

**PENERAPAN DATA MINING UNTUK MEMPREDIKSI
JUMLAH PRODUKSI KELAPA SAWIT MENGGUNAKAN
METODE REGRESI LINEAR BERGANDA**

(Studi Kasus : PT. SGSR Mujur Sinar Gunung Sawit Raya)

SKRIPSI

OLEH

REDI SEBASTIAN SITUMORANG

188160030



**FAKULTAS TEKNIK
PROGRAM STUDI TEKNIK INFORMATIKA
UNIVERSITAS MEDAN AREA
MEDAN
2023**

UNIVERSITAS MEDAN AREA

© Hak Cipta Di Lindungi Undang-Undang

1. Dilarang Mengutip sebagian atau seluruh dokumen ini tanpa mencantumkan sumber
2. Pengutipan hanya untuk keperluan pendidikan, penelitian dan penulisan karya ilmiah
3. Dilarang memperbanyak sebagian atau seluruh karya ini dalam bentuk apapun tanpa izin Universitas Medan Area

Document Accepted 13/12/23

Access From (repository.uma.ac.id)13/12/23

**PENERAPAN DATA MINING UNTUK MEMPREDIKSI
JUMLAH PRODUKSI KELAPA SAWIT MENGGUNAKAN
METODE REGRESI LINEAR BERGANDA**

(Studi Kasus : PT. SGSR Mujur Sinar Gunung Sawit Raya)

SKRIPSI

Diajukan Untuk Memenuhi Salah Satu Syarat
Guna Memperoleh gelar Sarja Komputer (S.Kom)
Program Studi Teknik Informatika

OLEH :

REDI SEBASTIAN SITUMORANG

188160030

FAKULTAS TEKNIK

PROGRAM STUDI TEKNIK INFORMATIKA

UNIVERSITAS MEDAN AREA

MEDAN

2023

UNIVERSITAS MEDAN AREA

© Hak Cipta Di Lindungi Undang-Undang

1. Dilarang Mengutip sebagian atau seluruh dokumen ini tanpa mencantumkan sumber
2. Pengutipan hanya untuk keperluan pendidikan, penelitian dan penulisan karya ilmiah
3. Dilarang memperbanyak sebagian atau seluruh karya ini dalam bentuk apapun tanpa izin Universitas Medan Area

Document Accepted 13/12/23

Access From (repository.uma.ac.id)13/12/23

PALAMAN PENGESAHAN

Judul Skripsi : Penerapan Data Mining Untuk Memprediksi Jumlah Produksi Kelapa Sawit Menggunakan Metode Regresi Linear Berganda
Nama : Redi Sebastian Situmorang
NPM : 188160030
Fakultas : Teknik

Disetujui Oleh
Komisi Pembimbing

Dr. Rahmat Syah, S.Kom. M.Kom
Pembimbing I

Susilawati, S.Kom. M.Kom
Pembimbing II

Dr. Rahmat Syah, S.Kom. M.Kom
Dekan

Rizki Nur Hafidha, S.Kom. M.Kom
Prodi

Tanggal Lulus : 29 Agustus 2023

LEMBAR PERNYATAAN ORISINILITAS

Saya yang bertanda tangan di bawah ini :

Nama : REDI SEBASTIAN SITUMORANG
Tempat dan tanggal Lahir : Tumba Jae, 24 November 2000
Program Studi : Teknik Informatika
NIM : 188160030

Menyatakan dengan sesungguhnya bahwa :

1. Seluruh data, informasi integritas serta dalam pembahasan dan kesimpulan yang disajikan dalam skripsi ini, kecuali yang disebutkan sumbernya ditulis dalam daftar pustaka adalah merupakan hasil pengamatan, penelitian, pengolahan, serta sepemikiran saya dengan pengarahan dari para pembimbing yang ditetapkan.
2. Skripsi yang saya tulis ini asli, bukan jiplakan dan belum pernah diajukan untuk mendapatkan gelar akademik, baik di Universitas Medan Area maupun di perguruan tinggi lainnya.
3. Apabila dikemudian hari ditemukan adanya bukti ketidak benaran dalam pernyataan tersebut diatas, maka saya bersedia menerima sanksi akademis sesuai ketentuan di Universitas Medan Area.

Demikian pernyataan ini dibuat dengan penuh kesadaran dan dapat dipertanggung jawabkan.

Medan, 29 Agustus 2023

Yang membuat pernyataan,



Redi Sebastian Situmorang

NIM. 188160030

HALAMAN PERNYARAAN PERSETUJUAN PUBLIKASI TUGAS AKHIR/SKRIPSI/TESIS UNTUK KEPENTINGAN AKADEMI

Sebagai sivitas akademik Universitas Medan Area, saya yang bertanda tangan dibawah ini :

Nama : Redi Sebastian Situmorang
NPM : 188160030
Program Studi : Teknik Informatika
Fakultas : Teknik
Jenis Karya : Skripsi

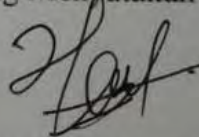
Demi pengembangan ilmu pengetahuan, menyetujui untuk memberikan kepada Universitas Medan Area Hak Bebas Royalti Noneksklusif (*Non-exclusive Royalty-Free Right*) atas karya ilmiah saya yang berjudul : Penerapan Data Mining Untuk Memprediksi Jumlah Produksi Kelapa Sawit Menggunakan Metode Regresi Linear Berganda (Studi Kasus : PT. SGSR Mujur Sinar Gunung Sawit Raya). Dengan Hak Bebas Royalti Nonesklusif ini Universitas Medan area berhak menyimpan, mengalih media/format-kan, mengelola dalam bentuk pangkalan data (*database*), dan mempublikasikan skripsi saya selama mencantumkan nama saya sebagai penulis/pencipta dan sebagai pemilik Hak Cipta.

Demikian pernyataan ini saya buat dengan sebenarnya.

Dibuat di : Medan

Pada tanggal : 29 Agustus 2023

Yang Menyatakan



(Redi Sebastian Situmorang)

ABSTRAK

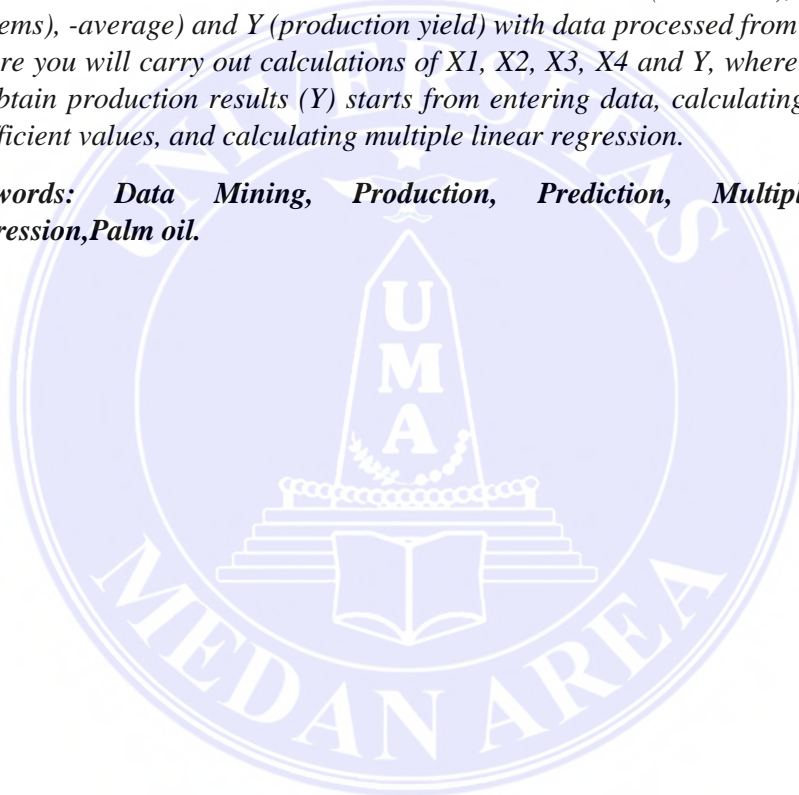
PT. SGSR Mujur Sinar Gunung Sawit Raya merupakan salah satu perusahaan kelapa sawit yang besar di Kabupaten Tapanuli Tengah Provinsi Sumatera Utara. Jumlah hasil produksi kelapa sawit setiap bulannya akan mengalami kenaikan ataupun penurunan yang tidak stabil sehingga sulit untuk memprediksi hasil produksi kelapa sawit di PT SGSR. Hal ini dilakukan untuk peningkatan akurasi dan menganalisis perkembangan hasil produksi kelapa sawit dilapangan sebagai bahan kebijakan bagi pemimpin dalam usaha mengatasi penurunan jumlah hasil produksi kelapa sawit. Metode regresi linear berganda merupakan metode yang digunakan dalam mengatasi permasalahan yang ada. Analisis regresi dapat digunakan dalam hal memprediksi dan juga memiliki kelebihan pada penguraian data yang terstruktur dan mudah untuk dipahami. Variabel yang akan digunakan yaitu X_1 (luas lahan), X_2 (banyak batang), X_3 (tandan/pokok), X_4 (berat janjang rata-rata) dan Y (hasil produksi) dengan data yang diproses dari tahun 2019-2022. Dimana akan melakukan perhitungan dari X_1 , X_2 , X_3 , X_4 dan Y , yang mana proses untuk mendapatkan hasil produksi (Y) dimulai dari memasukkan data, menghitung nilai koefisien regresi, dan perhitungan regresi linear berganda.

Kata Kunci : Data Mining, Produksi, Prediksi, Regresi Linear Berganda, Kelapa Sawit.

ABSTRACT

PT. SGSR Mujur Sinar Gunung Sawit Raya is one of the largest palm oil companies in Central Tapanuli Regency, North Sumatra Province. The amount of palm oil production every month will experience an unstable increase or decrease, making it difficult to predict palm oil production results at PT SGSR. This is done to increase accuracy and analyze developments in palm oil production results in the field as policy material for leaders in efforts to overcome the decline in the amount of palm oil production. The multiple linear regression method is a method used to overcome existing problems. Regression analysis can be used in terms of prediction and also has the advantage of decomposing data that is structured and easy to understand. The variables that will be used are X1 (land area), X2 (number of stems), -average) and Y (production yield) with data processed from 2019-2022. Where you will carry out calculations of X1, X2, X3, X4 and Y, where the process to obtain production results (Y) starts from entering data, calculating regression coefficient values, and calculating multiple linear regression.

Keywords: *Data Mining, Production, Prediction, Multiple Linear Regression, Palm oil.*



RIWAYAT HIDUP

Penulis dilahirkan di Provinsi Sumatera Utara, Kabupaten Tapanuli Tengah yaitu di Desa Tumbajae pada tanggal 24 November 2000 merupakan anak ke 3 dari 4 bersaudara dari ayah Alm. Robinson Situmorang dan ibu Florentina Manurung. Tahun 2018 penulis lulus dari SMA Sw. Budi Mulia Tumbajae dan pada tahun 2018 terdaftar sebagai mahasiswa Fakultas Teknik Universitas Medan Area. Penulis melaksanakan Kerja Praktek (KP) di UPT. Samsat Medan Utara.



KATA PENGANTAR

Ucapan syukur ke hadirat Tuhan Yang Maha Esa karena atas berkat dan rahmatnya, maka penulis dapat menyelesaikan tugas akhir ini tepat pada waktunya. Adapun tujuan penulis dalam menyusun tugas akhir ini adalah untuk memenuhi salah satu syarat mendapatkan gelar sarjana Program Studi Teknik Informatika Universitas Medan Area. Pada kesempatan ini penulis juga menyampaikan ucapan terima kasih kepada Ibu saya yang telah memberi semangat, doa dan bantuan baik secara moril maupun materil. Selama penyusunan skripsi ini, penulis banyak menerima bimbingan, nasihat dan bantuan dari berbagai pihak maka pada kesempatan ini penulis ingin menyampaikan terimakasih kepada :

1. Bapak Prof. Dr. Dadan Ramdan, M.Eng., Msc, selaku Rektor Universitas Medan Area
2. Bapak Dr. Rahmad Syah, S.Kom, M.Kom., selaku Dosen Pembimbing I dan Dekan Fakultas Teknik Universitas Medan Area.
3. Bapak Rizki Muliono, S.Kom, M.Kom., selaku Ketua Prodi Teknik Informatika Universitas Medan Area
4. Ibu Susilawati, S.Kom, M.Kom., selaku Dosen Pembimbing II yang sudah memberikan masukan dan saran dalam tiap proses pengerjaan Tugas Akhir ini hingga selesai.
5. Seluruh Staff Fakultas Teknik Universitas Medan Area, yang telah banyak memberikan bantuan dan informasi kepada penulis.
6. Kepada Pak Ruben Sitingjak yang telah menizinkan penulis untuk melakukan riset atau penelitian pada PT. SGSR Mujur

7. Kepada Pak Yans Hutagalung yang telah membantu dan juga menemani penulis selama melakukan riset atau penelitian di Afdeling 1.
8. Kepada Orang Tua saya yang selalu memberi dukungan dan semangat dalam segala hal, memberi motivasi dan dorongan semangat serta juga perhatian dalam segala kebutuhan yang diperlukan.
9. Kepada Kakak, Abang, dan Adik yang telah memberikan semangat, dukungan dan bantuan yang sangat luar biasa sehingga penulis dapat menyelesaikan tugas akhir ini.
10. Kepada Junaidi, Carmenita, Fordinan, Philipus, Jhon, Febriady, Ronal dan juga teman-teman Teknik Informatika 2018 yang juga memberikan semangat, dukungan dan bantuan dalam segala bidang.

Dalam tugas akhir ini, penulis menyadari bahwa masih terdapat banyak kekurangan karena keterbatasan pengetahuan dan pengalaman yang dimiliki. Oleh karena itu, penulis mengharapkan saran dan kritik yang bersifat membangun dari para pembaca. Serta beberapa kelebihan yang sepantasnya untuk dipertahankan bahkan dikembangkan kedepannya.

Demikianlah kata pengantar dari penulis, dan penulis berharap semoga penulisan tugas akhir ini dapat memberikan manfaat bagi semua orang yang memerlukannya.

Medan, 29 Agustus 2023

Penulis,



Redi Sebastian Situmorang

DAFTAR ISI

ABSTRAK.....	vi
ABSTRACT.....	vii
RIWAYAT HIDUP	viii
KATA PENGANTAR.....	ix
DAFTAR ISI.....	xi
DAFTAR TABEL	xiii
DAFTAR GAMBAR	xiv
BAB I PENDAHULUAN	2
1.1. Latar Belakang	2
1.2. Rumusan Masalah.....	6
1.3. Batasan Masalah	6
1.4. Tujuan Penelitian	6
1.5. Manfaat Penelitian.....	7
1.6. Statistika Penulisan.....	7
BAB II LANDASAN TEORI	9
2.1. Data Mining	9
2.2. Proses Data Mining.....	11
2.3. Prediksi.....	13
2.4. Jenis-Jenis Prediksi	14
2.5. Produksi	14
2.6. Kelapa Sawit	15
2.6.1 Luas Lahan	16
2.6.2 Jumlah Batang	16
2.6.3 Tandan Perpokok.....	16
2.6.4 Berat Janjang rata-rata.....	16
2.6.5 Hasil Produksi	16
2.7. Penelitian Terkait.....	17
2.8. <i>Python</i>	19
2.9. <i>Mean Absolute Error (MAE)</i>	20
2.10. Regresi.....	21

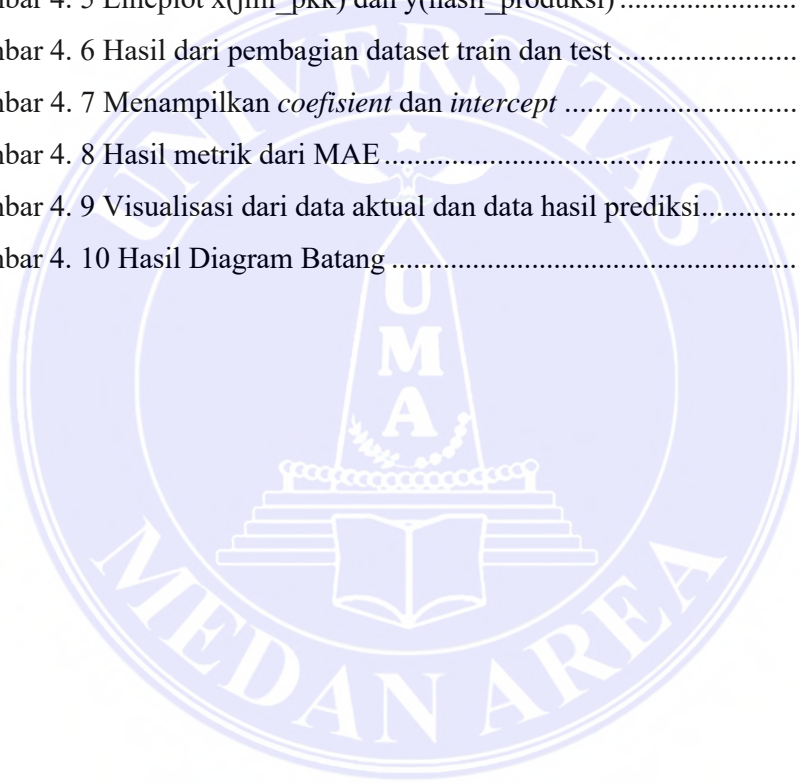
2.11. Regresi Linear Berganda	21
BAB III ANALISIS DAN PERANCANGAN	26
3.1. Analisis	26
3.1.1. Kerangka Kerja Penelitian	26
3.1.2. Analisis	27
3.1.3. Bahan Penelitian.....	29
3.1.4. Pengumpulan Data	29
3.1.5. Langkah-Langkah Perhitungan Metode Regresi Linear Berganda..	30
3.1. Perancangan.....	36
BAB IV HASIL DAN PEMBAHASAN	38
4.1. Hasil	38
4.1. Pembahasan	43
BAB V KESIMPULAN DAN SARAN.....	50
5.1. Kesimpulan.....	50
5.2. Saran	50
DAFTAR PUSTAKA.....	51
LAMPIRAN.....	54

DAFTAR TABEL

Tabel 3. 1 Data Hasil Produksi.....	30
Tabel 3. 2 Koefisien Regresi	31
Tabel 3. 3 Jumlah Data Prediksi.....	32
Tabel 3. 4 Hasil Determinan Matriks	33
Tabel 3. 5 Hasil Perhitungan Regresi.....	35
Tabel 4. 1 Hasil Pengujian Tahun 2019	38
Tabel 4. 2 Hasil Pengujian Tahun 2020	39
Tabel 4. 3 Hasil Pengujian Tahun 2021	40
Tabel 4. 4 Hasil Pengujian Tahun 2022	41
Tabel 4. 5 Prosedur yang dilakukan	42
Tabel 4. 6 Data Keseluruhan.	43
Tabel 4. 7 Cek data yang kosong	43
Tabel 4. 8 Cek data dari masing masing variabel	44
Tabel 4. 9 Cek statistik	44
Tabel 4. 10 Data 2022	45
Tabel 4. 11 Data Aktual dan Data Prediksi.....	48

DAFTAR GAMBAR

Gambar 2. 1 Proses Data mining.....	12
Gambar 2. 2 Kelapa Sawit	17
Gambar 3. 1 Alur Kerja Penelitian.....	26
Gambar 4. 1 Kolerasi Matriks	42
Gambar 4. 2 Visualisasi <i>input</i> (luas_lahan) dan <i>output</i> (hasil_produksi).....	45
Gambar 4. 3 Visualisasi <i>input</i> (jml_pkk) dan <i>output</i> (hasil_produksi)	46
Gambar 4. 4 Lineplot x(luas_ha) dan y(hasil_produksi).....	46
Gambar 4. 5 Lineplot x(jml_pkk) dan y(hasil_produksi)	47
Gambar 4. 6 Hasil dari pembagian dataset train dan test	47
Gambar 4. 7 Menampilkan <i>coefisient</i> dan <i>intercept</i>	47
Gambar 4. 8 Hasil metrik dari MAE.....	47
Gambar 4. 9 Visualisasi dari data aktual dan data hasil prediksi.....	48
Gambar 4. 10 Hasil Diagram Batang	49



BAB I

PENDAHULUAN

1.1. Latar Belakang

Perkebunan kelapa sawit mencakup jutaan hektar di seluruh dunia, yang mencakup sebagian besar perdagangan global. Pohon kelapa sawit adalah genus tanaman monokotil seperti pohon tanpa batang yang tumbuh subur di daerah tropis dan sangat berharga bagi manusia dan ekosistem. Kelapa sawit Afrika atau *Elaeis guineensis* adalah spesies palem asli Afrika Barat yang paling menonjol, yang mana buahnya sangat kaya minyak dan dibudidayakan sebagai sumber makanan selama lebih dari 7000 tahun. Pohon itu menghasilkan banyak tandan buah setiap tahun dengan masing-masing berisi antara 1000 dan 3000 buah. Olahan buah kelapa sawit merupakan sumber minyak yang signifikan bagi masyarakat dan merupakan turunan industri yang tidak terpisahkan, yaitu sabun, deterjen, dan kosmetik. Oleh karena itu, industri ini secara signifikan berdampak pada penduduk lokal dan keanekaragaman hayati yang lebih luas di wilayah asal mereka. (Nadhirah dkk., 2022)

Sektor pertanian di Indonesia dibedakan menjadi tiga jenis yaitu perkebunan, sawah dan ladang. Dari ketiga jenis sektor pertanian, sektor perkebunan yang lebih banyak diminati dikarenakan pertanian jenis perkebunan cenderung memiliki nilai jual yang tinggi, pembudidayaan dalam skala besar, serta daya tariknya yang terus meningkat. Sektor tanaman perkebunan di Indonesia banyak didominasi oleh tanaman kelapa sawit, kakao, karet, tebu dan kopi. Dari kelima tanaman ini kelapa sawit yang paling menguntungkan. Kelapa sawit lebih menguntungkan dibandingkan tebu dan karet dalam hal biaya produksi. Untuk perkebunan kelapa sawit dalam satu hektar biaya produksi yang dibutuhkan sebesar

Rp.9,7 juta/hektar/tahun dengan nilai produksi mencapai Rp.17 juta/hektar/tahun. Sedangkan untuk satu hektar karet per tahun membutuhkan biaya produksi sebesar Rp.9,2 juta dengan hasil produksi mencapai Rp.12,97 juta/hektar. Untuk tanaman tebu, dari awal proses penanaman hingga panen membutuhkan biaya produksi Rp.24,2 juta dengan nilai produksi Rp 31 juta. Pada kategori biaya upah pekerja, tanaman karet yang paling tinggi biayanya dengan rincian 31% untuk sawit, 57,09% untuk karet dan 26,21% untuk tebu. (Agasta dkk., 2018)

Pada revolusi industri 4.0 saat ini, teknologi merupakan hal yang sangat penting. Hal ini dikarenakan teknologi menjadi salah satu faktor pendukung dalam meningkatkan dan mengembangkan aktivitas yang dilakukan di berbagai bidang kehidupan. Perkembangan teknologi tersebut banyak juga dimanfaatkan sebagai teknologi informasi dibidang perkebunan, khususnya perkebunan Kelapa Sawit (No, 2020). Data mining penting pada perkembangan teknologi dan informasi saat ini karena ada sejumlah besar data yang dapat digunakan untuk menghasilkan informasi dan pengetahuan yang berguna. Di sektor perkebunan dan pabrik kelapa sawit (PKS) yang berciri khas padat karya, selama ini perkebunan dan PKS di Indonesia dikelola secara konvensional, dan sudah saatnya beradaptasi dan bertransformasi sejalan dengan modernisasi dan memanfaatkan kemajuan teknologi digital. Informasi dan pengetahuan yang diperoleh dapat digunakan di banyak bidang, seperti manajemen bisnis, manajemen produksi, dan kesehatan (Syairozi, 2021).

Prediksi hasil panen merupakan isu kritis namun menarik karena kebutuhannya untuk jangka panjang dan pemanfaatan sumber daya alam secara optimal. Banyak pemangku kepentingan dalam rantai pangan pertanian, termasuk

ahli agronomi, petani, eksportir produk, dan pembuat kebijakan, mendapat manfaat dari prakiraan hasil panen. Berbagai karakteristik khusus tanaman, kondisi lingkungan, dan praktik pengelolaan yang memengaruhi produksi tanaman adalah beberapa faktor pembaur untuk mengembangkan model prediksi. (Kamaruddin dkk., 2022)

Pada data mining untuk melakukan prediksi tentunya menggunakan sebuah metode, salah satu metode untuk melakukan prediksi ini menggunakan metode regresi linear. Regresi linear digunakan sebagai metode prediksi dengan kuantiti atau jumlah properti yang terjual sebagai variabel akibat dan periode penjualan properti sebagai variabel penyebab (Ayuni & Fitriyah, 2019). Metode regresi linear berganda merupakan salah satu teknik analisis data yang sering digunakan untuk mengkaji hubungan antar beberapa variabel dan prediksi satu variabel. Regresi linear berganda ini biasanya digunakan untuk mengetahui arah hubungan antara variabel independen dengan variabel dependen apakah masing-masing variabel independen berhubungan positif atau negatif. (Basri, 2019) Penelitian yang akan dilakukan ini menggunakan metode regresi linear berganda karena metode ini sesuai dengan studi kasus yang sedang diteliti.

PT. SGSR Mujur Sinar Gunung Sawit Raya merupakan sebuah perusahaan terbesar ke 3 di Tapanuli Tengah yang bergerak di bidang perkebunan kelapa sawit. Dikarenakan PT ini terlalu luas, maka dibagi menjadi 11 Afdeling. Afdeling adalah nama tempat yang dibuat dalam pembagian tempat di PT tersebut. Ada 11 Afdeling dibagi menjadi dua yang dipimpin oleh 2 orang manager rayon. Sedangkan di bawah Manager ada asisten manager yang memimpin di Afdeling-Afdeling tersebut. Dalam proses produksi kelapa sawit di PT. SGSR ataupun di afdeling-

afdeling pada PT tersebut belum dilengkapi dengan adanya suatu prediksi produksi. Dengan tidak adanya Prediksi produksi kelapa sawit untuk masa mendatang mengakibatkan sulitnya menentukan minimal jumlah target produksi. Jumlah target produksi itu biasanya ditentukan untuk mengetahui proses apa yang harus dilakukan dalam meningkatkan produksi kelapa sawit tersebut.

Prediksi yang akan dilakukan pada umumnya akan berdasarkan pada data masa lampau yang dianalisis dengan menggunakan cara-cara tertentu. Data masa lampau dikumpulkan, dipelajari, dan dianalisis lalu dihubungkan seiring dengan perjalanan waktu. Jelas dalam hal yang berhadapan dengan ketidakpastian sehingga ada faktor akurasi atau keseksamaan yang harus diperhitungkan. Akurasi suatu prediksi berbeda untuk tiap persoalan dan tergantung pada berbagai faktor, yang jelas tidak akan selalu didapatkan hasil Prediksi dengan ketepatan seratus persen. Tetapi bukan berarti prediksi ini menjadi percuma. Tetapi sebaliknya terbukti, bahwa prediksi banyak digunakan dan membantu dengan baik dalam berbagai manajemen sebagai dasar-dasar perencanaan, pengawasan, dan pengambilan keputusan. Salah satunya adalah prediksi produksi. (Santosa dkk., 2018). Salah satu cara yang bisa digunakan dalam memprediksikan jumlah produksi minyak kelapa sawit adalah dengan menggunakan metode regresi linear berganda, Karena terdapat beberapa data yang bisa digunakan melakukan perhitungan untuk mendapatkan ramalah jumlah produksi kelapa sawit. Dalam kasus ini, masalah yang timbul adalah bagaimana cara menerapkan metode regresi linear berganda untuk meramalkan produksi kelapa sawit berdasarkan jumlah permintaan. Pada PT. SGSR Mujur Sinar Gunung Sawit Raya ini berada di Kabupaten Tapanuli Tengah, Kecamatan Sirandorung. Peneliti menggunakan data produksi periode tahun 2019

– 2022 dengan parameter luas lahan, jumlah pokok, tandan/pokok, berat janjang rata-rata dan hasil produksi. Berdasarkan uraian latar belakang tersebut maka penulis bermaksud melakukan sebuah penelitian yang berjudul **“Penerapan Data Mining Untuk Memprediksi Jumlah Produksi Kelapa Sawit Menggunakan Metode Regresi Linear Berganda”**

1.1. Rumusan Masalah

Adapun yang menjadi rumusan masalah dalam penelitian ini yaitu bagaimana menerapkan prediksi jumlah produksi kelapa sawit dengan menggunakan metode regresi linear berganda di PT. SGSR Mujur Sinar Gunung Sawit Raya.

1.3. Batasan Masalah

Adapun batasan masalah dalam penelitian ini, yaitu:

1. Data yang digunakan adalah data produksi kelapa sawit PT. SGSR Mujur Sinar Gunung Sawit Raya pada periode 2019 – 2022.
2. Menganalisis data hasil produksi kelapa sawit pada PT. SGSR Mujur Sinar Gunung Sawit Raya.
3. Menerapkan prediksi produksi kelapa sawit ini menggunakan metode regresi linear berganda.
4. Atribut yang digunakan untuk memprediksi adalah luas lahan, jumlah batang, tandan/pokok, berat janjang rata-rata, dan hasil produksi.

1.1. Tujuan Penelitian

Adapun tujuan dari penelitian itu adalah:

1. Menerapkan prediksi jumlah produksi kelapa sawit dengan menggunakan metode regresi linear berganda

2. Mengetahui prediksi jumlah produksi yang meningkat atau menurun pada hasil panen.

1.5. Manfaat Penelitian

Adapun manfaat dari penelitian ini adalah :

Bagi Penulis :

1. Sebagai bentuk pemahaman penulis dalam memecahkan permasalahan dalam pekerjaan ataupun prakteknya.
2. Memberikan gambaran nyata kepada penulis terkait teori dan prakteknya.
3. Menambah pengetahuan dibidang tanaman perkebunan kelapa sawit.

Bagi Jurusan Teknik Informatika :

1. Menambah referensi mengenai penelitian menggunakan metode pembelajaran regresi linear berganda.
2. Menambah mitra Universitas Medan Area dan implementasi kerja sama prodi teknik informatika dengan PT. SGSR Mujur Sinar Gunung Sawit Raya.

Bagi PT. SGSR Mujur Sinar Gunung Sawit Raya

1. Dapat dijadikan acuan dalam menentukan kebijakan baru oleh perusahaan.
2. Perusahaan dapat memperkirakan biaya yang harus dikeluarkan pada satu periode.

1.5. Statistika Penulisan

Penelitian dilakukan dengan tahapan metode sebagai berikut :

BAB I : PENDAHULUAN

Pada bab ini berisi tentang gambaran umum dalam penulisan penelitian yang meliputi latar belakang, rumusan masalah, batasan masalah, tujuan penelitian, manfaat penelitian, dan sistematika penulisan.

BAB II : LANDASAN TEORI

Pada bab ini akan membahas seluruh dasar-dasar teori yang digunakan dalam penelitian, mulai dari definisi buku, sistem rekomendasi, regresi linear berganda, dan semua tahapan analisis serta perancangan.

BAB III : METODOLOGI PENELITIAN

Pada bab ini akan membahas mengenai tahap-tahap penyelesaian masalah dengan metode regresi linear berganda yang akan diterapkan dalam penelitian ini.

BAB IV : HASIL DAN PEMBAHASAN

Pada bab ini membahas tentang langkah-langkah hasil penelitian dan pembahasan yang memperoleh output penelitian

BAB V : KESIMPULAN DAN SARAN

Membahas tentang kesimpulan yang diperoleh dari hasil penelitian yang dilakukan berdasarkan tujuan penelitian dengan menetapkan metode penelitian. Kemudian saran yang perlu diperbaiki untuk pengembangan penelitian ini.

BAB II

LANDASAN TEORI

2.1. Data Mining

Data Mining (DM) adalah serangkaian proses untuk menggali nilai tambah dari suatu kumpulan data berupa pengetahuan yang selama ini tidak diketahui secara manual. Patut diingat bahwa kata mining sendiri berarti usaha untuk mendapatkan sedikit barang berharga dari sejumlah besar material dasar. Karena itu data mining sebenarnya memiliki akar yang panjang dari bidang ilmu seperti kecerdasan buatan (*artificial intelligent*), *machine learning*, statistik dan *database* (Agung & Putri, 2023). Data Mining adalah analisis otomatis dari data yang berjumlah besar atau kompleks dengan tujuan untuk menemukan pola atau kecenderungan yang penting yang biasanya tidak disadari keberadaannya. Alasan utama untuk menggunakan data mining adalah untuk membantu dalam analisis koleksi pengamatan perilaku. Data tersebut rentan terhadap collinearity karena diketahui keterkaitan. Fakta yang tak terelakkan data mining adalah bahwa subset/set data yang dianalisis mungkin tidak mewakili seluruh domain, dan karenanya tidak boleh berisi contoh-contoh hubungan kritis tertentu dan perilaku yang ada di bagian lain dari domain. Untuk mengatasi masalah semacam ini, analisis dapat ditambah menggunakan berbasis percobaan dan pendekatan lain, seperti *Choice Modelling* untuk data yang dihasilkan manusia. Dalam situasi ini, yang melekat dapat berupa korelasi dikontrol untuk, atau dihapus sama sekali, selama konstruksi desain eksperimental (Ishaq Hasibuan dkk., 2021). Beberapa teknik yang sering disebut-sebut dalam literatur Data Mining dalam penerapannya antara lain *clustering*, *classification*, *association rule mining*, *neural network*, *genetic algorithm* dan lain-lain. Yang membedakan persepsi terhadap Data Mining

adalah perkembangan teknik-teknik Data Mining untuk aplikasi pada database skala besar. Sebelum populernya Data Mining, teknik-teknik tersebut hanya dapat dipakai untuk data skala kecil saja. (Purwadi dkk., 2019)

Pengelompokan Data Mining dibagi menjadi beberapa kelompok yaitu (Ishaq Hasibuan dkk., 2021) :

1. Deskripsi

Deskripsi merupakan cara untuk menggambarkan pola dan kecenderungan yang terdapat dalam data yang dimiliki.

2. Estimasi

Estimasi hampir sama dengan klasifikasi, kecuali variable target estimasi lebih ke arah numerik daripada ke arah kategori. Model yang dibangun menggunakan record lengkap yang menyediakan nilai variable target sebagai nilai prediksi.

3. Prediksi

Prediksi menerka sebuah nilai yang belum diketahui dan juga memperkirakan nilai untuk masa mendatang.

4. Klasifikasi

Dalam klasifikasi terdapat target variable kategori, misal penggolongan pendapatan dapat dipisahkan dalam tiga kategori, yaitu tinggi, sedang, dan rendah.

5. Pengklasteran

Merupakan pengelompokan record, pengamatan, atau memperhatikan dan membentuk kelas objek-objek yang memiliki kemiripan.

6. Asosiasi

Asosiasi bertugas menemukan atribut yang muncul dalam satu waktu. Dalam dunia bisnis lebih umum.

2.2. Proses Data Mining

Data mining adalah proses mencari pola atau informasi menarik dalam data terpilih dengan menggunakan teknik atau metode tertentu. Teknik, metode, atau algoritma dalam data mining sangat bervariasi. Pemilihan metode atau algoritma yang tepat sangat bergantung pada tujuan dan proses KDD secara keseluruhan (Yuli Mardi, 2019).

Istilah *Knowledge Discovery in Database (KDD)* dan data mining seringkali digunakan secara bergantian untuk menjelaskan proses penggalian informasi tersembunyi dalam suatu basis data yang besar. Sebenarnya kedua istilah tersebut memiliki konsep yang berbeda, tetapi berkaitan satu sama lain, dan salah satu tahapan dalam keseluruhan proses KDD adalah data mining (Purnomo et al., 2018). Proses KDD secara garis besar dapat dijelaskan sebagai berikut (Asroni dkk., 2018):

1. Data Selection

Pemilihan (seleksi) data dari sekumpulan data operasional perlu dilakukan sebelum tahap penggalian informasi dalam *Knowledge Discovery in Database (KDD)* dimulai. Data hasil seleksi yang akan digunakan untuk proses data mining, disimpan dalam suatu berkas terpisah dari basis data operasional.

2. Pre-processing / Cleaning

Sebelum proses data mining dapat dilaksanakan, perlu dilakukan proses *cleaning* pada data yang menjadi fokus *Knowledge Discovery in Database (KDD)*. Proses *cleaning* mencakup antara lain membuang duplikasi data, memeriksa data yang inkonsisten, dan memperbaiki kesalahan pada data, seperti kesalahan cetak. Juga dilakukan proses *enrichment*, yaitu proses “memperkaya” data yang sudah ada dengan data atau informasi lain yang relevan dan diperlukan untuk *Knowledge*

Discovery in Database (KDD), seperti data atau informasi eksternal lainnya yang diperlukan.

3. *Transformation*

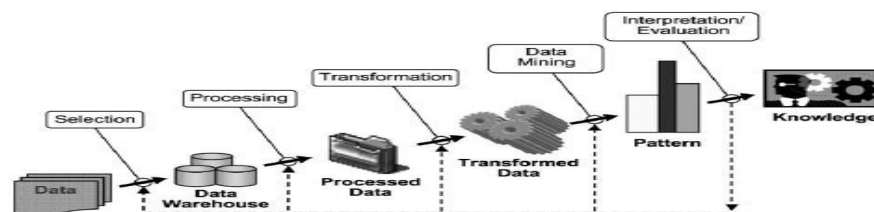
Coding adalah proses transformasi pada data yang telah dipilih, sehingga data tersebut sesuai untuk proses data mining. Proses *coding* dalam *Knowledge Discovery in Database (KDD)* merupakan proses kreatif dan sangat tergantung pada jenis atau pola informasi yang akan dicari dalam basis data.

4. *Data Mining*

Data mining adalah proses mencari pola atau informasi menarik dalam data terpilih dengan menggunakan teknik atau metode tertentu. Teknik-teknik, metode-metode, atau algoritma dalam data mining sangat bervariasi. Pemilihan metode atau algoritma yang tepat sangat bergantung pada tujuan dan proses *Knowledge Discovery in Database (KDD)* secara keseluruhan.

5. *Interpretation / Evaluation*

Pola informasi yang dihasilkan dari proses data mining perlu ditampilkan dalam bentuk yang mudah dimengerti oleh pihak yang berkepentingan. Tahap ini merupakan bagian dari proses *Knowledge Discovery in Database (KDD)* yang disebut *interpretation*. Tahap ini mencakup pemeriksaan apakah pola atau informasi yang ditemukan bertentangan dengan fakta atau hipotesis yang ada sebelumnya.



Gambar 2. 1 Proses Data mining (Yuli Mardi, 2019).

2.3. Prediksi

Prediksi adalah suatu proses memperkirakan secara sistematis tentang sesuatu yang paling mungkin terjadi di masa depan berdasarkan informasi masa lalu dan sekarang yang dimiliki, agar kesalahannya (selisih antara sesuatu yang terjadi dengan hasil perkiraan) dapat diperkecil. Prediksi tidak harus memberikan jawaban secara pasti kejadian yang akan terjadi, melainkan berusaha untuk mencari jawaban sedekat mungkin yang akan terjadi. Pengertian prediksi sama dengan prediksi atau perkiraan (Haryadi dkk., 2021). Menurut kamus besar bahasa Indonesia, prediksi adalah hasil dari kegiatan memprediksi atau meramal atau memperkirakan nilai pada masa yang akan datang dengan menggunakan data masa lalu. Prediksi menunjukkan apa yang akan terjadi pada suatu keadaan tertentu dan merupakan input bagi proses perencanaan dan pengambilan keputusan. Prediksi bisa berdasarkan metode ilmiah ataupun subjektif belaka. Ambil contoh, prediksi cuaca selalu berdasarkan data dan informasi terbaru yang didasarkan pengamatan termasuk oleh satelit. Permulaan awal, walaupun pengkajian yang mendalam mengenai alternatif masa depan adalah suatu disiplin baru, barangkali orang telah menaruh perhatian besar tentang apa yang akan terjadi kemudian semenjak manusia mulai mengetahui sesuatu. Populasi tukang ramal dan tukang nujum pada zaman kuno dan abad pertengahan merupakan satu manifestasi dari keinginan tahu orang tentang masa depannya. Perhatian tentang masa depan ini berlangsung terus bahkan berkembang menjadi kolom astrologi yang disindikatkan pada tahun 1973 (Haryadi dkk., 2021).

Prediksi adalah suatu prosedur untuk membuat informasi factual tentang situasi sosial masa depan atas dasar informasi yang telah ada tentang masalah

kebijakan. Prediksi mempunyai tiga bentuk utama : proyeksi, prediksi, dan perkiraan.

1. Suatu proyeksi adalah ramalan yang di dasarkan pada ekstrapolasi atas kecenderungan masa lalu maupun masa kini ke masa depan. Proyeksi membuat pertanyaan yang tegas berdasarkan argument yang diperoleh dari metode tertentu dan kasus yang paralel.

2. Sebuah prediksi adalah Ramalan yang di dasarkan pada asumsi teoritik yang tegas. Asumsi ini dapat berbentuk hokum teoretis (misalnya hokum berkurangnya nilai uang), proposisi teoritis (misalnya proposisi bahwa pecahnya masyarakat sipil di akibatkan oleh kesenjangan antara harapan dan kemampuan), atau analogi (misalnya analogi antara pertumbuhan organisasi pemerintah dengan pertumbuhan organisme biologis).

3. Suatu perkiraan (*conjecture*) adalah Ramalan yang di dasarkan pada penilaian yang informative atau penilaian pakar tentang situasi masyarakat masa depan. (Kafil & Industri, 2019)

2.4. Jenis-Jenis Prediksi

Masalah dalam prediksi sangatlah komplek dimulai dari bentuk pola datanya, faktor penentu perubahan, penentuan teknik penyelesaiannya dan berbagai aspek lainnya. Untuk menghadapi kondisi yang seperti itu, maka timbul metode baru melalui proses penelitian yang dilakukan oleh para ahli. Prediksi pada dasarnya dibedakan berdasarkan beberapa jenis tergantung dari segi pandang (Agasta dkk., 2018).

2.5. Produksi

Pengertian produksi adalah suatu kegiatan mengubah faktor-faktor produksi atau *input* menjadi produk atau *output*. Faktor-faktor produksi (*input*) merupakan

elemen yang harus ada untuk menghasilkan suatu produksi. Faktor-faktor produksi yang dimaksud adalah tenaga kerja, modal, dan manajemen, Teori produksi modern menambahkan unsur teknologi sebagai salah satu bentuk dari elemen *input*.

Teori produksi membahas bagaimana penggunaan *input* untuk menghasilkan sejumlah *output* tertentu. Teori produksi menjelaskan mengenai fungsi produksi, yaitu suatu fungsi atau persamaan yang menunjukkan hubungan fisik atau teknis antara faktor-faktor yang dipergunakan dengan jumlah produk yang dihasilkan persatuan waktu, tanpa memperhatikan harga, baik harga faktor-faktor produksi maupun harga produk. (Agustina & Kartika, 2018)

2.6. Kelapa Sawit

Tanaman Kelapa Sawit merupakan tumbuhan tropis golongan palma yang termasuk tanaman tahunan. Hasil panen kelapa sawit dikenal dengan istilah Tandan Buah Segar (TBS). Kelapa sawit merupakan komoditas perkebunan dengan luas lahan terluas dan produksi tertinggi di Indonesia karena perkebunan kelapa sawit saat ini telah berkembang dengan pesat tidak hanya perkebunan milik negara tetapi juga milik perusahaan swasta perkebunan rakyat. Tanaman kelapa sawit akan berbunga dan berbuah setelah umur 2-3 tahun.

Analisis produktivitas dengan melihat faktor-faktor yang mempengaruhi pertumbuhan kelapa sawit diperlukan dalam upaya peningkatan produktivitas kelapa sawit. Analisis faktor yang mempengaruhi produktivitas kelapa sawit tidak dapat dilakukan secara mudah mengingat banyak faktor yang mempengaruhi. Faktor-faktor yang mempengaruhi produktivitas kelapa sawit yaitu faktor lingkungan, faktor genetik, dan teknik budidaya.

Analisis produktivitas tanaman dengan melihat faktor-faktor yang mempengaruhinya diharapkan dapat memberikan masukan dalam upaya peningkatan produksi dan melatih mengembangkan kemampuan analisis (Yohansyah & Lubis, 2019). Berikut faktor penentu produktivitas kelapa sawit pada umumnya yaitu :

2.6.1 Luas Lahan

Ketersediaan luas lahan yang cukup luas dapat menghasilkan tanaman kelapa sawit lebih banyak sehingga untuk menghasilkan produktivitas sangat mempengaruhi. Adapun luas lahan rata-rata kelapa sawit dapat mencakup sekitar 130 tanaman/ha.

2.6.2 Jumlah Batang

Faktor produksi yang penting dan perlu diperhitungkan dalam produksi adalah pupuk karena merupakan salah satu kegiatan perawatan tanaman yang bertujuan untuk mendapatkan target produksi Tandan Buah Segar (TBS) yang optimal dan mendapatkan kualitas minyak lebih baik. (Ariyanto dkk., 2020)

2.6.3 Tandan Perpokok

Tandan perpokok adalah banyaknya tandan/janjang dalam satu batang Kelapa Sawit

2.6.4 Berat Janjang rata-rata

Berat Janjang Rata-rata (BJR) adalah rata-rata berat dari satu janjang Kelapa Sawit yang ditimbang per periode tertentu.

2.6.5 Hasil Produksi

Hasil produksi digunakan untuk membentuk pola dalam melakukan prediksi. Hasil produksi juga digunakan untuk melakukan analisis produktivitas, ketika perusahaan

melakukan perubahan pada faktor yang berpengaruh tinggi, misalnya pada faktor drainase air, maka hasil produksi bisa

dijadikan acuan apakah kebijakan perubahan drainase air tetap bisa dipertahankan atau tidak. (Agasta dkk., 2018)



Gambar 2. 2 Kelapa Sawit

2.7. Penelitian Terkait

Dalam penyusunan skripsi ini, penulis sedikit banyak terinspirasi dan mereferensi dari penelitian-penelitian sebelumnya yang berkaitan dengan latar belakang masalah pada skripsi ini. Berikut ini penelitian terdahulu yang berhubungan dengan skripsi ini antara lain:

1. Penelitian yang dilakukan oleh (Agasta dkk., 2018), “ Prediksi Jumlah Produksi Kelapa Sawit Dengan Menggunakan Metode *Extreme Learning Machine* (ELM) (Studi Kasus: PT. Sandabi Indah Lestari Kota Bengkulu)”. Penelitian tersebut bertujuan untuk memprediksi jumlah produksi kelapa sawit. Dengan menggunakan data produksi kelapa sawit 2 tahun terakhir yaitu tahun 2015-2017.

Variabel yang digunakan adalah Umur Tanam, Luas Areal, Jumlah Pokok dan Hasil Produksi. Peneliti melakukan tiga pengujian, Dari ketiga pengujian yang sudah dilakukan didapatkan hasil MAPE untuk masing – masing pengujian adalah 21,25% untuk pengujian jumlah *neuron* pada *hidden layer*, 20,42% untuk pengujian jumlah fitur, 20.19% untuk pengujian pola data. Berdasarkan ketiga pengujian tersebut, maka penelitian ini menghasilkan nilai prediksi yang cukup bagus, karena nilai MAPE berada diantara 20 - 50%.

2. Penelitian yang dilakukan oleh (Sidik Rahmatulla., 2018) . “Prediksi Alokasi Jumlah Produksi Minyak Sawit Dengan Metode Regresi Linier Berganda Pada PT. Palm Lampung Persada”. Berdasarkan Penelitian yang dilakukan bertujuan untuk memprediksi alokasi jumlah produksi minyak sawit sehingga dapat menjadi solusi bagi perusahaan dalam mengatasi masalah ketidakstabilan dalam hal produksi minyak sawit. Peneliti menggunakan data satu tahun sebelumnya yaitu 2017 dan variabel yang digunakan adalah data yang mempengaruhi produksi seperti bahan baku, jumlah tenaga kerja, curah hujan, dan luas lahan. Perhitungan prediksi dengan menggunakan metode regresi linier berganda dalam menentukan jumlah produksi minyak sawit menghasilkan rata-rata memiliki tingkat kebenaran prediksi sebagai berikut : bahan baku (x_1) adalah sebesar 0,424, jumlah tenaga kerja (x_2) adalah sebesar 0, 073, curah hujan (x_3) adalah sebesar 0,002 sedangkan luas lahan adalah sebesar 0,065. Jadi dari ke empat variable tersebut yang paling berpengaruh pada produksi minyak sawit adalah curah hujan (x_3) sebesar 0,002.

3. Penelitian yang dilakukan oleh (Yunus dkk., 2020). “Data Mining Untuk Memprediksi Hasil Produksi Buah Sawit Pada PT. Bumi Sawit Sukses (BSS) Menggunakan Metode *K-Nearest Neighbor*”. Berdasarkan penelitian yang

dilakukan bertujuan untuk memprediksi hasil produksi buah sawit. Data yang digunakan adalah data hasil produksi kelapa sawit dari 50 kelompok tani pada periode Juli-Desember 2011 pada tanaman tahun 2015-2017. Hasil tingkat akurasi dengan ROC (Receiver Operating Characteristic) Curve menunjukkan akurasi yang tergolong *Good Classification* yaitu nilai AUC (*Area Under Curve*) sebesar 0,888 dengan klastering data $K=5$.

4. Penelitian yang dilakukan oleh (Agusta dkk., 2023). “Peramalan Pendapatan dari Penjualan Bawang Merah Menggunakan Metode Regresi Linear Berganda”. Berdasarkan penelitian yang dilakukan bertujuan untuk meramalkan jumlah panen bawang merah. Data yang digunakan adalah data primer. Data primernya yaitu dengan mewawancarai langsung salah satu petani dikota kediri. Peneliti menggunakan data 8 tahun terakhir yaitu dari tahun 2015-2022 dan variabel yang digunakan adalah jumlah panen ($X1$), harga perkilogram ($X2$), dan total pendapatan (Y). Skor akurasi 60%, jadi bisa disimpulkan bahwa model yang telah dibuat bisa digunakan untuk melakukan peramalan. Hasil dari peramalan pada masa panen selanjutnya menunjukkan total pendapatan sebesar Rp. 15.482.489 dengan memperoleh nilai MAPE sebesar 13,65 atau 13%.

2.8. *Python*

Python adalah bahasa pemrograman interpretatif multiguna dengan filosofi perancangan yang berfokus pada tingkat keterbacaan kode. *Python* diklaim sebagai bahasa yang menggabungkan kapabilitas, kemampuan, dengan sintaksis kode yang sangat jelas, dan dilengkapi dengan fungsionalitas pustaka standar yang besar serta komprehensif.

Python mendukung multi paradigma pemrograman, utamanya; namun tidak dibatasi; pada pemrograman berorientasi objek, pemrograman imperatif, dan pemrograman fungsional. Salah satu fitur yang tersedia pada *python* adalah sebagai bahasa pemrograman dinamis yang dilengkapi dengan manajemen memori otomatis. Seperti halnya pada bahasa pemrograman dinamis lainnya, *python* umumnya digunakan sebagai bahasa skrip meski pada praktiknya penggunaan bahasa ini lebih luas mencakup konteks pemanfaatan yang umumnya tidak dilakukan dengan menggunakan bahasa skrip. *Python* dapat digunakan untuk berbagai keperluan pengembangan perangkat lunak dan dapat berjalan di berbagai *platform* sistem operasi (Effendi dkk., 2019).

2.9. Mean Absolute Error (MAE)

Mean Absolute Error (MAE) adalah salah satu metode yang digunakan untuk mengukur tingkat keakuratan model peramalan. Nilai MAE menunjukkan rata – rata kesalahan (*error*) *absolut* antara hasil peramalan/prediksi dengan nilai riil. Secara rumus MAE dijelaskan sebagai berikut, (Suryanto, 2019).

$$MAE = \frac{1}{n} \sum_{i=1}^n |y_i - \hat{y}_i| \quad \dots\dots (2.1)$$

Dimana :

\hat{y}_i : Nilai hasil prediksi

y_i : Nilai sebenarnya, dan

n : Jumlah data

Berdasarkan rumus diatas MAE menghitung rata – rata *error* dengan memberikan bobot yang sama untuk seluruh data ($i= 1 \dots n$) secara intuitif untuk evaluasi model peramalan, MAE lebih intuitif memberikan rata-rata *error* dari

keseluruhan data Dalam kasus ini pemilihan MAE menjadi tepat karena seluruh data diberikan bobot yang sama.

Dalam kasus klasifikasi biner, dimana hanya terdapat dua kelas dengan label kelas 1 dan 0, dalam penggunaan MAE. Hal ini karena nilai *error* hanya mempunyai dua kemungkinan, 1 jika prediksi kelas berbeda dengan kelas sebenarnya dan 0 jika prediksi benar.

2.10. Regresi

Regresi adalah hubungan yang dapat dan dinyatakan dalam bentuk persamaan matematik yang menyatakan hubungan fungsional antar variabel-variabel. Analisis regresi merupakan kajian terhadap hubungan satu variabel yang disebut sebagai variabel yang diterangkan dengan satu atau dua variabel yang dapat menerangkan. Variabel yang diterangkan selanjutnya disebut sebagai variabel terikat (variabel dependen), sedangkan variabel yang menerangkan biasa disebut variabel bebas (variabel independen) yaitu variabel yang menjelaskan atau mempengaruhi variable lain.

Analisis regresi dengan satu variabel bebas X disebut sebagai regresi linear sederhana, sedangkan jika terdapat lebih dari satu variabel bebas X, Disebut sebagai regresi linear berganda. (Panggabean dkk., 2020)

2.11. Regresi Linear Berganda

Regresi linear berganda adalah analisis regresi yang menjelaskan hubungan antara perubah respon (variabel dependen) dengan faktor-faktor yang mempengaruhi lebih dari satu prediktor (variabel independen). Ketika suatu hasil/keluaran, atau kelas berupa numerik, dan semua atribut adalah numerik, regresi linear adalah teknik yang tepat untuk menyelesaikan. Ini adalah metode

pokok di dalam ilmu statistik (Triyanto dkk., 2019). Gunanya adalah untuk mengekspresikan kelas sebagai kombinasi linear dari atribut, dengan bobot yang telah di tentukan, dengan rumus sebagai berikut:

$$Y = a + b_1 \cdot X_1 + b_2 \cdot X_2 \dots + b_n \cdot X_n \quad (2.2)$$

Di mana Y adalah kelas; X_1, X_2, \dots, X_n adalah nilai atribut; dan a, b_1, \dots, b_n adalah bobot. Bobot dihitung dari data sampel.

Y = Variabel tidak bebas (nilai yang diprediksikan)

X = Variabel bebas

a = Konstanta (nilai Y apabila $X_1, X_2 \dots X_n = 0$)

b = Koefisien regresi (nilai peningkatan ataupun penurunan)

Dimana nilai $a, b_1, b_2 \dots b_n$ dapat dihitung dengan metode persamaan normal yaitu :

$$\begin{aligned} \sum Y &= a \sum 1 + b_1 \sum X_1 + b_2 \sum X_2 \\ \sum Y \cdot X_1 &= a \sum X_1 + b_1 \sum X_1^2 + b_2 \sum X_1 X_2 \\ \sum Y \cdot X_2 &= a \sum X_2 + b_1 \sum X_1 X_2 + b_2 \sum X_2^2 \end{aligned} \quad (2.3)$$

Selain dihitung dengan persamaan normal diatas, nilai $a, b_1, b_2 \dots b_n$ dapat juga dihitung dengan metode kuadran terkecil, yaitu:

$$\begin{aligned} b_1 &= \frac{(\sum X_2^2)(\sum Y) - (\sum X_1 X_2)(\sum Y)}{(\sum X_1^2)(\sum X_2^2) - (\sum X_1 X_2)^2} \\ b_2 &= \frac{(\sum X_1^2)(\sum Y) - (\sum X_1 X_2)(\sum Y)}{(\sum X_1^2)(\sum X_2^2) - (\sum X_1 X_2)^2} \\ a &= \bar{Y} - b_1 \bar{X}_1 - b_2 \bar{X}_2 \end{aligned} \quad (2.4)$$

Dimana :

$$\begin{aligned}
 \sum x_1^2 &= \sum x_1^2 - \frac{(\sum x_1)^2}{n} \\
 \sum x_2^2 &= \sum x_2^2 - \frac{(\sum x_2)^2}{n} \\
 \sum x_1 x_2 &= \sum x_1 x_2 - \frac{(\sum x_1)(\sum x_2)}{n} \\
 \sum x_1 x_1 &= \sum x_1^2 - \frac{(\sum x_1)^2}{n} \\
 \sum x_2 x_2 &= \sum x_2^2 - \frac{(\sum x_2)^2}{n} \\
 \sum x_1 x_2 &= \sum x_1 x_2 - \frac{(\sum x_1)(\sum x_2)}{n}
 \end{aligned} \tag{2.5}$$

Ketika variable bebas lebih dari 2, nilai konstanta dan koefisien regresi setiap variabel bebas dapat diperoleh dengan menggunakan matriks determinan (Triyanto dkk., 2019). Contohnya adalah ketika terdapat 3 persamaan dengan 3 variabel yang tidak diketahui nilainya, yaitu a , b_1 , b_2 dan b_3 , persamaan tersebut dapat dinyatakan dalam persamaan matriks sebagai berikut:

$$A = \begin{vmatrix} \sum x_1^2 & \sum x_1 x_2 \\ \sum x_1 x_2 & \sum x_2^2 \end{vmatrix} \begin{matrix} x_1 \\ x_2 \end{matrix} = \begin{matrix} \sum x_1 a \\ \sum x_2 a \end{matrix}$$

3

|

$$\begin{aligned}
 & (2.6) \\
 & \sum_{i=1}^n x_{i1} \sum_{j=1}^n x_{i2} \sum_{k=1}^n x_{i3} \\
 & \sum_{i=1}^n x_{i1} \sum_{j=1}^n x_{i2} \sum_{k=1}^n x_{i3} \\
 & \sum_{i=1}^n x_{i1} \sum_{j=1}^n x_{i2} \sum_{k=1}^n x_{i3}
 \end{aligned}$$

$$\begin{aligned}
 A0 = & \left| \begin{array}{cccc}
 \sum_{i=1}^n x_{i1} & \sum_{i=1}^n x_{i2} & \sum_{i=1}^n x_{i3} & \sum_{i=1}^n x_{i4} \\
 \sum_{i=1}^n x_{i1} & \sum_{i=1}^n x_{i2} & \sum_{i=1}^n x_{i3} & \sum_{i=1}^n x_{i4} \\
 \sum_{i=1}^n x_{i1} & \sum_{i=1}^n x_{i2} & \sum_{i=1}^n x_{i3} & \sum_{i=1}^n x_{i4} \\
 \sum_{i=1}^n x_{i1} & \sum_{i=1}^n x_{i2} & \sum_{i=1}^n x_{i3} & \sum_{i=1}^n x_{i4}
 \end{array} \right| \quad (2.7)
 \end{aligned}$$

$$\begin{aligned}
 A1 = & \left| \begin{array}{cccc}
 \sum_{i=1}^n x_{i1} & \sum_{i=1}^n x_{i2} & \sum_{i=1}^n x_{i3} & \sum_{i=1}^n x_{i4} \\
 \sum_{i=1}^n x_{i1} & \sum_{i=1}^n x_{i2} & \sum_{i=1}^n x_{i3} & \sum_{i=1}^n x_{i4} \\
 \sum_{i=1}^n x_{i1} & \sum_{i=1}^n x_{i2} & \sum_{i=1}^n x_{i3} & \sum_{i=1}^n x_{i4} \\
 \sum_{i=1}^n x_{i1} & \sum_{i=1}^n x_{i2} & \sum_{i=1}^n x_{i3} & \sum_{i=1}^n x_{i4}
 \end{array} \right| \quad (2.8)
 \end{aligned}$$

$$A_2 = \begin{vmatrix} \sum_{i=1}^2 x_{11} & \sum_{i=1}^2 x_{12} & \sum_{i=1}^2 x_{13} \\ \sum_{i=1}^2 x_{21} & \sum_{i=1}^2 x_{22} & \sum_{i=1}^2 x_{23} \\ \sum_{i=1}^2 x_{31} & \sum_{i=1}^2 x_{32} & \sum_{i=1}^2 x_{33} \end{vmatrix} \quad (2.9)$$

$$A_3 = \begin{vmatrix} \sum_{i=1}^3 x_{11} & \sum_{i=1}^3 x_{12} & \sum_{i=1}^3 x_{13} \\ \sum_{i=1}^3 x_{21} & \sum_{i=1}^3 x_{22} & \sum_{i=1}^3 x_{23} \\ \sum_{i=1}^3 x_{31} & \sum_{i=1}^3 x_{32} & \sum_{i=1}^3 x_{33} \end{vmatrix} \quad (2.10)$$

Kemudian dapat dilakukan perhitungan untuk determinasi matriks A , A_0 , A_1 , A_2 , dan A_3 sebagai berikut:

$$\begin{aligned} \det(A) &= \sum_{i=1}^3 \sum_{j=1}^3 x_{ij} \sum_{k=1}^3 x_{kj} \sum_{l=1}^3 x_{li} \\ &= \sum_{i=1}^3 \sum_{j=1}^3 \sum_{k=1}^3 \sum_{l=1}^3 x_{ij} x_{kj} x_{li} \\ &= \sum_{i=1}^3 \sum_{j=1}^3 \sum_{k=1}^3 \sum_{l=1}^3 x_{ij} x_{kj} x_{li} \\ &= \sum_{i=1}^3 \sum_{j=1}^3 \sum_{k=1}^3 \sum_{l=1}^3 x_{ij} x_{kj} x_{li} \\ &= \sum_{i=1}^3 \sum_{j=1}^3 \sum_{k=1}^3 \sum_{l=1}^3 x_{ij} x_{kj} x_{li} \\ &= \sum_{i=1}^3 \sum_{j=1}^3 \sum_{k=1}^3 \sum_{l=1}^3 x_{ij} x_{kj} x_{li} \\ &= \sum_{i=1}^3 \sum_{j=1}^3 \sum_{k=1}^3 \sum_{l=1}^3 x_{ij} x_{kj} x_{li} \\ &= \sum_{i=1}^3 \sum_{j=1}^3 \sum_{k=1}^3 \sum_{l=1}^3 x_{ij} x_{kj} x_{li} \\ &= \sum_{i=1}^3 \sum_{j=1}^3 \sum_{k=1}^3 \sum_{l=1}^3 x_{ij} x_{kj} x_{li} \\ &= \sum_{i=1}^3 \sum_{j=1}^3 \sum_{k=1}^3 \sum_{l=1}^3 x_{ij} x_{kj} x_{li} \\ &= \sum_{i=1}^3 \sum_{j=1}^3 \sum_{k=1}^3 \sum_{l=1}^3 x_{ij} x_{kj} x_{li} \end{aligned}$$

$$\sum_{i=1}^n \sum_{j=1}^n \sum_{k=1}^n (x_{ij} \cdot x_{jk} \cdot x_{ki})$$

$$\sum_{i=1}^n \sum_{j=1}^n \sum_{k=1}^n (x_{ij} \cdot x_{jk} \cdot x_{ki}) \quad (2.11)$$



Dengan cara yang sama menghitung $\text{Det}(A)$, dapat diperoleh pula $\text{Det}(A0)$, $\text{Det}(A1)$, $\text{Det}(A2)$ dan $\text{Det}(A3)$. Nilai a , $b1$, $b2$, dan $b3$ dapat diperoleh dengan cara sebagai berikut :

$$\diamond = \frac{0}{\quad} \quad (2.12)$$

$$\diamond_1 = \frac{1}{\quad} \quad (2.13)$$



BAB III

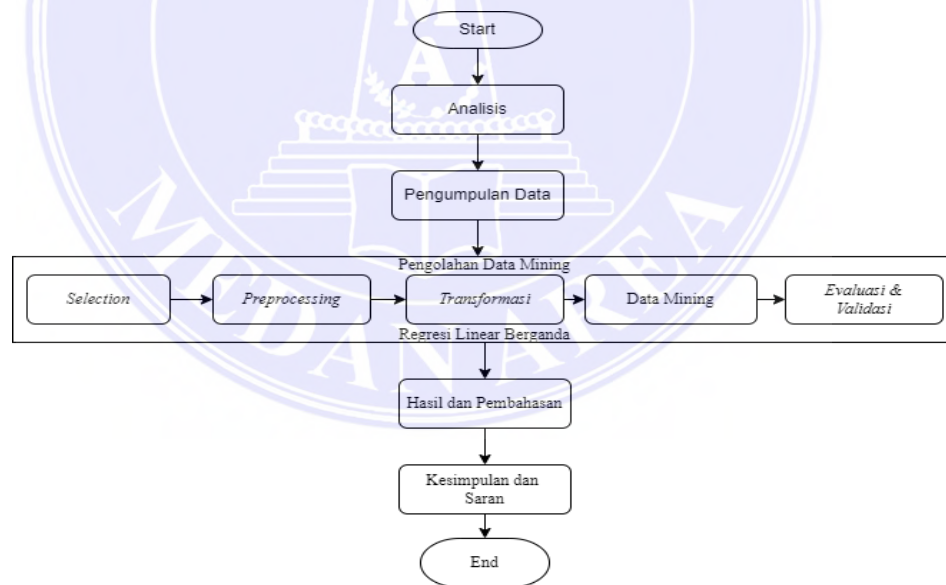
ANALISIS DAN PERANCANGAN

3.1. Analisis

Analisis adalah suatu kegiatan dalam mempelajari serta mengevaluasi suatu bentuk permasalahan atau kasus yang terjadi. Sedangkan perancangan adalah suatu kegiatan membuat desain teknis berdasarkan evaluasi yang telah dilakukan pada kegiatan analisis.

3.1.1. Kerangka Kerja Penelitian

Penelitian yang dilakukan dapat digambarkan dalam suatu alur kegiatan kerja penelitian seperti terlihat pada gambar 3.1. Sehingga alur kegiatan tersebut dapat menjadi pedoman dalam pemecahan masalah yang akan dihadapi dalam melakukan penelitian.



Gambar 3. 1 Alur Kerja Penelitian

Tahapan penelitian yang dituangkan pada Gambar 3.1 menggambarkan proses penelitian yang akan ditempuh sekaligus menggambarkan penelitian secara keseluruhan. Tahapan yang akan ditempuh yaitu:

1. Analisis. Sebuah proses untuk memeriksa data untuk menemukan informasi yang bermanfaat sehingga dapat memberikan petunjuk bagi peneliti untuk mengambil keputusan.
2. Pengumpulan data. Pengumpulan data dilakukan dengan cara wawancara kepada karyawan, observasi dan dokumen.
3. Pengolahan data mining. Data yang dikumpulkan diolah sesuai dengan tahapan *Knowledge Discovery in Database* (KDD).
4. Hasil pembahasan. Pembahasan pada tahapan ini menjelaskan hasil dari proses data mining yang dilakukan dengan menggunakan metode Regresi Linear Berganda.
5. Kesimpulan dan saran. Membuat Kesimpulan dan hasil penelitian dan memberikan saran untuk pihak perusahaan agar dapat menjadi lebih baik lagi.

3.1.2. Analisis

Pada analisis sistem terdiri dari Analisis Sistem yang sedang berjalan, analisis kebutuhan sistem, analisis sistem yang diusulkan, metode pengembangan sistem, tahapan penelitian dan langkah-langkah pengerjaan metode regresi linier berganda,

3.1.2.1. Analisis yang Berjalan

PT. SGSR (Sinar Gunung Sawit Raya) belum memiliki aplikasi untuk memprediksi Hasil produksi setiap tahunnya, selama ini proses prediksi hanya menggunakan perhitungan manual yang dihitung di aplikasi Microsoft Exel. Proses dalam melakukan prediksi seharusnya dapat dilakukan untuk membuat langkah pasti dalam menentukan target keperiode berikutnya seperti menambah pegawai/pekerja jika dibutuhkan agar dapat mencapai target prediksi tersebut.

3.1.2.2. Analisis Kebutuhan

Analisis yang berjalan saat ini, terdapat beberapa hal yang harus dipenuhi yaitu diperlukan adanya penerapan yang mampu untuk memprediksi Hasil Produksi berdasarkan variabel-variabel yang ada.

A. Kebutuhan Non-Fungsional

Kebutuhan non fungsional terdiri dari kebutuhan perangkat keras dan kebutuhan perangkat lunak.

1. Spesifikasi Kebutuhan *Hardware*

Perangkat Keras (*Hardware*) yang digunakan pada sistem prediksi produksi kelapa sawit adalah :

- a. *Processor* : *Intel Core i3-1005G1*
- b. *Ram* : 4,00 GB
- c. *SSD* : 500 GB
- d. *System tipe* : *64-bit Operating System*

2. Spesifikasi Kebutuhan *Software*

Perangkat Lunak (*Software*) yang digunakan pada sistem prediksi kelapa sawit adalah :

- a. *Sistem Operasi* : *Microsoft windows 11*
- b. *Bahasa Pemrograman* : *Python*

B. Kebutuhan Fungsional

Berikut adalah kebutuhan fungsional pada aplikasi yang dibangun terkait untuk memprediksi hasil produksi Afdeling 1.

1. Penerapan dapat memproses data untuk menghasilkan nilai variabel a , b_1 , b_2 , b_3 , dan b_4
2. Penerapan dapat memprediksi hasil produksi setelah admin menginput luas lahan (X_1), banyak batang (X_2), tandan/pokok (X_3) dan berat janjang rata-rata (X_4).
3. Admin dapat mencetak laporan hasil prediksi pendapatan.

3.1.2.3. Analisis Yang Diusulkan

Untuk memprediksi jumlah hasil produksi menggunakan bahasa pemrograman Python serta metode yang digunakan dalam memprediksi adalah Regresi Linear. Pada sistem tersebut terdapat *form* regresi yang digunakan untuk menentukan koefisien regresi untuk menghasilkan nilai variabel a , b_1 , b_2 , b_3 dan b_4 agar dapat digunakan untuk memprediksi data *form* regresi.

3.1.3. Bahan Penelitian

Adapun bahan penelitian yang akan digunakan dalam penelitian ini menggunakan data primer. Data primer yang didapatkan berasal dari PT. SGSR Mujur Sinar Gunung Sawit Raya. Data yang diperoleh ialah data hasil produksi kelapa sawit dengan rentang waktu 2019-2022.

3.1.4. Pengumpulan Data

Berikut teknik-teknik pengumpulan data yang akan dilakukan dalam pembuatan skripsi :

a. Studi Pustaka

Penulis mencari berbagai sumber referensi ke pustaka sebagai acuan penelitian yang mendukung pada permasalahan kasus skripsi yang dibuat. Data tersebut akan mendukung penulis untuk melakukan penelitian dan mendukung penulis dalam penyusunan skripsi ini.

b. Observasi

Penulis juga menggunakan metode observasi untuk penelitian ini. Metode pengumpulan data observasi digunakan untuk pengamatan lapangan. Observasi dilakukan dengan cara datang langsung ke PT. SGSR Mujur Sinar Gunung Sawit Raya. Data dari melakukan observasi adalah untuk mengetahui langsung keadaan kebun perusahaan.

3.1.5. Langkah-Langkah Perhitungan Metode Regresi Linear Berganda

Dalam penelitian ini, peneliti menggunakan metode Regresi Linear Berganda Berganda untuk memprediksi produksi kelapa sawit pada PT. SGSR Mujur Sinar Gunung Sawit Raya. Variabel yang digunakan dalam memprediksi pendapatan adalah Luas lahan, banyak batang, tandan/pokok, berat janjang rata-rata dan jumlah produksi. Berikut adalah alur tahapan dari metode regresi linear berganda dalam memprediksi hasil produksi berdasarkan variabel yang digunakan, yaitu :

- a. Menentukan data
- b. Menghitung nilai koefisien regresi
- c. Perhitungan regresi linear berganda

A. Menentukan Data

Berikut ini adalah tabel data yang diperoleh dari PT. SGSR Mujur Sinar Gunung Sawit Raya dimana variabel independen/bebas yaitu Luas lahan (X_1), Banyak Batang (X_2), Tandan/Pokok (X_3), Berat janjang rata-rata (X_4) dan variabel dependen/terikat adalah Hasil Produksi (Y) yang akan diprediksikan :

Tabel 3. 1 Data Hasil Produksi

No	Bulan	Luas lahan(X_1)	Banyak Batang(X_2)	Tandan perpokok(X_3)	BJR (X_4)	Hasil produksi(Y)
1	Jan	24.59	2628	0.54	21.23	24600
2	Jan	25.07	2810	0.60	24.81	31280

3	Jan	23.02	2578	0.43	23.54	25920
...
372	Des	19.74	2165	0.37	13.45	10680
Total		8350.20	1121148	515.44	4171.49	13572300

Data hasil produksi secara lengkap ada di lampiran

Keterangan :

X_1 = Luas Lahan

X_2 = Banyak Batang

X_3 = Tandan/Pokok

X_4 = Berat janjang rata-rata

Y = Hasil Produksi

B. Menghitung Nilai Koefisien Regresi

Metode regresi linear berganda ini adalah regresi yang bisa memprediksi hubungan antara 2 variabel yang bebas ataupun lebih dengan variabel yang terikat, dimana ditentukan hubungan antara Y dan $X_1, X_2, X_3, \dots, X_n$. Metode ini adalah salah satu jenis regresi linear pada ilmu statistika yang bisa mengadopsi data mining untuk mengetahui pengaruh dalam variabel-variabel bebas tersebut. Untuk dapat memprediksi Y atau tabel terikat, jika semua nilai variabel bebas diketahui, maka :

Tabel 3. 2 Koefisien Regresi

No	$X_1.Y$	$X_2.Y$	$X_3.Y$	$X_4.Y$...	X^2
1	604914	64648800	13309.09	522139.78	...	450.508
2	784189.6	87896800	187680	775922.60	...	615.323
3	596678.4	66821760	11069.80	610214.71	...	554.236
...
372	210823.2	23122200	3916.822	143655.41	...	180.92
Total	388917942.7	52865162100	21269519.05	142095297.8	...	58166.4814

Setelah data prediksi didapatkan maka langkah selanjutnya adalah menjumlahkan data tersebut berdasarkan kolom masing-masing.

Tabel 3. 3 Jumlah Data Prediksi

Variabel	Jumlah Data
$X1$	8350,20
$X2$	1121148
$X3$	514,74
$X4$	4170,49
Y	13572300
$X1.Y$	388917942
$X2.Y$	52865162100
$X3.Y$	21269519
$X4.Y$	124095297
$X1.X2$	32426481,72
$X1.X3$	11302,34
$X1.X4$	89203
$X2.X3$	1555489,46
$X2.X4$	11314735,03
$X3.X4$	5167,44
$X1^2$	239447,57
$X2^2$	4440700692
$X3^2$	871,79
$X4^2$	58116,48

Karena variable bebas lebih dari 2, nilai konstanta dan koefisien regresi setiap variabel bebas dapat diperoleh dengan menggunakan matriks determinan yaitu pada persamaan (2.5). Maka :

$$A = \begin{vmatrix} 372 & 8350 & 1121148 & 515 & 4170 \\ 8350 & 239448 & 32426482 & 11302 & 89203 \\ 1121148 & 32426482 & 44407000692 & 1555489 & 11314735 \\ 515 & 11302 & 1555489 & 872 & 5167 \\ 4170 & 89203 & 11314735 & 5167 & 58166 \end{vmatrix}$$

$$A0 = \begin{vmatrix} 13572300 & 8350 & 1121148 & 515 & 4170 \\ 388917942 & 239447 & 32426481 & 11302 & 89203 \\ 52865162100 & 32426481 & 4440700692 & 1555489 & 11314735 \\ 21269519 & 11302 & 1555489 & 871 & 5167 \\ 142095297 & 89203 & 11314735 & 5167 & 58166 \end{vmatrix}$$

$$A1 = \begin{vmatrix} 372 & 13572300 & 1121148 & 515 & 4170 \\ 8350 & 388917942 & 32426481 & 11302 & 89203 \\ 1121148 & 52865162100 & 4440700692 & 1555489 & 11314735 \\ 515 & 21269519 & 1555489 & 871 & 5167 \\ 4170 & 142095297 & 11314735 & 5167 & 58166 \end{vmatrix}$$

$$A2 = \begin{vmatrix} 372 & 8350 & 13572300 & 515 & 4170 \\ 8350 & 239447 & 388917942 & 11302 & 89203 \\ 1121148 & 32426481 & 52865162100 & 1555489 & 11314735 \\ 514 & 11302 & 21269519 & 871 & 5167 \\ 4170 & 89203 & 142095297 & 5167 & 58166 \end{vmatrix}$$

$$\begin{matrix}
 & \begin{matrix} 372 & 8350 & 1121148 & 13572300 & 4170 \\
 & 8350 & 239448 & 32426482 & 388917942 & 89203 \\
 A3 = & \begin{matrix} 1121148 & 32426482 & 44407000692 & 52865162100 & 11314735 \\
 & 515 & 11302 & 1555489 & 21269519 & 5167 \\
 & 4170 & 89203 & 11314735 & 142095297 & 58166 \end{matrix} \\
 \\
 & \begin{matrix} 372 & 8350 & 1121148 & 515 & 13572300 \\
 & 8350 & 239448 & 32426482 & 11302 & 388917942 \\
 A4 = & \begin{matrix} 1121148 & 32426482 & 44407000692 & 1555489 & 52865162100 \\
 & 515 & 11302 & 1555489 & 872 & 21269519 \\
 & 4170 & 89203 & 11314735 & 5167 & 142095297 \end{matrix}
 \end{matrix}$$

Selanjutnya adalah menentukan determinan dari matrik tersebut dengan menggunakan metode sesuai dengan persamaan (2.10). Berikut ini hasil dari perhitungan determinan dari A, A0, A1, A2, A3 dan A4

Tabel 3.4 Hasil Determinan Matriks

A	295208772749435000000
A0	-12977923846365500000000000
A1	6793737689264660000000000
A2	-10707252745753000000000
A3	6587525454214200000000000
A4	2328535258005980000000000

Dari nilai determinan yang telah dihitung maka nilai a, b1, b2, dan b3 dapat dicari dengan persamaan (2.11) sebagai berikut:

$$\begin{aligned}
 \diamond & \frac{-12977923846365500000000000}{295208772749435000000} = -44014.8502 \\
 \diamond 1 & \frac{6793737689264660000000000}{295208772749435000000} = 2333.333265 \\
 \diamond 2 & = \frac{10707252745753000000000}{295208772749435000000} = -3.827010351 \\
 \diamond 3 & = \frac{6587525454214200000000000}{295208772749435000000} = 22339.80248 \\
 \diamond 4 & = \frac{2328535258005980000000000}{295208772749435000000} = 776.7757658
 \end{aligned}$$

Setelah nilai dari a , b_1 , b_2 , dan b_3 telah dihitung maka dapat diperoleh persamaan regresi linear berganda untuk data hasil produksi tersebut :

$$Y = -44014.8502 + 2333.333265.X^1 + 3.827010351.X^2 + 22339.80248.X^3 + 776.7757658.X^4$$

C. Perhitungan Regresi Linear Berganda

Jika data awal dibuktikan dengan persamaan (2.1), maka dapat diperoleh hasil sebagai berikut :

$$Y_1 = -44014.8502 + 2333.333265.(24,59) + -3.827010351.(2628) + 22339.80248.(0,54) + 776.7757658.(21,23) = 31911$$

$$Y_2 = -44014.8502 + 2333.333265.(25,07) + -3.827010351.(2810) + 22339.80248.(0,6) + 776.7757658.(24,81) = 40080$$

$$Y_3 = -44014.8502 + 2333.333265.(23,02) + -3.827010351.(2578) + 22339.80248.(0,43) + 776.7757658.(23,54) = 29048$$

$$Y_4 = -44014.8502 + 2333.333265.(16,89) + -3.827010351.(1479) + 22339.80248.(0,64) + 776.7757658.(23,86) = 24943$$

$$Y_5 = -44014.8502 + 2333.333265.(11,29) + -3.827010351.(905) + 22339.80248.(1,01) + 776.7757658.(21,09) = 20652$$

$$Y_6 = -44014.8502 + 2333.333265.(26,51) + -3.827010351.(3231) + 22339.80248.(0,90) + 776.7757658.(12,71) = 33262$$

Proses perhitungan yang sama dilakukan terhadap semua data. Setelah dihitung terhadap semua data, maka hasilnya sebagai berikut :

Tabel 3.5 Hasil Perhitungan Regresi

No.	Bulan	Luas Lahan(X1)	Banyak batang(X2)	Tandan perpokok(X3)	BJR(X4)	Prediksi
1	Jan	24.59	2628	0.54	21.23	31911
2	Jan	25.07	2810	0.60	24.81	39937
3	Jan	23.02	2578	0.43	23.54	27764
...
372	Des	19.74	2165	0.37	13.45	12408

Jika terjadi data baru, diketahui luas lahan 500 Ha, banyak batang 20.000 batang, tandan/pokok 2 dan berat janjang rata-ratanya 15 Kg. Maka hasil prediksi produksinya sebagai berikut :

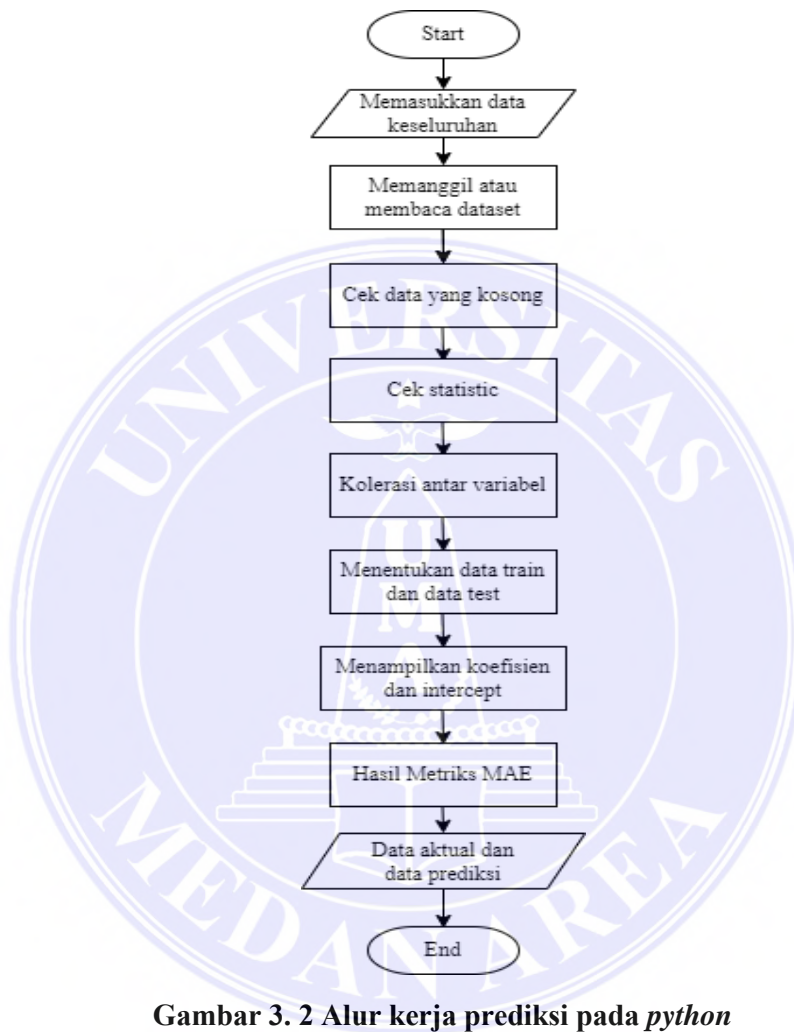
$$Y = -44014.8502 + 2333.333265.(500) + -3.827010351.(20000) + 22339.80248.(2) + 776.7757658.(15) = \underline{1090626 \text{ Kg}}$$

Hasil prediksi data baru adalah 1090626 Kg

3.1. Perancangan

3.2.1. Flowchart Langkah-langkah Kerja Prediksi menggunakan Python

Berikut adalah tahapan-tahapan dalam untuk memprediksi pada *python*:



Gambar 3. 2 Alur kerja prediksi pada *python*

3.2.2. Data Selection

Data yang digunakan dalam penelitian ini yaitu data produksi manual yang ada di Microsoft excel berdasarkan data produksi 4 tahun terakhir dari tahun 2019, 2020, 2021, dan 2022 yang berasal dari PT. SGSR Sinar Gunung Sawit Raya Mujur. Kemudian data tersebut diseleksi secara acak menggunakan *python*. Adapun

variabel yang digunakan dalam memprediksi hasil produksi adalah luas lahan, banyak pokok, tandan/pokok, berat janjang rata-rata dan hasil produksi.

3.2.3. *Preprocessing/Cleaning*

Tahap *preprocessing* ini adalah data yang menjadi fokus dari hasil data seleksi. Proses *cleaning* mencakup antara lain membuang duplikasi data, memeriksa data yang inkonsisten dan memperbaiki kesalahan pada data.

3.2.4. *Data Mining*

Tahapan ini merupakan proses mencari pola atau informasi menarik dalam data terpilih dengan menggunakan teknik atau metode tertentu berdasarkan proses KDD secara keseluruhan. Metode yang digunakan pada penelitian ini adalah metode regresi linear berganda dimana metode ini memiliki atribut yang diinisialisasikan sebagai y , yaitu hasil produksi yang dijadikan sebagai acuan pada regresi.

3.2.5. *Interpretation/Evaluasi*

Pola informasi yang dihasilkan dari proses data mining perlu ditampilkan dalam bentuk yang mudah dimengerti oleh pihak yang berkepentingan. Tahapan ini mencakup pemeriksaan apakah pola atau informasi yang ditemukan bertentangan dengan fakta yang ada sebelumnya.

BAB V

KESIMPULAN DAN SARAN

5.1. Kesimpulan

Berdasarkan sistem Penerapan Data Mining Untuk Memprediksi Jumlah Produksi Kelapa Sawit Menggunakan Metode Regresi Linear Berganda (Studi Kasus : PT. SGSR Mujur Sinar Gunung Sawit raya) maka ditarik kesimpulan bahwa Perhitungan prediksi dengan menggunakan metode regresi linier berganda dalam menentukan jumlah produksi kelapa sawit menghasilkan rata-rata memiliki tingkat kebenaran prediksi sebagai berikut : luas lahan ($X1$) sebesar 0.76, banyak batang ($X2$) sebesar 0.75, tandan/pokok ($X3$) sebesar 0.37, dan berat janjang rata-rata ($X4$) sebesar -0.16. Maka variabel yang paling berpengaruh pada hasil produksi kelapa sawit adalah luas lahan ($X1$) sebesar 0.76.

5.2. Saran

Penerapan metode Regresi Linear Berganda untuk memprediksi jumlah produksi kelapa sawit ini masih bisa menambahkan variabel-variabel lain seperti variabel pupuk, pemanen dan lain lain. Penerapan metode Regresi Linear Berganda untuk memprediksi jumlah produksi kelapa sawit ini juga merupakan simulasi yang belum sempurna sehingga dibutuhkan penyempurnaan yang lebih baik lagi.

DAFTAR PUSTAKA

- Agasta, E., Cholissodin, I., & Ratnawati, D. E. (2018). Prediksi Jumlah Produksi Kelapa Sawit Dengan Menggunakan Metode Extreme Learning Machine (ELM) (Studi kasus : PT . Sandabi Indah Lestari Kota Bengkulu). *Jurnal Pengembangan Teknologi Informasi Dan Ilmu Komputer*, 2(11), 5751–5759.
- Agung, A., & Putri, A. (2023). *Penerapan Data Mining Untuk Mengestimasi Laju Data Mining Usage To Estimate Civil Growth In Denpasar*. 6(1), 37–44.
- Agusta, A. S., Munir, M., & Daffa Rizki, Z. A. (2023). Peramalan Pendapatan dari Penjualan Bawang Merah Menggunakan Metode Regresi Linier Berganda. *Stains (Seminar Nasional Teknologi & Sains)*, 2(1), 383–389.
- Agustina, I. M., & Kartika, I. N. (2018). Pengaruh Tenaga Kerja, Modal dan Bahan Baku terhadap Produksi Industri Kerajinan Patung Kayu di Kecamatan Tegallalang. *E-Jurnal EP UNUD*, 6(7), 1302–1331.
- Ariyanto, A., Nizar, R., Mutryarny, D. E., Lancang, U., & Pekanbaru, K. (2020). *Analisis Faktor-Faktor Yang Mempengaruhi Produksi Kelapa Sawit Rakyat Pola Swadaya Di Kabupaten Kampar-Riau. Gambar 1*.
- Aroni, A., Fitri, H., & Prasetyo, E. (2018). Penerapan Metode Clustering dengan Algoritma K-Means pada Pengelompokan Data Calon Mahasiswa Baru di Universitas Muhammadiyah Yogyakarta (Studi Kasus: Fakultas Kedokteran dan Ilmu Kesehatan, dan Fakultas Ilmu Sosial dan Ilmu Politik). *Semesta Teknik*, 21(1), 60–64.
- Ayuni, G. N., & Fitriyah, D. (2019). Penerapan metode Regresi Linear untuk prediksi penjualan properti pada PT XYZ. *Jurnal Telematika*, 14(2), 79–86.
- Basri, H. (2019). Pemodelan Regresi Berganda Untuk Data Dalam Studi Kecerdasan Emosional. *DIDAKTIKA : Jurnal Kependidikan*, 12(2), 103–116.
- Effendi, M. R., Ahmad, E., Hamidi, Z., & Suhardi, A. A. (2019). *Rancang Bangun Sistem Pendeteksi Manusia Pada Ruangan Menggunakan Raspberry Pi 3 Type B Dan Internet*. 15–16.

- Haryadi, D., Marini Umi Atmaja, D., Rahman Hakim, A., & Suwaryo, N. (2021). Identifikasi Tingkat Resiko Penyakit Stroke Menggunakan Algoritma Regresi Linear Berganda. *Deny Haryadi, SNTEM*, 1(November), 1198–1207.
- Ishaq Hasibuan, Y., Sari Ramadhan, P., Yetri, M., Studi Sistem Informasi, P., & Triguna Dharma, S. (2021). Implementasi Data Mining Untuk Mengelompokkan Korban Penyalahgunaan Narkoba Menggunakan Metode K-Means Clustering Pada BNN Provinsi Sumatera Utara. *Jurnal CyberTech*, 4(8), 1–12.
- Kafil, M., & Industri, F. T. (2019). *Penerapan Metode K-Nearest Neighbors*. 3(2), 59–66.
- Kamaruddin, M. A., Syekh, U. U., & Zawawi, M. H. (2022). *Prediksi Hasil Kelapa Sawit Menggunakan Machine Learning dalam Perspektif Kondisi Cuaca dan Kelembaban Tanah yang Berfluktuasi : Evaluasi Alur Kerja Generik*.
- Nadhirah, F., Nain, M., Hashimah, N., Hassain, A., Abdullah, R., Rahim, F. A., Azinuddin, M., Mokhtar, A., & Mohamad, S. (2022). *algoritma Tinjauan Kerangka Kecerdasan Buatan untuk Mengidentifikasi Prediksi Minyak Sawit yang Paling Efektif*. 1–34.
- Meliati, Peny. (2020). Pengembangan Podcast sebagai Media Suplemen Pembelajaran Berbasis Digital pada Perguruan Tinggi. *Jurnal Sosial Humaniora Terapan*, 2(2), 107–116.
- Panggabean, D. S. O., Buulolo, E., & Silalahi, N. (2020). Penerapan Data Mining Untuk Memprediksi Pemesanan Bibit Pohon Dengan Regresi Linear Berganda. *JURIKOM (Jurnal Riset Komputer)*, 7(1), 56.
- Sidik, & Destia, P. (2018). *Dengan Metode Regresi Linier Berganda*. 2, 61–69.
- Purnomo, E., Najib, A., & Nyura, Y. (2018). Penerapan Metode Trend Moment Untuk Forecast Penjualan Barang di Indomaret. *Prosiding Seminar Ilmu Komputer Dan Teknologi Informasi*, 3(1), 98–102.
- Purwadi, P., Ramadhan, P. S., & Safitri, N. (2019). Penerapan Data Mining Untuk Mengestimasi Laju Pertumbuhan Penduduk Menggunakan Metode Regresi Linier Berganda Pada BPS Deli Serdang. *Jurnal SAINTIKOM (Jurnal Sains Manajemen*

Informatika Dan Komputer), 18(1), 55.

Santosa, B., Suharyanto, S., & Legono, D. (2020). Penerapan Optimasi Parameter pada Metode Exponential Smoothing untuk Perkiraan Debit. In *Media Komunikasi Teknik Sipil* (Vol. 18, Issue 1, pp. 73–79).

Suryanto, A. A. (2019). Penerapan Metode Mean Absolute Error (Mea) Dalam Algoritma Regresi Linear Untuk Prediksi Produksi Padi. *Saintekbu*, 11(1), 78–83.

Syairozi, M. I. (2021). Analisis Kemiskinan Di Sektor Pertanian (Studi Kasus Komoditas Padi di Kabupaten Malang). *Media Ekonomi*, 28(2), 113–128.

Triyanto, E., Sismoro, H., & Laksito, A. D. (2019). Implementasi Algoritma Regresi Linear Berganda Untuk Memprediksi Produksi Padi Di Kabupaten Bantul. *Rabit : Jurnal Teknologi Dan Sistem Informasi Univrab*, 4(2), 66–75.

Yohansyah, W. M., & Lubis, I. (2019). Analisis Produktivitas Kelapa Sawit (*Elaeis guineensis* Jacq.) di PT. Perdana Inti Sawit Perkasa I, Riau. *Buletin Agrohorti*, 2(1), 125.

Yuli Mardi. (2019). Data Mining : Klasifikasi Menggunakan Algoritma C4 . 5 Data mining merupakan bagian dari tahapan proses Knowledge Discovery in Database (KDD) . *Jurnal Edik Informatika*. *Jurnal Edik Informatika*, 2(2), 213–219.

Yunus, A., Akbar, M., & Andri. (2020). *Data Mining Untuk Memprediksi Hasil Produksi Buah Sawit Pada Pt Bumi Sawit Sukses (Bss) Menggunakan Metode K-Nearest Neighbor*. 198–207.

Purnama, Y., Ismail, I., Noviadri, D., Hendriyani, Y., Phong, N. T., & Darmawan, I. (2019). Expert System in Detecting Children's Intelligence using Certainty Factor. Expert System in Detecting Children's Intelligence using Certainty Factor (October 5, 2019). *Journal of Critical Reviews*, 7(1).

Lubis, A. H., Ikhwan, A., & Kan, P. L. E. (2018). Combination of levenshtein distance and rabin-karp to improve the accuracy of document equivalence level. *International Journal of Engineering & Technology*, 7(2.27), 17-21.

Lubis, A. H., Syed Idrus, S. Z., & Sarji, A. (2018). ICT usage amongst lecturers and its impact towards learning process quality. *Jurnal Komunikasi Malaysian Journal of Communication*, 34(1), 284-299.

UNIVERSITAS MEDAN AREA ~~STUDIAN S &~~ Lubis, A. H. (2022). A Survey on Online Computer Practicum during the

- COVID-19 Pandemic: Students' Perceptions. *Int. J. Res. Rev*, 9(9), 372-379.
- Lubis, A. H., & Lubis, Z. (2020). Parent's perceptions on e-learning during COVID-19 pandemic in Indonesia. *Journal of critical reviews*, 7(18), 3599-3607.
- Rahman, S., & Dafitri, H. (2022). Pengembangan Convolutional Neural Network untuk Klasifikasi Ketersediaan Ruang Parkir. *Explorer*, 2(1), 1-6.
- Rahman, S., & Ulfayani, M. (2017). Perancangan Aplikasi Identifikasi Biometrika Telapak Tangan Menggunakan Metode Freeman Chain Code. *CESS (Journal of Computer Engineering, System and Science)*, 2(2), 64-73.
- Rahman, S., & Dafitri, H. (2019). Aplikasi Simulasi Deteksi Lokasi Parkir Kosong Menggunakan Ekstraksi Ciri Objek. *InfoTekJar: Jurnal Nasional Informatika dan Teknologi Jaringan*, 4(1), 99-104.
- Rahman, S., Liza, R., Aulia, R., & Dafitri, H. (2022). Pelatihan Youtuber Pemula Bagi Remaja STM Al-Muhajirin Rumah Pondok 6. *Jurnal ABDIMAS Budi Darma*, 2(2), 155-159.
- Shidqi, S., Rahman, S., & Sembiring, A. (2022). Rancang Bangun Miniatur Sistem Pemantauan Kondisi Lahan Pertanian Dengan Sms Gateway Berbasis Arduino. *Jurnal Indonesia Sosial Teknologi: p-ISSN*, 2723, 6609.
- Zuhanda, M. K. (2013). Optimasi Program Linier Pecahan Dengan Fungsi Tujuan Berkoefisien Interval (Doctoral dissertation, Universitas Sumatera Utara).
- Zuhanda, M. K., Efendi, M., & Mardiningsih, M. (2016). OPTIMIZATION LINEAR FRACTIONAL PROGRAMMING PROBLEMS WITH INTERVAL COEFFICIENTS IN THE OBJECTIVE FUNCTION. *Bulletin of Mathematics*, 8(01), 55-68.
- Zuhanda, M. K. (2017). TRANSFORMASI LINIER UNTUK PERSOALAN PROGRAM KUADRATIK NOL-SATU. *Educational Building: Jurnal Pendidikan Teknik Bangunan dan Sipil*, 3(2 DESEMBER), 68-72.
- Harahap, M. K., & Khairina, N. (2017). Pencarian Jalur Terpendek dengan Algoritma Dijkstra. *Sinkron: jurnal dan penelitian teknik informatika*, 2(2), 18-23.
- Khairina, N. (2016). Analisis Perbandingan Metode Steganografi Two Sided Side Match Dengan Four Sided Side Match Pada Citra Multilayer TIFF (Doctoral dissertation).

LAMPIRAN

1. Data Hasil Prediksi Keseluruhan

Berikut adalah tabel hasil pengujian dari tahun 2019 sampai tahun 2022 :

**Tabel 5. 1 Seluruh Hasil Pengujian Tahun 2019
untuk Memprediksi Tahun 2020**

No.	Date	Data_Aktual	Data_Prediksi
4	2019-10-03	25140	32297
5	2019-09-26	5872	-2907
6	2019-08-17	33982	33578
11	2019-06-30	25724	35979
13	2019-02-07	9914	3779
17	2019-12-31	11477	1314
18	2019-02-22	64098	74966
41	2019-08-31	36907	35718
62	2019-10-15	30716	36027
65	2019-08-24	23791	27790
67	2019-11-08	8076	3229
78	2019-10-27	53588	52415
81	2019-09-27	69497	65971
82	2019-11-08	41696	48819
89	2019-06-02	51466	55107
90	2019-03-20	33802	33053
95	2019-09-10	30363	31791
98	2019-10-08	9205	12025
102	2019-10-28	93526	80799
106	2019-02-15	13990	14311
107	2019-01-30	10193	11698
120	2019-07-22	50179	52713
124	2019-12-08	34535	39333
125	2019-04-21	44112	43834
131	2019-03-06	9793	8922
132	2019-05-04	101282	88832
138	2019-07-30	16591	36981
150	2019-03-15	28150	34378
160	2019-12-19	6990	3378
161	2019-03-14	86190	61226
167	2019-03-21	10930	246
171	2019-01-17	34400	39965
173	2019-05-11	46690	49564

175	2019-12-12	74790	77841
179	2019-10-18	12730	68840
180	2019-06-07	58360	70150
189	2019-03-29	15885	22121
191	2019-09-24	14271	39735
222	2019-07-14	8641	6683
223	2019-06-24	70967	55898
230	2019-04-10	21705	40962
233	2019-12-24	56786	57199
234	2019-06-06	14507	8449
236	2019-05-31	100396	87367
242	2019-02-05	58098	61447
248	2019-05-19	32316	37749
249	2019-01-02	31773	35924
261	2019-03-29	10557	7726
267	2019-08-28	83174	78295
268	2019-01-06	60108	63880
277	2019-04-30	41001	35577
278	2019-08-09	27647	24565
289	2019-01-24	44097	47050
293	2019-08-21	7723	667
295	2019-03-06	33763	36550
296	2019-12-15	18629	27321
299	2019-10-07	56559	65255
304	2019-06-18	62252	58526
312	2019-04-24	41036	37931
321	2019-12-21	43504	45841
324	2019-08-09	9623	13038
333	2019-04-28	69822	62849
336	2019-10-27	57473	57214
337	2019-05-17	39254	37541
340	2019-11-28	21676	18945
341	2019-02-21	30186	37454
342	2019-08-31	27387	33099
346	2019-06-08	6431	-914
347	2019-04-07	56533	50273
348	2019-12-16	7433	909
352	2019-01-31	26932	34412
358	2019-03-27	14582	14646
359	2019-05-31	43184	48489
363	2019-06-23	96022	78177
369	2019-06-04	50517	53331
Total		2811191.21	2886145.037

**Tabel 5. 2 Seluruh Hasil Pengujian Tahun 2020
untuk Memprediksi Tahun 2021**

No.	Date	Data_Aktual	Data_Prediksi
4	2020-12-20	25824	33287
5	2020-02-02	5997	-5105
6	2020-10-23	39968	37879
11	2020-09-23	31117	41039
13	2020-07-10	8161	-4539
17	2020-08-29	12562	5516
18	2020-07-26	67891	73725
41	2020-11-28	47436	45972
62	2020-07-01	38491	41085
65	2020-07-11	21498	26018
67	2020-03-18	10035	12962
78	2020-05-04	34391	37871
81	2020-04-08	62415	60756
82	2020-11-26	46682	53401
89	2020-03-10	46765	50094
90	2020-05-28	33161	33571
95	2020-07-11	36534	35744
98	2020-11-01	7457	2598
102	2020-06-26	80230	74903
106	2020-10-09	11841	6361
107	2020-06-02	9338	7728
120	2020-07-05	49175	51008
124	2020-03-06	41749	43386
125	2020-01-23	34896	38109
131	2020-03-25	9257	6386
132	2020-06-08	82512	82211
138	2020-05-08	13398	26269
150	2020-12-15	35027	42177
160	2020-01-24	6990	13141
161	2020-11-25	86190	68452
167	2020-05-17	10930	892
171	2020-05-13	34400	37051
173	2020-10-24	46690	54133
175	2020-12-04	74790	81805
179	2020-01-03	12730	61143
180	2020-07-16	58360	71174
189	2020-09-18	22906	27736
191	2020-04-17	12622	31077

222	2020-06-26	9053	9608
223	2020-05-11	70554	56272
230	2020-04-29	21730	41970
233	2020-08-28	57115	58222
234	2020-02-27	14299	9005
236	2020-08-28	96378	87946
242	2020-10-26	54994	56667
248	2020-02-14	33481	38284
249	2020-09-06	32074	36167
261	2020-10-09	10961	6622
267	2020-01-09	78404	73538
268	2020-01-09	60291	64332
277	2020-04-05	37919	33101
278	2020-09-25	19940	17543
289	2020-11-10	48215	51418
293	2020-05-15	8301	4719
295	2020-05-28	31601	34959
296	2020-08-29	18928	31042
299	2020-03-12	50905	60175
304	2020-01-26	59440	53879
312	2020-04-23	41703	38447
321	2020-07-31	43504	46622
324	2020-02-12	9623	13972
333	2020-05-06	70869	63983
336	2020-04-30	64748	59594
337	2020-07-21	42129	39227
340	2020-01-04	19596	17109
341	2020-12-20	34693	38898
342	2020-12-26	36515	38554
346	2020-06-10	7268	4269
347	2020-12-24	51119	46787
348	2020-06-05	7576	2184
352	2020-08-16	27956	35598
358	2020-04-24	13278	9294
359	2020-10-22	46142	52307
363	2020-12-06	85468	72857
369	2020-10-12	48459	51263
Total		2775638	2863450

**Tabel 5. 3 Hasil Pengujian Tahun 2021
untuk Memprediksi Tahun 2022**

No.	Date	Data_Aktual	Data_Prediksi
4	2021-01-17	21510	26960
5	2021-09-05	6908	2476
6	2021-07-30	39244	37089
11	2021-02-22	31585	41036
13	2021-08-12	9271	859
17	2021-07-05	12035	4399
18	2021-03-12	75261	79210
41	2021-10-21	50231	48879
62	2021-05-22	43776	43250
65	2021-11-14	21305	25737
67	2021-01-14	9456	11752
78	2021-07-03	30789	34925
81	2021-05-19	67563	63791
82	2021-01-19	42998	49914
89	2021-04-17	53050	56524
90	2021-01-17	30370	28510
95	2021-04-22	37040	36082
98	2021-01-29	6952	1475
102	2021-01-29	74910	70339
106	2021-04-28	11223	5046
107	2021-07-12	8885	6352
120	2021-11-23	49890	51287
124	2021-03-11	34535	38796
125	2021-08-28	34269	37209
131	2021-06-07	8405	4565
132	2021-01-09	77250	77589
138	2021-08-29	13181	25882
150	2021-12-05	32012	39139
160	2021-09-27	6990	12157
161	2021-08-12	86190	61170
167	2021-09-02	10930	-362
171	2021-07-12	34400	37811
173	2021-11-02	46690	50969
175	2021-08-01	74790	78092
179	2021-02-10	12730	71122
180	2021-05-04	58360	68331
189	2021-04-30	21742	26380
191	2021-07-07	12146	29566
222	2021-05-09	8346	7963

223	2021-01-01	66085	52825
230	2021-10-25	21349	40317
233	2021-07-24	55518	55283
234	2021-01-14	13572	7706
236	2021-03-03	93787	84319
242	2021-11-10	53267	54050
248	2021-10-13	29400	34224
249	2021-12-15	26370	33263
261	2021-08-12	10557	5431
267	2021-10-11	76560	70257
268	2021-02-26	62645	64846
277	2021-08-13	38380	34004
278	2021-04-17	19940	17705
289	2021-05-12	44676	48097
293	2021-08-24	7618	2953
295	2021-07-26	25529	31025
296	2021-01-23	18410	29458
299	2021-03-22	52500	57628
304	2021-08-23	54863	50766
312	2021-06-04	34526	33698
321	2021-09-18	42930	46268
324	2021-04-01	7772	4903
333	2021-09-04	53816	51018
336	2021-03-22	57760	55976
337	2021-12-14	47934	46314
340	2021-01-15	22573	20356
341	2021-11-27	29330	35917
342	2021-06-08	34546	37129
346	2021-12-22	7012	4465
347	2021-07-02	51495	47001
348	2021-05-21	7142	2331
352	2021-11-07	28231	36036
358	2021-01-09	13793	13721
359	2021-09-23	42359	48568
363	2021-03-07	81037	69407
369	2021-06-23	46855	48875
Total		2687355	2768412

**Tabel 5. 4 Hasil Pengujian Tahun 2022
untuk Memprediksi Tahun 2023**

No.	Date	Data_Aktual	Data_Prediksi
4	2022-09-03	19320	22996
5	2022-12-19	6390	1478
6	2022-10-04	36885	34861
11	2022-01-26	30421	40172
13	2022-03-23	8623	-1555
17	2022-04-07	11340	2107
18	2022-05-12	68280	74852
41	2022-08-02	46859	45642
62	2022-08-15	39810	39301
65	2022-12-14	20340	23500
67	2022-12-30	7536	2990
78	2022-12-03	26770	31806
81	2022-07-04	71190	66434
82	2022-10-21	40010	46651
89	2022-02-25	53050	56424
90	2022-10-06	30370	28766
95	2022-07-27	29950	30816
98	2022-09-17	6505	-475
102	2022-03-29	74910	69120
106	2022-01-20	9724	-464
107	2022-02-04	7941	1067
120	2022-07-09	49890	51234
124	2022-11-30	21610	30284
125	2022-10-27	28120	33342
131	2022-11-14	8405	4966
132	2022-02-21	77250	76383
138	2022-10-15	14590	27679
150	2022-03-07	26730	32975
160	2022-04-09	6990	3123
161	2022-02-25	86190	63843
167	2022-01-20	10930	152
171	2022-03-03	34400	37552
173	2022-07-04	46690	50817
175	2022-08-22	74790	77907
179	2022-09-05	12730	71092
180	2022-09-06	58360	71069
189	2022-10-02	19240	24065
191	2022-04-13	14030	39211
222	2022-05-27	7810	5487

223	2022-07-04	66950	52387
230	2022-08-12	21470	40613
233	2022-04-12	54820	54867
234	2022-08-20	14160	8865
236	2022-12-25	92700	83572
242	2022-05-09	51180	53573
248	2022-07-31	24100	30313
249	2022-01-18	26370	32476
261	2022-03-14	9871	2954
267	2022-10-11	80270	72836
268	2022-05-31	59330	61433
277	2022-09-04	38380	34182
278	2022-02-14	19940	17315
289	2022-01-09	42900	45176
293	2022-09-12	7117	456
295	2022-05-24	21840	28093
296	2022-04-18	18440	27375
299	2022-03-17	52500	57331
304	2022-04-02	56160	51055
312	2022-12-27	39935	36196
321	2022-02-24	43448	46244
324	2022-08-26	7572	2836
333	2022-07-06	46010	44831
336	2022-07-22	54530	52822
337	2022-07-30	44800	43260
340	2022-03-31	19627	16909
341	2022-09-04	24470	33452
342	2022-11-29	21380	28736
346	2022-03-13	6521	1901
347	2022-01-26	43550	40268
348	2022-03-15	6979	786
352	2022-01-26	24270	33068
358	2022-03-02	13740	14195
359	2022-06-02	34140	39461
363	2022-04-14	100210	78637
369	2022-04-24	45320	46032
Total		2579978	2632177

2. Source Code

Panggil libraries yang digunakan

```
import pandas as pd # sebagai dataframe
import matplotlib.pyplot as plt # sebagai visualisasi
import seaborn as sns # sebagai visualisasi
from sklearn.model_selection import train_test_split # sebagai pembagi dataset untuk
train dan test
from sklearn.linear_model import LinearRegression # metode yang digunakan
```

```
# metrik yang digunakan
```

```
from sklearn.metrics import mean_absolute_error
from google.colab import drive
drive.mount('/content/drive')
```

```
# Membaca dan memanggil dataset
```

```
df = pd.read_csv('/content/drive/MyDrive/dataReg.csv')
df.head()
```

Preprocessing Data

```
# Melihat dimensi data
```

```
df.shape
```

```
#Menghitung jumlah data kosong
```

```
df.isna().sum()
```

```
#Melihat tipe data dari tiap variabel
```

```
df.dtypes
```

```
#Melihat Statistic dari dataframe
```

```
df.describe()
```

```
# membuat korelasi antar variabel menggunakan heatmap
```

```
corr = df.corr()
```

```
plt.figure(figsize=(15,12))
```

```
palette = sns.diverging_palette(20, 220, n=256)
```

```
sns.heatmap(corr, annot=True, fmt=".2f", cmap=palette, center=0,
            square=True, linewidths=.5, cbar_kws={"shrink": .5})
```

```
plt.title("Correlation Matrix",size=15, weight="bold")
```

Filter yang digunakan di tahun 2019

```
#Membaca dan Memanggil Data Tahun 2019
```

```
data_19 = df.loc[(df['date'] >= '2019-01-01') & (df['date'] <= '2019-12-31')]
```

```
data_19.head()
```

```
# visualisasi persebaran data input(luas_ha) dan output(hasil_produksi)
data_19.plot(kind='scatter', x='luas_ha', y='hasil_produksi')

# visualisasi persebaran data input(jml_pkk) dan output(hasil_produksi)
data_19.plot(kind='scatter', x='jml_pkk', y='hasil_produksi')

#Grafik antara luas lahan dengan hasil produksi
plt.figure(figsize=(12, 7))
sns.lineplot(data=data_19, x='luas_ha', y='hasil_produksi')
plt.show()

#Grafik antara jumlah pokok dengan hasil produksi
plt.figure(figsize=(12, 7))
sns.lineplot(data=data_19, x='jml_pkk', y='hasil_produksi')
plt.show()

# menentukan X dan Y untuk dilakukan pembagian dataset berupa train 80% dan test 20%
X = data_19[['luas_ha', 'jml_pkk', 'tdn/pkk', 'bjr']] # variabel input
y = data_19['hasil_produksi'] # variabel output
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2, random_state=1)

# menampilkan jumlah dari pembagian dataset train dan test
print('Jumlah data train = ', X_train.shape)
print('Jumlah data test = ', X_test.shape)

# Membuat model regresi dan melatih berdasarkan x_train dan y_train
linreg = LinearRegression().fit(X_train, y_train)

# menampilkan coefficient dan intercept
print('Koefisien : ', linreg.coef_)
print('Intercept : ', linreg.intercept_)

# menentukan y train pada model yang sudah dibuat
y_train_pred = linreg.predict(X_train)

# menentukan y test pada model yang sudah dibuat
y_test_pred = linreg.predict(X_test)

# Hasil Metrik dari Mean Absolute Error
print("MAE Training : ", mean_absolute_error(y_train, y_train_pred))
print("MAE Testing : ", mean_absolute_error(y_test, y_test_pred))

# menampilkan data aktual dan data hasil prediksi
y_pred = linreg.predict(X_test)
dataframe = pd.DataFrame({'Data_Aktual' : y_test, 'Data_Prediksi' : y_pred})
```

```
dataframe
```

```
# Menampilkan visualisasi dari data aktual dan data hasil prediksi  
sns.regplot(x=dataframe['Data_Aktual'], y=dataframe['Data_Prediksi'])
```

```
#Memasukkan 25 data teratas variabel graph  
graph = dataframe.head(25)
```

```
#Menampilkan 25 data teratas  
graph.plot(kind='bar', legend=True, figsize=(12,7))  
plt.title('Top 25 teratas', size=14)
```

Filter Data Tahun 2020

```
#Membaca dan memanggil data tahun 2020  
dt_20 = df.loc[(df['date'] >= '2020-01-01') & (df['date'] <= '2020-12-31')]  
data_20 = dt_20.reset_index(drop=True)  
data_20.head()
```

```
# visualisasi persebaran data input(luas_ha & jml_pkk) dan output(hasil_produksi)  
data_20.plot(kind='scatter', x='luas_ha', y='hasil_produksi')
```

```
# visualisasi persebaran data input(jml_pkk) dan output(hasil_produksi)  
data_20.plot(kind='scatter', x='jml_pkk', y='hasil_produksi')
```

```
#Grafik antara luas lahan dengan hasil produksi  
plt.figure(figsize=(12, 7))  
sns.lineplot(data=data_20, x='luas_ha', y='hasil_produksi')  
plt.show()
```

```
# menentukan X dan Y untuk dilakukan pembagian dataset berupa train 80% dan test 20%
```

```
X = data_20[['luas_ha', 'jml_pkk', 'tdn/pkk', 'bjr']] # variabel input  
y = data_20['hasil_produksi'] # variabel output  
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2, random_state=1)
```

```
# Membuat model regresi dan melatih berdasarkan x_train dan y_train  
linreg = LinearRegression().fit(X_train,y_train)
```

```
# menampilkan coefficient dan intercept  
print('Koefisien : ', linreg.coef_)  
print('Intercept : ', linreg.intercept_)
```

```
# menentukan y train pada model yang sudah dibuat  
y_train_pred = linreg.predict(X_train)
```

```
# menentukan y test pada model yang sudah dibuat
```

```

y_test_pred = linreg.predict(X_test)

# Hasil Metrik dari Mean Absolute Error
print("MAE Training : ", mean_absolute_error(y_train,y_train_pred))
print("MAE Testing : ", mean_absolute_error(y_test,y_test_pred))

# menampilkan data aktual dan data hasil prediksi
y_pred = linreg.predict(X_test)
dataframe = pd.DataFrame({'Data_Aktual' : y_test, 'Data_Prediksi' : y_pred})
dataframe

# Menampilkan visualisasi dari data aktual dan data hasil prediksi
sns.regplot(x=dataframe['Data_Aktual'], y=dataframe['Data_Prediksi'])

#Memasukkan 25 data teratas variabel graph
graph = dataframe.head(25)

#Menampilkan 25 data teratas
graph.plot(kind='bar', legend=True,figsize=(12,7))
plt.title('Top 25 teratas', size=14)

Filter data berdasarkan tahun 2021
#Membaca dan memanggil data tahun 2021
dt_21 = df.loc[(df['date'] >= '2021-01-01') & (df['date'] <= '2021-12-31')]
data_21 = dt_21.reset_index(drop=True)
data_21.head()

# visualisasi persebaran data input(luas_ha & jml_pkk) dan output(hasil_produksi)
data_21.plot(kind='scatter', x='luas_ha', y='hasil_produksi')

# visualisasi persebaran data input(jml_pkk) dan output(hasil_produksi)
data_21.plot(kind='scatter', x='jml_pkk', y='hasil_produksi')

#Grafik antara luas lahan dengan hasil produksi
plt.figure(figsize=(12, 7))
sns.lineplot(data=data_21, x='luas_ha', y='hasil_produksi')
plt.show()

#Grafik antara jumlah pokok dengan hasil produksi
plt.figure(figsize=(12, 7))
sns.lineplot(data=data_21, x='jml_pkk', y='hasil_produksi')
plt.show()

# menentukan X dan Y untuk dilakukan pembagian dataset berupa train 80% dan test
20%
X = data_21[['luas_ha', 'jml_pkk', 'tdn/pkk', 'bjr']] # variabel input

```

```
y = data_21['hasil_produksi'] # variabel output
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2, random_state=1)

# Membuat model regresi dan melatih berdasarkan x_train dan y_train
linreg = LinearRegression().fit(X_train,y_train)

# menampilkan coeficient dan intercept
print('Koefisien : ', linreg.coef_)
print('Intercept :', linreg.intercept_)

# menentukan y train pada model yang sudah dibuat
y_train_pred = linreg.predict(X_train)

# menentukan y test pada model yang sudah dibuat
y_test_pred = linreg.predict(X_test)

# Hasil Metrik dari Mean Absolute Error
print("MAE Training : ", mean_absolute_error(y_train,y_train_pred))
print("MAE Testing : ", mean_absolute_error(y_test,y_test_pred))

# menampilkan data aktual dan data hasil prediksi
y_pred = linreg.predict(X_test)
dataframe = pd.DataFrame({'Data_Aktual' : y_test, 'Data_Prediksi' : y_pred})
dataframe

# Menampilkan visualisasi dari data aktual dan data hasil prediksi
sns.regplot(x=dataframe['Data_Aktual'], y=dataframe['Data_Prediksi'])

# Memasukkan 25 data teratas variabel graph
graph = dataframe.head(25)

#Menampilkan 25 data teratas
graph.plot(kind='bar', legend=True,figsize=(12,7))
plt.title('Top 25 teratas', size=14)

Filter data tahun 2022
##Membaca dan memanggil data tahun 2022
dt_22 = df.loc[(df['date'] >= '2022-01-01') & (df['date'] <= '2022-12-31')]
data_22 = dt_22.reset_index(drop=True)
data_22.head()

# visualisasi input(luas_ha) dan output(hasil_produksi)
data_22.plot(kind='scatter', x='luas_ha', y='hasil_produksi')
```

```
# visualisasi input(tdn/pkk) dan output(hasil_produksi)
data_22.plot(kind='scatter', x='jml_pkk', y='hasil_produksi')

# Grafik antara jumlah pokok dengan hasil produksi
plt.figure(figsize=(12, 7))
sns.lineplot(data=data_22, x='luas_ha', y='hasil_produksi')
plt.show()

# Grafik antara jumlah pokok dengan hasil produksi
plt.figure(figsize=(12, 7))
sns.lineplot(data=data_22, x='jml_pkk', y='hasil_produksi')
plt.show()

# menentukan X dan Y untuk dilakukan pembagian dataset berupa train 80% dan test 20%
X = data_22[['luas_ha', 'jml_pkk', 'tdn/pkk', 'bjr']] # variabel input
y = data_22['hasil_produksi'] # variabel output
X_train, X_test, y_train, y_test = train_test_split(X,y,test_size=0.2, random_state=1)

# menampilkan hasil dari pembagian dataset train dan test
print('Jumlah data train = ',X_train.shape)
print('Jumlah data test = ',X_test.shape)

# Membuat model regresi dan melatih berdasarkan x_train dan y_train
linreg = LinearRegression().fit(X_train,y_train)

# menampilkan coeficient dan intercept
print(linreg.coef_)
print(linreg.intercept_)

# menentukan y train pada model yang sudah dibuat
y_train_pred = linreg.predict(X_train)

# menentukan y test pada model yang sudah dibuat
y_test_pred = linreg.predict(X_test)

# Hasil Metrik dari Mean Absolute Error
print("MAE Training : ", mean_absolute_error(y_train,y_train_pred))
print("MAE Testing : ", mean_absolute_error(y_test,y_test_pred))

# menampilkan data aktual dan data hasil prediksi
y_pred = linreg.predict(X_test)
dataframe = pd.DataFrame({'Data_Aktual' : y_test, 'Data_Prediksi' : y_pred})
dataframe

# Menampilkan visualisasi dari data aktual dan data hasil prediksi
sns.regplot(x=dataframe['Data_Aktual'], y=dataframe['Data_Prediksi'])
```

```
# Memasukkan 25 data teratas variabel graph
graph = dataframe.head(25)

# Menampilkan 25 data teratas
graph.plot(kind='bar', legend=True,figsize=(12,7))
plt.title('Top 25 teratas', size=14)
```



3. Hasil Plagiat



The image shows a Turnitin Similarity Report for a document titled "188160030_Redi Sebastian Situmorang.pdf". The report includes the following details:

- PAPER NAME:** 188160030_Redi Sebastian Situmorang.pdf
- AUTHOR:** Redi Sebastian Situmorang
- WORD COUNT:** 8545 Words
- CHARACTER COUNT:** 51210 Characters
- PAGE COUNT:** 55 Pages
- FILE SIZE:** 911.0KB
- SUBMISSION DATE:** Nov 9, 2023 8:04 PM GMT+7
- REPORT DATE:** Nov 9, 2023 8:05 PM GMT+7

The report indicates a **20% Overall Similarity**, which is the combined total of all matches, including overlapping sources, for each database. The breakdown of matches is as follows:

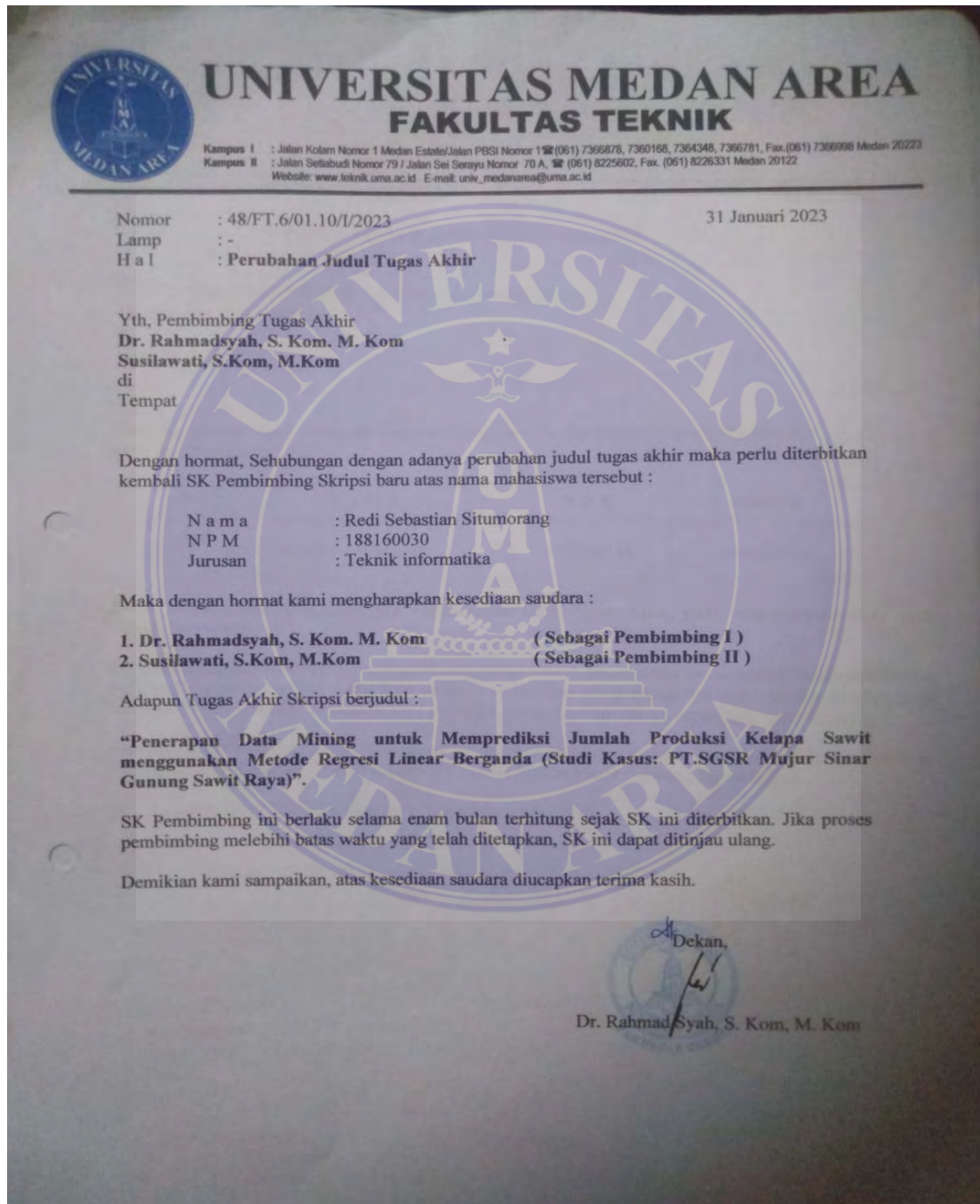
- 16% Internet database
- 10% Publications database
- Crossref database
- Crossref Posted Content database
- 13% Submitted Works database

Additionally, the report notes that certain matches were **Excluded from Similarity Report**, specifically **Small Matches (Less than 10 words)**.

The background of the report features a large, faint watermark of the Universitas Medan Area logo, which includes a central emblem with a book and a star, surrounded by the text "UNIVERSITAS MEDAN AREA".

At the bottom right of the report, there is a link labeled "Summary".

4. Surat Keterangan Dosen Pembimbing



UNIVERSITAS MEDAN AREA
FAKULTAS TEKNIK

Kampus I : Jalan Kolam Nomor 1 Medan Estate/Jalan PBSI Nomor 1 (061) 7366878, 7360168, 7364348, 7366781, Fax. (061) 7366998 Medan 20223
Kampus II : Jalan Selisabadi Nomor 79 / Jalan Sei Serayu Nomor 70 A, (061) 8225602, Fax. (061) 8226331 Medan 20122
Website: www.teknik.uma.ac.id E-mail: univ_medanarea@uma.ac.id

Nomor : 48/FT.6/01.10/I/2023
Lamp : -
Hal : **Perubahan Judul Tugas Akhir** 31 Januari 2023

Yth, Pembimbing Tugas Akhir
Dr. Rahmadsyah, S. Kom. M. Kom
Susilawati, S. Kom, M. Kom
di
Tempat

Dengan hormat, Sehubungan dengan adanya perubahan judul tugas akhir maka perlu diterbitkan kembali SK Pembimbing Skripsi baru atas nama mahasiswa tersebut :

Nama : Redi Sebastian Situmorang
N P M : 188160030
Jurusan : Teknik informatika

Maka dengan hormat kami mengharapkan kesediaan saudara :

1. **Dr. Rahmadsyah, S. Kom. M. Kom** (Sebagai Pembimbing I)
2. **Susilawati, S. Kom, M. Kom** (Sebagai Pembimbing II)

Adapun Tugas Akhir Skripsi berjudul :

“Penerapan Data Mining untuk Memprediksi Jumlah Produksi Kelapa Sawit menggunakan Metode Regresi Linear Berganda (Studi Kasus: PT.SGSR Mujur Sinar Gunung Sawit Raya)”.

SK Pembimbing ini berlaku selama enam bulan dihitung sejak SK ini diterbitkan. Jika proses pembimbing melebihi batas waktu yang telah ditetapkan, SK ini dapat ditinjau ulang.

Demikian kami sampaikan, atas kesediaan saudara diucapkan terima kasih.

Dekan,
Dr. Rahmad Syah, S. Kom, M. Kom

5. Surat Riset dan Pengambilan Data Tugas Akhir

UNIVERSITAS MEDAN AREA
FAKULTAS TEKNIK

Kampus I : Jalan Kolam Nomor 1 Medan Estate/Jalan PBSI Nomor 1 ☎ (061) 7366876, 7360166, 7364348, 7366781, Fax (061) 7366996 Medan 20223
Kampus II : Jalan Setiabudi Nomor 79 / Jalan Sei Serayu Nomor 70 A, ☎ (061) 8225602, Fax. (061) 8226331 Medan 20122
Website: www.teknik.uma.ac.id E-mail: univ_medanarea@uma.ac.id

Nomor : 291/FT.6/01.10/X/2022
Lamp : -
Hal : Penelitian Dan Pengambilan Data Tugas Akhir

5 Oktober 2022

Yth. Pimpinan PT. SGSR Mujur Sinar Gunung Sawit Raya
Bajamas Kec. Sirandorung
Di
Tapanuli Tengah

Dengan hormat,
Kami mohon kesediaan Bapak/Ibu berkenan untuk memberikan izin dan kesempatan kepada mahasiswa kami tersebut dibawah ini :

NO	N A M A	N P M	PRODI
1	Redi Sebastian Situmorang	188160030	informatika

Untuk melaksanakan Penelitian dan Pengambilan Data Tugas Akhir pada perusahaan/Instansi yang Bapak/Ibu Pimpin.

Perlu kami jelaskan bahwa Pengambilan Data tersebut adalah semata-mata untuk tujuan ilmiah dan Skripsi yang merupakan salah satu syarat bagi mahasiswa tersebut untuk mengikuti ujian sarjana lengkap pada Fakultas Teknik Universitas Medan Area dan tidak untuk dipublikasikan, dengan judul penelitian :

Penerapan Data Mining untuk Memprediksi Jumlah Produksi Kelapa Sawit menggunakan Metode Regresi Linear Berganda (Studi Kasus: PT.SGSR Mujur Sinar Gunung Sawit Raya)

Atas perhatian dan kerja sama yang baik diucapkan terima kasih.

Dekan,
Dr. Rahmad Satri, S. Kom, M. Kom

Tembusan :
1. Ka. BAMAI
2. Mahasiswa
3. File

6. Surat Selesai Riset

