham *et al.* 2019)

4.2 Uji Coba

Pengujian pada metode Random Forest memakai 4 skenario. Masing-masing skenario menggunakan perbandingan data training dan data testing yang berbeda-beda. Pada setiap skenario akan dilakukan *costumize* metode Random Forest dengan *Tree State* yang berpatokan pada nilai : 50, 100, 500 dan 1000.

Menurut Kaur *et al.*(2022), pengujian prediksi dengan menggunakan perbandingan data training dan data testing yang terdiri atas 90:10, 80:20, 70:30 dan 60:40 sangat efektif untuk karakteristik data yang mempunyai feature attribut lebih dari 10 dan dataset merupakan data yang imbalanced. Model pengujian tersebut bisa menghasilkan nilai yang efektif pada teknik dengan pendekatan *Bagging* atau *Boosting*.

Uji coba didukung dengan penggunaan library *scikit-learn* dan dikombinasikan dengan library *tkinter*, *matplotlib* pada python sehingga menjadi aplikasi desktop sederhana yang berjalan pada sistem operasi *Windows* dengan tampilan interface sederhana. Tampilan interface tersebut ditunjukkan pada gambar 4.3.



Gambar 4.3 : Interface Prediksi Metode Random Forest

Pada gambar 4.3, interface prediksi metode Random Forest pada kotak warna merah dengan angka nomor 1 menunjukkan inputan untuk menentukan data testing yang memakai parameter tanggal mulai dan tanggal berakhir dengan format

tanggal yyyy-mm-dd. Kotak merah dengan angka nomor 2 menunjukkan grafik prediksi dari metode Random Forest dan kotak merah dengan angka nomor 3 menunjukkan hasil detail pertanggal kesesuaian antara prediksi dan kondisi nyata. Kotak merah dengan angka 4 menunjukkan nilai performance dari metode Random Forest yang berisi *MAE*, *MAPE*, *Accuracy* dan *Duration*. Sedangkan kotak merah dengan angka nomor 5 menunjukkan skor *variable important* masing-masing atribut dari dataset sesuai dengan perhitungan metode Random Forest.

Proses *customize* metode Random Forest dengan mengubah *n_estimators* dilakukan dengan mengubah source code pada *method* Random Forest yang ada pada python. Perubahan tersebut hanya merubah angka yang terdiri dari 4 macam yaitu : 50,100,500 dan 1000. Dibawah ini merupakan penggalan source code yang diubah untuk *n_estimators* sesuai dengan skenario pengujian.

```
# deklarasi metode random forest
def random_forest_forecast(train, testX):
    # ubah data train menjadi array
    train = asarray(train)
    # pisah data train menjadi input dan output
    trainX, trainy = train[:, :-1], train[:, -1]

    # Proses model Random forest
    model = RandomForestRegressor(n_estimators=50)
    model.fit(trainX, trainy)
    # Lakukan prediksi data test
    yhat = model.predict([testX])

    importance = model.feature_importances_

    return yhat[0],importance
```

Penggalan source code di atas dengan warna yang tercetak merah, angka yang terdapat pada kurung akan diubah sesuai dengan skenario. Setelah dirubah aplikasi akan dijalankan ulang dan proses training dan testing data akan dilakukan kembali dengan performance yang berbeda. Evaluasi yang akan dilakukan mengambil hasil paling baik dari *customize* metode Random Forest dan skenario

uji coba dengan hasil terbaik akan dibandingkan dengan metode Logistic Regression.

4.2.1 Skenario 1

Skenario 1 dilakukan uji coba dengan menggunakan perbandingan data training dan data testing 90:10. Pengujian dilakukan sebanyak $4\times$ dengan $n_estimators$ dari metode Random Forest yang berbeda-beda. Performance yang diambil dari pengujian ini mencakup nilai *MAE*, *MAPE*, *accuracy* dan *duration* dari masing-masing $n_estimators$. Nilai $n_estimators$ terdiri atas : 50, 100, 500 dan 1000. Hasil dari pengujian dengan nilai $n_estimators = 50$ metode Random Forest ditunjukkan pada Tabel 4.1.

Tabel 4.1 Hasil Prediksi Random Forest $n_estimators = 50$ skenario 1

index	Tanggal	Expected	Predicted	Dev (predict - expect)
0	2023-01-01	3	3.0	0
1	2023-01-02	1	1.0	0
2	2023-01-03	3	2.88	-0.12
3	2023-01-04	2	2.01	0.01
4	2023-01-05	2	1.98	-0.02
...
16	2023-01-17	2	2.0	0
17	2023-01-18	2	2.02	0.02

Hasil uji coba skenario 1 dengan $n_estimators = 50$, menunjukkan hasil prediksi level omset dengan kondisi nyata omset hampir akurat. Pada Tabel 4.1 di tanggal 2023-01-03 hasil prediksi menunjukkan nilai lebih rendah -0.12 dari kondisi nyata, begitupun sebaliknya ditanggal 2023-01-18 hasil prediksi menunjukkan nilai yang lebih tinggi dari 0.02 dari kondisi nyata. Pengujian dengan $n_estimators = 100$ ditunjukkan pada Tabel 4.2.

Tabel 4.2 Hasil Prediksi Random Forest $n_estimators = 100$ skenario 1

index	Tanggal	Expected	Predicted	Dev (predict - expect)
0	2023-01-01	3	3.0	0
1	2023-01-02	1	1.0	0
2	2023-01-03	3	2.94	-0.06
...
17	2023-01-18	2	2.0	0

Pengujian dengan $n_estimators = 100$ menunjukkan hasil yang lebih stabil. Tabel 4.2 selisih antara kondisi nyata dengan kondisi prediksi ditunjukkan pada tanggal 2023-01-03 dengan selisih kondisi prediksi lebih rendah -0.06. Hasil prediksi ditanggal yang lain sangat akurat. Tabel 4.3 merupakan hasil pengujian dengan $n_estimators = 500$.

Tabel 4.3 Hasil Prediksi Random Forest $n_estimators = 500$ skenario 1

index	Tanggal	Expected	Predicted	Dev (predict - expect)
0	2023-01-01	3	3.0	0
1	2023-01-02	1	1.004	0.004
2	2023-01-03	3	2.896	-0.104
3	2023-01-04	2	1.988	-0.012
4	2023-01-05	2	2.0	0
...
16	2023-01-17	2	2.0	0
17	2023-01-18	2	1.986	-0.014

Pada Tabel 4.3 nilai prediksi yang telah didapat menunjukkan nilai yang beragam dengan kondisi nyata, setiap tanggal terdapat selisih antara nilai dari kondisi prediksi dengan kondisi nyata. Hal ini ditunjukkan pada tanggal 2023-01-02 nilai prediksi lebih tinggi 0.004 dengan kondisi nyata, tanggal 2023-01-03 nilai prediksi lebih rendah -0.104 dengan kondisi nyata, dan beberapa tanggal yang lain juga terdapat selisih. Pengujian selanjutnya memakai $n_estimators = 1000$ ditunjukkan pada Tabel 4.4.

Tabel 4.4 Hasil Prediksi Random Forest $n_estimators = 1000$ skenario 1

index	Tanggal	Expected	Predicted	Dev (predict - expect)
0	2023-01-01	3	2.966	-0.034
1	2023-01-02	1	1.0	0
2	2023-01-03	3	2.889	-0.111
3	2023-01-04	2	2.0	0
4	2023-01-05	2	2.0	0
...
16	2023-01-17	2	2.0	0
17	2023-01-18	2	1.991	-0.009

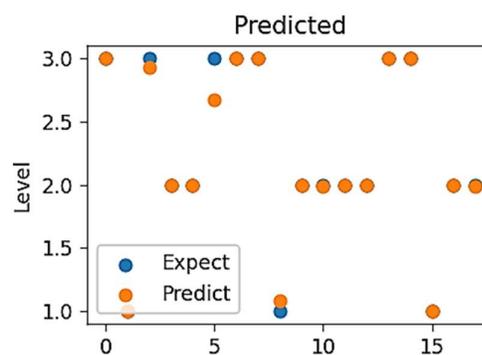
Tabel 4.4 menunjukkan hasil yang hampir sama dengan Tabel 4.3, bahwa kondisi prediksi lebih beragam dibandingkan dengan kondisi nyata. Hasil pengujian ini nilai selisih dari prediksi dengan kondisinya nyata lebih banyak negatif daripada

positif, hal ini menunjukkan nilai prediksi lebih rendah seperti pada tanggal 2023-01-01 selisih -0.034, tanggal 2023-0103 selisih -0.111 dan tanggal 2023-01-18 selisih -0.009. Proses perhitungan metode Random Forest pada skenario 1 menghasilkan *Variable Important* seperti ditunjukkan pada Tabel 4.5.

Tabel 4.5 Rank *Variable Important* Random Forest skenario 1

Rank	Variable Name	Value
1	Teh Hanget / Dingin	0.7148
2	Soto Ayam Besar	0.1026
3	Soto Ayam Kecil	0.0509
4	Soto Daging Besar	0.0372
5	Tempe Mendoan	0.0344
...
30	Dawet	0.0002
31	Black Garlic 30k	0.0000

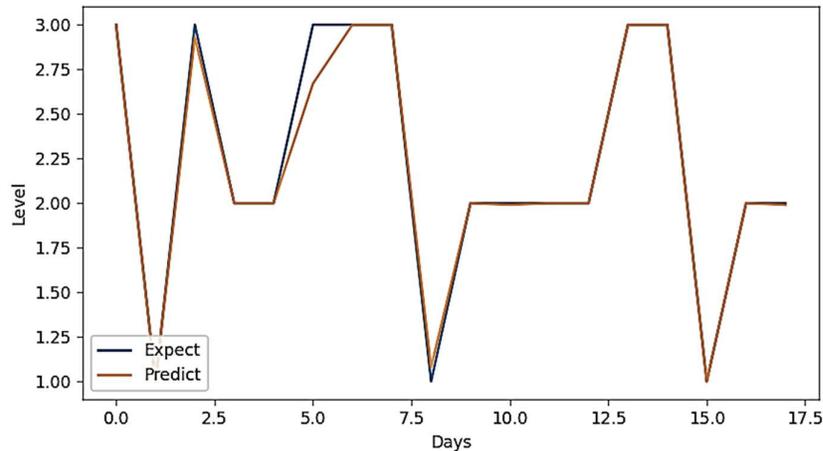
Tabel 4.5 menunjukkan evaluasi ranking dari *Variable Important* metode Random Forest dalam melakukan prediksi. Variabel paling penting ditunjukkan oleh “Teh Hanget / Dingin” dengan nilai 0.7148 dan variabel paling penting peringkat ke-2 ditunjukkan oleh “Soto Ayam Besar” dengan nilai 0.1026. peringkat paling bawah ditunjukkan oleh “Black Garlic 30K” dengan nilai 0 dan nomor ke-2 dari bawah ditunjukkan oleh “Dawet” dengan nilai 0.0002. Hasil diagram prediksi dengan metode Random Forest ditunjukkan pada Gambar 4.4.



Gambar 4.4 : Diagram plot prediksi Random Forest skenario 1

Gambar 4.4 menunjukkan visualisasi hasil prediksi metode Random Forest dengan $n_estimators = 100$ yang merupakan hasil terbaik dari nilai $n_estimators$

yang lain. Diagram plot menunjukkan plot warna biru yang merupakan kondisi nyata bersinggungan dengan kondisi prediksi pada hari ke 3 ke 6 dan ke 8. Diagram yang menunjukkan hubungan kondisi nyata dan prediksi lebih mendetail ditunjukkan pada gambar 4.5.



Gambar 4.5 : Diagram line prediksi Random Forest skenario 1

Diagram line pada Gambar 4.5 menunjukkan hasil nyata dan prediksi dengan digambar garis kuning sebagai prediksi dan garis biru merupakan kondisi nyata. Hubungan garis kuning dan garis biru yg tidak sejajar ada pada hari ke-3, ke 6 dan ke 8. Garis kuning yang sangat jauh dengan garis biru ditunjukkan pada hari ke 6. Percobaan yang telah dilakukan dengan *customize* metode Random Forest yang mengacu pada nilai $n_estimators$ dirangkum pada Tabel 4.6.

Tabel 4.6 Performance metode Random Forest Skenario 1

$n_estimators$	<i>MAE</i>	<i>MAPE</i>	<i>Accuracy</i>	<i>Duration</i>
50	0.23222	22.1852	98.78	26s
100	0.24111	22.3982	98.78	32s
500	0.23822	22.3260	98.78	44s
1000	0.23989	22.3620	98.78	68s

Tabel 4.6 menunjukkan pada skenario 1 performance hasil terbaik terletak pada kondisi *customize* metode Random Forest dengan menggunakan $n_estimators = 50$ dengan perolehan *MAE* 0.23222, *MAPE* 22.1852, *accuracy* 98,78% dan

duration 26s. Bila dibandingkan dengan $n_estimators$ yang lain, $n_estimators = 50$ mendapat akurasi yang baik, waktu proses cepat dan *MAE* dan *MAPE* yang rendah.

4.2.2 Skenario 2

Skenario 2 menggunakan perbandingan data training dan data testing 80:20. Pengujian tetap menggunakan 4 macam $n_estimators$ dari metode Random Forest yang terdiri atas : 50, 100, 500 dan 1000. Performance yang diambil adalah nilai *MAE*, *MAPE*, *accuracy* dan *duration* dari masing-masing $n_estimators$. Hasil dari pengujian dengan nilai $n_estimators = 50$ ditunjukkan pada Tabel 4.7.

Tabel 4.7 Hasil Prediksi Random Forest $n_estimators = 50$ skenario 2

index	Tanggal	Expected	Predicted	Dev (predict - expect)
0	2023-01-01	3	2.96	-0.04
1	2023-01-02	1	1.0	0
2	2023-01-03	3	2.9	-0.1
3	2023-01-04	2	2.0	0
4	2023-01-05	2	2.0	0
5	2023-01-06	3	2.74	-0.26
...
33	2023-02-03	3	2.74	-0.26
34	2023-02-04	3	3.0	0
35	2023-02-05	3	3.0	0

Hasil uji coba skenario 2 dengan $n_estimators = 50$, menunjukkan hasil prediksi level omset dengan kondisi nyata mempunyai rentan selisih yang besar. Hal ini ditunjukkan pada tanggal 2023-01-06 hasil prediksi lebih rendah -0.26 dan tanggal 2023-02-03 hasil prediksi lebih rendah -0.26. Pengujian dengan $n_estimators = 100$ ditunjukkan pada Tabel 4.8.

Tabel 4.8 Hasil Prediksi Random Forest $n_estimators = 100$ skenario 2

index	Tanggal	Expected	Predicted	Dev (predict - expect)
0	2023-01-01	3	3.0	0
1	2023-01-02	1	1.01	0.01
2	2023-01-03	3	2.92	-0.08
3	2023-01-04	2	2.0	0
4	2023-01-05	2	2.0	0
5	2023-01-06	3	2.73	-0.27
...
34	2023-02-04	3	3.0	0
35	2023-02-05	3	3.0	0

Pengujian dengan $n_estimators = 100$ menunjukkan hasil yang lebih baik dari pengujian sebelumnya. Pada Tabel 4.8 selisih antara kondisi nyata dengan kondisi prediksi ditunjukkan pada tanggal 2023-01-03 dengan selisih kondisi prediksi lebih rendah -0.08, tanggal 2023-01-06 lebih rendah -0.27 dan tanggal 2023-02-03 lebih rendah -0.19. Tabel 4.9 merupakan hasil pengujian dengan $n_estimators = 500$.

Tabel 4.9 Hasil Prediksi Random Forest $n_estimators = 500$ skenario 2

index	Tanggal	Expected	Predicted	Dev (predict - expect)
0	2023-01-01	3	2.998	-0.002
1	2023-01-02	1	1.0	0
2	2023-01-03	3	2.924	-0.076
3	2023-01-04	2	2.0	0
4	2023-01-05	2	2.0	0
5	2023-01-06	3	2.732	-0.268
...
33	2023-02-03	3	2.778	-0.222
34	2023-02-04	3	3.0	0
35	2023-02-05	3	3.0	0

Tabel 4.9 hasil prediksi menunjukkan nilai yang tidak jauh berbeda dengan $n_estimators = 100$, akan tetap waktu proses pada pengujian ini lebih lama. Selisih kondisi nyata dan kondisi prediksi ditunjukkan pada tanggal 2023-01-01 nilai prediksi lebih rendah - 0.002 dengan kondisi nyata, tanggal 2023-01-03 nilai prediksi lebih rendah -0.076 dengan kondisi nyata, dan beberapa tanggal yang lain juga terdapat selisih. Pengujian selanjutnya memakai $n_estimators = 1000$ ditunjukkan pada Tabel 4.10.

Tabel 4.10 Hasil Prediksi Random Forest $n_estimators = 1000$ skenario 2

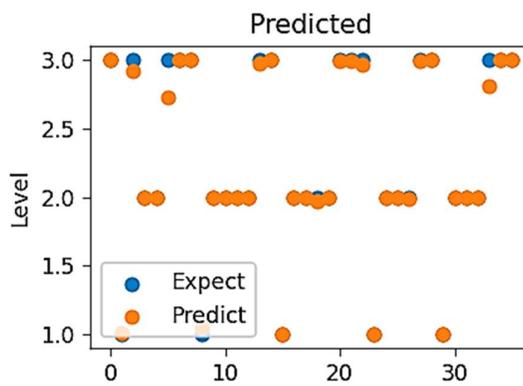
index	Tanggal	Expected	Predicted	Dev (predict - expect)
0	2023-01-01	3	2.999	-0.001
1	2023-01-02	1	1.001	0.001
2	2023-01-03	3	2.888	-0.112
3	2023-01-04	2	2.0	0
4	2023-01-05	2	2.0	0
5	2023-01-06	3	2.702	-0.298
...
33	2023-02-03	3	2.792	-0.208
34	2023-02-04	3	3.0	0
35	2023-02-05	3	3.0	0

Tabel 4.10 menunjukkan hasil kondisi prediksi lebih beragam dibandingkan dengan kondisi nyata. Hasil dari pengujian ini nilai selisih dari prediksi dengan kondisinya nyata lebih banyak bersinggungan, hal ini menunjukkan nilai prediksi lebih rendah seperti pada tanggal 2023-01-01 selisih -0.001, tanggal 2023-01-02 lebih tinggi 0.001 dan tanggal 2023-01-03 selisih -0.112, dan lain sebagainya. Evaluasi peringkat *Variable Important* metode Random Forest pada skenario 2 ditunjukkan pada Tabel 4.11.

Tabel 4.11 Rank *Variable Important* Random Forest skenario 2

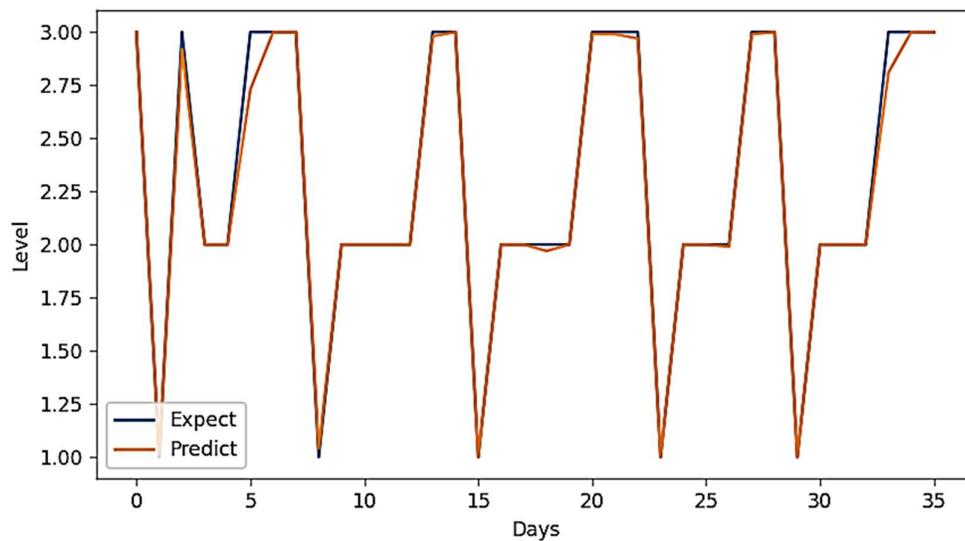
Rank	Variable Name	Value
1	Teh Hanget / Dingin	0.7385
2	Soto Ayam Kecil	0.0981
3	Tempe Mendoan	0.0539
4	Soto Ayam Besar	0.0409
5	Soto Daging Besar	0.0149
....
30	Soto Ayam	0.0000
31	Kripik Kentang	0.0000

Tabel 4.11 menunjukkan evaluasi ranking dari *Variable Important* paling tinggi ditunjukkan oleh “Teh Hanget / Dingin” dengan nilai 0.7385 dan variabel paling penting peringkat ke-2 ditunjukkan oleh “Soto Ayam Kecil” dengan nilai 0.0981, berbeda dengan skenario 1. peringkat paling bawah ditunjukkan oleh “Kripik kentang” dengan nilai 0 dan nomor ke-2 dari bawah ditunjukkan oleh “Soto Ayam” dengan nilai 0. Visualisasi prediksi dengan metode Random Forest ditunjukkan pada Gambar 4.6.



Gambar 4.6 : Diagram plot prediksi Random Forest skenario 2

Gambar 4.6 adalah visualisasi dengan diagram plot yang merupakan hasil prediksi metode Random Forest nilai terbaik. Plot warna biru adalah kondisi nyata dan plot warna kuning adalah kondisi prediksi. Plot yang bersinggungan terdapat pada hari ke-2, ke-4, ke-9, ke-12, ke-18, ke-22, ke-27 dan ke-33. Visualisasi prediksi lebih mendetail ditunjukkan pada gambar 4.7.



Gambar 4.7 : Diagram line prediksi Random Forest skenario 2

Diagram line skenario 2 pada Gambar 4.7 menunjukkan persinggungan antara garis biru dengan kuning lebih sedikit. Terlihat pada diagram tersebut singgungan garis terdapat pada hari ke-2, ke-4, ke-9, ke-18 da ke-33. Hasil rangkuman performance dari skenario 2 dengan *customize* metode Random Forest yang mengacu pada nilai $n_estimators$ ditunjukkan pada Tabel 4.12.

Tabel 4.12 Performance metode Random Forest Skenario 2

$n_estimators$	<i>MAE</i>	<i>MAPE</i>	<i>Accuracy</i>	<i>Duration</i>
50	0.22838	21.8469	98.78	32s
100	0.23324	21.9144	98.78	37s
500	0.23281	21.9703	98.78	71s
1000	0.23349	22.0149	98.78	108s

Pada Tabel 4.12, performance terbaik terletak pada $n_estimators = 50$ dengan perolehan MAE 0.22838, $MAPE$ 21.8469, $accuracy$ 98,78% dan $duration$ 32s. Sama seperti skenario 1, $n_estimators = 50$ mendapat akurasi baik, waktu proses cepat dan MAE dan $MAPE$ yang rendah dibandingkan dengan $n_estimators$ yang lain pada skenario 2.

4.2.3 Skenario 3

Percobaan yang dilakukan di Skenario 3 memakai perbandingan 70:30, dengan $n_estimators$ dari metode Random Forest yang terdiri atas : 50, 100, 500 dan 1000. Evaluasi mendapatkan hasil dari nilai MAE , $MAPE$, $accuracy$ dan $duration$ dari masing-masing $n_estimators$. Hasil dari pengujian dengan nilai $n_estimators = 50$ ditunjukkan pada Tabel 4.13.

Tabel 4.13 Hasil Prediksi Random Forest $n_estimators = 50$ skenario 3

index	Tanggal	Expected	Predicted	Dev (predict - expect)
0	2023-01-01	3	3.0	0
1	2023-01-02	1	1.0	0
2	2023-01-03	3	2.88	-0.12
3	2023-01-04	2	2.0	0
4	2023-01-05	2	2.0	0
5	2023-01-06	3	2.7	-0.3
6	2023-01-07	3	3.0	0
...
51	2023-02-21	2	1.98	-0.02
52	2023-02-22	2	1.98	-0.02
53	2023-02-23	2	2.0	0
54	2023-02-24	2	2.0	0

Hasil skenario 3 dengan $n_estimators = 50$, menunjukkan prediksi dengan kondisi nyata mempunyai rentan selisih yang banyak. Hal ini ditunjukkan pada tanggal 2023-01-03, 2023-01-06, 2023-02-21, 2023-02-22 dan lain sebagainya. Pengujian dengan $n_estimators = 100$ ditunjukkan pada Tabel 4.14.

Tabel 4.14 Hasil Prediksi Random Forest $n_estimators = 100$ skenario 3

index	Tanggal	Expected	Predicted	Dev (predict - expect)
0	2023-01-01	3	3.0	0
1	2023-01-02	1	1.0	0
2	2023-01-03	3	2.91	-0.09
3	2023-01-04	2	2.0	0

Lanjutan Tabel Hasil Prediksi Random Forest $n_{estimators} = 100$ skenario 3

index	Tanggal	Expected	Predicted	Dev (predict - expect)
4	2023-01-05	2	2.0	0
5	2023-01-06	3	2.74	-0.26
6	2023-01-07	3	2.99	-0.01
...
51	2023-02-21	2	2.0	0
52	2023-02-22	2	1.97	-0.03
53	2023-02-23	2	2.02	0.02
54	2023-02-24	2	2.0	0

Tabel 4.14 merupakan hasil uji coba dengan $n_{estimators} = 100$ menunjukkan hasil yang sangat efektif. Hal ini karena kondisi nyata dengan kondisi prediksi lebih sedikit selisihnya, yang ditunjukkan pada tanggal 2023-01-03 dengan selisih kondisi prediksi lebih rendah -0.09, tanggal 2023-01-06 lebih rendah -0.26 dan tanggal 2023-02-22 lebih rendah -0.03. Hasil prediksi pengujian dengan $n_{estimators} = 500$ ditunjukkan pada Tabel 4.15.

Tabel 4.15 Hasil Prediksi Random Forest $n_{estimators} = 500$ skenario 3

index	Tanggal	Expected	Predicted	Dev (predict - expect)
0	2023-01-01	3	2.998	-0.002
1	2023-01-02	1	1.0	0
2	2023-01-03	3	2.908	-0.092
3	2023-01-04	2	2.0	0
4	2023-01-05	2	2.0	0
5	2023-01-06	3	2.688	-0.312
6	2023-01-07	3	3.0	0
...
51	2023-02-21	2	1.992	-0.008
52	2023-02-22	2	1.98	-0.02
53	2023-02-23	2	2.006	0.006
54	2023-02-24	2	2.0	0

Pada Tabel 4.15 hasil prediksi dengan kondisi nyata sangat beragam, sehingga mempengaruhi nilai akurasi. Waktu eksekusi dari proses prediksi juga lebih lama dan hasil yang didapatkan kurang maksimal hal ini ditunjukkan pada selisih prediksi yang lebih rendah pada tanggal 2023-01-01, 2023-01-03, 2023-01-06, 2023-02-21, 2023-02-22. Uji coba dengan nilai $n_{estimators} = 1000$ ditunjukkan pada Tabel 4.16.

Tabel 4.16 Hasil Prediksi Random Forest $n_estimators = 1000$ skenario 3

index	Tanggal	Expected	Predicted	Dev (predict - expect)
0	2023-01-01	3	2.998	-0.002
1	2023-01-02	1	1.001	0.001
2	2023-01-03	3	2.889	-0.111
3	2023-01-04	2	2.0	0
4	2023-01-05	2	2.0	0
5	2023-01-06	3	2.708	-0.292
6	2023-01-07	3	3.0	0
...
51	2023-02-21	2	1.994	-0.006
52	2023-02-22	2	1.974	-0.026
53	2023-02-23	2	2.006	0.006
54	2023-02-24	2	2.0	0

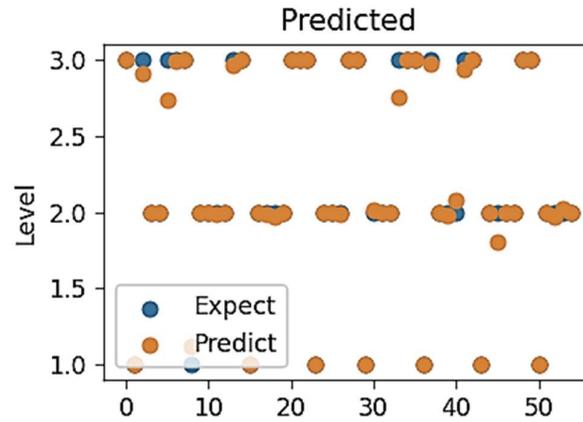
Tabel 4.16 menunjukkan hasil kondisi prediksi lebih banyak selisih dengan kondisi nyata. Hasil dari pengujian ini nilai selisih dari prediksi dengan kondisinya nyata lebih banyak bersinggungan, hal ini ditunjukkan pada tanggal 2023-01-01, 2023-01-02, 2023-01-03, 2023-01-06, 2023-02-21 dan lain sebagainya. Evaluasi peringkat *Variable Important* metode Random Forest pada skenario 3 ditunjukkan pada Tabel 4.17.

Tabel 4.17 Rank *Variable Important* Random Forest skenario 3

Rank	Variable Name	Value
1	Teh Hanget / Dingin	0.7129
2	Soto Ayam Kecil	0.1574
3	Soto Ayam Besar	0.0665
4	Tahu Bakso	0.0158
5	Teh Tawar	0.0069
...
30	Soto Ayam Bungkus 20k	0.0000
31	Black Garlic 30k	0.0000

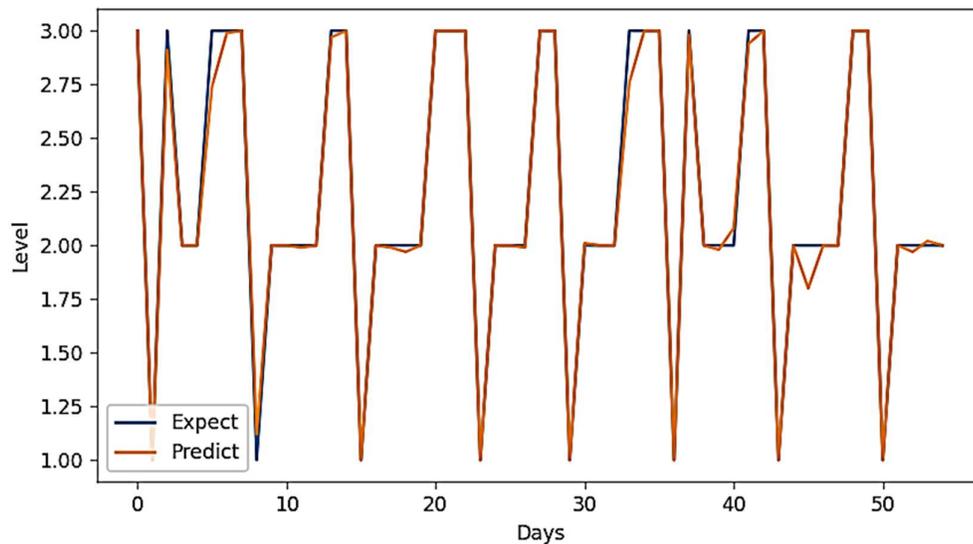
Tabel 4.17 menunjukkan evaluasi ranking dari *Variable Important* paling tinggi ditunjukkan oleh “Teh Hanget / Dingin” dengan nilai 0.7129 dan variabel paling penting peringkat ke-2 ditunjukkan oleh “Soto Ayam Kecil” dengan nilai 0.1574, hampir sama dengan skenario 2 tetapi peringkat paling bawah ditunjukkan oleh “Black Garlic 30k” dengan nilai 0 dan nomor ke-2 dari bawah ditunjukkan oleh “Soto Ayam bungkus 20k” dengan nilai 0 berbeda dengan skenario-skenario

sebelumnya. Visualisasi prediksi dengan metode Random Forest ditunjukkan pada Gambar 4.8.



Gambar 4.8 : Diagram plot prediksi Random Forest skenario 3

Hasil pemetaan metode Random Forest divisualisasikan pada Gambar 4.8 yang menunjukkan kondisi nyata dan kondisi prediksi berisikan pada hari ke-2, ke 6, ke-9, ke-11, ke-32, ke-37, ke-46, ke-52 dan ke-53. Visualisasi prediksi lebih mendetail ditunjukkan pada gambar 4.9.



Gambar 4.9 : Diagram line prediksi Random Forest skenario 3

Diagram line hasil prediksi di skenario 3 pada Gambar 4.9 menunjukkan tingkat akurasi yang bagus. Hal tersebut dapat dilihat dari persinggungan garis prediksi dengan garis nyata terjadi pada hari ke-3, ke-6, ke-16, ke-18, ke-33, ke 38, dan ke-46 dalam kategori sedikit. Untuk hasil rangkuman performance dari skenario 3 dengan *customize* metode Random Forest yang mengacu pada nilai $n_estimators$ ditunjukkan pada Tabel 4.18.

Tabel 4.18 Performance metode Random Forest Skenario 3

$n_estimators$	<i>MAE</i>	<i>MAPE</i>	<i>Accuracy</i>	<i>Duration</i>
50	0.23145	21.9576	98.78	34s
100	0.23382	22.0849	98.78	58s
500	0.23393	22.0909	98.78	102s
1000	0.23398	220809	98.78	334s

Performance terbaik metode Random Forest pada skenario 3 seperti pada Tabel 4.18 terletak pada $n_estimators = 50$ dengan perolehan *MAE* 0.23145, *MAPE* 21.9576, *accuracy* 98,78% dan *duration* 34s. Sama seperti skenario 1 dan 2, $n_estimators = 50$ mendapat akurasi baik, waktu proses cepat dan *MAE* dan *MAPE* yang rendah dibandingkan dengan $n_estimators$ yang lain pada skenario 3.

4.2.4 Skenario 4

Skenario 4 metode Random Forest dilakukan percobaan dengan memakai perbandingan 60:40, dan *customize* metode Random Forest memakai $n_estimators$ yang terdiri atas : 50, 100, 500 dan 1000. Performance yang akan dievaluasi adalah nilai *MAE*, *MAPE*, *accuracy* dan *duration* dari masing-masing $n_estimators$. Pengujian dengan nilai $n_estimators = 50$ pada skenario 4 ditunjukkan pada Tabel 4.19.

Tabel 4.19 Hasil Prediksi Random Forest $n_estimators = 50$ skenario 4

index	Tanggal	Expected	Predicted	Dev (predict - expect)
0	2023-01-01	3	3.0	0
1	2023-01-02	1	1.0	0
2	2023-01-03	3	2.92	-0.08
3	2023-01-04	2	2.0	0
4	2023-01-05	2	2.0	0
5	2023-01-06	3	2.7	-0.3
6	2023-01-07	3	3.0	0

Lanjutan Tabel Hasil Prediksi Random Forest $n_estimators = 50$ skenario 4

index	Tanggal	Expected	Predicted	Dev (predict - expect)
7	2023-01-08	3	3.0	0
...
68	2023-03-09	2	1.98	-0.02
69	2023-03-10	2	2.16	0.16
70	2023-03-11	3	3.0	0
71	2023-03-12	3	3.0	0
72	2023-03-13	1	1.0	0

Hasil skenario 4 dengan $n_estimators = 50$, menghasilkan nilai prediksi yang baik. Hal ini ditunjukkan pada tanggal 2023-01-03, 2023-01-06, 2023-03-09, 2023-02-10 dan lain sebagainya. Pengujian dengan $n_estimators = 100$ ditunjukkan pada Tabel 4.20.

Tabel 4.20 Hasil Prediksi Random Forest $n_estimators = 100$ skenario 4

index	Tanggal	Expected	Predicted	Dev (predict - expect)
0	2023-01-01	3	3.0	0
1	2023-01-02	1	1.0	0
2	2023-01-03	3	2.87	-0.13
3	2023-01-04	2	2.0	0
4	2023-01-05	2	2.0	0
5	2023-01-06	3	2.75	-0.25
6	2023-01-07	3	3.0	0
7	2023-01-08	3	3.0	0
...
68	2023-03-09	2	1.96	-0.04
69	2023-03-10	2	2.26	0.26
70	2023-03-11	3	3.0	0
71	2023-03-12	3	3.0	0
72	2023-03-13	1	1.0	0

Hasil uji coba ditunjukkan pada Tabel 4.20 dengan $n_estimators = 100$ sangat baik. Hal ini karena kondisi nyata dengan kondisi prediksi lebih sedikit selisihnya, yang ditunjukkan pada tanggal 2023-01-03 dengan selisih kondisi prediksi lebih rendah -0.13, tanggal 2023-01-06 lebih rendah -0.25 dan tanggal 2023-03-09 lebih rendah -0.04. Hasil prediksi pengujian dengan $n_estimators = 500$ ditunjukkan pada Tabel 4.21.

Tabel 4.21 Hasil Prediksi Random Forest $n_estimators = 500$ skenario 4

index	Tanggal	Expected	Predicted	Dev (predict - expect)
0	2023-01-01	3	2.998	-0.002
1	2023-01-02	1	1.004	0.004
2	2023-01-03	3	2.898	-0.102
3	2023-01-04	2	2.0	0
4	2023-01-05	2	2.0	0
5	2023-01-06	3	2.714	-0.286
6	2023-01-07	3	2.998	-0.002
7	2023-01-08	3	3.0	0
...
68	2023-03-09	2	1.964	-0.036
69	2023-03-10	2	2.208	0.208
70	2023-03-11	3	3.0	0
71	2023-03-12	3	3.0	0
72	2023-03-13	1	1.0	0

Pada Tabel 4.21 hasil prediksi dengan kondisi nyata sangat beragam, sehingga mempengaruhi nilai akurasi. Waktu eksekusi dari proses prediksi juga lebih lama dan hasil yang didapatkan kurang maksimal hal ini ditunjukkan pada selisih prediksi yang lebih rendah pada tanggal 2023-01-01, 2023-01-02, 2023-01-03, 2023-01-06, 2023-03-09, 2023-03-10. Uji coba dengan nilai $n_estimators = 1000$ ditunjukkan pada Tabel 4.22.

Tabel 4.22 Hasil Prediksi Random Forest $n_estimators = 1000$ skenario 4

index	Tanggal	Expected	Predicted	Dev (predict - expect)
0	2023-01-01	3	2.998	-0.002
1	2023-01-02	1	1.0	0
2	2023-01-03	3	2.896	-0.104
3	2023-01-04	2	2.0	0
4	2023-01-05	2	2.0	0
5	2023-01-06	3	2.736	-0.264
6	2023-01-07	3	3.0	0
7	2023-01-08	3	3.0	0
...
68	2023-03-09	2	1.956	-0.044
69	2023-03-10	2	2.222	0.222
70	2023-03-11	3	3.0	0
71	2023-03-12	3	3.0	0
72	2023-03-13	1	1.001	0.001

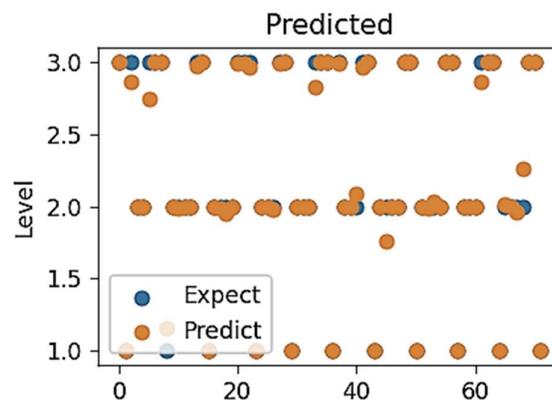
Tabel 4.22 menunjukkan hasil kondisi prediksi lebih baik dari $n_estimators = 500$, tetapi masih belum maksimal. Hasil dari pengujian ini nilai selisih dari prediksi dengan kondisinya nyata ditunjukkan pada tanggal 2023-01-01, 2023-01-

03, 2023-01-06, 2023-03-09, 2023-03-10 dan lain sebagainya. Evaluasi peringkat *Variable Important* metode Random Forest pada skenario 4 ditunjukkan pada Tabel 4.23.

Tabel 4.23 Rank *Variable Important* Random Forest skenario 4

Rank	Variable Name	Value
1	Teh Hanget / Dingin	0.7092
2	Soto Ayam Kecil	0.1386
3	Soto Ayam Besar	0.0507
4	Tempe Mendoan	0.0246
5	Soto Daging Besar	0.0143
....
30	Soto Ayam	0.0000
31	Black Garlic 30k	0.0000

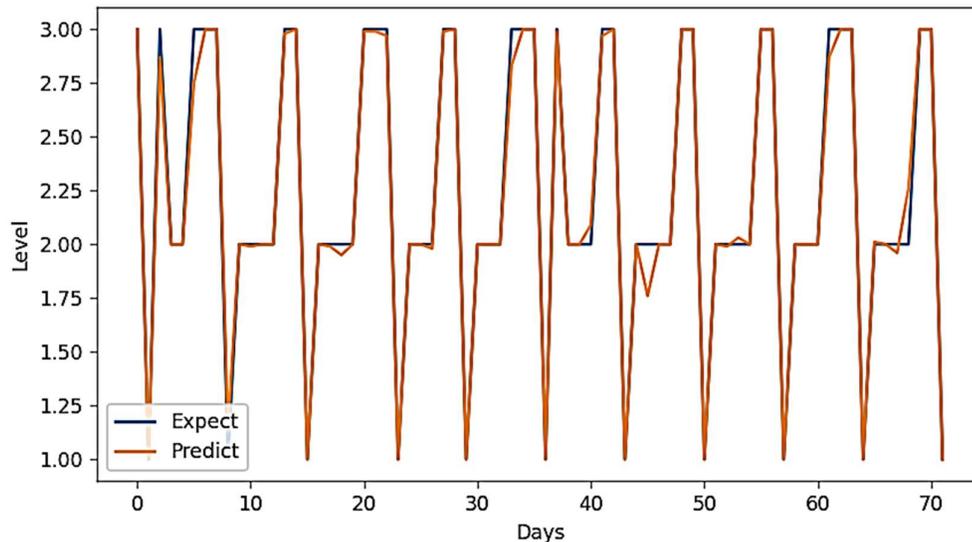
Tabel 4.23 menunjukkan evaluasi ranking dari *Variable Important* paling tinggi ditunjukkan oleh “Teh Hanget / Dingin” dengan nilai 0.7029 dan variabel paling penting peringkat ke-2 ditunjukkan oleh “Soto Ayam Kecil” dengan nilai 0.1386, hampir sama dengan skenario 3 tetapi peringkat paling bawah ditunjukkan oleh “Black Garlic 30k” dengan nilai 0 dan nomor ke-2 dari bawah ditunjukkan oleh “Soto Ayam”. Visualisasi prediksi dengan metode Random Forest ditunjukkan pada Gambar 4.10.



Gambar 4.10 : Diagram plot prediksi Random Forest skenario 4

Hasil pemetaan metode Random Forest divisualisasikan pada Gambar 4.10 yang menunjukkan kondisi nyata dan kondisi prediksi berisikan pada hari ke-2, ke

3, ke-6, ke-22, ke-32, ke-37, ke-42, ke-62 dan ke-70. Visualisasi prediksi lebih mendetail ditunjukkan pada gambar 4.11.



Gambar 4.11 : Diagram line prediksi Random Forest skenario 4

Diagram line hasil prediksi di skenario 4 pada Gambar 4.11 menunjukkan tingkat akurasi yang bagus. Hal tersebut dapat dilihat dari persinggungan garis prediksi dengan garis nyata terjadi pada hari ke-3, ke-6, ke-18, ke-33, ke-42, ke 36, ke-46, ke-54, ke-61 dan ke-70. Untuk hasil rangkuman performance dari skenario 4 metode Random Forest yang mengacu pada nilai $n_estimators$ ditunjukkan pada Tabel 4.24.

Tabel 4.24 Performance metode Random Forest Skenario 4

$n_estimators$	<i>MAE</i>	<i>MAPE</i>	<i>Accuracy</i>	<i>Duration</i>
50	0.23222	21.9907	98.78	36s
100	0.23403	22.0902	98.78	46s
500	0.23267	22.0092	98.78	102s
1000	0.22354	22.0446	98.78	337s

Performance terbaik metode Random Forest pada skenario 4 seperti pada Tabel 4.24 terletak pada $n_estimators = 50$ dengan perolehan *MAE* 0.23222, *MAPE* 21.9907, *accuracy* 98,78% dan *duration* 36s. Sama seperti skenario 1,2 dan 3,

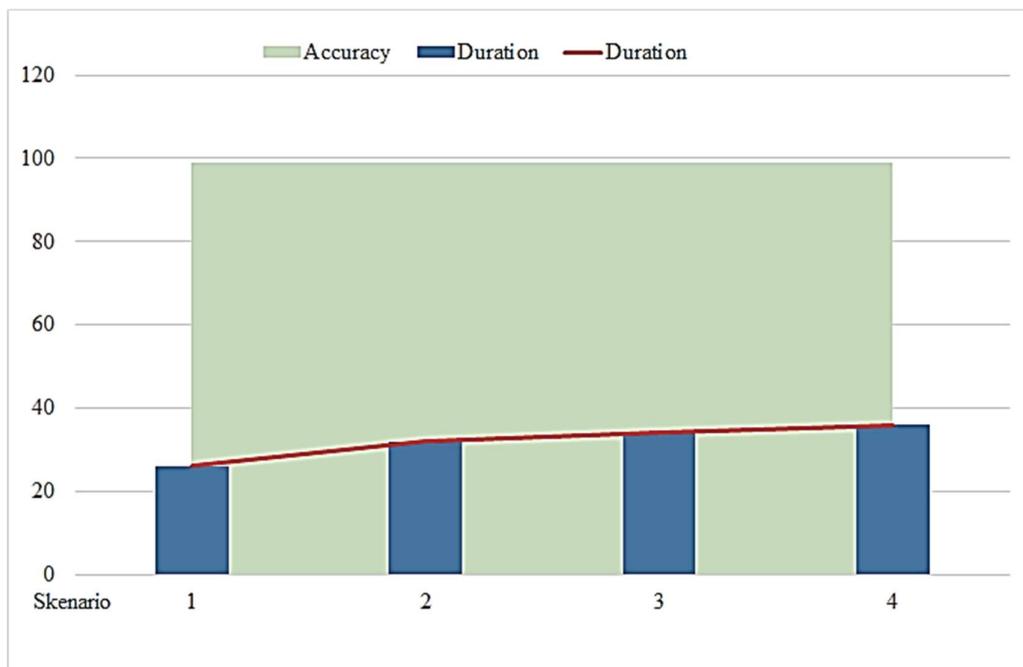
skenario 4 dengan $n_estimators = 50$ mendapat akurasi yang baik, waktu proses cepat dan MAE dan $MAPE$ yang rendah dibandingkan dengan $n_estimators$ yang lain pada skenario 4.

4.3 Evaluasi

Tahap evaluasi dalam pengujian metode Random Forest mengambil dari hasil prediksi dengan performance terbaik dari masing-masing skenario. Hasil yang paling baik dari 4 skenario akan menjadi bahan perbandingan dengan metode yang lain. Evaluasi dari performance terbaik masing-masing skenario ditunjukkan pada Tabel 4.25.

Tabel 4.25 Evaluasi Performance Metode Random Forest

Skenario	Perbandingan data	$n_estimators$	MAE	$MAPE$	$Accuracy$	$Duration$
1	90:10	50	0.23222	22.1852	98.78	26s
2	80:20	50	0.22838	21.8469	98.78	32s
3	70:30	50	0.23145	21.9576	98.78	34s
4	60:40	50	0.23222	21.9907	98.78	36s



Gambar 4.12 : Diagram Evaluasi Performance Metode Random Forest

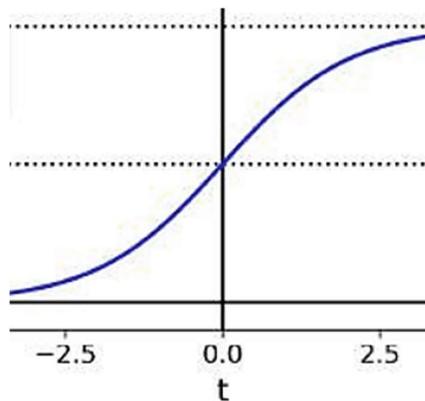
Hasil dari evaluasi yang ditunjukkan pada Tabel 4.25 dan gambar 4.12 bahwa Prediksi dengan dataset bisnis restoran Soto Kwali Pak Wasis dengan menggunakan metode Random Forest dapat mencapai hasil yang maksimal dengan parameter perbandingan data training dan data testing 90:10 dan *customize* metodenya menggunakan $n_estimators = 50$. Menurut Sahinbas *et al.* (2022) metode yang menghasilkan nilai *accuracy* paling tinggi dan *duration* terkecil mempunyai performance yang lebih bagus, Sehingga hasil nilai *duration* yang paling kecil yaitu 26s, dan nilai *accuracy* tertinggi 98,78%, karena semakin rendah nilai *duration* maka akan sangat berpengaruh dengan efisiensi suatu metode sesuai dengan penelitian Kureljusic *et al.* (2022).

BAB V

METODE LOGISTIC REGRESSION

5.1 Desain

Pembelajaran pada metode Logistic Regression membutuhkan variabel independen dan variabel dependen. Menurut Chen *et al.* (2021) algoritma Logistic Regression merupakan transformasi dari Regresi Linier menjadi tidak Linier dengan menggunakan parameter probabilitas. Dalam menentukan prediksi metode Logistic Regression menghitung dengan menggunakan teknik Regresi Linier dan hasil dari perhitungan tersebut akan dikombinasikan dengan nilai threshold supaya bisa menjadi tidak linier dan selanjutnya hasilnya diproses menjadi logistic sehingga membentuk curva seperti gambar 5.1.

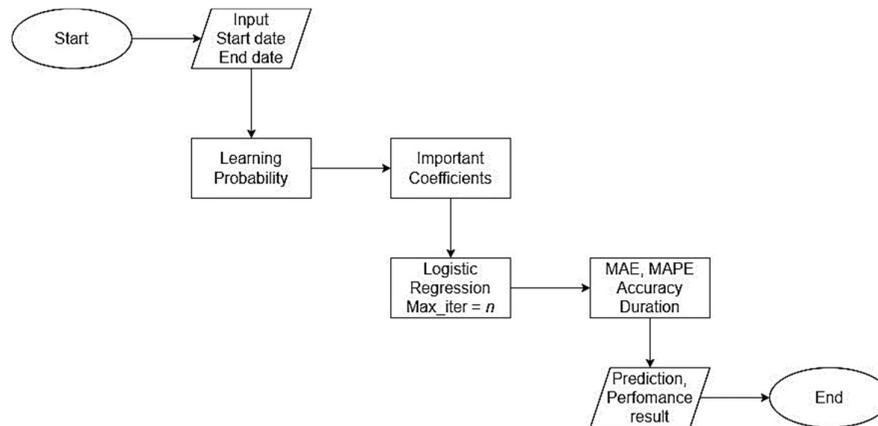


Gambar 5.1 : Curva Logistic Regression
(Sumber : diolah dari Farnham *et al.* 2019)

Gambar 5.1 merupakan ilustrasi curva yang terbentuk dari hasil probability untuk memprediksikan sesuatu dari metode Logistic Regression, semakin menyerupai bentuk curva seperti gambar 5.1 maka hasil prediksi juga akan semakin baik. Farnham *et al.* (2019) menjelaskan bahwa Logistic Regression dibangun sesuai dengan penentuan nilai threshold sesuai dengan persamaan (5.1).

$$\sigma(t) = \frac{1}{1 + \exp(-t)} \quad (5.1)$$

Di mana $\sigma(t) < 0,5$ ketika $t < 0$ dan $\sigma(t) > 0,5$ ketika $t > 0$, Nilai dari threshold dari *Logistic Regression* 1 jika nilai prediksi positif dan 0 jika nilai prediksi negatif, nilai t sering disebut nilai logit. Desain flowchart metode *Logistic Regression* ditunjukkan pada gambar 5.2.



Gambar 5.2 : Flowchart Prediksi Model Metode *Logistic Regression*
(Sumber : diolah dari Gai *et al.* 2023)

Flowchart pada gambar 5.2 dimulai dengan memasukkan tanggal mulai dan tanggal berakhir sebagai parameter penentuan test data untuk prediksi. Langkah selanjutnya setelah memasukkan parameter tanggal sistem akan melakukan *Learning* dengan parameter bobot probability. Dikutip dalam buku Gaso. (2019) formula *posterior probabilities* metode *Logistic Regression* ditunjukkan pada persamaan (5.2).

$$\hat{y} = \operatorname{argmax}_{k \in \{0, \dots, k-1\}} P(y = k | x) \quad (5.2)$$

Di mana k merupakan nilai skor dari variabel independen.

Langkah selanjutnya metode *Logistic Regression* dalam proses prediksi yaitu menghitung nilai *coefficients* yang berfungsi untuk melakukan estimasi hubungan antara variabel independen dengan variabel dependen seperti yang dijelaskan Schober *et al.*(2021). Formula untuk mendapatkan nilai *coefficient* pada metode *Logistic Regression* ditunjukkan pada persamaan (5.3).

$$\ln\left(\frac{p}{1-p}\right) = b_0 + b_1X \quad (5.3)$$

Di mana p merupakan probability dan X merupakan variabel independen.

Tahap proses prediksi Logistic Regression dengan menetapkan parameter *iteration* dengan insialisasi $\text{max_iter} = n$ yang terdiri atas 100, 500 dan 1000. Pemilihan nilai *iteration* metode Logistic Regression sesuai dengan yang artikel Rani *et al.* (2023) yang bertujuan untuk mengurangi resiko kegagalan dalam melakukan proses perhitungan keterkaitan antara variabel independen dan variabel dependen. Dengan menggunakan nilai *iteration* tersebut hasil prediksi yang diperoleh akan lebih efektif.

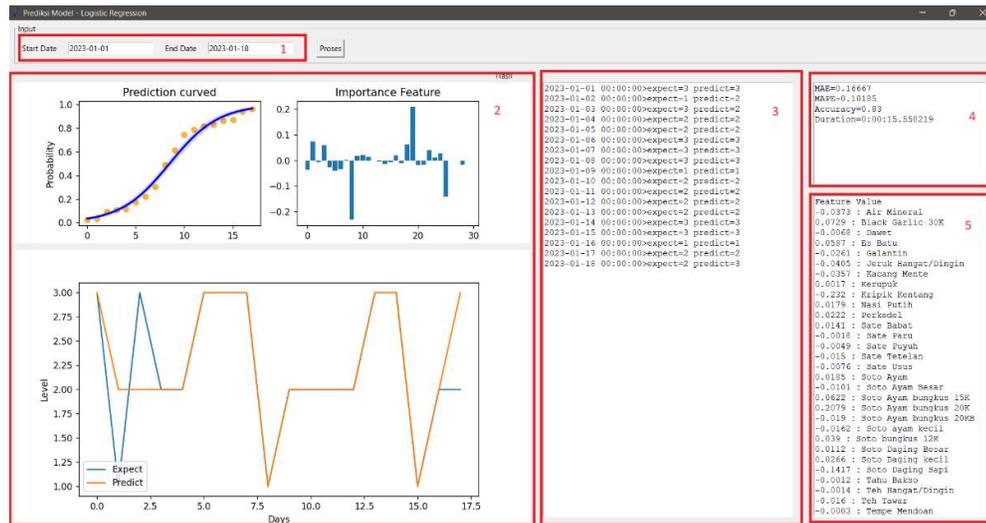
Proses dari flowchart Logistic Regression merupakan mendapatkan nilai performance dengan melakukan proses perhitungan *MAE, MAPE, accuracy* dan *duration*. Hasil akhir dari sistem akan menampilkan nilai prediksi yang dibandingkan dengan nilai nyata, nilai performance yang terdiri atas *MAE, MAPE, accuracy* dan *duration*, serta nilai *coefficient* dari variabel independen dalam hubungan dengan variabel dependen

5.2 Uji Coba

Pengujian pada metode Logistic Regression memakai 4 skenario. Masing-masing skenario menggunakan perbandingan data training dan data testing yang berbeda-beda. Pada setiap skenario akan diterapkan inisialisasi *iteration* metode Logistic regression yang berpatokan pada nilai : 100, 500 dan 1000.

Menurut Kaur *et al.*(2022), pengujian prediksi dengan menggunakan perbandingan data training dan data testing yang terdiri atas 90:10, 80:20, 70:30 dan 60:40 sangat efektif untuk karakteristik data yang mempunyai feature attribut lebih dari 10 dan dataset merupakan data yang imbalanced.

Uji coba didukung dengan penggunaan library *scikit-learn* dan dikombinasikan dengan library *tkinter, matplotlib* pada python sehingga menjadi aplikasi desktop sederhana yang berjalan pada sistem operasi *Windows* dengan tampilan interface sederhana. Tampilan interface tersebut ditunjukkan pada gambar 5.3.



Gambar 5.3 : Interface Prediksi Metode Logistic Regression

Pada gambar 5.3, interface prediksi metode Logistic Regression pada kotak warna merah dengan angka nomor 1 menunjukkan inputan untuk menentukan data testing yang memakai parameter tanggal mulai dan tanggal berakhir dengan format tanggal yyyy-mm-dd. Kotak merah dengan angka nomor 2 menunjukkan grafik prediksi dari metode Logistic Regression dan kotak merah dengan angka nomor 3 menunjukkan hasil detail pertanggal kesesuaian antara prediksi dan kondisi nyata. Kotak merah dengan angka 4 menunjukkan nilai performance dari metode Logistic Regression yang berisi *MAE*, *MAPE*, *Accuracy* dan *Duration*. Sedangkan kotak merah dengan angka nomor 5 menunjukkan estimasi nilai *coefficient* variabel independen.

Proses metode Logistic Regression dengan mengubah *max_iter* yang dilakukan dengan mengubah source code pada *method* Logistic Regression yang ada pada python. Perubahan tersebut hanya merubah angka yang terdiri dari 3 macam yaitu : 100,500 dan 1000. Dibawah ini merupakan penggalan source code yang diubah untuk *max_iter* sesuai dengan skenario pengujian.

```
# deklarasi metode Logistic Regression
def logistic_regression_forecast(train, testX):
    # ubah data menjadi array
```

```

train = asarray(train)

# pisah data menjadi input dan output
trainX, trainy = train[:, :-1], train[:, -1]

# proses metode
model = LogisticRegression(solver='liblinear',max_iter=100)
model.fit(trainX, trainy)

# hasilkan prediksi
yhat = model.predict([testX])
ylog = model.predict_proba([testX])

# hasilkan nilai coefficient
importance = model.coef_[0]

return yhat[0],ylog,importance

```

Penggalan source code di atas dengan warna yang tercetak merah, angka yang terdapat pada kurung akan diubah sesuai dengan skenario. Setelah dirubah aplikasi akan dijalankan ulang dan proses training dan testing data akan dilakukan kembali dengan performance yang berbeda. Evaluasi yang akan dilakukan mengambil hasil paling baik dari nilai *iteration* metode Logistic Regression yang berbeda-beda. Skenario uji coba dengan hasil yang terbaik akan dibandingkan dengan metode Random Forest.

5.2.1 Skenario 1

Skenario 1 dilakukan uji coba dengan menggunakan perbandingan data training dan data testing 90:10. Pengujian dilakukan sebanyak $3\times$ dengan *max_iter* dari metode Logistic Regression yang berbeda-beda. Performance yang diambil dari pengujian ini mencakup nilai *MAE*, *MAPE*, *accuracy* dan *duration* dari masing-masing *max_iter*. Nilai *max_iter* terdiri atas : 100, 500 dan 1000. Hasil dari pengujian dengan nilai *max_iter* = 100 metode Logistic Regression ditunjukkan pada Tabel 5.1.

Tabel 5.1 Hasil Prediksi Logistic Regression $max_iter = 100$ skenario 1

index	Tanggal	Expected	Predicted	Dev (predict - expect)
0	2023-01-01	3	3	0
1	2023-01-02	1	2	1
2	2023-01-03	3	2	-1
3	2023-01-04	2	2	0
4	2023-01-05	2	2	0
...
16	2023-01-17	2	2	0
17	2023-01-18	2	3	1

Hasil uji coba skenario 1 dengan $max_iter = 100$, menunjukkan hasil prediksi level omset dengan kondisi nyata omset hampir akurat. Pada Tabel 5.1 di tanggal 2023-01-03 hasil prediksi menunjukkan nilai lebih rendah -1 dari kondisi nyata, begitupun sebaliknya ditanggal 2023-01-18 hasil prediksi menunjukkan nilai yang lebih tinggi dari 1 dari kondisi nyata. Pengujian dengan $max_iter = 500$ ditunjukkan pada Tabel 5.2.

Tabel 5.2 Hasil Prediksi Logistic Regression $max_iter = 500$ skenario 1

index	Tanggal	Expected	Predicted	Dev (predict - expect)
0	2023-01-01	3	3	0
1	2023-01-02	1	2	1
2	2023-01-03	3	2	-1
3	2023-01-04	2	2	0
4	2023-01-05	2	2	0
...
16	2023-01-17	2	2	0
17	2023-01-18	2	3	1

Pengujian dengan $max_iter = 500$ menunjukkan hasil yang sama dengan sebelumnya. Tabel 5.2 selisih antara kondisi nyata dengan kondisi prediksi ditunjukkan pada tanggal 2023-01-03 dengan selisih kondisi prediksi lebih rendah -1. Hasil prediksi ditanggal yang lain sangat akurat. Tabel 5.3 merupakan hasil pengujian dengan $max_iter = 1000$.

Tabel 5.3 Hasil Prediksi Logistic Regression $max_iter = 1000$ skenario 1

index	Tanggal	Expected	Predicted	Dev (predict - expect)
0	2023-01-01	3	3	0
1	2023-01-02	1	2	1
2	2023-01-03	3	2	-1
3	2023-01-04	2	2	0
4	2023-01-05	2	2	0

Lanjutan Tabel Hasil Prediksi Logistic Regression *max_iter* = 1000 skenario 1

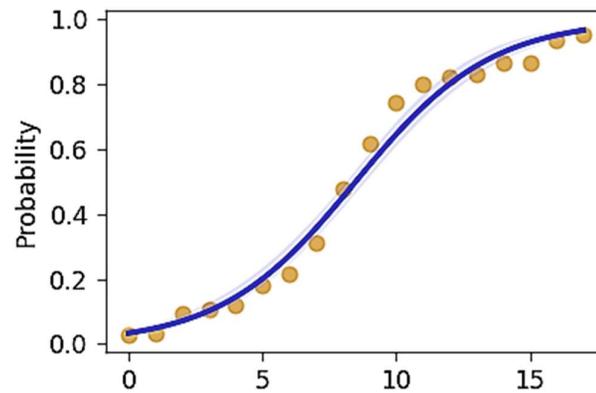
index	Tanggal	Expected	Predicted	Dev (predict - expect)
...
16	2023-01-17	2	2	0
17	2023-01-18	2	3	1

Pada Tabel 5.3 nilai prediksi masih tetap sama dengan nilai sebelumnya, yaitu tanggal 2023-01-03 nilai prediksi lebih rendah -1 dengan kondisi nyata, dan beberapa tanggal yang lain juga terdapat selisih. Proses perhitungan metode Logistic Regression pada skenario 1 menghasilkan nilai *Coefficient* sebagai estimasi hubungan antara variabel independen dengan variabel dependen, seperti ditunjukkan pada Tabel 5.4.

Tabel 5.4 Rank *Coefficient* Logistic Regression skenario 1

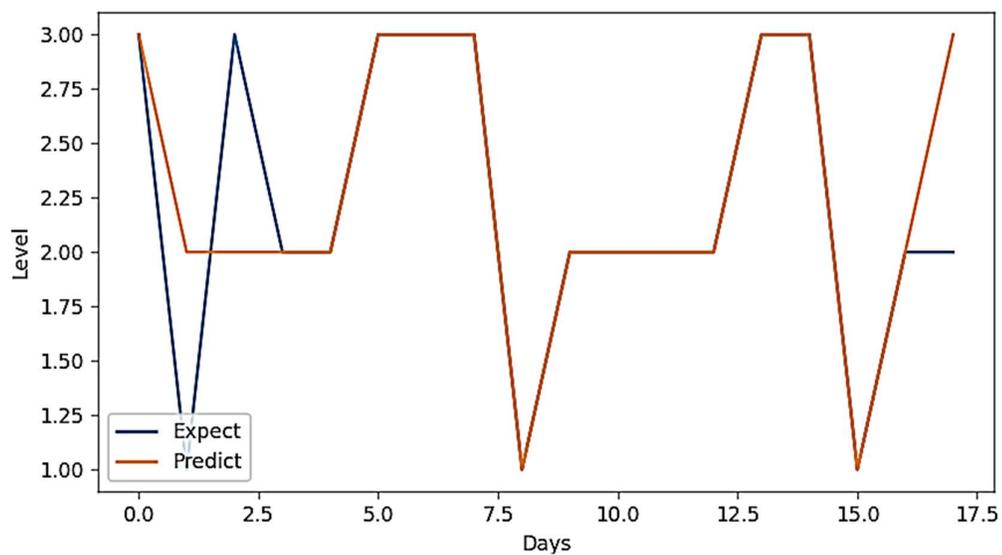
Rank	Variable Independen	Value
1	Soto Ayam Bungkus 20k	0.2079
2	Black Garlick 30k	0.0729
3	Soto Ayam Bungkus 15k	0.0622
4	Es Batu	0.0587
5	Jeruk Hangat/Dingin	0.0405
...
30	Soto Daging Sapi	-0.1417
31	Kripik Kentang	-0.2320

Tabel 5.4 menunjukkan evaluasi ranking dari nilai *coefficient* metode Logistic Regression dalam melakukan prediksi. Variabel independen yang paling berhubungan erat dengan dependen ditunjukkan oleh “Soto Ayam Bungkus 20k” dengan nilai 0.2079 dan variabel selanjutnya pada peringkat ke-2 ditunjukkan oleh “Black Garlick 30k” dengan nilai 0.0729. peringkat paling bawah atau variabel yang paling tidak berhubungan ditunjukkan oleh “Kripik Kentang” dengan nilai -0.2320 dan nomor ke-2 dari bawah ditunjukkan oleh “Soto Daging Sapi” dengan nilai -0.1417. Hasil diagram curva prediksi dengan metode Logistic Regression ditunjukkan pada Gambar 5.4.



Gambar 5.4 : Curva prediksi Logistic Regression skenario 1

Gambar 5.4 menunjukkan visualisasi tingkat akurasi prediksi yang baik metode Logistic Regression. Curva yang melengkung dengan titik yang menyimpang dari garis menunjukkan kondisi prediksi dan kondisi nyata tidak sesuai pada bobot probability urutan ke 10, 11 dan 15. Diagram yang menunjukkan hubungan kondisi nyata dan prediksi lebih mendetail ditunjukkan pada gambar 5.5.



Gambar 5.5 : Diagram line prediksi Logistic Regression skenario 1

Diagram line pada Gambar 5.5 menunjukkan hasil nyata dan prediksi dengan digambar garis kuning sebagai prediksi dan garis biru merupakan kondisi

nyata. Hubungan garis kuning dan garis biru yg tidak sejajar ada pada hari ke-1, ke-2, ke-3 dan Ke-18. Percobaan yang telah dilakukan dengan nilai *iteration* yang berbeda-beda metode Logistic Regression dirangkum pada Tabel 5.5.

Tabel 5.5 Performance metode Logistic Regression Skenario 1

<i>max_iter</i>	<i>MAE</i>	<i>MAPE</i>	<i>Accuracy</i>	<i>Duration</i>
100	0.31667	28.1851	88	10s
500	0.31667	28.1851	88	12s
1000	0.31667	28.1851	88	14s

Tabel 5.5 menunjukkan pada skenario 1 performance sangat identik sama dengan perolehan *MAE* 0.31667, *MAPE* 28.1851, *accuracy* 88%. Pada hasil skenario 1 yang bisa dievaluasi terletak pada *duration*, dengan hasil terbaik 10s pada kondisi *max_iter* = 100.

5.2.2 Skenario 2

Perbandingan data training dan data testing yang akan diuji pada skenario 2 80:20. Pengujian tetap dilakukan sebanyak 3× dengan nilai *max_iter* terdiri atas : 100, 500 dan 1000. Evaluasi performance yang akan dirangkum terdiri atas : *MAE*, *MAPE*, *Accuracy* dan *Duration*. Hasil dari pengujian dengan nilai *max_iter* = 100 metode Logistic Regression ditunjukkan pada Tabel 5.6.

Tabel 5.6 Hasil Prediksi Logistic Regression *max_iter* = 100 skenario 2

index	Tanggal	Expected	Predicted	Dev (predict - expect)
0	2023-01-01	3	3	0
1	2023-01-02	1	2	1
2	2023-01-03	3	2	-1
3	2023-01-04	2	2	0
4	2023-01-05	2	2	0
5	2023-01-06	3	3	0
...
33	2023-02-03	3	3	0
34	2023-02-04	3	3	0
35	2023-02-05	3	3	0

Hasil uji coba skenario 2 dengan *max_iter* = 100, menunjukkan hasil prediksi level omset sangat baik. Pada Tabel 5.6 di tanggal 2023-01-02 hasil prediksi menunjukkan nilai lebih tinggi 1 dari kondisi nyata, begitupun sebaliknya

ditanggal 2023-01-3 hasil prediksi menunjukkan nilai yang lebih rendah dari -1 dari kondisi nyata. Pengujian dengan $max_iter = 500$ ditunjukkan pada Tabel 5.7.

Tabel 5.7 Hasil Prediksi Logistic Regression $max_iter = 500$ skenario 2

index	Tanggal	Expected	Predicted	Dev (predict - expect)
0	2023-01-01	3	3	0
1	2023-01-02	1	2	1
2	2023-01-03	3	2	-1
3	2023-01-04	2	2	0
4	2023-01-05	2	2	0
5	2023-01-06	3	3	0
...
34	2023-02-04	3	3	0
35	2023-02-05	3	3	0

Pengujian dengan $max_iter = 500$ menunjukkan hasil yang sama dengan sebelumnya. Tabel 5.7 selisih antara kondisi nyata dengan kondisi prediksi ditunjukkan pada tanggal 2023-01-03 dengan selisih kondisi prediksi lebih rendah -1. Hasil prediksi ditanggal yang lain sangat akurat. Tabel 5.8 merupakan hasil pengujian dengan $max_iter = 1000$.

Tabel 5.8 Hasil Prediksi Logistic Regression $max_iter = 1000$ skenario 2

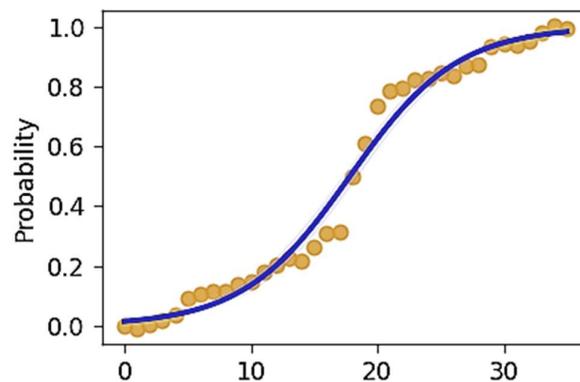
index	Tanggal	Expected	Predicted	Dev (predict - expect)
0	2023-01-01	3	3	0
1	2023-01-02	1	2	1
2	2023-01-03	3	2	-1
3	2023-01-04	2	2	0
4	2023-01-05	2	2	0
5	2023-01-06	3	3	0
...
33	2023-02-03	3	3	0
34	2023-02-04	3	3	0
35	2023-02-05	3	3	0

Pada Tabel 5.8 nilai prediksi masih tetap sama dengan nilai sebelumnya, yaitu tanggal 2023-01-03 nilai prediksi lebih rendah -1 dengan kondisi nyata, dan beberapa tanggal yang lain juga terdapat selisih. Proses perhitungan metode Logistic Regression pada skenario 2 menghasilkan nilai *Coefficient* sebagai estimasi hubungan antara variabel independen dengan variabel dependen, seperti ditunjukkan pada Tabel 5.9.

Tabel 5.9 Rank *Coefficient* Logistic Regression skenario 2

Rank	Variable Independen	Value
1	Soto Ayam Bungkus 20k	0.2079
2	Es Batu	0.0624
3	Black Garlick 30k	0.0582
4	Soto Ayam Bungkus 15k	0.0569
5	Soto bungkus 12K	0.0385
....
30	Soto Daging Sapi	-0.1543
31	Kripik Kentang	-0.2215

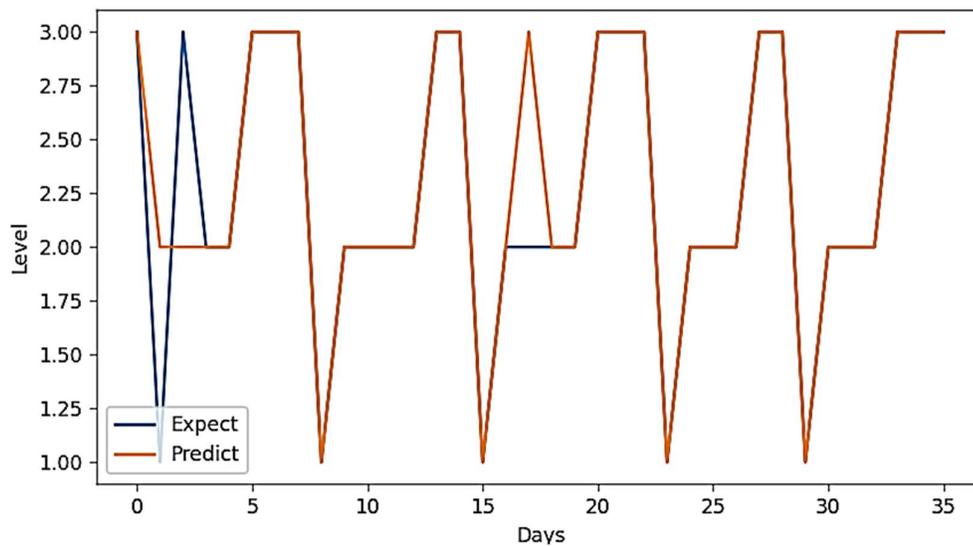
Tabel 5.9 menunjukkan evaluasi ranking dari nilai *coefficient* metode Logistic Regression dalam melakukan prediksi. Variabel independen yang paling berhubungan erat dengan dependen ditunjukkan oleh “Soto Ayam Bungkus 20k” dengan nilai 0.2079 dan variabel selanjutnya pada peringkat ke-2 ditunjukkan oleh “Es Batu” dengan nilai 0.0624. peringkat paling bawah atau variabel yang paling tidak berhubungan ditunjukkan oleh “Kripik Kentang” dengan nilai -0.2215 dan nomor ke-2 dari bawah ditunjukkan oleh “Soto Daging Sapi” dengan nilai -0.1543. Hasil diagram curva prediksi dengan metode Logistic Regression ditunjukkan pada Gambar 5.6.



Gambar 5.6 : Curva prediksi Logistic Regression skenario 2

Gambar 5.6 menunjukkan visualisasi tingkat akurasi prediksi yang baik metode Logistic Regression. Curva yang melengkung dengan titik yang menyimpang dari garis menunjukkan kondisi prediksi dan kondisi nyata tidak

sesuai pada bobot probability urutan ke 15-18 dan 21-22. Diagram yang menunjukkan hubungan kondisi nyata dan prediksi lebih mendetail ditunjukkan pada gambar 5.7.



Gambar 5.7 : Diagram line prediksi Logistic Regression skenario 2

Diagram line pada Gambar 5.7 menunjukkan hasil nyata dan prediksi dengan digambar garis kuning sebagai prediksi dan garis biru merupakan kondisi nyata. Hubungan garis kuning dan garis biru yg tidak sejajar ada pada hari ke-1, ke-2, ke-3 dan Ke-18. Percobaan yang telah dilakukan dengan nilai *iteration* yang berbeda-beda metode Logistic Regression dirangkum pada Tabel 5.10.

Tabel 5.10 Performance metode Logistic Regression Skenario 2

<i>max iter</i>	<i>MAE</i>	<i>MAPE</i>	<i>Accuracy</i>	<i>Duration</i>
100	0.23108	22.9549	97	10s
500	0.23108	22.9549	97	12s
1000	0.23108	22.9549	97	14s

Tabel 5.10 menunjukkan pada skenario 2 performance sangat identik sama dengan perolehan *MAE* 0.23108, *MAPE* 22.9549, *accuracy* 97%. Pada hasil skenario 2 yang bisa dievaluasi terletak pada *Duration*, dengan hasil terbaik 10s pada kondisi *max_iter* = 100.

5.2.3 Skenario 3

Skenario 3 memakai perbandingan data training dan data testing yang akan diuji coba yaitu 70:30. Pengujian masih dilakukan sebanyak $3\times$ dengan nilai *max_iter* terdiri atas : 100, 500 dan 1000. Evaluasi performance yang akan dirangkum terdiri atas : *MAE*, *MAPE*, *Accuracy* dan *Duration*. Hasil dari pengujian dengan nilai *max_iter* = 100 metode Logistic Regression ditunjukkan pada Tabel 5.11.

Tabel 5.11 Hasil Prediksi Logistic Regression *max_iter* = 100 skenario 3

index	Tanggal	Expected	Predicted	Dev (predict - expect)
0	2023-01-01	3	3	0
1	2023-01-02	1	2	1
2	2023-01-03	3	2	-1
3	2023-01-04	2	2	0
4	2023-01-05	2	2	0
5	2023-01-06	3	3	0
6	2023-01-07	3	3	0
...
52	2023-02-22	2	2	0
53	2023-02-23	2	2	0
54	2023-02-24	2	3	-0

Hasil uji coba skenario 3 dengan *max_iter* = 100, menunjukkan hasil prediksi level omset sedikit lebih baik dari skenario 2. Pada Tabel 5.11 di tanggal 2023-01-02 dan 2023-02-24 hasil prediksi menunjukkan nilai lebih tinggi 1 dari kondisi nyata, begitupun sebaliknya ditanggal 2023-01-3 hasil prediksi menunjukkan nilai yang lebih rendah dari -1 dari kondisi nyata. Pengujian dengan *max_iter* = 500 ditunjukkan pada Tabel 5.12.

Tabel 5.12 Hasil Prediksi Logistic Regression *max_iter* = 500 skenario 3

index	Tanggal	Expected	Predicted	Dev (predict - expect)
0	2023-01-01	3	3	0
1	2023-01-02	1	2	1
2	2023-01-03	3	2	-1
3	2023-01-04	2	2	0
4	2023-01-05	2	2	0
5	2023-01-06	3	3	0
6	2023-01-07	3	3	0
...
51	2023-02-21	2	2	0
52	2023-02-22	2	2	0

Lanjutan Tabel Hasil Prediksi Logistic Regression $max_iter = 500$ skenario 3

index	Tanggal	Expected	Predicted	Dev (predict - expect)
53	2023-02-23	2	2	0
54	2023-02-24	2	3	-0

Pengujian dengan $max_iter = 500$ menunjukkan hasil yang sama dengan sebelumnya. Tabel 5.12 selisih antara kondisi nyata dengan kondisi prediksi ditunjukkan pada tanggal 2023-01-03 dengan selisih kondisi prediksi lebih rendah -1. Hasil prediksi ditanggal yang lain sangat akurat. Tabel 5.13 merupakan hasil pengujian dengan $max_iter = 1000$.

Tabel 5.13 Hasil Prediksi Logistic Regression $max_iter = 1000$ skenario 3

index	Tanggal	Expected	Predicted	Dev (predict - expect)
0	2023-01-01	3	3	0
1	2023-01-02	1	2	1
2	2023-01-03	3	2	-1
3	2023-01-04	2	2	0
4	2023-01-05	2	2	0
5	2023-01-06	3	3	0
6	2023-01-07	3	3	0
...
51	2023-02-21	2	2	0
52	2023-02-22	2	2	0
53	2023-02-23	2	2	0
54	2023-02-24	2	3	-0

Pada Tabel 5.13 nilai prediksi masih tetap sama dengan nilai sebelumnya, yaitu tanggal 2023-01-03 nilai prediksi lebih rendah -1 dengan kondisi nyata, dan beberapa tanggal yang lain juga terdapat selisih. Proses perhitungan metode Logistic Regression pada skenario 3 menghasilkan nilai *Coefficient* sebagai estimasi hubungan antara variabel independen dengan variabel dependen, seperti ditunjukkan pada Tabel 5.14.

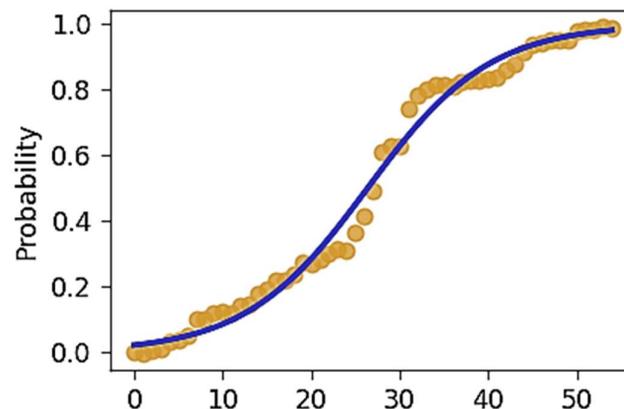
Tabel 5.14 Rank *Coefficient* Logistic Regression skenario 3

Rank	Variable Independen	Value
1	Soto Ayam Bungkus 20k	0.2078
2	Soto Ayam Bungkus 15k	0.0629
3	Es Batu	0.0624
4	Black Garlick 30k	0.0504
5	Soto bungkus 12K	0.0416
...
30	Soto Daging Sapi	-0.1663

Lanjutan Tabel Rank *Coefficient* Logistic Regression skenario 3

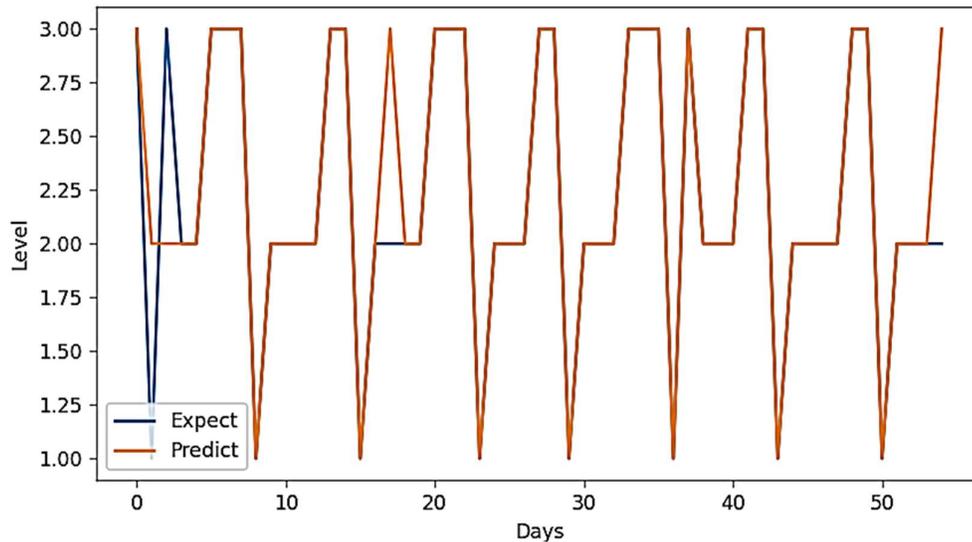
Rank	Variable Independen	Value
31	Kripik Kentang	-0.2251

Tabel 5.14 menunjukkan evaluasi ranking dari nilai *coefficient* metode Logistic Regression dalam melakukan prediksi. Variabel independen yang paling berhubungan erat dengan dependen ditunjukkan oleh “Soto Ayam Bungkus 20k” dengan nilai 0.2078 dan variabel selanjutnya pada peringkat ke-2 ditunjukkan oleh “Soto Ayam Bungkus 15k” dengan nilai 0.0629. peringkat paling bawah atau variabel yang paling tidak berhubungan ditunjukkan oleh “Kripik Kentang” dengan nilai -0.2251 dan nomor ke-2 dari bawah ditunjukkan oleh “Soto Daging Sapi” dengan nilai -0.1663. Hasil diagram curva prediksi dengan metode Logistic Regression ditunjukkan pada Gambar 5.8.



Gambar 5.8 : Curva prediksi Logistic Regression skenario 3

Gambar 5.8 menunjukkan visualisasi tingkat akurasi prediksi yang baik metode Logistic Regression. Curva yang melengkung dengan titik yang menyimpang dari garis menunjukkan kondisi prediksi dan kondisi nyata tidak sesuai pada bobot probability urutan ke 21-26 dan 33-36. Diagram yang menunjukkan hubungan kondisi nyata dan prediksi lebih mendetail ditunjukkan pada gambar 5.9.



Gambar 5.9 : Diagram line prediksi Logistic Regression skenario 3

Diagram line pada Gambar 5.9 menunjukkan hasil nyata dan prediksi dengan digambar garis kuning sebagai prediksi dan garis biru merupakan kondisi nyata. Hubungan garis kuning dan garis biru yg tidak sejajar ada pada hari ke-1, ke-2, ke-3, Ke-18 dan ke-52. Percobaan yang telah dilakukan dengan nilai *iteration* yang berbeda-beda metode Logistic Regression dirangkum pada Tabel 5.15.

Tabel 5.15 Performance metode Logistic Regression Skenario 3

<i>max iter</i>	<i>MAE</i>	<i>MAPE</i>	<i>Accuracy</i>	<i>Duration</i>
100	0.22273	22.24242	97	10s
500	0.22273	22.24242	97	12s
1000	0.22273	22.24242	97	14s

Tabel 5.15 menunjukkan pada skenario 3 performance sangat identik sama dengan perolehan *MAE* 0.22273, *MAPE* 22.24242, *accuracy* 97%. Pada hasil skenario 3 yang bisa dievaluasi terletak pada *Duration*, dengan hasil terbaik 10s pada kondisi *max_iter* = 100.

5.2.4 Skenario 4

Pada Skenario 4 uji coba akan dilakukan dengan memkai perbandingan 60:40. Pengujian selalu dilakukan sebanyak 3× dengan nilai *max_iter* terdiri atas :

100, 500 dan 1000. Evaluasi performance yang akan dirangkum terdiri atas : *MAE*, *MAPE*, *Accuracy* dan *Duration*. Hasil dari pengujian dengan nilai *max_iter* = 100 metode Logistic Regression ditunjukkan pada Tabel 5.16.

Tabel 5.16 Hasil Prediksi Logistic Regression *max_iter* = 100 skenario 4

index	Tanggal	Expected	Predicted	Dev (predict - expect)
0	2023-01-01	3	3	0
1	2023-01-02	1	2	1
2	2023-01-03	3	2	-1
3	2023-01-04	2	2	0
4	2023-01-05	2	2	0
5	2023-01-06	3	3	0
6	2023-01-07	3	3	0
7	2023-01-08	3	3	0
...
68	2023-03-09	2	2	0
69	2023-03-10	2	2	0
70	2023-03-11	3	2	-1
71	2023-03-12	3	3	0
72	2023-03-13	1	1	0

Hasil uji coba skenario 4 dengan *max_iter* = 100, menunjukkan hasil prediksi level omset hampir sama dengan hasil skenario 3. Pada Tabel 5.16 di tanggal 2023-01-02 hasil prediksi menunjukkan nilai lebih tinggi 1 dari kondisi nyata, begitupun sebaliknya ditanggal 2023-01-3 dan 2023-03-11 hasil prediksi menunjukkan nilai yang lebih rendah dari -1 dari kondisi nyata. Pengujian dengan *max_iter* = 500 ditunjukkan pada Tabel 5.17

Tabel 5.17 Hasil Prediksi Logistic Regression *max_iter* = 500 skenario 4

index	Tanggal	Expected	Predicted	Dev (predict - expect)
0	2023-01-01	3	3	0
1	2023-01-02	1	2	1
2	2023-01-03	3	2	-1
3	2023-01-04	2	2	0
4	2023-01-05	2	2	0
5	2023-01-06	3	3	0
6	2023-01-07	3	3	0
7	2023-01-08	3	3	0
...
68	2023-03-09	2	2	0
69	2023-03-10	2	2	0
70	2023-03-11	3	2	-1
71	2023-03-12	3	3	0
72	2023-03-13	1	1	0

Pengujian dengan $max_iter = 500$ menunjukkan hasil yang sama dengan sebelumnya. Tabel 5.17 selisih antara kondisi nyata dengan kondisi prediksi ditunjukkan pada tanggal 2023-01-03 dengan selisih kondisi prediksi lebih rendah -1. Hasil prediksi ditanggal yang lain sangat akurat. Tabel 5.18 merupakan hasil pengujian dengan $max_iter = 1000$.

Tabel 5.18 Hasil Prediksi Logistic Regression $max_iter = 1000$ skenario 4

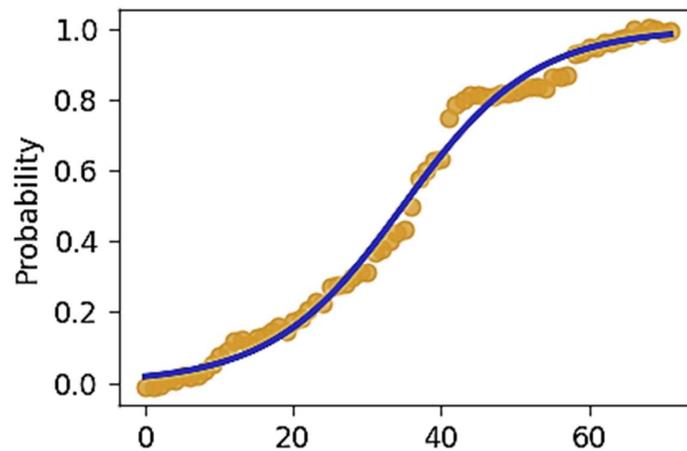
index	Tanggal	Expected	Predicted	Dev (predict - expect)
0	2023-01-01	3	3	0
1	2023-01-02	1	2	1
2	2023-01-03	3	2	-1
3	2023-01-04	2	2	0
4	2023-01-05	2	2	0
5	2023-01-06	3	3	0
6	2023-01-07	3	3	0
7	2023-01-08	3	3	0
...
68	2023-03-09	2	2	0
69	2023-03-10	2	2	0
70	2023-03-11	3	2	-1
71	2023-03-12	3	3	0
72	2023-03-13	1	1	0

Pada Tabel 5.18 nilai prediksi masih tetap sama dengan nilai sebelumnya, yaitu tanggal 2023-01-03 nilai prediksi lebih rendah -1 dengan kondisi nyata, dan beberapa tanggal yang lain juga terdapat selisih. Proses perhitungan metode Logistic Regression pada skenario 4 menghasilkan nilai *Coefficient* sebagai estimasi hubungan antara variabel independen dengan variabel dependen, seperti ditunjukkan pada Tabel 5.19.

Tabel 5.19 Rank *Coefficient* Logistic Regression skenario 4

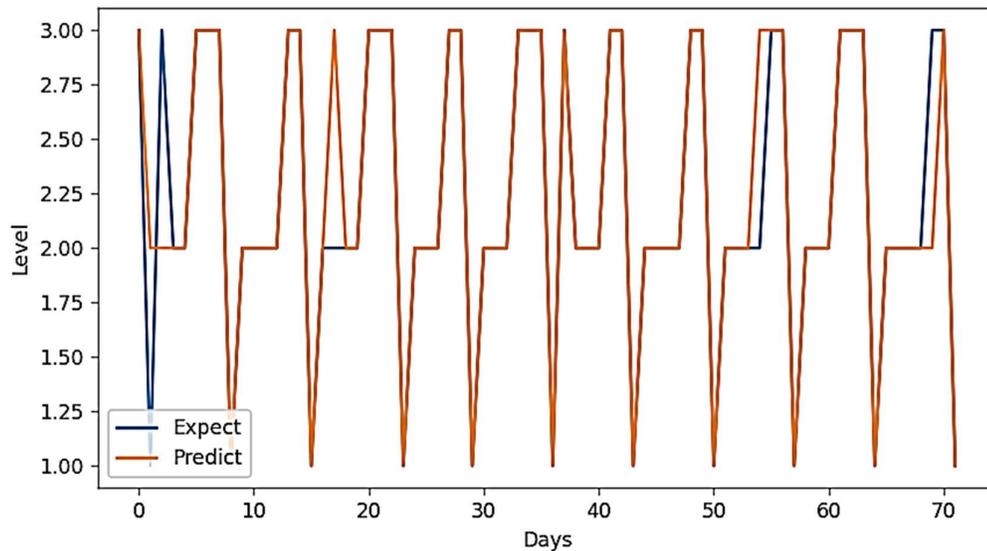
Rank	Variable Independen	Value
1	Soto Ayam Bungkus 20k	0.1922
2	Es Batu Soto	0.0648
3	Ayam Bungkus 15k	0.0623
4	Black Garlick 30k	0.0479
5	Soto bungkus 12K	0.0412
...
30	Soto Daging Sapi	-0.1721
31	Kripik Kentang	-0.2271

Tabel 5.19 menunjukkan evaluasi ranking dari nilai *coefficient* metode Logistic Regression dalam melakukan prediksi. Variabel independen yang paling berhubungan erat dengan dependen ditunjukkan oleh “Soto Ayam Bungkus 20k” dengan nilai 0.1922 dan variabel selanjutnya pada peringkat ke-2 ditunjukkan oleh “Es Batu” dengan nilai 0.0648. peringkat paling bawah atau variabel yang paling tidak berhubungan ditunjukkan oleh “Kripik Kentang” dengan nilai -0.2271 dan nomor ke-2 dari bawah ditunjukkan oleh “Soto Daging Sapi” dengan nilai -0.1721. Hasil diagram curva prediksi dengan metode Logistic Regression ditunjukkan pada Gambar 5.10.



Gambar 5.10 : Curva prediksi Logistic Regression skenario 4

Gambar 5.10 menunjukkan visualisasi tingkat akurasi prediksi yang baik metode Logistic Regression. Curva yang melengkung dengan titik yang menyimpang dari garis menunjukkan kondisi prediksi dan kondisi nyata tidak sesuai pada bobot probability urutan ke 41-46 dan 55-60. Diagram yang menunjukkan hubungan kondisi nyata dan prediksi lebih mendetail ditunjukkan pada gambar 5.11.



Gambar 5.11 : Diagram line prediksi Logistic Regression skenario 4

Diagram line pada Gambar 5.11 menunjukkan hasil nyata dan prediksi dengan digambar garis kuning sebagai prediksi dan garis biru merupakan kondisi nyata. Hubungan garis kuning dan garis biru yg tidak sejajar ada pada hari ke-1, ke-2, ke-3, Ke-18, ke-52 dan ke-70. Percobaan yang telah dilakukan dengan nilai *iteration* yang berbeda-beda metode Logistic Regression dirangkum pada Tabel 5.20.

Tabel 5.20 Performance metode Logistic Regression Skenario 4

<i>max iter</i>	<i>MAE</i>	<i>MAPE</i>	<i>Accuracy</i>	<i>Duration</i>
100	0.21944	21.7037	97	14s
500	0.21944	21.7037	97	18s
1000	0.21944	21.7037	97	22s

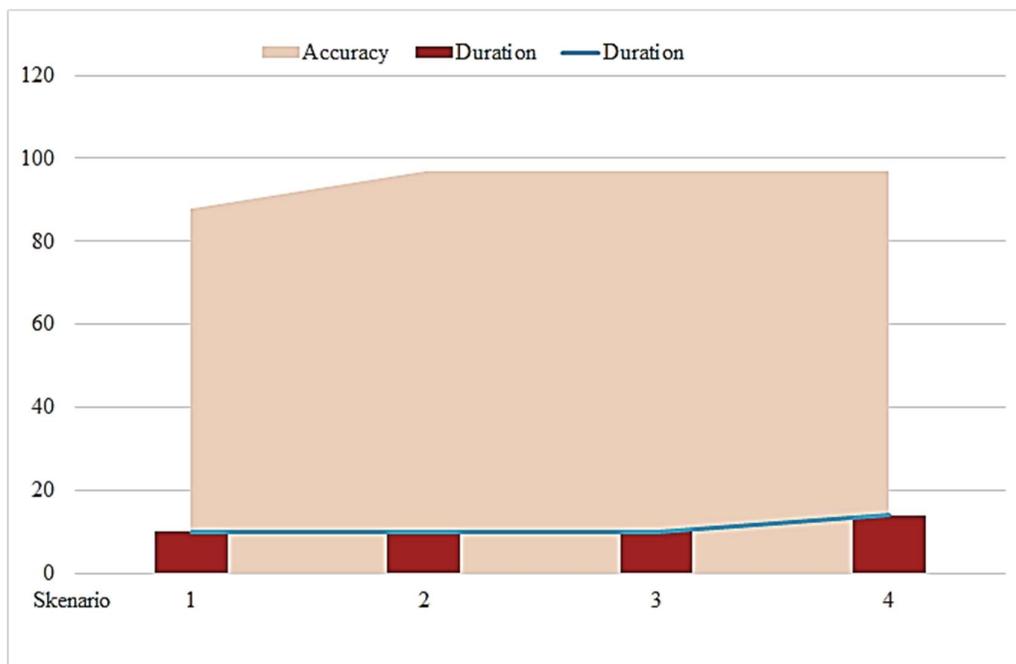
Tabel 5.20 menunjukkan pada skenario 4 performance sangat identik sama dengan perolehan *MAE* 0.1944, *MAPE* 21.7037, *accuracy* 97%. Pada hasil skenario 4 yang bisa dievaluasi terletak pada *Duration*, dengan hasil terbaik 14s pada kondisi *max_iter* = 100.

5.3 Evaluasi

Tahap evaluasi dalam pengujian metode Logistic Regression mengambil dari hasil prediksi dengan performance terbaik dari masing-masing skenario. Hasil yang paling baik dari 4 skenario akan menjadi bahan perbandingan dengan metode yang lain. Evaluasi dari performance terbaik masing-masing skenario ditunjukkan pada Tabel 5.21.

Tabel 5.21 Evaluasi Performance Metode Logistic Regression

Skenario	Perbandingan data	<i>max_iter</i>	<i>MAE</i>	<i>MAPE</i>	<i>Accuracy</i>	<i>Duration</i>
1	90:10	100	0.31667	28.1851	88	10s
2	80:20	100	0.23108	22.9549	97	10s
3	70:30	100	0.22273	22.2424	97	10s
4	60:40	100	0.21944	21.7037	97	14s



Gambar 5.12 : Diagram Evaluasi Performance Metode Logistic Regression

Hasil dari evaluasi yang ditunjukkan pada Tabel 5.21 dan gambar 5.12 bahwa Prediksi model dengan dataset bisnis restoran Soto Kwali Pak Wasis dengan menggunakan metode Logistic Regression dapat mencapai hasil yang maksimal dengan parameter perbandingan data training dan data testing 70:30 dan *iteration*

metode-nya menggunakan $max_iter = 100$. Menurut Alam *et al.* (2019) metode yang menghasilkan nilai *accuracy* tertinggi dengan *duration* yang lebih kecil mempunyai kalkulasi perhitungan yang efisien, hasil uji coba pada tabel 5.21 dan gambar 5.12 *Accuracy* 97% dan *duration* 10s merupakan hasil terbaik.

BAB VI

PEMBAHASAN

6.1. Perbandingan Performance Random Forest dan Logistic Regression

Dalam proses mendapatkan prediksi model dengan metode *Random Forest* (RF) atau *Logistic Regression* (LR), metode dengan hasil terbaik memerlukan pembahasan evaluasi dari hasil uji coba masing-masing metode dengan skenario dan modifikasi metode yang telah diterapkan. Evaluasi performance metode mengacu pada hasil utama yaitu *accuracy*. Dalam membandingkan metode dengan memakai performance *accuracy*, bila terjadi hasil yang sama maka akan dipertimbangkan untuk membandingkan nilai *duration*.

Untuk mendukung tingkat efisiensi dari proses kalkulasi metode dalam memprediksikan, evaluasi metode juga mengacu pada *duration*. Evaluasi *duration* merupakan penilaian metode dengan mempertimbangkan lama waktu setelah tombol proses ditekan sampai hasil prediksi dan grafik evaluasi prediksi ditampilkan. Satuan yang dipakai dalam evaluasi *duration* memakai satuan *second(s)* atau detik, semakin kecil nilai *duration*-nya maka semakin cepat metode dalam memproses prediksi menjadikan metode tersebut sangat efisien.

Evaluasi dengan memakai acuan *MAE* dan *MAPE* menggunakan kondisi bahwa nilai yang paling kecil dari uji coba metode merupakan hasil yang paling baik, seperti yang dipaparkan Kureljusic *et al.* (2022) bahwa semakin rendah nilai *MAE* dan *MAPE* maka akan sangat berpengaruh dengan hasil prediksi dari suatu metode.

6.1.1. Perbandingan Metode Menggunakan Accuracy

Rangkuman dari hasil uji coba dari metode *Random Forest* (RF) dan *Logistic Regression* (LR) dengan masing-masing skenario menunjukkan hasil prediksi yang bagus. Dari hasil yang telah didapat dari experiment akan diambil nilai *accuracy* paling tinggi dengan satuan persen (%), sehingga nilai yang mendekati 100% merupakan metode yang terbaik.

Metode *Random Forest* (RF) dalam uji coba menggunakan parameter atau modifikasi metode dengan $n_estimator$ yang terdiri atas : 50,100,500 dan 1000. 4 macam skenario yang diterapkan dalam uji coba dengan perbedaan dari dataset yang ditest menghasilkan nilai *accuracy* paling tinggi yaitu 98,78% dengan parameter $n_estimator = 50$ dan perbandingan dataset 90:10 dengan total jumlah dataset 182. Nilai *accuracy* ini diproses dengan membandingkan hasil prediksi dengan kondisi nyata yang dihitung memakai rumus *confusion matrix*. Bila dirata-rata nilai *accuracy* dari metode *Random Forest* (RF) mendapatkan nilai 98,78%, dan nilai min dari *accuracy* yaitu 98,78% dan nilai max dari *accuracy* adalah 98,78%, sehingga standart deviasi (std) mendapatkan nilai 0 sehingga konsistensi *accuracy*-nya bagus.

Metode *Logistic Regression* (LR) diuji coba dengan skenario 4 macam dan menggunakan nilai *iteration* yang terdiri atas max_iter yaitu : 100,500 dan 1000 menghasilkan nilai *accuracy* paling tinggi yaitu 97% pada kondisi max_iter bernilai 100 dan perbandingan dataset 70:30 dari total dataset 182. Nilai rata dari *accuracy* metode *Logistic Regression* (LR) yaitu 94,75% dan mendapatkan nilai min dari *accuracy* yaitu : 88% sedangkan nilai max 97%, sehingga nilai dari standart deviasi (std) mendapatkan nilai 3,89711.

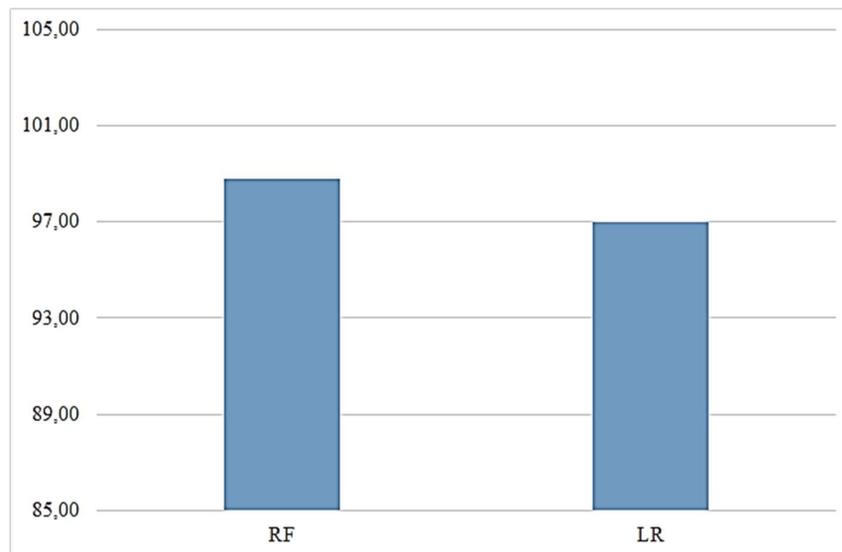
Perbandingan antara metode *Random Forest* (RF) dan *Logistic Regression* (LR) dari nilai *accuracy* bertujuan untuk mendapatkan nilai *accuracy* paling tinggi dari metode yang telah dilakukan uji coba. Hasil *accuracy* masing-masing metode dirangkum pada Tabel 6.1.

Tabel 6.1 Perbanding *accuracy* Metode RF dan LR

Metode	<i>accuracy</i>	Min	Max	Std	Avg
RF	98,78%	98,78	98,78	0	98,78%
LR	97%	88%	97%	3,897114	94,75%

Pada tabel 6.1 nilai *accuracy* metode *Random Forest* (RF) lebih baik daripada metode *Logistic Regression* (LR) untuk melakukan prediksi omset dari bisnis restaurant dengan nilai 98,78%. Nilai std dari metode *Random Forest* (RF) sangat kecil yaitu 0 sehingga keberagaman hasil *accuracy* yang didapat tidak terlalu banyak, hal ini menunjukkan bahwa metode *Random Forest* (RF) mampu

menghasilkan hasil prediksi dengan tingkat *accuracy* yang stabil. Begitupun rentang antara nilai min dan max *accuracy* yang tidak terlalu jauh sehingga lebih memperkuat bahwa metode *Random Forest* (RF) mempunyai nilai *accuracy* yang baik dalam proses prediksi. Visualisasi perbandingan metode *Random Forest* (RF) dan *Logistic Regression* (LR) yang mengacu pada nilai *accuracy* ditunjukkan pada Gambar 6.1.



Gambar 6.1 : Perbandingan *Accuracy* RF dan LR

Gambar 6.1 menunjukkan rentang hasil *accuracy* metode *Random Forest* (RF) dengan metode *Logistic Regression* (LR) sangat tipis yaitu kurang lebih 1,5%. Unggulnya metode *Random Forest* (RF) dibandingkan dengan *Logistic Regression* (LR) pada nilai *accuracy* juga dibuktikan dengan nilai standart deviasi (std) dari metode *Random Forest* (RF) juga sangat rendah yang berada pada rentang 0. Sedangkan metode *Logistic Regression* (LR) mempunyai nilai standart deviasi (std) *accuracy* diatas 3. Kesimpulan ini untuk *accuracy* metode *Random Forest* (RF) sedikit lebih baik dalam melakukan prediksi omset bisnis restoran dibandingkan dengan metode *Logistic Regression* (LR).

6.1.2. Perbandingan Metode Menggunakan *Duration*

Evaluasi performance metode selanjutnya adalah membahas tentang *duration* yang bertujuan untuk mengetahui seberapa cepat metode dalam melakukan proses prediksi. Evaluasi *duration* menggunakan satuan second (s) / detik yang menunjukkan lama dari proses prediksi yang akan dilakukan. Semakin kecil nilai *duration* per second (s) maka semakin baik metode tersebut dalam melakukan perhitungan yang dapat menghasilkan prediksi terbaik.

Metode *Random Forest* (RF) dengan melakukan uji coba sebanyak 4 kali dengan perbedaan perbandingan dataset dan menggunakan parameter *n_estimators* yang terdiri atas : 50,100,500 dan 1000 mendapatkan hasil tercepat 26s pada *accuracy* tertinggi dan error terendah. Rata-rata dari *duration* yaitu 32s dan min-nya adalah 26s dan max-nya adalah 36s. sehingga standart deviasi-nya adalah 3,74165. Dari hasil tersebut metode *Random Forest* (RF) memerlukan waktu untuk melakukan proses prediksi karena banyaknya variasi perhitungan yang dilakukan.

Metode *Logistic Regression* (LR) dengan parameter *max_iter* yang terdiri atas : 100,500 dan 1000 dengan dilakukan 4 kali uji coba yang menggunakan perbandingan dataset bervariasi menghasilkan nilai *duration* tercepat 10s pada *accuracy* tertinggi dan error terendah. Rata-rata *duration* yaitu 11s dan min-nya adalah 10s sedangkan max-nya adalah 14s. standart deviasi-nya adalah 1,73205. Waktu eksekusi metode *Logistic Regression* (LR) lebih cepat atau lebih baik dari metode *Random Forest* (RF) sehingga dapat disimpulkan bahwa metode *Logistic Regression* (LR) merupakan metode yang ringan dan cepat sehingga sangat cocok untuk penerapan dalam bisnis yang membutuhkan waktu eksekusi yang efisien.

Perbandingan antara metode *Random Forest* (RF) dan *Logistic Regression* (LR) dari nilai *duration* bertujuan untuk mendapatkan metode tercepat dalam melakukan prediksi. Hasil *duration* masing-masing metode dirangkum pada Tabel 6.3.

Tabel 6.2 Perbandingan *duration* Metode RF dan LR

Metode	<i>Duration</i>
RF	26s
LR	10s

Berdasarkan tabel 6.2 metode *Logistic Regression* (LR) lebih cepat dengan perolehan *duration* 10s dibandingkan dengan *Random Forest* (RF) dengan perolehan 26s. Hal ini menunjukkan bahwa metode *Logistic Regression* (LR) merupakan metode yang ringan dan cepat dan nilai *accuracy* yang masih baik. Visualisasi perbandingan *duration* ditunjukkan pada gambar 6.5.



Gambar 6.2 : Perbandingan nilai *duration* RF dan LR

6.1.3. Perbandingan Metode Menggunakan *MAE* dan *MAPE*

Pembuktian metode yang terbaik dengan melakukan tingkat rendahnya dari nilai error hasil prediksi dari masing-masing metode dengan menggunakan *MAE* dan *MAPE*. Nilai *MAE* dan *MAPE* yang paling mendekati 0 merupakan hasil yang paling baik untuk menentukan perbandingan antar metode, *MAE* dan *MAPE* menggunakan perhitungan yang memakai selisih dari nilai nyata dengan nilai prediksi.

Evaluasi *MAE* dan *MAPE* pada metode *Random Forest* (RF) menggunakan hasil dari uji coba 4 skenario dengan modifikasi metode yang terdiri atas parameter $n_estimators = 50, 100, 500$ dan 1000. Nilai *MAE* dan *MAPE* terbaik ada pada kondisi skenario dengan dataset 90:10 dengan parameter $n_estimators = 50$ mendapatkan nilai *MAE* 0,23222 dan *MAPE* mendapatkan nilai 22,1852. Rata-rata dari nilai error yang didapat yaitu *MAE* 0,231068 dan *MAPE* 21,9951. Sedangkan

untuk nilai min dan max yaitu *MAE* min 0,22838, max 0,23222 dan *MAPE* min 21,8469, max 22,1852. Sehingga standart deviasi-nya adalah *MAE* 0,001583, *MAPE* 0,1220 sehingga performace error metode *Random Forest* (RF) tergolong stabil.

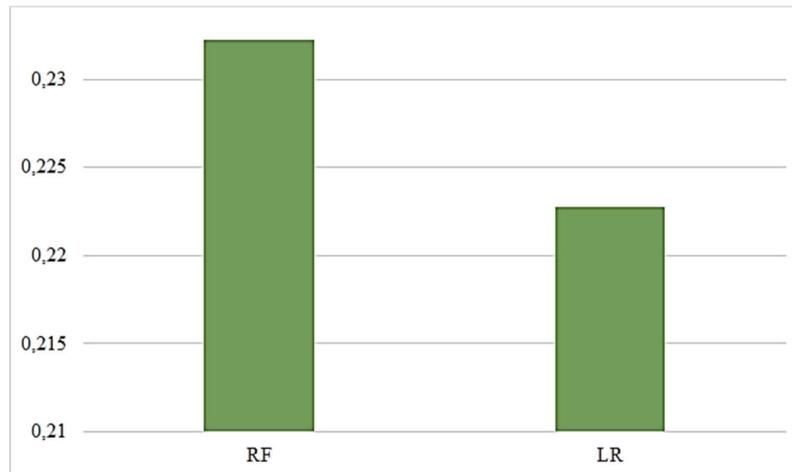
Evaluasi pada metode *Logistic Regression* (LR) nilai *MAE* dan *MAPE* menggunakan 4 skenario dengan dataset yang berbeda-beda dan diterapkan menggunakan parameter *max_iter* yang terdiri atas : 100,500 dan 1000. Nilai terbaik *MAE* dan *MAPE* terdapat pada nilai perbandingan dataset 70:30 dengan parameter *max_iter* = 100 dengan perolehan *MAE* 0,22273, *MAPE* 22,2424. Rata-rata dari *MAE* dan *MAPE* yaitu 0,24748, 23,7715. Sedangkan nilai min-max dari evaluasi error yaitu *MAE* min 0,21944, max 0,31667, *MAPE* min 21,7037, max 0,281851. Sehingga standart deviasi-nya adalah *MAE* 0,040172, *MAPE* 2,5865 yang menunjukkan performace error kategori stabil.

Perbandingan antara metode *Random Forest* (RF) dan *Logistic Regression* (LR) dari nilai *MAE* dan *MAPE* bertujuan untuk mendapatkan nilai error yang paling rendah dari metode yang telah dilakukan uji coba. Hasil *MAE* dan *MAPE* masing-masing metode dirangkum pada Tabel 6.2.

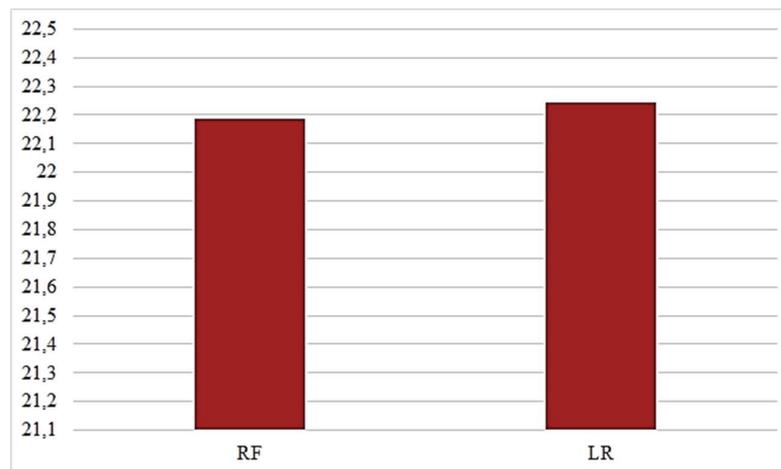
Tabel 6.3 Perbandingan *MAE* dan *MAPE* Metode RF dan LR

Metode	<i>MAE</i>	<i>MAPE</i>
RF	0,23222	22,1852
LR	0,22273	22,2424

Berdasarkan tabel 6.3 metode *Random Forest* (RF) dan *Logistic Regression* (LR) mempunyai performace *MAE* dan *MAPE* yang mirip. Nilai *Logistic Regression* (LR) perolehan *MAE* 0,22273 dan *MAPE* 22,1852 mempunyai nilai yang sama dengan perolehan *Random Forest* (RF) yang memperoleh *MAE* 0,23222 dan *MAPE* 22,1852. Hal ini menunjukkan metode *Logistic Regression* (LR) dan *Random Forest* (RF) dalam evaluasi error mempunyai kesalahan yang rendah dalam melakukan prediksi omset bisnis restoran karena hampir tidak ada selisih antara 2 metode tersebut. Visualisasi perbandingan *MAE* dan *MAPE* ditunjukkan pada gambar 6.3 dan 6.4



Gambar 6.3 : Perbandingan nilai *MAE* RF dan LR



Gambar 6.4 : Perbandingan nilai *MAPE* RF dan LR

6.2. Implementasi Metode Terbaik

Hasil yang telah diperoleh dari perbandingan metode *Random Forest* (RF) dan *Logistic Regression* (LR) membuktikan bahwa metode *Logistic Regression* (LR) merupakan metode paling cocok dalam melakukan prediksi omset bisnis restoran soto-kwali pak wasis karena durasinya lebih unggul. Untuk mempertegas dari hasil tersebut maka akan dilakukan implementasi dari metode terbaik.

Data yang akan diimplementasikan pada metode *Logistic Regression* (LR) berjumlah 335 dengan menggunakan perbandingan dataset 70:30 dan parameter

yang akan diterapkan dalam metode memakai $max_iter = 100$. Omset yang akan dievaluasi mencakup kategori tinggi, standart dan rendah. Hasil dari implementasi akan dilakukan evaluasi dari selisih kondisi nyata dengan kondisi prediksi masing-masing kategori dengan performance nilai rata-rata, min, max dan standart deviasi (std). hasil implementasi ditunjukkan pada tabel 6.4.

Tabel 6.4 Performance Rata-rata, Min, Max dan Std berdasarkan omset

Omset	Rata-rata	Min	Max	Std
Tinggi	-0,125	-1	0	0,330719
Standart	0,046512	-1	1	0,30143
Rendah	0,066667	0	1	0,249444

Evaluasi pada tabel 6.4 menunjukkan bahwa prediksi yang paling baik terletak pada omset Standart dengan nilai rata-rata paling kecil yaitu 0,046 dan Rendah dengan nilai 0,06. Sehingga prediksi untuk omset Standart dan Rendah mempunyai tingkat akurasi yang baik. Sedangkan nilai prediksi yang kurang sesuai dengan kondisi nyata terdapat pada omset Tinggi yang memperoleh Std 0,33.

Prediksi yang telah diimplementasikan memperoleh hasil yang sangat baik sehingga prediksi ini sangat membantu untuk analisa omset dari bisnis restoran tanpa perlu menghitung secara manual dari nominal uang dari hasil makanan yang terjual. Harapan setelah dilakukan implementasi prediksi model dapat digunakan menjadi bahan pertimbangan dalam mempersiapkan pengembangan bisnis restoran dimasa yang akan datang sehingga bisnis restoran dapat menjadi lebih besar dan membuka lapangan pekerjaan yang lebih banyak.

6.3. Prediksi Omset Bisnis Restoran dalam Pandangan Islam

Bisnis restoran dan setiap kegiatannya dalam pandangan Islam diawali dengan terjalinnya hubungan antara Allah SWT dengan manusia atau yang sering disebut (حَبْلٌ مِّنَ اللَّهِ), di mana ketetapan Allah SWT yang berupa takdir telah sudah ditetapkan sebelum dunia diciptakan, tetapi Allah memberi jaminan takdir tersebut dapat berubah bila manusia mau berusaha dalam melakukan kegiatannya dimuka bumi, sesuai dengan surah Ar-Ra'd ayat 11 yang berbunyi :

لَهُ مُعَقِّبَاتٌ مِّنْ بَيْنِ يَدَيْهِ وَمِنْ خَلْفِهِ يَحْفَظُونَهُ مِنْ أَمْرِ اللَّهِ إِنَّ اللَّهَ لَا يُغَيِّرُ مَا بِقَوْمٍ حَتَّىٰ يُغَيِّرُوا مَا بِأَنْفُسِهِمْ ۗ وَإِذَا أَرَادَ اللَّهُ بِقَوْمٍ سُوءًا فَلَا مَرَدَّ لَهُ ۗ وَمَا لَهُمْ مِنْ دُونِهِ مِنْ وَالٍ

Artinya: Bagi manusia ada malaikat-malaikat yang selalu mengikutinya bergiliran, di muka dan di belakangnya, mereka menjaganya atas perintah Allah. Sesungguhnya Allah tidak merubah keadaan sesuatu kaum sehingga mereka merubah keadaan yang ada pada diri mereka sendiri. Dan apabila Allah menghendaki keburukan terhadap sesuatu kaum, maka tak ada yang dapat menolaknya; dan sekali-kali tak ada pelindung bagi mereka selain Dia.

Dalam tafsir imam al-qurthubi menyatakan bahwa Allah tidak akan mengubah keadaan suatu kaum sehingga mereka berusaha untuk mengubah keadaan mereka sendiri. Sudut pandang dalam prediksi omset bisnis restoran bisa dianalisa dengan berusaha untuk melakukan evaluasi omset dengan melakukan prediksi sebagai bentuk ikhtiar *hablum minallah* sebagai hamba yang ingin merubah kondisi bisnis supaya menjadi lebih baik.

Dalam tuntunan agama Islam tentang *hablum minallah* disebutkan setelah melakukan ikhtiar maka sebagai hamba Allah diharuskan untuk tawakal kepada Allah seperti dalam surah Ali-Imran ayat 160 yang berbunyi :

إِن يَنْصُرْكُمُ اللَّهُ فَلَا غَالِبَ لَكُمْ وَإِن يَخْذَلْكُمْ فَمَنْ ذَا الَّذِي يَنْصُرُكُمْ مِنْ بَعْدِهِ ۗ وَعَلَى اللَّهِ فَلْيَتَوَكَّلِ الْمُؤْمِنُونَ

Artinya: Jika Allah menolong kamu, maka tidak ada yang dapat mengalahkanmu, tetapi jika Allah membiarkan kamu (tidak memberi pertolongan), maka siapa yang dapat menolongmu setelah itu? Karena itu, hendaklah kepada Allah saja orang-orang mukmin bertawakal.

Dalam tafsir Quraish Shihab yaitu bila Allah mendukung kalian dengan pertolongan-Nya, seperti yang terjadi dalam perang Badar, tidak akan ada yang dapat mengalahkan kalian, Begitu pula sebaliknya. Sesungguhnya hanya kepada Allahlah, orang-orang mukmin boleh bersandar dan menyerahkan urusan. Dalam perpektif Islam melakukan usaha dengan langkah memprediksikan omset dari bisnis restoran sangat dianjurkan untuk tetap bertawakal kepada Allah apapun hasil yang telah tercapai, bila tidak sesuai antara prediksi dengan kondisi nyata maka berprasangka baik kepada Allah, karena hanya Allah yang bisa menolong manusia

dari segala urusan duniawi dan manusia tetap bersyukur atas takdir yang telah terjadi.

Bila jalinan manusia dengan Allah telah dijalankan dengan baik, maka selanjutnya adalah jalinan antar sesama manusia perlu dijalin dengan baik. Tuntunan dakwah yang baik dalam memberikan pelayanan dan memudahkan sesama manusia supaya menjadi lebih produktif. Bisnis restoran soto-kwali pak wasis yang merupakan bisnis dengan skala kecil menyajikan pilihan harga makanan yang terjangkau dengan kualitas yang baik. Dalam ajaran agama Islam dalam menjalankan bisnis harus berdasarkan kepercayaan dan menepati janji sesuai dengan harga dan hak-kewajiban dalam bisnis tersebut seperti yang dibahas oleh Pratama *et al.*(2022). Landasan hukum Islam dalam menjalankan bisnis terdapat pada surah Asy-Syu'ara ayat 181-183 yang berbunyi :

أَوْفُوا الْكَيْلَ وَلَا تَكُونُوا مِنَ الْمُخْسِرِينَ ﴿١٨١﴾ وَزِنُوا بِالْقِسْطِاسِ الْمُسْتَقِيمِ ﴿١٨٢﴾ وَلَا تَبْخَسُوا النَّاسَ أَشْيَاءَهُمْ وَلَا تَعْنُوا فِي الْأَرْضِ مُفْسِدِينَ ﴿١٨٣﴾

Artinya: 181. Sempurnakanlah takaran dan janganlah kamu merugikan orang lain; 182. dan timbanglah dengan timbangan yang benar. 183. Dan janganlah kamu merugikan manusia dengan mengurangi hak-haknya, dan janganlah membuat kerusakan di bumi; (Q. S. Asy-Syu'ara (26): 181-183).

Dalam tafsir Quraish Shihab menyatakan dan timbanglah sesuatu dengan timbangan yang sempurna, sehingga orang-orang dapat mengambil haknya secara adil dan benar. Dalam hadis rosul At Tirmidzi no.1209 yang berbunyi :

عَنْ أَبِي سَعِيدٍ
عَنْ النَّبِيِّ صَلَّى اللَّهُ عَلَيْهِ وَسَلَّمَ قَالَ التَّاجِرُ الصَّنُوقُ الْأَمِينُ مَعَ النَّبِيِّينَ وَالصِّدِّيقِينَ وَالشُّهَدَاءِ

*Dari Abu Sa'id dari Nabi shallallahu 'alaihi wasallam, beliau bersabda:
"Seorang pedagang yang jujur dan dipercaya akan bersama dengan para Nabi, shiddiqun dan para syuhada."*

Dari hadis tersebut seperti sahabat nabi, Abdurrahman bin Auf melakukan bisnis dengan menganut aspek kejujuran dan ketelitian sehingga barang yang dijual ke konsumen dijaga kualitasnya dan bila terdapat kekurangan maka akan segera

dilakukan pembenahan dan evaluasi, karena dalam melakukan bisnis perlu berdasarkan atas kejujuran dan ketelitian serta perlu menanamkan sifat gigih dan bekerja keras. Dalam hubungan dengan melakukan prediksi omset bisnis restoran dapat diketahui pendapatan dari restoran sehingga langkah untuk tetap menjaga kualitas dari makanan yang dijual tetap bisa dipertahankan bahkan ditingkatkan.

Dari pandangan agama Islam tentang prediksi omset bisnis restoran sangat dianjurkan untuk dilakukan dengan konsep tetap berusaha dengan giat dan baik supaya hasil prediksi yang telah didapat dapat diubah menjadi lebih baik, serta perlu dalam menerapkan konsep kejujuran dan ketelitian supaya kualitas dari bisnis tetap baik dan bisa berkembang dengan baik.

BAB VII

KESIMPULAN

7.1. Kesimpulan

Dari penelitian yang telah dilakukan penulis menyimpulkan menjadi 2 poin penting, yaitu:

1. Prediksi omset bisnis restoran soto-kwali pak wasis dengan hasil terbaik adalah metode *Logistic Regression* (LR) yang penggunaannya memakai perbandingan dataset 70:30 dan modifikasi metode menggunakan parameter $max_iter = 100$. Hasil ini diperoleh dari fakta berikut :
 - a. Uji coba metode *Random Forest* (RF) diterapkan pada 4 skenario memakai modifikasi metode dengan parameter $n_estimators$ yang terdiri atas : 50,100,500 dan 1000. Hasil terbaik dari uji coba yang telah dilakukan pada metode *Random Forest* (RF) terletak pada skenario 2 memakai 80:20 dengan modifikasi metode memakai parameter $n_estimators = 100$ mendapatkan nilai *accuracy* 98,78% dengan *duration* 26s dan level error yang didapat paling rendah mendapat nilai *MAE* 0,23222, *MAPE* 22,1852.
 - b. Uji coba metode *Logistic Regression* (LR) diterapkan pada 4 skenario memakai modifikasi metode dengan parameter max_iter yang terdiri atas : 100,500 dan 1000. Hasil terbaik dari uji coba yang telah dilakukan pada metode *Logistic Regression* (LR) terletak pada skenario 3 memakai 70:30 dengan modifikasi metode memakai parameter $max_iter = 100$ mendapatkan nilai *accuracy* 97% dengan *duration* 10s dan level error yang didapat paling rendah mendapat nilai *MAE* 0,22273, *MAPE* 22,2424.
 - c. Implementasi metode *Logistic Regression* (LR) dengan menggunakan dataset yang berjumlah 335 mendapatkan akurasi dan durasi yang terbaik dengan selisih rata-rata antara prediksi dengan kondisi nyata yaitu 0,046, sedangkan nilai standart deviasinya mendapatkan nilai terbaik 0,0249.
2. Hasil Evaluasi akurasi dan durasi dari 2 metode yang paling cocok dalam melakukan prediksi model bisnis restoran soto-kwali pak wasis yaitu metode *Logistic Regression* (LR) yang mendapatkan nilai akurasi 97% dengan error

berkisar antara 0,2 dan waktu eksekusi 10s dibandingkan dengan metode *Random Forest* (RF) yang memang memiliki akurasi tinggi tetapi waktu eksekusi yang lebih lama, sehingga dalam bisnis kecepatan eksekusi menjadi pertimbangan paling utama. Hasil tersebut dari fakta yaitu :

- a. Perbandingan dari metode *Random Forest* (RF) dengan *Logistic Regression* (LR) dalam prediksi omset bisnis restoran soto kwali pak wasis bila ditinjau dari *accuracy* maka *Random Forest* (RF) mendapatkan *accuracy* lebih tinggi 1 point yaitu 98,78% bila dibandingkan dengan metode *Logistic Regression* (LR) dengan perolehan 97%.
- b. Hasil dari durasi kecepatan masing-masing metode dalam melakukan prediksi menunjukkan bahwa metode *Logistic Regression* (LR) lebih cepat dan lebih efektif dengan perolehan *duration* 10s dibandingkan dengan perolehan metode *Random Forest* (RF) dengan perolehan *duration* 26s.
- c. Perbandingan metode dengan mempertimbangkan dari hasil error yang didapat dari 2 metode tidak mampu dievaluasi karena selisih antara 2 metode tersebut hampir mirip yaitu metode *Random Forest* (RF) mendapatkan nilai error paling terendah dengan perolehan *MAE* 0,23222, dan *MAPE* 22,1852 dan metode *Logistic Regression* (LR) dengan perolehan error *MAE* 0,22273, dan *MAPE* 22,2424.

Dari evaluasi yang dipertimbangkan dari segi *accuracy* dan *duration*, metode *Logistic Regression* (LR) merupakan metode yang sangat cocok dibandingkan dengan metode *Random Forest* (RF) dalam melakukan prediksi omset bisnis restoran soto-kwali pak wasis. Karena kebutuhan bisnis memerlukan segi kecepatan metode *Logistic Regression* (LR) dengan tingkat akurasi yang baik dan error yang rendah. Hasil prediksi dengan akurasi yang baik dapat dijadikan tolak ukur dan evaluasi administrasi bisnis restoran soto-kwali pak wasis sehingga perkembangan bisnis dapat lebih efektif dan dapat membuka lapangan pekerjaan yang lebih banyak.

7.2. Saran

Penelitian yang telah dilakukan harapannya agar tetap dilakukan penelitian berikutnya supaya menjadi inspirasi bagi penulis lainnya, saran dari penulis penelitian ini sebagai berikut :

- a. Penelitian berikutnya diharapkan dapat memakai pada studi kasus dengan bisnis yang berskala lebih besar, sehingga variasi data lebih interaktif dan dataset yang telah didapatkan menjadi lebih banyak.
- b. Metode yang diterapkan diharapkan ada penggabungan metode atau model hybrid dari metode terbaik, sehingga hasil dari prediksi diharapkan menjadi lebih baik dan performance kecepatan dalam proses prediksi dapat menjadi lebih cepat.

DAFTAR PUSTAKA

- A. K. Umam, R. T. Ratnasari, and S. Herianingrum, "The Effect Of Macroeconomic Variables In Predicting Indonesian Sharia Stock Index," *JEBIS*, vol. 5, no. 2, p. 223, Dec. 2019, doi: 10.20473/jebis.v5i2.15031.
- Azimovna, M. S., Shokhrukhovich, U. F., & Sodirovich, U. B. (2022). Problems in the Implementation of Quality Management Systems in Small Business Enterprises. *Eurasian Research Bulletin*, 7, 54-57.
- B. Farnham, S. Tokyo, B. Boston, F. Sebastopol, and T. Beijing, "Hands-On Machine Learning With Scikit-Learn, Keras, And Tensorflow Concepts, Tools, And Techniques To Build Intelligent Systems Second Edition," Sebastopol, CA: O'Reilly Media, 2019, pp. 192–200.
- Bagui, S., Li, K. Resampling imbalanced data for network intrusion detection datasets. *J Big Data* 8, 6 (2021). <https://doi.org/10.1186/s40537-020-00390-x>.
- Chaibi, M., Tarik, L., Berrada, M., & El Hmaidi, A. (2022). Machine Learning Models Based on Random Forest Feature Selection and Bayesian Optimization for Predicting Daily Global Solar Radiation. *International Journal of Renewable Energy Development*, 11(1), 309.
- Emmanuel, T., Maupong, T., Mpoeleng, D. et al. A survey on missing data in machine learning. *J Big Data* 8, 140 (2021). <https://doi.org/10.1186/s40537-021-00516-9>.
- G. S. Rani, S. Jayan and B. Alatas, "Analysis of Chaotic Maps for Global Optimization and a Hybrid Chaotic Pattern Search Algorithm for Optimizing the Reliability of a Bank," in *IEEE Access*, vol. 11, pp. 24497-24510, 2023, doi: 10.1109/ACCESS.2023.3253512.
- Gajewar A, Bansal G. Revenue forecasting for enterprise products. arXiv preprint arXiv:1701.06624. 2016 Nov 21.
- Gai, R., Zhang, H. Prediction model of agricultural water quality based on optimized logistic regression algorithm. *EURASIP J. Adv. Signal Process.* 2023, 21 (2023). <https://doi.org/10.1186/s13634-023-00973-9>.
- Gasso G. Logistic regression. INSA Rouen-ASI Departement Laboratory: Saint-Etienne-du-Rouvray, France. 2019.pp.1-30.
- H. P. Singh and H. N. Alhulail, "Predicting Student-Teachers Dropout Risk and Early Identification: A Four-Step Logistic Regression Approach," in *IEEE Access*, vol. 10, pp. 6470-6482, 2022, doi: 10.1109/ACCESS.2022.3141992.

- Hongyang Xie, Jianjun Dong, Yong Deng, Yiwen Dai, "Prediction Model of the Slope Angle of Rocky Slope Stability Based on Random Forest Algorithm", *Mathematical Problems in Engineering*, vol. 2022, Article ID 9441411, 10 pages, 2022. <https://doi.org/10.1155/2022/9441411>.
- K. Saetia and J. Yokrattanasak, "Stock Movement Prediction Using Machine Learning Based on Technical Indicators and Google Trend Searches in Thailand," *IJFS*, vol. 11, no. 1, p. 5, Dec. 2022, doi: 10.3390/ijfs11010005.
- K. Sahinbas, "Price Prediction Model for Restaurants In Istanbul By Using Machine Learning Algorithms," *Ekonomi İşletme ve Maliye Araştırmaları Dergisi*, vol. 4, no. 2, pp. 159–171, Aug. 2022, doi: 10.38009/ekimad.1148216.
- K. Singh, P. M. Booma, and U. Eaganathan, "E-Commerce System for Sale Prediction Using Machine Learning Technique," *J. Phys.: Conf. Ser.*, vol. 1712, no. 1, p. 012042, Dec. 2020, doi: 10.1088/1742-6596/1712/1/012042.
- Kaur, R., Kumar, R. & Gupta, M. Predicting risk of obesity and meal planning to reduce the obese in adulthood using artificial intelligence. *Endocrine* 78, 458–469 (2022). <https://doi.org/10.1007/s12020-022-03215-4>.
- Kumar, Indrajeet, Jyoti Rawat, Noor Mohd, and Shahnawaz Husain. "Opportunities of Artificial Intelligence and Machine Learning in the Food Industry." Edited by Rijwan Khan. *Journal of Food Quality* 2021 (July 12, 2021): 1–10. <https://doi.org/10.1155/2021/4535567>.
- M. Irfan Farraz Haecal and Dadan Rusmana, "Historical Conception According To The Natural Malay Tafsir Of Turjuman Al-Mustafad: A Tafsir Study Of The Nusantara," *Al-Risalah*, pp. 78–92, May 2022, doi: 10.24252/al-risalah.vi.28360.
- M. Kureljusic and L. Reisch, "Revenue forecasting for European capital market-oriented firms: A comparative prediction study between financial analysts and machine learning models," *COC*, vol. 19, no. 2, pp. 159–178, 2022, doi: 10.22495/cocv19i2art13.
- Mr. T. Phase, "Predict the Level of Income using Random Forest Classifier," *IJRASET*, vol. 7, no. 12, pp. 558–561, Dec. 2019, doi: 10.22214/ijraset.2019.12090.
- P. Bajaj, R. Ray, S. Shedge, S. Vidhate, and D. Nikhilkumar, "Sales Prediction Using Machine Learning Algorithms," vol. 07, no. 06, 2020.
- Pratama, Noor Aditya, and Abdul Gafur. "Manajemen Strategi Usaha Rumah Makan Dalam Menghadapi Pandemi Covid-19 Berdasarkan Konsep Abdurrahman Bin Auf (Studi Kasus Rumah Makan Dandito Balikpapan)." *Jurnal Ekonomi Syariah Mulawarman (JESM)* 1, no. 4 (2022): 251-259.

- R. Ghorbani and R. Ghousi, "Comparing Different Resampling Methods in Predicting Students' Performance Using Machine Learning Techniques," in *IEEE Access*, vol. 8, pp. 67899-67911, 2020, doi: 10.1109/ACCESS.2020.2986809
- R. Gai and H. Zhang, "Prediction model of agricultural water quality based on optimized logistic regression algorithm," *EURASIP J. Adv. Signal Process.*, vol. 2023, no. 1, p. 21, Feb. 2023, doi: 10.1186/s13634-023-00973-9.
- R. Russell, *Machine Learning: Step-by-Step Guide To Implement Machine Learning Algorithms with Python*, First Edition. Independent Publishing Platform : United States of America, 2018. pp.19-20
- Raul, Nataasha, Yash Shah, and Mehul Devganiya. "Restaurant revenue prediction using machine learning." *Int. J. Eng. Sci* 6.4 (2016): 91-94.
- S. D. A. Bujang et al., "Multiclass Prediction Model for Student Grade Prediction Using Machine Learning," in *IEEE Access*, vol. 9, pp. 95608-95621, 2021, doi: 10.1109/ACCESS.2021.3093563.
- S. Kuhle et al., "Comparison of logistic regression with machine learning methods for the prediction of fetal growth abnormalities: a retrospective cohort study," *BMC Pregnancy Childbirth*, vol. 18, no. 1, p. 333, Dec. 2018, doi: 10.1186/s12884-018-1971-2.
- S. Siddamsetty, R. R. Vangala, L. Reddy, and P. R. Vattipally, "Restaurant Revenue Prediction using Machine Learning," vol. 08, no. 12, 2021.
- Schober P, Vetter TR. Logistic Regression in Medical Research. *Anesth Analg*. 2021 Feb 1;132(2):365-366. doi: 10.1213/ANE.0000000000005247
- T. Alam, C. F. Ahmed, S. A. Zahin, M. A. H. Khan and M. T. Islam, "An Effective Recursive Technique for Multi-Class Classification and Regression for Imbalanced Data," in *IEEE Access*, vol. 7, pp. 127615-127630, 2019, doi: 10.1109/ACCESS.2019.2939755.
- U. M. Sirisha, M. C. Belavagi and G. Attigeri, "Profit Prediction Using ARIMA, SARIMA and LSTM Models in Time Series Forecasting: A Comparison," in *IEEE Access*, vol. 10, pp. 124715-124727, 2022, doi: 10.1109/ACCESS.2022.3224938.
- Xiao Li, Yu Wang, Sumanta Basu, Karl Kumbier, and Bin Yu. 2019. A debiased MDI feature importance measure for random forests. *Proceedings of the 33rd International Conference on Neural Information Processing Systems*. Curran Associates Inc., Red Hook, NY, USA, Article 723, 8049–8059.

- Y. Lin et al., "Revenue prediction for integrated renewable energy and energy storage system using machine learning techniques," *Journal of Energy Storage*, vol. 50, p. 104123, Jun. 2022, doi: 10.1016/j.est.2022.104123.
- Y. -R. Chen, J. -S. Leu, S. -A. Huang, J. -T. Wang and J. -I. Takada, "Predicting Default Risk on Peer-to-Peer Lending Imbalanced Datasets," in *IEEE Access*, vol. 9, pp. 73103-73109, 2021, doi: 10.1109/ACCESS.2021.3079701.
- V. K. Gupta, A. Gupta, D. Kumar and A. Sardana, "Prediction of COVID-19 confirmed, death, and cured cases in India using random forest model," in *Big Data Mining and Analytics*, vol. 4, no. 2, pp. 116-123, June 2021, doi: 10.26599/BDMA.2020.9020016.
- Zahoor, Kanwal, Narmeen Zakaria Bawany, and Soomaiya Hamid. "Sentiment Analysis and Classification of Restaurant Reviews Using Machine Learning." In *2020 21st International Arab Conference on Information Technology (ACIT)*, 1–6. Giza, Egypt: IEEE, 2020. <https://doi.org/10.1109/ACIT50332.2020.9300098>.

Lampiran

Tahapan Metode Random Forest

Dalam tahap perhitungan metode Random Forest dibantu dengan menggunakan excel serta dataset yang digunakan untuk menguji perhitungan manual hanya memuat 10 dataset sederhana. Berikut tahapannya :

1. Dataset

Level	Soto Ayam	Air Mineral	Kerupuk
2	39	129	494
1	14	39	157
2	15	77	427
2	15	78	444
1	19	92	297
1	26	91	345
2	23	123	544
2	37	131	635
1	15	33	188
2	17	84	252

2. Menghitung nilai gini

Gini Var Kerupuk

	level=2	level=1	Total
>345	5	1	6
<=345	1	3	4
	grand total		10

gini (5,1) 0,277777778
 gini (1,3) 0,375
 gini index 0,316666667

Gini Var Air mineral

	level =2	level =1	total
>= 91	3	2	5
<91	3	2	5
	grand total		10

gini (3,2) 0,48
 gini (3,2) 0,48
 gini index 0,48

Gini Var Soto Ayam

	level =2	level =1	total
>15	4	2	6
<=15	2	2	4
	grand total		10

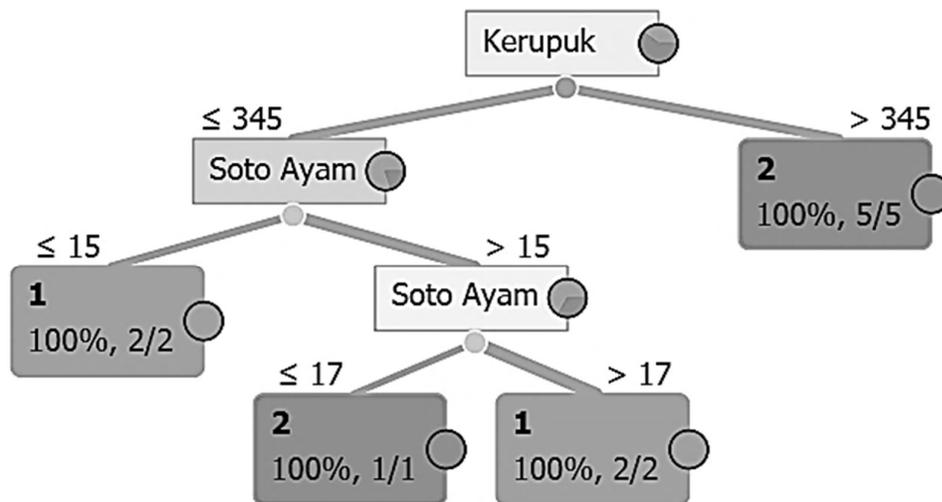
gini (4,2) 0,444444444
 gini (2,2) 0,5
 gini index 0,466666667

Gini Var Soto Ayam

	level =2	level =1	total
>17	3	1	4
<=17	3	3	6
	grand total		10

gini (3,1) 0,375
 gini (3,3) 0,5
 gini index 0,45

3. Membuat pohon keputusan



4. Menentukan prediksi

Level	Soto Ayam	Air Mineral	Kerupuk	Prediction
2	39	129	494	2,00775
1	14	39	157	1,00769
2	15	77	427	2,01298
2	15	78	444	1,46565
1	19	92	297	1,00326
1	26	91	345	1,00329
2	23	123	544	2,00813
2	37	131	635	2,00763
1	15	33	188	1,74004
2	17	84	252	2,00357

5. Membuat confusion matrix dan menghitung nilai akurasi

		prediction	
		1	2
actual	1	3	1
	2	1	5

4

6

10

accuracy 0,8 80%

Tahapan Metode Logistic Regression

Dalam tahap perhitungan metode Random Forest dibantu dengan menggunakan excel serta dataset yang digunakan untuk menguji perhitungan manual hanya memuat 10 dataset sederhana. Berikut tahapannya :

1. Dataset

Level	Soto Ayam	Air Mineral	Kerupuk
2	39	129	494
1	14	39	157
2	15	77	427
2	15	78	444
1	19	92	297
1	26	91	345
2	23	123	544
2	37	131	635
1	15	33	188
2	17	84	252

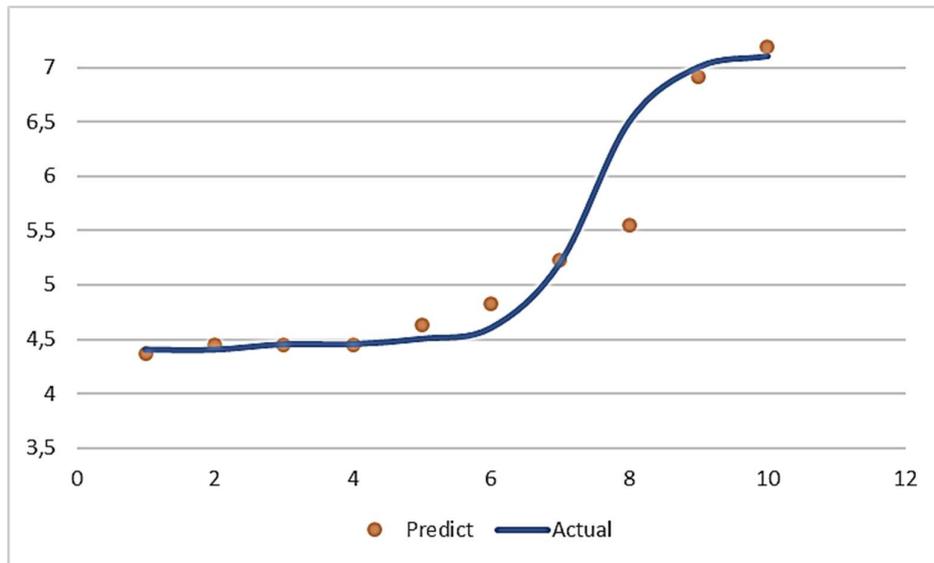
2. Menghitung nilai regresi dengan threshold 0,5

Initial Solver Decision Variables	b0 =	0,50
	b1 =	0,02
	b2 =	0,00
	b3 =	-2,38014644018999E-08

Level	Soto Ayam	Air Mineral	Kerupuk	Linier	exp^Linier
2	39	129	494	1,279986	3,596591
1	14	39	157	0,779996	2,181463
2	15	77	427	0,799988	2,225515
2	15	78	444	0,799988	2,225514
1	19	92	297	0,879991	2,410879
1	26	91	345	1,01999	2,773168
2	23	123	544	0,959985	2,611657
2	37	131	635	1,239983	3,455554
1	15	33	188	0,799995	2,22553
2	17	84	252	0,839993	2,31635

Nilai uji probability = 5

3. Menentukan nilai probabilitas



4. Melakukan prediksi

Level	Soto Ayam	Air Mineral	Kerupuk	Prediction
2	39	129	494	2
1	14	39	157	1
2	15	77	427	1
2	15	78	444	1
1	19	92	297	1
1	26	91	345	2
2	23	123	544	2
2	37	131	635	2
1	15	33	188	1
2	17	84	252	2

5. Membuat confusion matrix dan menghitung nilai akurasi

		prediction		
		1	2	
actual	1	3	1	4
	2	2	4	6
accuracy	0,7	70%	Total data	10