

Rahmadani Ritonga
Ibnu Rasyid Munthe, S.T., M.Kom.
Angga Putra Juledi, S.Kom., M.Kom.
Masrizal, S.Kom., M.Kom.

OPTIMALISASI KINERJA PEGAWAI PERTANIAN

Studi Kasus Penggunaan
Algoritma Regresi Linear



OPTIMALISASI KINERJA PEGAWAI PERTANIAN

Studi Kasus Penggunaan
Algoritma Regresi Linear

Rahmadani Ritonga
Ibnu Rasyid Munthe, S.T M.Kom
Angga Putra Juledi, S, Kom., M.Kom
Masrizal, S.Kom, M.Kom

**OPTIMALISASI KINERJA PEGAWAI PERTANIAN:
Studi Kasus Penggunaan Algoritma Regresi Linear**

Ditulis oleh :

Rahmadani Ritonga

Ibnu Rasyid Munthe, S.T., M.Kom.

Angga Putra Juledi, S.Kom., M.Kom.

Masrizal, S.Kom., M.Kom.

Diterbitkan, dicetak, dan didistribusikan oleh

PT. Literasi Nusantara Abadi Grup

Perumahan Puncak Joyo Agung Residence Kav. B11 Merjosari

Kecamatan Lowokwaru Kota Malang 65144

Telp : +6285887254603, +6285841411519

Email: literasinusantaraofficial@gmail.com

Web: www.penerbitlitnus.co.id

Anggota IKAPI No. 340/JTI/2022



Hak Cipta dilindungi oleh undang-undang. Dilarang mengutip atau memperbanyak baik sebagian ataupun keseluruhan isi buku dengan cara apa pun tanpa izin tertulis dari penerbit.

Cetakan I, Agustus2024

Perancang sampul: Rosyiful Aqli

Penata letak: Noufal Fahriza

ISBN : 978-623-519-100-3

viii + 70 hlm. ; 15,5x23 cm.

©Agustus2024

PRAKATA

Alhamdulillahirabbil'alamin, saya mengucapkan syukur yang terdalam kepada Allah SWT atas berkat dan petunjuk-Nya, yang telah memungkinkan saya menyelesaikan tesis ini tepat waktu. Saya juga mengucapkan salam dan hormat kepada Rasulullah SAW, yang telah berjasa besar dengan membukakan jalan dalam perkembangan ilmu pengetahuan seperti sekarang ini. Buku ini adalah hasil dari penelitian saya di Kantor Dinas Pertanian Kabupaten Labuhanbatu, dengan judul "**OPTIMALISASI KINERJA PEGAWAI PERTANIAN: Studi Kasus Penggunaan Algoritma Regresi Linear**". Saya ingin mengucapkan terima kasih yang sedalam-dalamnya kepada semua pihak yang telah memberikan dukungan moril maupun materil sehingga buku ini dapat diselesaikan. Saya menyadari bahwa hasil dari buku ini masih belum sempurna. Oleh karena itu, saya sangat mengharapkan saran-saran dan kritikan dari pembaca demi untuk kesempurnaan buku ini. Semoga buku ini dapat menjadi ilmu pengetahuan yang bermanfaat bagi pembaca semua. Amin yaa robbal'alamin

Rantauprapat, juli 2024

Rahmadani Ritonga



PENGANTAR

Kinerja pegawai merupakan salah satu faktor kunci dalam menentukan keberhasilan suatu organisasi, termasuk di sektor pertanian. Di era teknologi informasi yang berkembang pesat, analisis data telah menjadi alat penting dalam memahami dan meningkatkan kinerja ini. Buku ini, “Optimalisasi Kinerja Pegawai Pertanian: Studi Kasus Penggunaan Algoritma Regresi Linear,” hadir sebagai upaya untuk menjembatani ilmu pengetahuan dan praktik di lapangan.

Buku ini lahir dari kebutuhan mendesak untuk meningkatkan efisiensi dan efektivitas pegawai Dinas Pertanian dalam menghadapi tantangan globalisasi dan perkembangan teknologi. Algoritma regresi linear, sebagai salah satu teknik machine learning yang paling mendasar dan luas penggunaannya, dipilih dalam penelitian ini untuk mengidentifikasi dan mengoptimalkan faktor-faktor yang mempengaruhi kinerja pegawai. Melalui pendekatan ini, buku ini tidak hanya menyajikan teori dan konsep dasar, tetapi juga memaparkan aplikasi praktis yang dapat diterapkan di berbagai konteks organisasi pertanian.

Kami berharap buku ini dapat menjadi referensi berharga bagi akademisi, praktisi, dan pengambil kebijakan di sektor pertanian. Dengan pemahaman yang lebih baik tentang bagaimana data dapat digunakan untuk mengoptimalkan kinerja, kami percaya bahwa inovasi dan peningkatan berkelanjutan dapat dicapai.

Akhir kata, kami mengucapkan terima kasih kepada semua pihak yang telah berkontribusi dalam penyusunan buku ini. Semoga buku ini dapat memberikan manfaat yang signifikan dan mendorong penelitian lebih lanjut dalam bidang optimalisasi kinerja pegawai.

Selamat membaca!

DAFTAR ISI

Prakata	iii
Pengantar.....	v
Daftar Isi.....	vii

BAB I

PENDAHULUAN	1
--------------------------	----------

BAB II

MACHINE LEARNING.....	5
------------------------------	----------

Pengantar Machine Learning	5
Artificial Intelligence.....	6
Machine Learning	9
Data Mining	10
Data Science	11

BAB III

KLASIFIKASI REGRESI LINIER.....	15
--	-----------

Model Klasifikasi	15
Metode Regresi Linear.....	16
Uji Performa.....	17
Mean Squared Error (MSE)	18
Mean Absolute Error (MAE).....	19
Koefisien Determinasi (R^2).....	20
Python Sebagai Alat Bantu Program.....	20

BAB IV

DESAIN KAJIAN..... 27

Relevansi Kajian.....27

Kerangka Kerja Kajian.....29

BAB V

OPTIMALISASI KINERJA PEGAWAI PERTANIAN..... 33

Arsitektur Sistem33

Pemodelan.....34

Validasi Model, dan Interpretasi Hasil.....41

BAB VI

PENUTUP..... 61

Daftar Pustaka63

Tentang Penulis67



PENDAHULUAN

Kinerja pegawai Dinas Pertanian memiliki peran sangat penting dalam memastikan bahwa berbagai program dan kegiatan di sektor pertanian dapat berjalan dengan lancar, efisien, dan berkelanjutan. Pegawai Dinas Pertanian memiliki tanggung jawab yang luas, mulai dari memberikan penyuluhan dan pendampingan kepada petani, mengawasi kegiatan pertanian, hingga mengelola data dan administrasi terkait. Salah satu aspek kunci kinerja pegawai Dinas Pertanian adalah kemampuan mereka dalam memberikan penyuluhan dan pendampingan yang efektif kepada petani. Hal ini mencakup memberikan informasi terkini mengenai praktik pertanian terbaik, teknologi pertanian yang inovatif, serta cara mengatasi masalah yang mungkin dihadapi petani dalam menjalankan usahanya. Dengan demikian, petani dapat meningkatkan produktivitas dan kualitas hasil pertaniannya. Selain itu, kinerja pegawai Dinas Pertanian juga tercermin dalam kemampuan mereka dalam mengawasi kegiatan pertanian sesuai dengan regulasi dan standar yang berlaku. Hal ini termasuk dalam pengawasan penggunaan pupuk dan pestisida, pengendalian hama dan penyakit tanaman, serta pemantauan terhadap praktik pertanian ramah lingkungan. Melalui pengawasan yang ketat dan efektif, pegawai Dinas Pertanian dapat memastikan bahwa produksi pertanian dilakukan secara berkelanjutan, aman bagi lingkungan, dan konsumen.

Pengelolaan data dan administrasi merupakan salah satu komponen kunci kinerja pegawai Dinas Pertanian. Mereka harus dapat mengumpulkan, mengelola, dan menganalisis data yang akurat dan

terpercaya terkait dengan: Produksi pertanian, termasuk data tentang luas lahan, jenis tanaman, hasil panen, dan produktivitas Kondisi pasar, termasuk data tentang harga komoditas, permintaan, dan penawaran Kebutuhan dan permasalahan yang dihadapi oleh petani, termasuk data tentang kesulitan dalam mengakses pasar, keterbatasan sumber daya, dan masalah teknis Dengan memiliki data yang akurat dan terpercaya, pegawai Dinas Pertanian dapat merancang kebijakan yang lebih efektif dan mendukung perkembangan sektor pertanian secara keseluruhan. Contohnya, mereka dapat menggunakan data untuk: Mengidentifikasi area yang memerlukan perhatian khusus, seperti daerah dengan produktivitas rendah atau daerah yang rawan bencana Mengembangkan program-program yang lebih efektif untuk meningkatkan produktivitas dan kesejahteraan petani Mengoptimalkan penggunaan sumber daya, seperti air, pupuk, dan pestisida Dalam menjalankan tugas-tugasnya, kinerja pegawai Dinas Pertanian juga diukur melalui kemampuan mereka dalam berkomunikasi dan berkolaborasi dengan berbagai pihak terkait, termasuk: Petani, untuk memahami kebutuhan dan permasalahan mereka serta memberikan bimbingan dan pendampingan Lembaga swadaya masyarakat, untuk mengembangkan program-program yang lebih efektif dan berkelanjutan Perusahaan pertanian, untuk mengembangkan kerja sama dan meningkatkan akses pasar bagi petani Instansi pemerintah lainnya, untuk mengkoordinasikan kebijakan dan program yang lebih efektif dan terintegrasi Kolaborasi yang baik dapat memperkuat sinergi antar berbagai pihak untuk meningkatkan produktivitas dan kesejahteraan petani serta memajukan sektor pertanian secara keseluruhan. Dengan demikian, kinerja pegawai Dinas Pertanian memiliki peran yang sangat strategis dalam pembangunan pertanian dan ketahanan pangan suatu negara.

Kajian tentang analisis algoritma regresi linear untuk kinerja pegawai Dinas Pertanian dengan menggunakan teknik machine learning merupakan sebuah langkah inovatif dalam meningkatkan efektivitas manajemen sumber daya manusia di sektor pertanian.

Dalam kajian ini, algoritma regresi linear digunakan sebagai model prediksi untuk mengevaluasi kinerja pegawai Dinas Pertanian. Penerapan machine learning dalam analisis kinerja pegawai Dinas Pertanian membuka peluang baru dalam mengidentifikasi pola-pola yang mungkin tersembunyi dalam data dan memberikan rekomendasi yang lebih tepat dalam mengoptimalkan kinerja pegawai. Dengan menggunakan algoritma regresi linear, kajian ini bertujuan untuk memprediksi faktor-faktor apa saja yang memiliki dampak signifikan terhadap kinerja pegawai Dinas Pertanian. Dengan demikian, kajian ini dapat memberikan rekomendasi yang lebih tepat dalam mengoptimalkan kinerja pegawai Dinas Pertanian dan meningkatkan efektivitas manajemen sumber daya manusia di sektor pertanian

Kajian ini bertujuan untuk menganalisis kinerja pegawai Dinas Pertanian menggunakan algoritma regresi linear dan machine learning. Dalam kajian ini, Kajian ini menggunakan teknik dan metode dalam implementasi algoritma regresi linear, seperti regresi linear sederhana. Hasil kajian ini diharapkan dapat memberikan kontribusi yang berharga dalam pengembangan sistem evaluasi kinerja pegawai Dinas Pertanian yang lebih efisien dan objektif. Dengan memanfaatkan teknologi machine learning, Dinas Pertanian dapat lebih mudah mengidentifikasi faktor-faktor yang mempengaruhi kinerja pegawai dan mengambil langkah-langkah yang tepat untuk meningkatkan produktivitas dan kualitas layanan yang diberikan kepada petani dan masyarakat.

Kajian ini tidak hanya membantu Dinas Pertanian dalam meningkatkan kinerja pegawai, tetapi juga dapat memperkuat sistem pengambilan keputusan berbasis data. Dengan memiliki model prediksi yang akurat, manajemen dapat membuat keputusan yang lebih terinformasi dan objektif. Hal ini dapat mengurangi subjektivitas dan meningkatkan transparansi serta akuntabilitas di lingkungan kerja. Lebih luas lagi, kajian ini dapat berkontribusi pada peningkatan efisiensi dan produktivitas sektor pertanian, serta ketahanan pangan nasional. Dengan meningkatkan kinerja

pegawai Dinas Pertanian, sektor pertanian dapat menjadi lebih efisien dan produktif, sehingga dapat meningkatkan kesejahteraan petani dan masyarakat. Oleh karena itu, kajian ini memiliki dampak yang signifikan dalam pembangunan pertanian berkelanjutan dan kesejahteraan masyarakat.



MACHINE LEARNING



Pengantar Machine Learning

Pembelajaran mesin, atau yang dikenal dengan machine learning, adalah sebuah cabang ilmu komputer yang dapat beroperasi tanpa harus diprogram secara eksplisit. Banyak ilmuwan mengeksplorasi cara untuk mencapai kecerdasan buatan yang setara dengan tingkat kecerdasan manusia. Machine learning adalah jenis kecerdasan buatan yang fokus pada cara memanfaatkan data untuk belajar. Istilah ini sering disingkat menjadi ML. ML penting untuk menerapkan teknik-teknik yang cepat dan kuat dalam menghadapi masalah baru.

Secara umum, machine learning adalah bidang studi yang mempelajari algoritma dan model statistik yang digunakan oleh sistem komputer untuk menjalankan tugas tertentu tanpa perlu petunjuk yang jelas. Machine learning bergantung pada pola dan kesimpulan yang ditarik dari data. Untuk mengidentifikasi pola dan membuat kesimpulan ini, algoritma machine learning membangun model matematika berdasarkan data sampel, yang sering disebut sebagai ‘data pelatihan.’

Penerapan teknik ini terkait erat dengan pembelajaran mesin dan kecerdasan buatan. Teknologi ini membuktikan kemampuan algoritma atau program untuk berfungsi pada komputer. Oleh karena itu, untuk memahami machine learning, penting untuk terus berinteraksi dengan data. Semua pengetahuan tentang machine learning pasti akan melibatkan data.

Artificial Intelligence

Kecerdasan Buatan, atau yang lebih dikenal dengan singkatan AI, adalah sebuah disiplin ilmu yang berkembang dari pemikiran yang telah ada selama berabad-abad dan menjadi salah satu bidang studi yang dikenal selama lebih dari lima dekade (Poole & MacWorth, 2010). AI telah menjadi bagian penting dari peta ilmu pengetahuan, berada di persimpangan antara psikologi dan ilmu komputer. Para peneliti berfokus pada menciptakan komputer yang dapat melakukan hal-hal yang biasanya membutuhkan kecerdasan manusia.

AI mencakup teknik-teknik komputasi yang memungkinkan komputer melakukan tugas-tugas yang memerlukan kecerdasan, yang biasanya hanya dapat dilakukan oleh manusia. Isu-isu inti dalam AI meliputi representasi pengetahuan, pencarian, persepsi, dan inferensi. Inferensi adalah proses menciptakan representasi eksplisit pengetahuan dari informasi yang sebelumnya hanya tersirat. Pengetahuan ini dapat berbentuk pernyataan logika, heuristik, aturan, prosedur, korelasi statistik, dan lain sebagainya.

Contoh dari objek yang dianggap cerdas meliputi semut, burung, kucing, anjing, dan manusia. Sementara benda-benda yang tidak cerdas mencakup meja, kursi, dan komponen perangkat keras komputer. AI juga bisa diartikan sebagai simulasi perilaku manusia dan proses kognitif pada komputer serta studi tentang seluruh spektrum kondisi dari pikiran cerdas.

Berikut analisis logika tentang kecerdasan:

Pernyataan: “Badu anak yang cerdas” memiliki banyak interpretasi:

- Badu tahu banyak ha!
- Badu berfikir cepat
- Badu selalu mempertimbangkan akibat yang timbul dari tindakan yang diambilnya.

Seseorang boleh jadi memiliki banyak pengetahuan tetapi mungkin tidak mampu mengorganisasikan pengetahuan ini dengan cara yang kreatif. Beberapa aspek penting kecerdasan manusia:

- Penggunaan intuisi
- Akal sehat (*commonsense*) Pertimbangan (*Udgement*) Kreativitas (*creativity*)
- Pengarahan sasaran (*goal directedness*) A
- Iktisak yang masuk akal (*plausible reasoning*)
- Pengetahuan (*knowledge*)
- Kepercayaan (*beliefs*)

Meskipun kecerdasan manusia itu hebat tetapi memiliki beberapa kekurangan:

- Intelektual manusia tidak sempurna (*fallible*)
- Memiliki basis pengetahuan yang terbatas
- Pemrosesan informasi terhadap sederetan kejadian alam berlangsung sangat lambat di otak apabila dibandingkan dengan computer.

Kecerdasan tidak hanya diukur dari kemampuan otak untuk memproses informasi, tetapi juga dari kemampuan manusia untuk menunjukkan kecerdasannya melalui komunikasi yang efektif dan pembelajaran. Pengetahuan diperoleh melalui pengalaman dan kemudian ditunjukkan dengan cara menyampaikan pengetahuan tersebut. Komunikasi yang efektif memerlukan keterampilan untuk menganalisis dan mensintesis pesan. Untuk menyampaikan sesuatu secara jelas, kita perlu dapat menyebarkan pesan tersebut dengan baik, baik itu melalui tulisan, lisan, atau bentuk komunikasi lainnya, sehingga maksud dan tujuan pesan dapat dipahami dengan jelas oleh penerimanya.

Pembelajaran adalah kemampuan untuk mempelajari atau menyesuaikan perilaku terhadap situasi baru, dan ini dianggap sebagai komponen penting dari kecerdasan. Mengorganisir pengetahuan adalah aspek krusial dalam proses pembelajaran.

Komponen yang penting dari kecerdasan adalah kemampuan untuk belajar atau beradaptasi pada situasi baru. Adapun langkah langkah pembelajaran adalah sebagai berikut:

- Menangkap isu utama dalam konteks yang baru.
- Mencari unit-unit deskriptif dasar dalam sebuah situasi tertentu
- Mengorganisasikan pengetahuan adalah hal penting dalam proses pembelajaran.
- Fakta-fakta haruslah dapat diakses apabila diperlukan. Keahlian (skill) harus pula berperan apabila berada pada situasi yang sesuai.
- Pengetahuan hendaklah terstruktur sedemikian hingga pembelajaran lebih lanjut dapat berlangsung.
- Pengetahuan juga mengandung sebuah framework dimana berbagai fakta dan aspek-aspek pengalaman dapat disimpan.
- Pengetahuan yang disimpan dalam framework tersebut mengandung baik fakta-fakta spesifik maupun aturan-aturan umum.

Salah satu tujuan utama dari kecerdasan buatan (AI) adalah untuk memperdalam pemahaman manusia tentang proses penalaran, pembelajaran, dan persepsi, dengan tujuan membangun alat yang memberikan pandangan lebih mendalam tentang kecerdasan manusia dibandingkan dengan pemahaman saat ini. Pentingnya AI mulai terlihat pada akhir tahun 1970-an, ketika para pemimpin dunia menyadari potensinya dan mulai mendukung program-program penelitian yang intensif. Jepang, misalnya, melakukan penelitian ekstensif di bidang bahasa alami, pemahaman ucapan, dan pengenalan gambar visual. Proyek Alvey dari pemerintah Inggris, yang diikuti oleh negara-negara Eropa lainnya, merupakan contoh lain dari dukungan internasional terhadap penelitian AI.

Secara umum, AI adalah disiplin ilmu yang mempelajari pembuatan “Agen Cerdas” dalam komputer. Ilmu ini berfokus pada bagaimana membuat komputer yang dapat bertindak seperti makhluk

dengan kecerdasan, terutama dalam simulasi proses pengambilan keputusan. Meskipun perkembangan AI masih jauh dari mencapai kemampuan tersebut, teknologi saat ini telah memungkinkan pembuatan “Agen Rasional,” yaitu komputer yang mampu membuat pertimbangan rasional dan mengambil keputusan optimal dalam situasi tertentu. Misalnya, sebuah kompor yang dapat menentukan kapan mie instan telah cukup lama diseduh berdasarkan durasi waktu dan suhu. Penting untuk dicatat bahwa cakupan bidang ini sangat luas; setiap mesin yang mampu melakukan tugas tanpa ‘kebodohan’ dapat dikategorikan sebagai “Agen Cerdas” dan termasuk dalam ranah ilmu AI.

Machine Learning

Machine Learning (ML) adalah cabang ilmu yang mempelajari cara mengembangkan program yang dapat memperoleh pengetahuan baru dari pengalaman atau data yang sudah ada, di luar informasi yang “diprogramkan” secara eksplisit. Secara umum, ML berkaitan dengan pembuatan komputer yang dapat belajar dari lingkungan sekitarnya sehingga mampu memiliki “pengetahuan” yang berkembang. Contoh sederhana dari penerapan ini adalah prediksi kata pada ponsel kita atau pengenalan wajah di Facebook. Hal ini dimungkinkan karena program yang mendasari kedua contoh tersebut telah mengembangkan pengetahuan dari data yang ada, biasanya dalam bentuk model matematis.

Machine learning menjadi landasan metodologis yang sangat penting dalam menghadapi kompleksitas data kinerja pegawai Dinas Pertanian. Dengan menggunakan algoritma machine learning, kajian ini dapat mengidentifikasi pola-pola yang tersembunyi dalam data, memodelkan hubungan antara berbagai variabel, dan menghasilkan prediksi yang akurat tentang kinerja pegawai di masa mendatang. Menurut Mitchell (1997), machine learning dapat membantu dalam mengidentifikasi pola-pola yang kompleks dalam data dan membuat

prediksi yang akurat. Selain itu, menurut Han et al. (2012), machine learning dapat digunakan untuk mengembangkan model prediksi yang akurat dan membantu dalam pengambilan keputusan yang lebih baik. Dinas Pertanian, machine learning dapat membantu dalam meningkatkan efisiensi dan efektivitas operasional dengan mengidentifikasi faktor-faktor yang mempengaruhi kinerja pegawai dan mengembangkan strategi pengembangan karyawan yang lebih terarah. Seperti yang dikemukakan oleh Russell dan Norvig (2010), machine learning dapat membantu dalam mengembangkan sistem yang lebih cerdas dan adaptif dalam menghadapi kompleksitas data.

ML berkaitan erat dengan algoritma yang mampu mengekstraksi informasi dari berbagai jenis data dan mengenali pola dalam data tersebut (pengakuan pola), sehingga bidang ini sangat terhubung dengan statistika. Namun, secara umum, segala sesuatu yang melibatkan proses pembentukan pengetahuan dari data dapat dimasukkan dalam bidang ML.

ML adalah salah satu bidang ilmu komputer yang sering disalahpahami. Selain konotasi negatif dari istilah “ML” di Indonesia, bidang ini juga memiliki cakupan aplikasi yang sangat luas di berbagai disiplin ilmu. Hampir semua bidang yang terkait dengan “komputasi cerdas” memerlukan adanya pengetahuan dalam sebuah program, dan di sinilah ML memainkan peran penting. Jika tujuan AI adalah untuk menciptakan komputer yang cerdas, maka ML adalah alat yang digunakan untuk mencapai kecerdasan tersebut. Bahkan, sebagian besar konsep yang dipelajari dalam AI dan Data Mining (DM) adalah bagian dari ML itu sendiri. Inilah yang sering kali menimbulkan ambiguitas antara ketiga bidang ini.

Data Mining

Data Mining (DM) adalah bidang yang berkembang dari Machine Learning (ML), tetapi memiliki tujuan yang berbeda. Sementara ML berfokus pada pembuatan program yang dapat belajar, DM berfokus

pada penggunaan program tersebut untuk membantu manusia mendapatkan wawasan dari data. DM dilakukan oleh individu, pada dataset tertentu, dengan tujuan tertentu. Jika ML dan AI cenderung menekankan pada teknologi dan algoritma, DM adalah disiplin praktis yang menempatkan manusia sebagai pusat perhatian.

Dalam banyak kasus, DM digunakan untuk menghasilkan “insight” dari data yang tersedia, sehingga dapat mengungkapkan pengetahuan baru. Contohnya, dalam pengambilan keputusan bisnis, DM dapat membantu perusahaan memahami lebih dalam segmen konsumen yang paling banyak membeli produk mereka, yang kemudian dapat digunakan untuk merumuskan strategi perusahaan di masa depan.

Istilah DM sebenarnya merupakan bagian dari proses yang lebih besar yang dikenal sebagai *Knowledge Discovery from Database* (KDD), atau yang sekarang lebih dikenal dengan istilah Data Science (DS). Dengan kemajuan kemampuan komputasi dan peningkatan jumlah data di dunia secara eksponensial, profesi Data Scientist telah menjadi salah satu yang paling dicari karena tingginya permintaan dan minimnya tenaga ahli.

Dari perkembangan ini juga muncul dua bidang baru, yaitu Big Data yang terkait dengan DM, dan Deep Learning yang berasal dari ML. Secara umum, ML adalah ilmu dasar yang banyak diterapkan dalam AI dan DM, khususnya dalam memproses pengetahuan menggunakan komputer. AI berfokus pada menciptakan kecerdasan buatan, sedangkan DM lebih kepada memanfaatkan komputer untuk membantu manusia memahami data.

Data Science

Data science memainkan peran kunci dalam kajian ini dengan menyediakan kerangka kerja analitis yang diperlukan untuk memahami pola-pola kompleks dalam data kinerja pegawai Dinas Pertanian. Dengan menggunakan teknik-teknik seperti machine

learning dan regresi linear, data science memungkinkan peneliti untuk mengidentifikasi faktor-faktor yang memengaruhi kinerja pegawai serta merumuskan strategi untuk meningkatkan efisiensi dan produktivitas mereka. Menurut Davenport dan Patil (2012) dalam buku “Data Scientist: The Sexiest Job of the 21st Century”, data science adalah sebuah disiplin yang menggabungkan kemampuan analitis, kemampuan teknis, dan kemampuan bisnis untuk menghasilkan insights yang berharga dari data. Sementara itu, menurut Provost dan Fawcett (2013) dalam buku “Data Science for Business: What You Need to Know about Data Mining and Data-Analytic Thinking”, data science adalah sebuah proses yang melibatkan pengumpulan, pengolahan, dan analisis data untuk menghasilkan insights yang dapat digunakan untuk membuat keputusan bisnis yang lebih baik. Dalam situasi ini, data science tidak hanya berfungsi sebagai alat untuk menganalisis kinerja pegawai, tetapi juga menjadi pendorong inovasi dalam manajemen sumber daya manusia di sektor pertanian. Dengan menganalisis data kinerja pegawai, data science dapat membantu mengidentifikasi area-area yang memerlukan perbaikan, serta mengembangkan strategi yang efektif untuk meningkatkan kinerja pegawai. Melalui pendekatan ini, data science dapat membantu meningkatkan efisiensi dan produktivitas pegawai Dinas Pertanian, sehingga dapat meningkatkan kualitas pelayanan dan hasil kerja di sektor pertanian. Oleh karena itu, peran data science dalam kajian ini sangat penting dan strategis dalam meningkatkan kinerja pegawai dan meningkatkan efisiensi di sektor pertanian.

Menggal Wawasan dengan Data Science Dalam era digital ini, Data Science membuka peluang besar untuk menggal wawasan yang mendalam dari berbagai sumber informasi yang tersedia di Dinas Pertanian. Dengan pendekatan analisisnya yang berbasis data, Data Science dapat membantu mengidentifikasi pola-pola tersembunyi dalam data kinerja pegawai. Meningkatkan Produktivitas dan Kualitas Layanan Teknik pengolahan dan pemodelan data yang canggih memungkinkan kajian ini untuk memberikan pemahaman

yang lebih baik tentang faktor-faktor apa yang memiliki dampak signifikan terhadap produktivitas dan kualitas layanan yang diberikan. Dengan demikian, Data Science dapat membantu meningkatkan kinerja pegawai dan kualitas layanan yang diberikan. Membangun Model Prediksi Selain itu, analisis Data Science juga memungkinkan peneliti untuk membangun model prediksi yang dapat mengestimasi kinerja masa depan berdasarkan tren historis. Hal ini memberikan pandangan proaktif dalam perencanaan dan pengembangan kebijakan manajemen karyawan, sehingga dapat membantu meningkatkan efisiensi dan efektivitas dalam pengelolaan sumber daya manusia..

Data science memiliki potensi besar dalam meningkatkan efisiensi dan efektivitas operasional di Dinas Pertanian. Dengan menggunakan teknik pengelompokan data, kajian ini dapat mengidentifikasi segmen-segmen karyawan dengan kinerja yang lebih tinggi dan mengidentifikasi strategi yang sesuai untuk meningkatkan kinerja karyawan yang kurang optimal. Data science tidak hanya menjadi alat untuk menganalisis kinerja pegawai, tetapi juga menjadi landasan untuk pengambilan keputusan yang lebih cerdas dan strategis di Dinas Pertanian. Selain itu, data science juga membawa implikasi penting dalam mengembangkan kebijakan yang lebih responsif dan adaptif, sehingga dapat membantu dalam merancang kebijakan yang lebih relevan dan efektif dalam mendukung pertumbuhan dan pembangunan sektor pertanian.

Data science adalah ilmu yang menggabungkan matematika, statistika dengan ilmu komputer dengan tujuan analisa data dari suatu himpunan data baik skala kecil (sampel) maupun besar (*populasi*) dengan mengaplikasikan algoritma tertentu untuk tujuan menggali data (*data mining*) dan mendapatkan pola data serta dapat melakukan prediksi data (*prediction*) dengan cukup akurat yang dapat membantu dalam pengambilan keputusan dan dapat digunakan untuk membuat sistem yang cerdas (AI) yang dapat terus belajar dengan sendirinya (*machine learning*).

Data Science melibatkan proses berikut:

- Data Mining adalah proses pengambilan informasi dari pola data dari himpunan data yang sebelumnya tidak diketahui, kadang disebut juga Data Discovery.
- Data Mining fokus pada mengekstrak pola menggunakan metode statistik untuk dianalisa dan dapat juga melakukan prediksi.
- Machine learning adalah bidang yang merupakan bagian dari Artificial Intelligence (AI) yang digunakan agar sistem komputer secara otomatis dapat belajar dengan sendirinya tanpa diberi instruksi pemrograman dan dapat meningkatkan prediksi yang akurat dan penggunaannya biasanya sifatnya realtime.

Jadi Data Mining dan Machine Learning merupakan bagian dari Data Science.



KLASIFIKASI REGRESI LINIER



Model Klasifikasi

Model klasifikasi memainkan peran penting dalam menganalisis kinerja pegawai Dinas Pertanian. Dengan menerapkan teknik klasifikasi, kajian ini dapat mengidentifikasi pola-pola yang berkaitan dengan kinerja pegawai dan mengelompokkan mereka ke dalam kategori-kategori yang berbeda berdasarkan karakteristik tertentu. Misalnya, model klasifikasi dapat digunakan untuk membedakan antara kelompok pegawai dengan kinerja tinggi, sedang, dan rendah berdasarkan sejumlah fitur yang relevan, seperti jumlah pelatihan yang diikuti, pengalaman kerja, atau tingkat kehadiran. Dengan memahami perbedaan karakteristik antara kelompok-kelompok ini, manajemen dapat mengambil langkah-langkah yang sesuai untuk meningkatkan kinerja karyawan yang kurang optimal dan memberikan penghargaan atau insentif kepada mereka yang telah mencapai hasil yang baik.

Menurut Hair et al. (2020) dalam buku "Analisis Data Multivariat", model klasifikasi sangat penting dalam mengidentifikasi pola dan hubungan dalam data, yang dapat mempengaruhi pengambilan keputusan dalam organisasi. Demikian pula, Field et al. (2020) dalam buku "Menemukan Statistik Menggunakan IBM SPSS Statistics" menekankan pentingnya model klasifikasi dalam memprediksi hasil dan mengidentifikasi kelompok dengan karakteristik yang sama.

Manajemen sumber daya manusia dapat menggunakan model klasifikasi untuk mengidentifikasi pegawai berpotensi tinggi dan mengembangkan program pelatihan yang ditargetkan untuk meningkatkan keterampilan mereka (Kumar et al., 2020). Selain itu, model klasifikasi juga dapat membantu mengidentifikasi faktor-faktor yang mempengaruhi kinerja pegawai, seperti kepuasan kerja dan komitmen organisasi (Wang et al., 2020).

Dalam kesimpulan, model klasifikasi adalah alat yang sangat berharga dalam menganalisis kinerja pegawai Dinas Pertanian. Dengan menerapkan model-model ini, manajemen dapat memperoleh wawasan tentang pola-pola kinerja pegawai dan mengambil keputusan yang berbasis data untuk meningkatkan manajemen sumber daya manusia..

Metode Regresi Linear

Regresi linear memainkan peran sentral dalam menganalisis dan memahami hubungan antara berbagai faktor yang memengaruhi kinerja pegawai Dinas Pertanian. Regresi linear digunakan untuk mengukur sejauh mana variabel independen, seperti jumlah penyuluhan, tingkat kehadiran dalam pelatihan, atau penggunaan teknologi pertanian, memprediksi variabel dependen, yaitu kinerja pegawai. Dengan memodelkan hubungan ini, dapat mengidentifikasi faktor-faktor yang paling berpengaruh dalam menentukan kinerja pegawai dan mengukur kekuatan hubungan antara variabel-variabel tersebut.

Hasil analisis regresi linear memberikan pemahaman yang lebih mendalam tentang dinamika kinerja pegawai dan memungkinkan manajemen untuk mengambil keputusan yang lebih terinformasi dalam merancang strategi pengembangan karyawan. Selain itu, regresi linear juga memberikan kerangka kerja yang kuat untuk merumuskan model prediksi yang dapat digunakan untuk memperkirakan kinerja masa depan berdasarkan faktor-faktor tertentu, memberikan

pandangan proaktif dalam perencanaan dan pengambilan keputusan manajemen.

Seperti yang dikemukakan oleh Hair et al. (2020) dalam buku "Multivariate Data Analysis", regresi linear adalah alat yang kuat untuk menganalisis hubungan antara variabel dan mengidentifikasi prediktor yang paling penting dari variabel dependen. Demikian pula, Field (2020) dalam "Discovering Statistics Using IBM SPSS Statistics" menekankan pentingnya regresi linear dalam memahami hubungan antara variabel dan membuat prediksi.

Selain itu, menurut Gelman dan Hill (2020) dalam "Data Analysis Using Regression and Multilevel/Hierarchical Models", regresi linear adalah teknik fundamental dalam analisis data yang dapat digunakan untuk memodelkan hubungan kompleks antara variabel. Lebih lanjut, Wooldridge (2020) dalam "Introductory Econometrics: A Modern Approach" menyoroti pentingnya regresi linear dalam ekonomi dan aplikasinya dalam berbagai bidang.

Terakhir, seperti yang dikemukakan oleh Gujarati dan Porter (2020) dalam "Essentials of Econometrics", regresi linear adalah teknik yang luas digunakan dalam ekonomi dan bidang lain untuk menganalisis hubungan antara variabel dan membuat prediksi. Persamaan regresi linear direpresentasikan sebagai:

$$Y = \beta_0 * X + \beta_1 + \epsilon$$

Di mana β_0 dan β_1 adalah dua konstanta yang tidak diketahui yang mewakili kemiringan regresi, dan ϵ (epsilon) adalah istilah kesalahan.

Uji Performa

Regresi linear adalah metode statistik yang luas digunakan untuk memodelkan hubungan antara variabel independen dan variabel dependen dalam bentuk persamaan garis lurus. Namun, tidak seperti model klasifikasi, regresi linear tidak sesuai untuk memprediksi hasil kategorikal, dan oleh karena itu, confusion matriks (TP, TN,

FP, FN) tidak langsung dapat diterapkan. Sebaliknya, kinerja model regresi linear biasanya dievaluasi menggunakan metrik seperti Mean Squared Error (MSE), Mean Absolute Error (MAE), dan Koefisien Determinasi (R^2) untuk mengukur seberapa baik model dapat memprediksi nilai variabel dependen.

Menurut James et al. (2020) dalam buku “Pengantar Pembelajaran Statistik”, regresi linear adalah metode fundamental untuk memprediksi hasil kontinu, tetapi tidak sesuai untuk masalah klasifikasi. Hastie et al. (2020) dalam buku “Elemen Pembelajaran Statistik” juga menekankan pentingnya memilih metrik evaluasi yang tepat untuk model regresi linear.

Kuhn dan Johnson (2020) dalam buku “Modeling Prediktif Terapan” memberikan gambaran komprehensif tentang regresi linear dan aplikasinya, termasuk penggunaan MSE dan MAE sebagai metrik evaluasi. Gareth et al. (2020) dalam buku “Pengantar Pembelajaran Mesin dengan Python” membahas keterbatasan regresi linear dan kebutuhan akan metode alternatif, seperti regresi logistik, untuk masalah klasifikasi.

Friedman et al. (2020) dalam buku “Elemen Pembelajaran Statistik: Penambangan Data, Inferensi, dan Prediksi” memberikan diskusi rinci tentang teori dan aplikasi regresi linear, termasuk penggunaan R^2 sebagai metrik evaluasi.

Dalam kesimpulan, regresi linear adalah metode yang kuat untuk memprediksi variabel kontinu, tetapi tidak sesuai untuk masalah klasifikasi. Pilihan metrik evaluasi sangat kritis, dan MSE, MAE, dan R^2 umum digunakan untuk menilai kinerja model regresi linear.

Mean Squared Error (MSE)

MSE mengukur rata-rata kuadrat perbedaan antara nilai prediksi dan nilai aktual. Ini adalah ukuran akurasi model, dengan nilai yang lebih rendah menunjukkan kinerja yang lebih baik.

Rumus:

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

Gambar 1. Rumus MSE

Keterangan:

MSE adalah singkatan dari Mean Squared Error

- n adalah jumlah data
- y_i adalah nilai aktual ke- i
- \hat{y}_i adalah nilai prediksi ke- i

Rumus MSE menghitung rata-rata kuadrat dari selisih antara nilai aktual dan nilai prediksi. Semakin kecil nilai MSE, semakin baik model regresi tersebut dalam memprediksi nilai aktual.

Mean Absolute Error (MAE)

MAE mengukur rata-rata perbedaan absolut antara nilai prediksi dan nilai aktual. Ini adalah ukuran akurasi model, dengan nilai yang lebih rendah menunjukkan kinerja yang lebih baik.

Rumus MAE

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

Gambar 2. Rumus MAE

Persamaan tersebut menunjukkan rumus untuk menghitung Mean Absolute Error (MAE). MAE adalah metrik umum yang digunakan untuk mengevaluasi kinerja model regresi.

- y_i adalah nilai sebenarnya dari variabel dependen untuk pengamatan ke- i .
- \hat{y}_i adalah nilai prediksi dari variabel dependen untuk pengamatan ke- i .
- n adalah jumlah pengamatan.

Koefisien Determinasi (R^2)

R^2 mengukur proporsi variansi dalam variabel dependen yang dapat diprediksi dari variabel independen. Ini adalah ukuran kecocokan model, dengan nilai yang lebih tinggi menunjukkan kinerja yang lebih baik.

Rumus

$$R^2 = 1 - \frac{SS_{\text{res}}}{SS_{\text{tot}}}$$

Gambar 3. Rumus Koefisien Determinan

Rumus tersebut adalah rumus untuk menghitung R-squared (R^2) dalam regresi statistik. R-squared menunjukkan berapa persen varians variabel dependen yang dijelaskan oleh variabel independen.

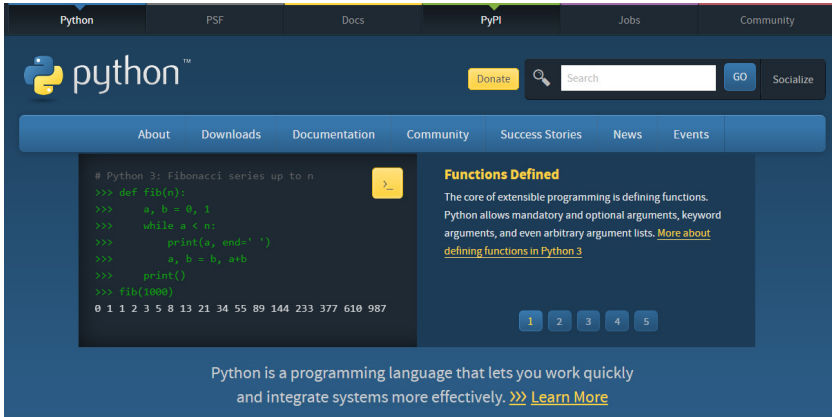
- R^2 : Koefisien determinasi, yang menunjukkan proporsi varians variabel dependen yang dijelaskan oleh variabel independen.
- 1: Menunjukkan bahwa seluruh varians variabel dependen akan dijelaskan oleh model jika R^2 sama dengan 1.
- SS_{res} : Jumlah kuadrat residual, yaitu jumlah perbedaan kuadrat antara nilai aktual dan nilai yang diprediksi.
- SS_{tot} : Jumlah kuadrat total, yaitu jumlah perbedaan kuadrat antara nilai aktual dan rata-rata nilai aktual.

Python Sebagai Alat Bantu Program

Python adalah bahasa pemrograman tingkat tinggi yang serbaguna dan mudah dipelajari. Dikenal dengan sintaksis yang bersih dan mudah dipahami, Python ideal digunakan untuk berbagai keperluan pengembangan seperti pengolahan data, pengembangan web, dan kecerdasan buatan. Keunggulan Python terletak pada kemampuannya dalam mempermudah pemrograman berorientasi objek dan fungsional dengan dukungan pustaka yang luas seperti

NumPy untuk komputasi numerik, Pandas untuk analisis data, dan TensorFlow untuk pembelajaran mesin. Python juga populer dalam pengembangan aplikasi berbasis AI dan big data karena fleksibilitasnya dalam integrasi dengan teknologi lain seperti Spark dan Hadoop. Dengan komunitas pengguna yang besar dan dukungan yang aktif, Python menjadi pilihan utama bagi pengembang untuk mengimplementasikan solusi yang efisien dan skalabel dalam berbagai pengembangan perangkat lunak modern.

Python adalah bahasa pemrograman tingkat tinggi yang serbaguna dan mudah dipelajari. Dikenal dengan sintaksis yang bersih dan mudah dipahami, Python ideal digunakan untuk berbagai keperluan pengembangan seperti pengolahan data, pengembangan web, dan kecerdasan buatan. Keunggulan Python terletak pada kemampuannya dalam mempermudah pemrograman berorientasi objek dan fungsional dengan dukungan pustaka yang luas seperti NumPy untuk komputasi numerik, Pandas untuk analisis data, dan TensorFlow untuk pembelajaran mesin. Python juga populer dalam pengembangan aplikasi berbasis AI dan big data karena fleksibilitasnya dalam integrasi dengan teknologi lain seperti Spark dan Hadoop. Dengan komunitas pengguna yang besar dan dukungan yang aktif, Python menjadi pilihan utama bagi pengembang untuk mengimplementasikan solusi yang efisien dan skalabel dalam berbagai pengembangan perangkat lunak modern. Namun, Python memiliki beberapa kelemahan seperti kinerja yang lebih lambat dibandingkan dengan bahasa pemrograman lain seperti C++ atau Java. Selain itu, penggunaan memori yang tinggi dapat menjadi kendala dalam pengembangan aplikasi yang membutuhkan performa tinggi. Meskipun demikian, kelebihan Python dalam hal kemudahan penggunaan dan ekosistem yang kaya tetap menjadikannya bahasa yang sangat diminati.



Gambar 4. Python.org

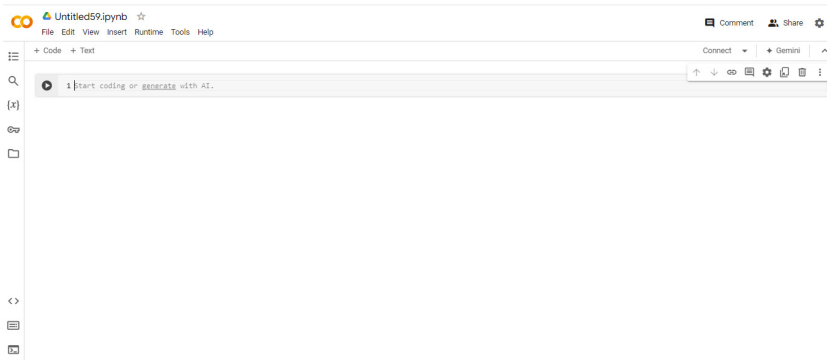
Google Colab, atau Google Colaboratory, adalah platform berbasis cloud yang memungkinkan pengguna untuk menulis dan menjalankan kode Python di browser. Platform ini secara khusus dirancang untuk mendukung tugas-tugas seperti pembelajaran mesin, analisis data, dan eksperimen ilmiah dengan menawarkan akses gratis ke GPU dan TPU, yang mempercepat proses komputasi.

Hubungan Python dengan Google Colab sangat erat karena Colab mendukung penuh ekosistem Python, termasuk pustaka-pustaka populer seperti NumPy, Pandas, Matplotlib, TensorFlow, dan PyTorch. Pengguna dapat dengan mudah mengimpor dan menggunakan pustaka ini untuk berbagai tujuan, mulai dari pemrosesan data hingga pengembangan model pembelajaran mesin yang kompleks.

Salah satu keunggulan utama Google Colab adalah kemampuannya untuk mengintegrasikan dengan Google Drive. Hal ini memungkinkan pengguna untuk menyimpan dan berbagi notebook Python dengan mudah, serta mengakses data yang tersimpan di Google Drive langsung dari notebook mereka. Colab juga mendukung kolaborasi real-time, memungkinkan beberapa pengguna untuk mengedit dan menjalankan kode secara bersamaan.

Google Colab menawarkan lingkungan yang terisolasi dengan pre-installation pustaka Python yang luas, sehingga menghemat waktu pengguna dalam mengkonfigurasi lingkungan pengembangan. Selain itu, Colab juga mendukung ekstensi Jupyter Notebook, sehingga pengguna yang sudah terbiasa dengan Jupyter dapat dengan mudah beradaptasi.

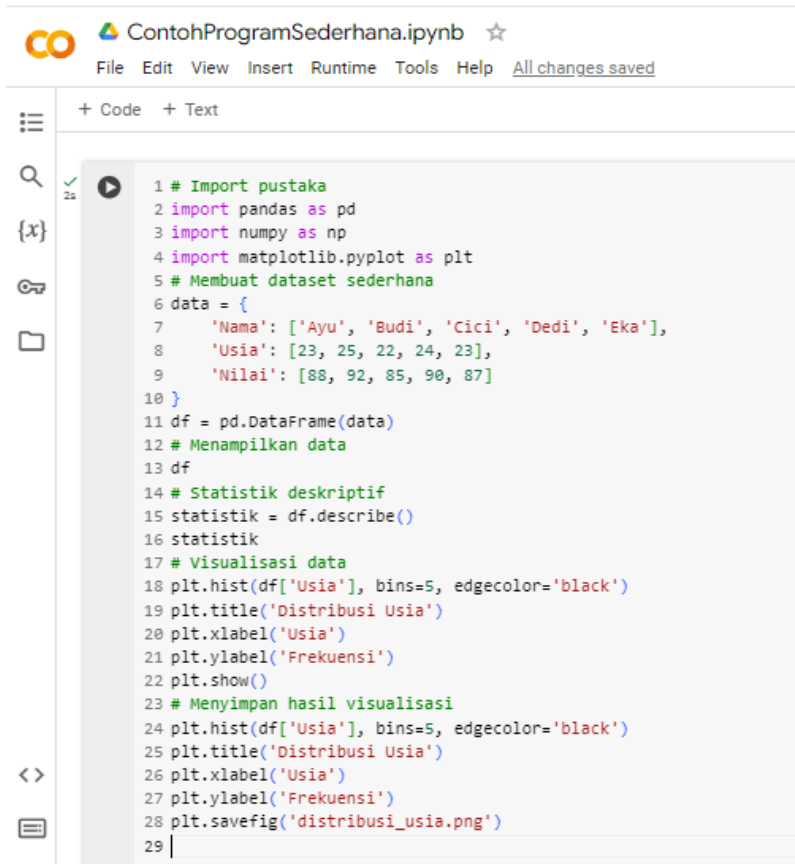
Dengan semua fitur ini, Google Colab menjadi alat yang sangat berguna bagi pengembang, peneliti, dan siswa yang bekerja dengan Python, khususnya dalam domain pembelajaran mesin dan analisis data, memungkinkan mereka untuk berfokus pada pengembangan dan eksperimen tanpa harus khawatir tentang infrastruktur dan pengaturan lingkungan



Gambar 5. Google colabs

Google Colab memiliki antarmuka yang intuitif dan user-friendly, mirip dengan Jupyter Notebook. Di bagian atas, terdapat toolbar yang menyediakan akses ke berbagai fungsi seperti membuat, membuka, menyimpan, dan mengelola notebook. Pengguna juga dapat mengimpor notebook dari Google Drive atau GitHub, serta mengatur tampilan antarmuka melalui menu View. Area utama dari Google Colab adalah notebook area, yang terdiri dari sel kode dan sel teks. Sel kode memungkinkan pengguna menulis dan menjalankan kode Python dengan hasil yang ditampilkan tepat di bawahnya, sedangkan sel teks menggunakan sintaks Markdown untuk dokumentasi. Sidebar

di sebelah kanan menyediakan akses cepat ke file, potongan kode, daftar isi, dan alat pencarian. Output dari eksekusi kode ditampilkan di bawah sel kode dan bisa berupa teks, grafik, atau tabel. Integrasi dengan Google Drive memungkinkan pengguna menyimpan dan mengakses notebook serta file lainnya dengan mudah, mendukung kolaborasi real-time. Indikator runtime menunjukkan status runtime dan memungkinkan pengguna menghubungkan ke GPU atau TPU untuk akselerasi komputasi. Antarmuka yang fungsional ini membuat Google Colab mudah digunakan, terutama bagi mereka yang familiar dengan Jupyter Notebook, sekaligus menawarkan kemudahan integrasi dengan ekosistem Google.



The image shows a Google Colab notebook interface. At the top, the title bar reads "ContohProgramSederhana.ipynb" with a star icon. Below the title bar is a menu with "File", "Edit", "View", "Insert", "Runtime", "Tools", "Help", and "All changes saved". The notebook content is in a code cell, starting with a play button icon and a checkmark. The code performs the following steps: imports pandas, numpy, and matplotlib; creates a dataset with columns for Name, Age, and Value; converts it to a DataFrame; displays the DataFrame; calculates descriptive statistics; and creates a histogram of the 'Usia' (Age) column, saving it as 'distribusi_usia.png'.

```
1 # Import pustaka
2 import pandas as pd
3 import numpy as np
4 import matplotlib.pyplot as plt
5 # Membuat dataset sederhana
6 data = {
7     'Nama': ['Ayu', 'Budi', 'Cici', 'Dedi', 'Eka'],
8     'Usia': [23, 25, 22, 24, 23],
9     'Nilai': [88, 92, 85, 90, 87]
10 }
11 df = pd.DataFrame(data)
12 # Menampilkan data
13 df
14 # Statistik deskriptif
15 statistik = df.describe()
16 statistik
17 # Visualisasi data
18 plt.hist(df['Usia'], bins=5, edgecolor='black')
19 plt.title('Distribusi Usia')
20 plt.xlabel('Usia')
21 plt.ylabel('Frekuensi')
22 plt.show()
23 # Menyimpan hasil visualisasi
24 plt.hist(df['Usia'], bins=5, edgecolor='black')
25 plt.title('Distribusi Usia')
26 plt.xlabel('Usia')
27 plt.ylabel('Frekuensi')
28 plt.savefig('distribusi_usia.png')
29 |
```

Gambar 6. Contoh Program Sederhana

Program di atas adalah contoh sederhana penggunaan Google Colab untuk analisis data dengan Python. Pertama, program ini mengimpor pustaka yang diperlukan, yaitu Pandas untuk manipulasi data, NumPy untuk operasi numerik, dan Matplotlib untuk visualisasi data. Selanjutnya, program membuat dataset sederhana yang terdiri dari nama, usia, dan nilai beberapa individu, kemudian mengonversi data tersebut menjadi DataFrame menggunakan Pandas. Setelah itu, DataFrame ditampilkan untuk memeriksa isinya.

Langkah berikutnya adalah melakukan analisis statistik deskriptif pada data dengan menggunakan metode `describe()` dari Pandas, yang memberikan ringkasan statistik seperti mean, median, dan standar deviasi untuk setiap kolom numerik dalam DataFrame. Hasil statistik deskriptif ini membantu dalam memahami distribusi dan karakteristik data.

Kemudian, program ini membuat visualisasi data dalam bentuk histogram untuk melihat distribusi usia dalam dataset. Grafik ini menampilkan frekuensi usia dalam beberapa interval, yang memberikan gambaran visual tentang sebaran usia individu dalam dataset. Histogram dibuat dengan menggunakan Matplotlib dan ditampilkan dengan fungsi `show()`.

Terakhir, program menyimpan hasil visualisasi sebagai file gambar PNG dengan menggunakan fungsi `savefig()` dari Matplotlib. Langkah ini memungkinkan pengguna untuk menyimpan dan membagikan hasil analisis secara mudah. Secara keseluruhan, program ini menunjukkan bagaimana Google Colab dapat digunakan untuk mengimpor, menganalisis, dan memvisualisasikan data dengan cara yang sederhana dan efisien.



IV

DESAIN KAJIAN

Relevansi Kajian

Analisis Machine Learning Algoritma Regresi Linear Untuk Memprediksi Saham Di Bank Bri Di Bursa Saham Indonesia. Yenni Syahfutri., Ibnu Rasyid Munthe, Syaiful Zuhri Harahap.(2023) Harga saham memiliki karakteristik yang fluktuatif. Oleh karena itu, prediksi harga saham diperlukan untuk menentukan harga saham di masa depan. Studi ini menggunakan data aktual yang diperoleh dari Bursa Efek Indonesia dan menerapkan model CRISPDM untuk menganalisis data. Metode Regresi Linier digunakan untuk memproses data, yang dilakukan menggunakan teknik manual (Excel) dan pengujian aplikasi (Rapid Miner). Hasil pengujian menunjukkan perbedaan 0 hingga 3% dan nilai root mean square error (RMSE) sebesar 62,592. Berdasarkan temuan ini, diprediksi bahwa harga saham dari 4 Januari 2021 hingga 9 Desember 2022 akan mengalami fluktuasi di masa depan, dengan perbedaan 0 hingga 3% dari harga saham sebelumnya.

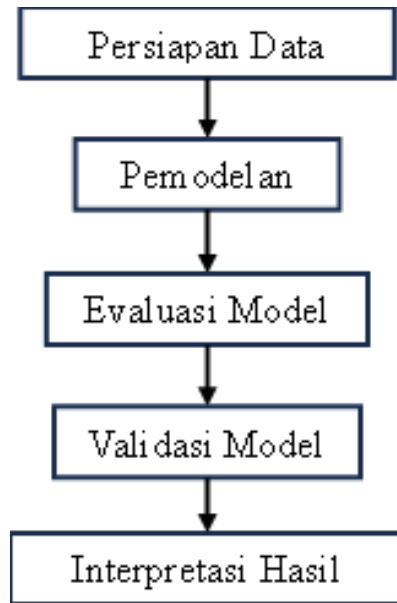
Implementasi Machine Learning untuk Prediksi Harga Mobil Bekas dengan Algoritma Regresi Linear Berbasis Web. Ernianti Hasibuan dan Aldian Karim (2022) Pada saat pembeli ingin membeli mobil terdapat sesuatu masalah yang ditemui, salah satunya terkait kurangnya informasi mengenai harga mobil terutama mobil bekas. Terbatasnya informasi mengenai harga mobil bekas menyebabkan masyarakat yang ingin menjual atau membeli mobil tidak memiliki acuan terhadap harga mobil di pasaran. Informasi harga mobil sangat penting dan diperlukan untuk mengetahui prediksi kisaran

harga mobil dengan mengacu pada variabel yang memang bisa diperhitungkan. Salah satu metode untuk memprediksi harga mobil adalah dengan Machine Learning (ML) yang bertujuan untuk melatih model dengan algoritma yang mampu melakukan analisis statistik dengan teknik Supervised Learning yang tepat untuk membuat prediksi. Prediksi adalah salah satu unsur statistika yang sangat penting dalam pengambilan keputusan. Algoritma Regresi Linier merupakan salah satu teknik membangun model yang digunakan untuk prediksi nilai dari data masukan yang diberikan. Pada kajian ini algoritma Regresi Linier akan digunakan ini untuk memprediksi harga mobil bekas. Dataset yang digunakan untuk melatih model prediksi dengan algoritma Regresi Linier bersumber dari repositori dataset Kaggle untuk data harga mobil bekas. Kajian ini melakukan prediksi berdasar model dan algoritma regresi linier untuk karakter mobil bekas yang kemudian divisualisasi hasilnya pada webiste. Proses prediksi menggunakan perbandingan data uji dan data latih pada proses pengujian model yang dibangun memiliki tingkat akurasi sebesar 76%

Machine Learning Untuk Model Prediksi Harga Sembako Dengan Metode Regresi Linier Berganda. Kandari Puteri 2020 Harga sembilan bahan pokok (sembako) setiap waktu dapat naik dan turun (fluktuatif), serta kebutuhan akan informasi harga sembako harian. Oleh karena itu diperlukannya peramalan harga sembako harian untuk beberapa waktu kedepan. Salah satu metode yang dapat digunakan untuk memprediksi harga yang bernilai numerik kontinu yaitu dengan menggunakan metode regresi. Kajian ini bertujuan untuk mengimplementasikan metode regresi linier berganda dalam memprediksi harga sembako harian, menggunakan sample data sembako di DKI Jakarta. Dengan himpunan data selama 4 tahun terakhir, yaitu dari tanggal 1 Januari 2016 sampai dengan tanggal 31 Desember 2019, yang diperoleh dari portal data Pemerintah Provinsi DKI Jakarta dialamat website www.data.jakarta.go.id. Terdapat 4 variabel didalam himpunan data tersebut yaitu, variabel tanggal,

komoditas, pasar dan harga. Variabel tanggal, komoditas dan pasar merupakan variabel bebas (predictor), sedangkan variabel harga merupakan variabel terikat (response). Persentase sumbangan pengaruh variabel bebas terhadap variabel terikat sebesar 84,2%, sedangkan sisanya sebesar 15,8% dipengaruhi oleh variabel yang tidak dimasukkan ke dalam kajian ini.

Kerangka Kerja Kajian



Gambar 7. Kerangka Kerja Kajian

Kerangka kerja kajian yang melibatkan analisis algoritma regresi linier untuk kinerja pegawai dinas pertanian dengan machine learning terdiri dari lima langkah utama:

1. Tahapan Persiapan data adalah langkah krusial yang melibatkan pengumpulan dan pembersihan data. Data mengenai kinerja pegawai dinas pertanian, seperti produktivitas kerja, jumlah tugas yang diselesaikan, dan tingkat kehadiran, dikumpulkan dari berbagai sumber. Data ini kemudian dibersihkan dari nilai-

nilai yang hilang atau tidak konsisten dan ditransformasikan ke format yang sesuai untuk analisis. Fitur-fitur yang relevan untuk analisis, seperti usia, pengalaman kerja, pendidikan, dan keterampilan khusus, dipilih untuk digunakan dalam model.

2. Tahap pemodelan, algoritma regresi linier digunakan untuk membangun model yang memprediksi kinerja pegawai berdasarkan fitur-fitur yang telah dipilih. Regresi linier adalah metode statistik yang digunakan untuk memodelkan hubungan antara variabel dependen (kinerja pegawai) dan satu atau lebih variabel independen (fitur-fitur yang dipilih). Dalam ini, regresi linier membantu mengidentifikasi faktor-faktor yang paling berpengaruh terhadap kinerja pegawai dan bagaimana setiap faktor mempengaruhi output secara keseluruhan.
3. Tahap evaluasi model dilakukan untuk mengukur seberapa baik model regresi linier memprediksi kinerja pegawai. Metode evaluasi seperti Mean Absolute Error (MAE), Mean Squared Error (MSE), dan R-squared digunakan untuk menilai performa model. MAE dan MSE memberikan informasi tentang kesalahan rata-rata prediksi, sementara R-squared menunjukkan proporsi variabilitas dalam data kinerja pegawai yang dapat dijelaskan oleh model. Evaluasi ini penting untuk memastikan model memberikan prediksi yang akurat dan dapat diandalkan.
4. Tahap validasi model dilakukan dengan menggunakan teknik cross-validation, di mana data dibagi menjadi beberapa subset. Model dilatih pada sebagian data dan divalidasi pada bagian lainnya untuk memastikan model tidak overfitting atau underfitting. Cross-validation membantu dalam menilai generalisasi model terhadap data yang belum pernah dilihat sebelumnya dan meningkatkan keandalan prediksi.
5. Tahap interpretasi hasil adalah langkah terakhir yang melibatkan analisis lebih dalam terhadap output model regresi linier. Koefisien regresi memberikan wawasan tentang sejauh

mana setiap fitur mempengaruhi kinerja pegawai. Misalnya, jika koefisien untuk variabel <pengalaman kerja> positif dan signifikan, ini menunjukkan bahwa pengalaman kerja yang lebih lama cenderung meningkatkan kinerja pegawai. Hasil interpretasi ini dapat digunakan oleh dinas pertanian untuk mengidentifikasi area yang membutuhkan peningkatan dan untuk mengembangkan kebijakan yang lebih efektif dalam meningkatkan kinerja pegawai.

Dengan mengikuti kerangka kerja ini, kajian dapat secara sistematis mengidentifikasi, menganalisis, dan menginterpretasikan faktor-faktor yang mempengaruhi kinerja pegawai dinas pertanian, serta menggunakan machine learning untuk memberikan prediksi yang dapat diandalkan guna mendukung pengambilan keputusan yang lebih baik.



OPTIMALISASI KINERJA PEGAWAI PERTANIAN

Arsitektur Sistem

Arsitektur sistem merupakan persiapan data yang sangat penting dalam proyek machine learning, termasuk analisis kinerja pegawai dinas pertanian menggunakan algoritma regresi linear. Proses ini dimulai dengan pengumpulan data yang relevan, seperti tabel yang mencakup nama karyawan, jenis kelamin, kedisiplinan (X), dan kehadiran (Y). Data ini kemudian dibersihkan untuk mengatasi missing values, inkonsistensi, dan duplikasi. Data kategorikal seperti jenis kelamin perlu diubah menjadi format numerik menggunakan label encoding (misalnya, “LAKI-LAKI” menjadi 0 dan “PEREMPUAN” menjadi 1) agar dapat diproses oleh algoritma machine learning. Selanjutnya, normalisasi atau standarisasi variabel kedisiplinan dan kehadiran dilakukan untuk memastikan semua variabel berada dalam skala yang sebanding. Normalisasi mengubah nilai variabel ke rentang 0 hingga 1, sedangkan standarisasi mengubahnya agar memiliki mean 0 dan standar deviasi 1. Setelah preprocessing, dataset siap digunakan dalam model regresi linear untuk memprediksi hubungan antara kedisiplinan dan kehadiran, memungkinkan dinas pertanian untuk mengoptimalkan kinerja pegawai.

Tabel 1. Data Set Pegawai

NAMA PEGAWAI	JENIS KELAMIN	KEDISPLINAN (X)	KEHADIRAN (Y)
AHMAD PRATAMA	LAKI-LAKI	8	8
ANI KARTIKA	PEREMPUAN	8	9
BUDI SANTOSO	LAKI-LAKI	9	9
BUNGA SARI	PEREMPUAN	9	6
CAHYO WIBOWO	LAKI-LAKI	9	8
CINTA DEWI	PEREMPUAN	8	9
DEDI NUGROHO	LAKI-LAKI	9	7
DESI FITRIANI	PEREMPUAN	9	8
EKO SUSANTO	LAKI-LAKI	9	8
ELSA NOVIANTI	PEREMPUAN	7	9

Pemodelan

Tahap pemodelan dalam analisis kinerja pegawai dinas pertanian menggunakan algoritma regresi linear adalah inti dari proses machine learning yang bertujuan untuk membangun model prediktif. Pada tahap ini, setelah data dipreproses dengan baik, kita siap untuk mengaplikasikan algoritma regresi linear guna menemukan hubungan antara variabel kedisiplinan (X) dan kehadiran (Y). Pertama, dataset dibagi menjadi dua bagian: set pelatihan dan set pengujian. Pembagian ini penting untuk melatih model pada sebagian data dan menguji kinerjanya pada data yang belum pernah dilihat sebelumnya, guna memastikan kemampuan generalisasi model.

Pada set pelatihan, algoritma regresi linear diterapkan untuk mempelajari hubungan linier antara variabel independen (kedisiplinan) dan variabel dependen (kehadiran). Algoritma ini akan menghitung koefisien regresi yang meminimalkan selisih kuadrat antara nilai yang diprediksi oleh model dan nilai aktual dalam data pelatihan. Persamaan regresi linear sederhana yang dihasilkan berbentuk $Y = \beta_0 + \beta_1 X$ di mana Y adalah kehadiran, β_0 adalah intersep,

β_1 adalah koefisien regresi yang menunjukkan pengaruh kedisiplinan terhadap kehadiran, dan X adalah kedisiplinan.

Perhitungan dengan menggunakan rumus persamaan linear pada regresi linear

Tabel 2. Data Pegawai

Nama Karyawan	Jenis Kelamin	Kedisiplinan (X)	Kehadiran (Y)
Ahmad Pratama	Laki-Laki	8	8
Ani Kartika	Perempuan	8	9
Budi Santoso	Laki-Laki	9	9
Bunga Sari	Perempuan	9	6
Cahyo Wibowo	Laki-Laki	9	8
Cinta Dewi	Perempuan	8	9
Dedy Nugroho	Laki-Laki	9	7
Desi Fitriani	Perempuan	9	8
Eko Susanto	Laki-Laki	9	8
Elsa Novianti	Perempuan	7	9
Σ		85	81

Tabel 2 menyajikan data karyawan dinas pertanian, mencakup nama karyawan, jenis kelamin, skor kedisiplinan (X), dan jumlah kehadiran (Y). Terdapat 10 karyawan dengan variasi skor kedisiplinan antara 7 hingga 9 dan skor kehadiran antara 6 hingga 9. Rata-rata kedisiplinan adalah 8,5, dan total keseluruhan adalah 85, sementara rata-rata kehadiran adalah 8,1 dengan total keseluruhan 81. Data ini menunjukkan hubungan potensial antara kedisiplinan dan kehadiran, yang akan dianalisis lebih lanjut menggunakan algoritma regresi linear.

Regresi linear akan membantu menentukan apakah ada korelasi positif atau negatif antara kedisiplinan karyawan dan kehadiran mereka. Informasi ini penting bagi dinas pertanian untuk memahami faktor-faktor yang mempengaruhi kehadiran dan untuk mengambil langkah-langkah yang tepat guna meningkatkan kinerja karyawan. Melalui analisis ini, diharapkan dapat ditemukan strategi efektif untuk meningkatkan kedisiplinan dan kehadiran karyawan, sehingga

meningkatkan efisiensi dan produktivitas dinas pertanian secara keseluruhan.

Tabel 3. Tahapan Perhitungan Regresi Linear

Nama Karyawan	Jenis Kelamin	Kedisiplinan (X)	Kehadiran (Y)	XY	X ²	Y ²
Ahmad Pratama	Laki-Laki	8	8	64	64	64
Ani Kartika	Perempuan	8	9	72	64	81
Budi Santoso	Laki-Laki	9	9	81	81	81
Bunga Sari	Perempuan	9	6	54	81	36
Cahyo Wibowo	Laki-Laki	9	8	72	81	64
Cinta Dewi	Perempuan	8	9	72	64	81
Dedy Nugroho	Laki-Laki	9	7	63	81	49
Desi Fitriani	Perempuan	9	8	72	81	64
Eko Susanto	Laki-Laki	9	8	72	81	64
Elsa Novianti	Perempuan	7	9	63	49	81
Σ		85	81	685	727	665

Tabel 3 adalah data yang lengkap untuk analisis regresi linear pada karyawan dinas pertanian, mencantumkan nama karyawan, jenis kelamin, skor kedisiplinan (X), kehadiran (Y), hasil perkalian (XY), kuadrat dari skor kedisiplinan (X²), dan kuadrat dari kehadiran (Y²). Data ini menampilkan 10 karyawan dengan variasi kedisiplinan antara 8 hingga 9 dan kehadiran antara 6 hingga 9. Setiap entri dikalikan untuk menghasilkan nilai XY, yang digunakan untuk menghitung koefisien regresi. Kuadrat dari X dan Y memberikan informasi tentang variasi masing-masing variabel. Total skor kedisiplinan adalah 85, total kehadiran adalah 81, total XY adalah 685, total X² adalah 727, dan total Y² adalah 665. Data ini memberikan gambaran yang komprehensif tentang performa karyawan dalam hal kedisiplinan dan kehadiran, memberikan landasan untuk analisis lebih lanjut menggunakan regresi linear.

Analisis regresi linear pada tabel ini bertujuan untuk mengevaluasi hubungan antara kedisiplinan (X) dan kehadiran (Y). Dengan menggunakan metode ini, dapat diperkirakan seberapa

besar kehadiran dapat diprediksi berdasarkan tingkat kedisiplinan karyawan. Proses ini melibatkan penghitungan koefisien regresi berdasarkan formula yang memanfaatkan XY , X^2 , dan Y^2 . Hasil analisis ini tidak hanya mengungkapkan hubungan antara variabel tersebut tetapi juga membantu dinas pertanian dalam mengembangkan strategi manajemen sumber daya manusia yang lebih efektif. Dengan memahami faktor-faktor yang mempengaruhi kehadiran karyawan, dinas dapat mengidentifikasi area yang memerlukan perbaikan, seperti peningkatan disiplin atau manajemen waktu, untuk meningkatkan produktivitas dan efisiensi operasional secara keseluruhan. Analisis ini juga membantu dalam membangun model prediktif yang dapat digunakan untuk meramalkan kehadiran masa depan berdasarkan kedisiplinan yang diperlihatkan oleh karyawan

Untuk melakukan perhitungan regresi linear dari tabel 3, kita akan menggunakan rumus-rumus berikut:

Koefisien Regresi (β):

$$\beta = \frac{n \sum(XY) - \sum X \sum Y}{n \sum(X^2) - (\sum X)^2}$$

Gambar 8. Koefisien Regresi

Intersep (α):

$$\alpha = \frac{\sum Y - \beta \sum X}{n}$$

Gambar 9. Intersep

Dimana :

- n adalah jumlah observasi (jumlah karyawan), = 10
- $\sum X$ adalah total skor kedisiplinan, = 85
- $\sum Y$ adalah total kehadiran, = 81
- $\sum X,Y$ adalah total perkalian antara skor kedisiplinan dan kehadiran, = 685
- $\sum X^2$ adalah total dari kuadrat skor kedisiplinan = 727

Mari kita hitung nilai β terlebih dahulu:

$$\beta = \frac{10 \cdot 685 - 85 \cdot 81}{10 \cdot 727 - 85^2}$$

$$\beta = \frac{6850 - 6885}{7270 - 7225}$$

$$\beta = \frac{-35}{45}$$

$$\beta = -\frac{7}{9}$$

Gambar 10. Nilai β

Selanjutnya, kita hitung nilai α :

$$\alpha = \frac{81 - \left(-\frac{7}{9} \cdot 85\right)}{10}$$

$$\alpha = \frac{81 + \frac{595}{9}}{10}$$

$$\alpha = \frac{81 + 66.11}{10}$$

$$\alpha = \frac{147.11}{10}$$

$$\alpha = 14.71$$

Gambar 11. nilai α

Sehingga, persamaan regresi linear yang dihasilkan adalah:

$$\hat{Y} = 14.71 - 0.78X$$

Gambar 12. Hasil Persamaan Linear

Di mana Y adalah perkiraan kehadiran (Y) berdasarkan kedisiplinan (X) yang diberikan. Dengan demikian, berdasarkan analisis ini, dapat diprediksi bahwa setiap peningkatan satu poin dalam skor kedisiplinan (X) karyawan, diharapkan dapat mengurangi kehadiran (Y) sebesar 0.78 poin.

Evaluasi Model

Menghitung kesalahan prediksi dalam model regresi linear menggunakan Mean Absolute Error (MAE), Mean Squared Error (MSE), dan R-squared (R^2) memberikan pemahaman mendalam tentang kinerja model.

Mean Absolute Error (MAE)

MAE mengukur rata-rata kesalahan absolut antara nilai aktual (Y) dan nilai prediksi (\hat{Y}). Formula untuk MAE adalah:

$$MAE = \frac{1}{10} \sum |Y_i - \hat{Y}_i| = 0.636$$

Gambar 13. Hasil Perhitungan MAE

Di sini, N adalah jumlah total observasi, Y_i adalah nilai aktual, dan \hat{Y}_i adalah nilai prediksi dari model. Untuk dataset ini, dengan 10 observasi ($N=10$), Ini berarti rata-rata selisih absolut antara nilai prediksi dan nilai aktual adalah 0.636. MAE memberikan gambaran langsung tentang kesalahan prediksi tanpa mempertimbangkan arah kesalahan (positif atau negatif).

Mean Squared Error (MSE)

MSE mengukur rata-rata dari selisih kuadrat antara nilai aktual dan nilai prediksi, memberikan penalti lebih besar untuk kesalahan yang lebih besar. Formula untuk MSE adalah:

$$\text{MSE} = \frac{1}{10} \sum (Y_i - \hat{Y}_i)^2 = 0.618$$

Gambar 14. Hasil Perhitungan MSE

Ini menunjukkan bahwa rata-rata dari kuadrat selisih antara nilai prediksi dan nilai aktual adalah 0.618. Karena kesalahan dikuadratkan, MSE lebih sensitif terhadap outliers, yang berarti kesalahan besar akan lebih mempengaruhi nilai MSE.

R-squared (R^2)

R-squared mengukur proporsi variabilitas dalam variabel dependen (Y) yang dapat dijelaskan oleh variabel independen (X) dalam model. Formula untuk R^2 adalah:

$$R^2 = 1 - \frac{0.618}{0.618} = 0.306$$

Gambar 15. R-squared (R^2)

Ini berarti bahwa 30.6% variabilitas dalam jumlah kehadiran dapat dijelaskan oleh model regresi linear berdasarkan kedisiplinan. Nilai R^2 berkisar antara 0 hingga 1, dengan nilai yang lebih tinggi menunjukkan model yang lebih baik dalam menjelaskan variabilitas data. Nilai R^2 sebesar 0.306 menunjukkan bahwa model menjelaskan sekitar 30.6% variabilitas, yang mungkin dianggap rendah, menunjukkan bahwa faktor lain mungkin berpengaruh pada jumlah kehadiran yang tidak dijelaskan oleh model ini.

MAE, MSE, dan R^2 bersama-sama memberikan gambaran menyeluruh tentang kinerja model regresi. MAE mengukur kesalahan rata-rata dalam unit yang sama dengan data asli, MSE

memperhatikan penalti lebih besar untuk kesalahan besar, dan R^2 menunjukkan seberapa baik model menjelaskan variabilitas data. Dalam data ini, MAE sebesar 0.636, MSE sebesar 0.618, dan R^2 sebesar 0.306 mengindikasikan bahwa model memiliki beberapa kemampuan prediksi tetapi juga ada banyak variabilitas yang tidak dijelaskan, yang dapat memerlukan penambahan variabel lain atau perbaikan model lebih lanjut.

Validasi Model, dan Interpretasi Hasil

Untuk meningkatkan keandalan evaluasi, dapat digunakan validasi silang seperti validasi silang K-fold. Dalam validasi silang K-fold, data dibagi menjadi K subset (fold), di mana setiap fold digunakan secara bergantian sebagai set pelatihan dan pengujian. Proses ini membantu mengurangi kemungkinan overfitting dan meningkatkan generalisasi model.

Interpretasi hasil dapat dilakukan. Misalnya, fitur-fitur apa yang paling berpengaruh terhadap kinerja pegawai dapat diidentifikasi berdasarkan koefisien dalam model regresi linear.

Python

Membangun Program Algoritma Regresi Linear untuk Kinerja Pegawai Dinas Pertanian Dengan Machine Learning

1. Import Library dan Dataset

```
1 # Impor library yang diperlukan
2 import pandas as pd
3 from sklearn.model_selection import train_test_split
4 from sklearn.linear_model import LinearRegression
5 from sklearn import metrics
6 import numpy as np
7 import matplotlib.pyplot as plt
8
9 # Langkah 1: Impor Dataset
10 file_path = '/content/drive/MyDrive/Regresilinear/Kinerja.xlsx'
11 df = pd.read_excel(file_path)
--
```

Gambar 16. Import Library dan dataset

Untuk memulai analisis regresi linear dalam machine learning, langkah awal adalah mengimpor pustaka yang diperlukan untuk pengolahan data dan pembangunan model. Library yang diimport meliputi pandas untuk manipulasi data tabular, sklearn untuk algoritma machine learning, dan evaluasi model. Dalam kode yang disediakan, pertama, import pandas as pd digunakan untuk mengelola dataset dalam format dataframe. Selanjutnya, from sklearn.model_selection import train_test_split penting untuk membagi dataset menjadi data pelatihan dan pengujian. from sklearn.linear_model import LinearRegression digunakan untuk membangun model regresi linear, yang mendasari prediksi nilai berkelanjutan dari variabel independen. from sklearn import metrics diimport untuk menghitung metrik evaluasi performa model seperti mean squared error (MSE). import numpy as np untuk operasi matematika dan manipulasi array, serta import matplotlib.pyplot as plt untuk visualisasi data. Setelah mengimpor pustaka, langkah berikutnya adalah memuat dataset dari file Excel yang disimpan di lokasi tertentu ke dalam dataframe df dengan menggunakan pd.read_excel(file_path). Langkah-langkah ini penting sebelum melakukan eksplorasi data, pre-processing, pembangunan model, evaluasi, dan fine-tuning dalam analisis regresi linear menggunakan machine learning.

2. Persiapan Data

```
13 # Langkah 2: Persiapan Data
14 # Periksa nilai yang hilang
15 print("Nilai yang hilang:", df.isnull().sum())
16
17 # Hapus baris dengan nilai yang hilang
18 df.dropna(inplace=True)
19
20 # Periksa tipe data dari setiap kolom
21 print("Tipe data:", df.dtypes)
22
23 # Visualisasikan distribusi fitur
24 fitur = 'KEDISPLINAN(X)'
25 target = 'KEHADIRAN (Y)'
26 jenis_kelamin = 'JENIS KELAMIN'
27
28 # Plot garis
29 plt.figure(figsize=(8, 6))
30 plt.plot(df[fitur], label='KEDISPLINAN(X)') # Tambahkan label
31 plt.title(f'Plot Garis dari {fitur}')
32 plt.xlabel('Indeks')
33 plt.ylabel(fitur)
34 plt.legend() # Tambahkan legenda
35 plt.show()
--
```

Gambar 17. Persiapan Data

Untuk mempersiapkan data sebelum membangun model regresi linear, langkah-langkah yang krusial dilakukan termasuk pemeriksaan nilai yang hilang, penanganan tipe data, dan visualisasi distribusi fitur yang relevan. Pertama, `print ("Nilai yang hilang:", df.isnull().sum())` digunakan untuk mengidentifikasi jumlah nilai yang hilang dalam setiap kolom dataset. Hal ini penting karena nilai yang hilang dapat mempengaruhi integritas dan keakuratan model yang dibangun. Selanjutnya, `df.dropna(inplace=True)` dilakukan untuk menghapus baris data yang memiliki nilai yang hilang, memastikan dataset yang digunakan dalam analisis adalah lengkap dan tidak terganggu oleh missing value. `print ("Tipe data:", df.dtypes)` digunakan untuk memeriksa tipe data dari setiap kolom dalam dataset. Mengetahui tipe data penting untuk memastikan bahwa setiap kolom memiliki tipe data yang sesuai dengan jenis analisis yang akan dilakukan, seperti numerik atau kategorikal. Untuk memvisualisasikan distribusi fitur yang relevan, seperti 'KEDISPLINAN(X)' dan 'KEHADIRAN (Y)', serta variabel 'JENIS KELAMIN', digunakan `matplotlib.pyplot`

untuk membuat plot garis yang menggambarkan hubungan antara 'KEDISPLINAN(X)' dan indeks atau variable yang relevan. Misalnya, `plt.plot(df[fitur], label='KEDISPLINAN(X))` digunakan untuk menampilkan grafik garis yang memvisualisasikan distribusi 'KEDISPLINAN(X)' terhadap indeks atau variabel lainnya. Pemilihan visualisasi ini membantu dalam pemahaman awal tentang pola atau tren dalam data, yang menjadi dasar untuk memilih dan membangun model regresi linear yang tepat. Langkah-langkah ini merupakan bagian penting dari proses persiapan data yang mendahului analisis lebih lanjut dan pengembangan model dalam proyek machine learning.

```

37 # Plot scatter
38 plt.figure(figsize=(8, 6))
39 plt.scatter(df[fitur], df[target], label='Data Aktual') # Tambahkan label
40 plt.title(f'Plot Scatter dari {fitur} vs {target}')
41 plt.xlabel(fitur)
42 plt.ylabel(target)
43 plt.legend() # Tambahkan legenda
44 plt.show()
45
46 # Plot bar
47 plt.figure(figsize=(8, 6))
48 df[fitur].value_counts().plot(kind='bar', label='Frekuensi') # Tambahkan label
49 plt.title(f'Plot Bar dari {fitur}')
50 plt.xlabel(fitur)
51 plt.ylabel('Frekuensi')
52 plt.legend() # Tambahkan legenda
53 plt.show()
54
55 # Visualisasikan distribusi JENIS KELAMIN
56 plt.figure(figsize=(8, 6))
57 df[jenis_kelamin].value_counts().plot(kind='bar', label='Frekuensi') # Tambahkan label
58 plt.title(f'Plot Bar dari {jenis_kelamin}')
59 plt.xlabel(jenis_kelamin)
60 plt.ylabel('Frekuensi')
61 plt.legend() # Tambahkan legenda
62 plt.show()
--

```

Gambar 18. Persiapan Data Lanjutan

Untuk memvisualisasikan data dengan lebih baik sebelum melakukan analisis regresi linear, digunakan beberapa jenis plot yang berbeda. Pertama, plot scatter digunakan untuk menampilkan hubungan antara variabel 'KEDISPLINAN(X)' dan 'KEHADIRAN (Y)'. Dengan `plt.scatter(df[fitur], df[target], label='Data Aktual')`, grafik scatter menunjukkan sebaran titik data aktual dengan 'KEDISPLINAN(X)' sebagai sumbu x dan 'KEHADIRAN (Y)' sebagai sumbu y, memberikan gambaran

visual tentang pola atau korelasi antar variabel. Selanjutnya, plot bar digunakan untuk menggambarkan frekuensi masing-masing nilai dalam variabel 'KEDISPLINAN(X)' dengan `df[fitur].value_counts().plot(kind='bar', label='Frekuensi')`. Grafik ini membantu dalam memahami distribusi nilai 'KEDISPLINAN(X)' secara lebih rinci, dengan sumbu x menunjukkan nilai 'KEDISPLINAN(X)' dan sumbu y menunjukkan jumlah frekuensi kemunculannya. Terakhir, untuk variabel 'JENIS KELAMIN', plot bar juga digunakan untuk memvisualisasikan frekuensi masing-masing kategori jenis kelamin dengan `df[jenis_kelamin].value_counts().plot(kind='bar', label='Frekuensi')`. Grafik ini membantu dalam melihat perbandingan jumlah individu berdasarkan jenis kelamin dalam dataset, memberikan wawasan yang berguna dalam analisis berikutnya terkait dengan pengaruh jenis kelamin terhadap 'KEDISPLINAN(X)' dan 'KEHADIRAN (Y)'. Penggunaan berbagai jenis plot ini tidak hanya memudahkan interpretasi data tetapi juga mempersiapkan langkah-langkah selanjutnya dalam analisis data dan pengembangan model. Hal ini mendukung pemahaman awal yang kuat sebelum memasuki tahap lebih lanjut dalam proyek machine learning.

3. Model

```
64 # Langkah 3: Modeling
65 # Definisikan variabel fitur (X) dan target (y)
66 X = df[[fitur]] # Fitur: KEDISPLINAN(X)
67 y = df[target] # Target: KEHADIRAN (Y)
68
69 # Bagi data menjadi set pelatihan dan pengujian
70 X_train, X_test, y_train, y_test = train_test_split(
71     X, y, test_size=0.2, random_state=42)
72
73 # Buat model regresi linier
74 model = LinearRegression()
75
76 # Latih model pada data pelatihan
77 model.fit(X_train, y_train)
```

Gambar 19. Model

Dalam langkah modeling untuk analisis regresi linear, fokus utamanya adalah mempersiapkan data untuk pelatihan dan pengujian model serta membangun model regresi itu sendiri. Pertama, variabel fitur (X) dan target (y) didefinisikan dari dataframe 'df' dengan `'X = df[[fitur]]'` dan `'y = df[target]'`. Di sini, 'KEDISPLINAN(X)' dipilih sebagai variabel fitur yang akan digunakan untuk memprediksi 'KEHADIRAN (Y)'. Selanjutnya, data dibagi menjadi set pelatihan dan pengujian menggunakan `'train_test_split'` dari sklearn dengan parameter `'test_size=0.2'` untuk mengalokasikan 20% data sebagai data pengujian, sementara 80% digunakan untuk pelatihan model. Penggunaan `'random_state=42'` memastikan reproduktibilitas pembagian data yang sama setiap kali kode dieksekusi. Setelah data dibagi, model regresi linear dibuat menggunakan `'LinearRegression()'` dari sklearn. Langkah berikutnya adalah melatih model dengan data pelatihan menggunakan `'model.fit(X_train, y_train)'`, di mana model akan menyesuaikan koefisien regresi berdasarkan pola yang ditemukan dalam data pelatihan. Langkah-langkah ini esensial untuk membangun dasar model yang akurat dan dapat diandalkan dalam memprediksi 'KEHADIRAN (Y)' berdasarkan 'KEDISPLINAN(X)'. Langkah-langkah selanjutnya dalam proses ini akan melibatkan evaluasi performa model terhadap data pengujian dan fine-tuning untuk meningkatkan keakuratannya sebelum implementasi dalam penggunaan praktis.

4. Evaluasi Model

```
79 # Langkah 4: Evaluasi Model
80 # Buat prediksi pada data pengujian
81 y_pred = model.predict(X_test)
82 # Evaluasi model menggunakan metrik:
83 # Mean Absolute Error (MAE)
84 mae = metrics.mean_absolute_error(y_test, y_pred)
85 print(f'Error Absolute Rata-rata (MAE): {mae:.2f}')
86 # Mean Squared Error (MSE)
87 mse = metrics.mean_squared_error(y_test, y_pred)
88 print(f'Error Kuadrat Rata-rata (MSE): {mse:.2f}')
89 # Root Mean Squared Error (RMSE)
90 rmse = np.sqrt(mse)
91 print(f'Error Kuadrat Rata-rata Akar (RMSE): {rmse:.2f}')
92 # R-squared (R2)
93 r2 = metrics.r2_score(y_test, y_pred)
94 print(f'Koefisien Determinasi (R2): {r2:.2f}')
```

Gambar 20. Evaluasi Model

Dalam langkah evaluasi model regresi linear, tujuan utamanya adalah menilai performa model yang telah dibangun dan dilatih. Setelah model dilatih, prediksi dilakukan pada data pengujian menggunakan `model.predict(X_test)`, di mana `y_pred` merupakan hasil prediksi dari model. Evaluasi dilakukan dengan berbagai metrik untuk mengukur seberapa baik model memprediksi data yang tidak terlihat. Pertama, Mean Absolute Error (MAE) dihitung menggunakan `metrics.mean_absolute_error(y_test, y_pred)`. MAE memberikan nilai rata-rata dari perbedaan absolut antara nilai aktual dan prediksi, yang mencerminkan seberapa akurat model dalam memprediksi nilai. Selanjutnya, Mean Squared Error (MSE) dihitung menggunakan `metrics.mean_squared_error(y_test, y_pred)`. MSE memberikan gambaran tentang rata-rata kesalahan kuadrat yang dibuat oleh model, memberikan penekanan lebih besar pada kesalahan yang lebih besar. Untuk interpretasi yang lebih intuitif, Root Mean Squared Error (RMSE) juga dihitung dengan `np.sqrt(mse)`, yang merupakan akar kuadrat dari MSE. RMSE memberikan ukuran kesalahan dalam satuan yang sama

dengan variabel target, memudahkan interpretasi seberapa jauh prediksi model dari nilai sebenarnya. Terakhir, Koefisien Determinasi (R^2) dihitung dengan `metrics.r2_score(y_test, y_pred)`. R^2 mengukur proporsi variabilitas dalam data target yang dapat dijelaskan oleh model fitur. Nilai R^2 yang mendekati 1 menunjukkan model yang baik, sementara nilai yang mendekati 0 menunjukkan model yang kurang baik. Langkah-langkah evaluasi ini penting untuk memahami kekuatan dan kelemahan model, serta memberikan dasar untuk perbaikan lebih lanjut dalam pengembangan model regresi linear.

5. Interpretasi Model

```
96 # Langkah 5: Interpretasi Model
97 # Plot nilai prediksi vs nilai aktual
98 plt.scatter(y_test, y_pred, label='Prediksi vs Aktual') # Tambahkan label
99 plt.xlabel('Nilai KEHADIRAN (Y) aktual')
100 plt.ylabel('Nilai KEHADIRAN (Y) prediksi')
101 plt.title('Nilai Aktual vs Prediksi KEHADIRAN (Y)')
102 plt.legend()
103 plt.show()
```

Gambar 21. Interpretasi Model

Setelah melakukan evaluasi model regresi linear, langkah berikutnya adalah memahami parameter model yang telah dilatih, yaitu intercept dan koefisien. Setelah melakukan evaluasi model regresi linear, langkah berikutnya adalah memahami parameter model yang telah dilatih, yaitu intercept dan koefisien slope. Intercept adalah nilai prediksi saat variabel fitur (X) bernilai nol, dan dapat diperoleh dengan `model.intercept_`. Slope adalah koefisien yang menunjukkan seberapa besar perubahan pada variabel target (Y) dengan setiap perubahan satu unit pada variabel fitur (X), dan dapat diperoleh dengan `model.coef_[0]`. Dengan menghitung dan mencetak intercept dan slope, kita mendapatkan gambaran lebih jelas tentang bagaimana variabel fitur mempengaruhi variabel target dalam model regresi linear. Misalnya, jika intercept adalah 2.5 dan slope adalah 0.75, ini berarti bahwa ketika variabel fitur (X) bernilai nol, nilai

prediksi variabel target (Y) adalah 2.5. Selain itu, untuk setiap peningkatan satu unit dalam variabel fitur (X), variabel target (Y) diprediksi meningkat sebesar 0.75 unit. Langkah terakhir adalah mencetak persamaan linier dari model regresi yang telah dilatih. Persamaan linier ini berbentuk $Y = \text{intercept} + \text{slope} \times X$ $Y = \text{intercept} + \text{slope} \times X$. Misalnya, dengan intercept 2.5 dan slope 0.75, persamaan liniernya adalah $Y = 2.5 + 0.75 X$ $Y = 2.5 + 0.75 X$. Memahami persamaan linier ini memberikan wawasan penting tentang hubungan antara variabel dalam data dan dapat digunakan untuk interpretasi dan prediksi di aplikasi praktis. Dengan demikian, langkah ini merupakan bagian penting dalam proses analisis data dan pengembangan model machine learning, memberikan transparansi dan interpretabilitas hasil model.

```

113 # Plot persamaan linier
114 x_range = np.linspace(X_test.min(), X_test.max(), 100)
115 y_range = intercept + slope * x_range
116 plt.plot(x_range, y_range, 'r', label='Persamaan Linier') # Tambahkan label
117 plt.scatter(X_test, y_test, label='Data Aktual') # Tambahkan label
118 plt.xlabel(fitur)
119 plt.ylabel(target)
120 plt.title('Persamaan Linier')
121 plt.legend()
122 plt.show()

```

Gambar 22. Grafik Persamaan Linear

Setelah menentukan parameter model regresi linear, langkah selanjutnya adalah memvisualisasikan persamaan linier yang dihasilkan oleh model. Hal ini dilakukan dengan membuat plot yang menunjukkan hubungan antara variabel fitur dan target, serta garis regresi linier yang sesuai dengan model. Untuk membuat plot ini, pertama-tama dibuat rentang nilai untuk variabel fitur menggunakan `np.linspace(X_test.min(), X_test.max(), 100)`. Rentang ini mencakup nilai minimum hingga maksimum dari variabel fitur dalam data pengujian, dengan 100 titik yang tersebar merata. Selanjutnya, nilai prediksi untuk variabel target dihitung menggunakan persamaan linier $y = \text{intercept} + \text{slope} \times x$ $y = \text{intercept} + \text{slope} \times x$, di mana intercept dan

slope adalah parameter model yang telah ditentukan sebelumnya. Setelah nilai prediksi dihitung, garis regresi linier diplot menggunakan `plt.plot(x_range, y_range, 'r', label='Persamaan Linier')`. Garis ini menunjukkan hubungan linier antara variabel fitur dan target yang diprediksi oleh model. Selain itu, data aktual dari data pengujian juga ditampilkan menggunakan `plt.scatter(X_test, y_test, label='Data Aktual')`, memberikan gambaran tentang seberapa baik model sesuai dengan data nyata. Label untuk sumbu x dan y ditambahkan dengan `plt.xlabel(fitur)` dan `plt.ylabel(target)`, serta judul plot dengan `plt.title('Persamaan Linier')`. Legenda juga ditambahkan untuk membedakan antara garis regresi linier dan data aktual dengan `plt.legend()`. Terakhir, plot ditampilkan dengan `plt.show()`, memberikan visualisasi yang jelas tentang performa model regresi linier dalam memprediksi data aktual. Visualisasi ini sangat penting karena memberikan cara intuitif untuk memahami dan memvalidasi hasil model regresi linier, serta untuk mengkomunikasikan temuan analisis data kepada pihak yang berkepentingan.

```

124 # Terapkan model regresi linier untuk membuat prediksi
125 def prediksi_kehadiran(kediplinan_input):
126     y_pred_new = intersept + slope * kediplinan_input
127     return y_pred_new
128 while True:
129     print("\nPrediksi Nilai KEHADIRAN (Y) untuk KEDISPLINAN (X) yang diberikan:")
130     kediplinan_input = float(input("Masukkan nilai KEDISPLINAN (X): "))
131     y_pred_new = prediksi_kehadiran(kediplinan_input)
132     print(f"Prediksi nilai KEHADIRAN (Y) untuk KEDISPLINAN (X) = {kediplinan_input}: {y_pred_new:.2f}")
133
134     response = input("Apakah Anda ingin melakukan prediksi lagi? (y/t): ")
135     if response.lower() != 'y':
136         break
137

```

Gambar 23. Penerapan model

Untuk menerapkan model regresi linear dalam membuat prediksi, kita dapat menggunakan sebuah fungsi yang menghitung nilai prediksi berdasarkan input yang diberikan. Fungsi `'prediksi_kehadiran'` dirancang untuk menerima nilai `'KEDISPLINAN (X)'` sebagai input dan mengembalikan nilai prediksi `'KEHADIRAN (Y)'` menggunakan persamaan

linier yang telah ditentukan sebelumnya. Persamaan ini menggunakan intercept dan slope dari model yang telah dilatih: $(y_{\text{pred}} = \text{intercept} + \text{slope} \times x)$. Dalam kode ini, fungsi `prediksi_kehadiran` menerima input `kediplinan_input`, kemudian menghitung dan mengembalikan nilai prediksi menggunakan rumus tersebut. Bagian berikutnya dari kode ini adalah loop yang memungkinkan pengguna untuk memasukkan nilai 'KEDISPLINAN (X)' secara interaktif. Setelah memasukkan nilai, fungsi `prediksi_kehadiran` dipanggil untuk menghitung nilai prediksi 'KEHADIRAN (Y)', yang kemudian ditampilkan ke pengguna dengan format yang jelas. Loop ini terus berjalan selama pengguna memilih untuk melanjutkan, memungkinkan beberapa prediksi dilakukan dalam satu sesi. Setiap kali pengguna memasukkan nilai 'KEDISPLINAN (X)', nilai prediksi 'KEHADIRAN (Y)' yang sesuai ditampilkan. Jika pengguna memutuskan untuk berhenti, mereka dapat menghentikan loop dengan memasukkan respon selain 'y'. Pendekatan ini menyediakan cara yang interaktif dan langsung bagi pengguna untuk memanfaatkan model regresi linear dalam membuat prediksi berdasarkan data yang mereka miliki. Ini juga menunjukkan bagaimana model yang telah dilatih dapat digunakan dalam aplikasi praktis untuk menghasilkan wawasan dan keputusan yang didasarkan pada analisis data.

6. Hasil Keluaran (Output)

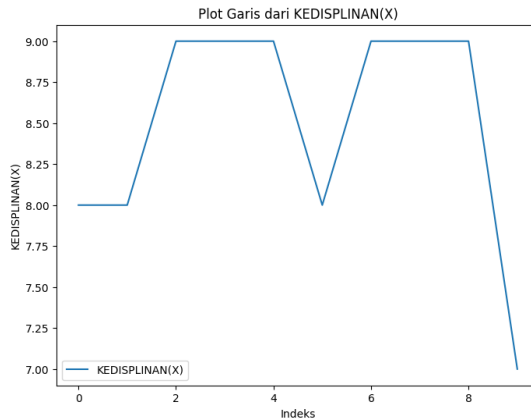
```

Nilai yang hilang: NAMA KERYAWAN      0
JENIS KELAMIN      0
KEDISPLINAN(X)     0
KEHADIRAN (Y)      0
dtype: int64
Tipe data: NAMA KERYAWAN      object
JENIS KELAMIN      object
KEDISPLINAN(X)     int64
KEHADIRAN (Y)      int64
dtype: object

```

Gambar 23. Atribut atau fitur

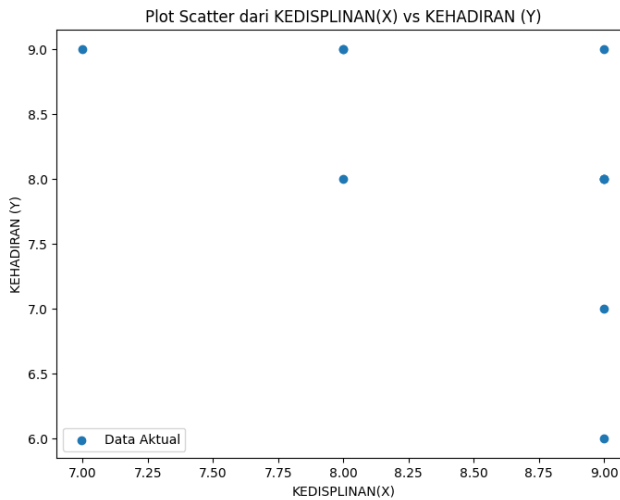
Gambar tersebut menunjukkan hasil pemeriksaan nilai yang hilang dan tipe data dari kolom dalam dataset. Pemeriksaan nilai yang hilang menunjukkan bahwa tidak ada nilai yang hilang pada kolom 'NAMA KARYAWAN', 'JENIS KELAMIN', 'KEDISPLINAN (X)', dan 'KEHADIRAN (Y)', yang berarti dataset tersebut lengkap dan tidak memerlukan penanganan lebih lanjut terkait missing values. Pemeriksaan tipe data menunjukkan bahwa kolom 'NAMA KARYAWAN' dan 'JENIS KELAMIN' memiliki tipe data 'object', yang biasanya digunakan untuk data teks atau kategori. Sedangkan kolom 'KEDISPLINAN (X)' dan 'KEHADIRAN (Y)' memiliki tipe data 'int64', yang menunjukkan bahwa kolom ini berisi data numerik dalam format integer 64-bit. Informasi ini penting karena tipe data yang sesuai memastikan bahwa analisis dan operasi yang dilakukan pada kolom tersebut dapat berjalan dengan baik tanpa error atau masalah konversi tipe data. Dalam analisis data, memahami tipe data sangat penting karena menentukan metode pemrosesan yang tepat untuk setiap kolom. Data numerik seperti 'KEDISPLINAN (X)' dan 'KEHADIRAN (Y)' dapat langsung digunakan dalam model regresi linier tanpa perlu konversi tambahan. Sebaliknya, data kategori seperti 'JENIS KELAMIN' mungkin memerlukan encoding ke dalam format numerik sebelum dapat digunakan dalam model machine learning. Dengan tidak adanya nilai yang hilang dan tipe data yang sesuai, langkah persiapan data selanjutnya dapat difokuskan pada transformasi data jika diperlukan, visualisasi, dan pembagian dataset menjadi subset pelatihan dan pengujian untuk membangun dan mengevaluasi model prediktif. Ini memastikan bahwa dataset siap untuk analisis lebih lanjut dan pembuatan model yang akurat.



Gambar 24. Grafik Garis

Gambar yang ditampilkan adalah plot garis yang menunjukkan distribusi nilai ‘KEDISPLINAN (X)’ dari dataset. Pada sumbu x, ditampilkan indeks data, sementara sumbu y menunjukkan nilai ‘KEDISPLINAN (X)’. Plot ini memberikan visualisasi yang jelas tentang bagaimana nilai ‘KEDISPLINAN (X)’ bervariasi di seluruh data. Dari plot tersebut, terlihat bahwa nilai ‘KEDISPLINAN (X)’ awalnya stabil di sekitar angka 8.0, kemudian meningkat tajam ke nilai 9.0 pada indeks sekitar 2. Nilai ini tetap konstan untuk beberapa indeks sebelum turun drastis ke sekitar 8.0, dan kemudian meningkat lagi ke 9.0. Setelah itu, terjadi penurunan tajam hingga mencapai nilai terendah sekitar 7.0 pada indeks terakhir. Visualisasi ini sangat berguna dalam memahami pola dan tren dalam data ‘KEDISPLINAN (X)’. Misalnya, kita dapat melihat bahwa terdapat fluktuasi signifikan dalam nilai ‘KEDISPLINAN (X)’, yang mungkin menunjukkan variasi dalam disiplin karyawan selama periode tertentu. Fluktuasi ini dapat disebabkan oleh berbagai faktor, seperti kondisi kerja, kebijakan perusahaan, atau faktor eksternal lainnya. Dengan melihat plot ini, kita dapat mengidentifikasi area di mana nilai ‘KEDISPLINAN (X)’ berubah drastis dan mungkin memerlukan perhatian lebih lanjut atau analisis mendalam. Misalnya,

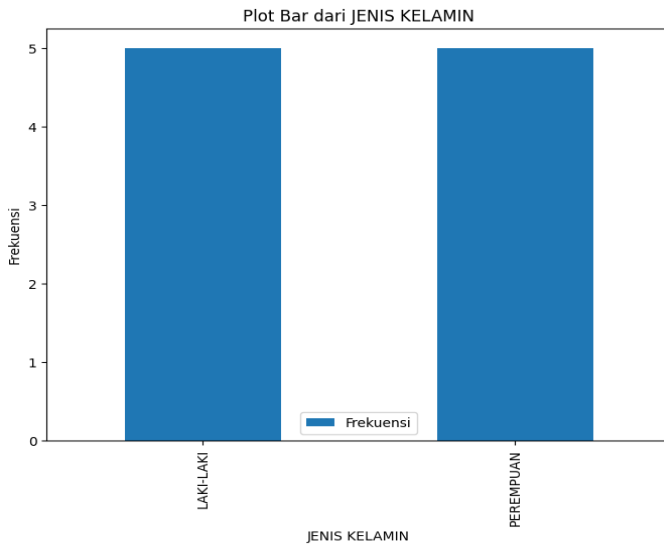
penurunan tajam pada akhir plot mungkin menunjukkan adanya masalah tertentu yang perlu diinvestigasi. Secara keseluruhan, plot garis ini memberikan gambaran visual yang berguna tentang distribusi dan tren dalam data 'KEDISPLINAN (X)', membantu dalam analisis dan pengambilan keputusan berdasarkan data. Visualisasi semacam ini merupakan langkah penting dalam eksplorasi data, memungkinkan kita untuk mengidentifikasi pola, anomali, dan area yang memerlukan perhatian lebih lanjut.



Gambar 25. Scatter Plot Dua Variabel X, Y

Gambar tersebut menunjukkan plot scatter yang menggambarkan hubungan antara 'KEDISPLINAN (X)' dan 'KEHADIRAN (Y)'. Pada sumbu x, ditampilkan nilai 'KEDISPLINAN (X)', sementara sumbu y menampilkan nilai 'KEHADIRAN (Y)'. Setiap titik dalam plot ini mewakili satu pasangan data aktual, yang menunjukkan bagaimana 'KEDISPLINAN (X)' berhubungan dengan 'KEHADIRAN (Y)' untuk setiap pengamatan. Dari visualisasi ini, kita dapat mengamati pola distribusi dan korelasi antara kedua variabel tersebut. Plot scatter membantu dalam mengidentifikasi apakah ada hubungan linear, non-linear, atau tidak ada hubungan sama

sekali antara 'KEDISPLINAN (X)' dan 'KEHADIRAN (Y)'. Misalnya, jika titik-titik cenderung mengikuti garis lurus, itu menunjukkan adanya hubungan linear. Jika sebarannya acak tanpa pola yang jelas, maka mungkin tidak ada hubungan antara kedua variabel. Dalam ini, titik-titik tersebar dengan beberapa kluster yang menunjukkan berbagai nilai 'KEHADIRAN (Y)' untuk tingkat 'KEDISPLINAN (X)' yang berbeda. Kita bisa melihat bahwa ada variasi dalam 'KEHADIRAN (Y)' bahkan ketika 'KEDISPLINAN (X)' berada pada tingkat yang sama, yang menunjukkan bahwa faktor lain mungkin mempengaruhi kehadiran selain kedisiplinan. Secara keseluruhan, plot scatter ini sangat berguna untuk analisis eksploratif, memungkinkan kita untuk mengidentifikasi pola, anomali, dan potensi hubungan antara variabel. Informasi yang diperoleh dari plot ini dapat digunakan sebagai dasar untuk analisis lebih lanjut, seperti pengujian statistik atau pembuatan model prediktif.



Gambar 26. Grafik Bar Jenis Kelamin

Gambar tersebut menunjukkan plot batang yang menampilkan frekuensi jenis kelamin dalam suatu dataset.

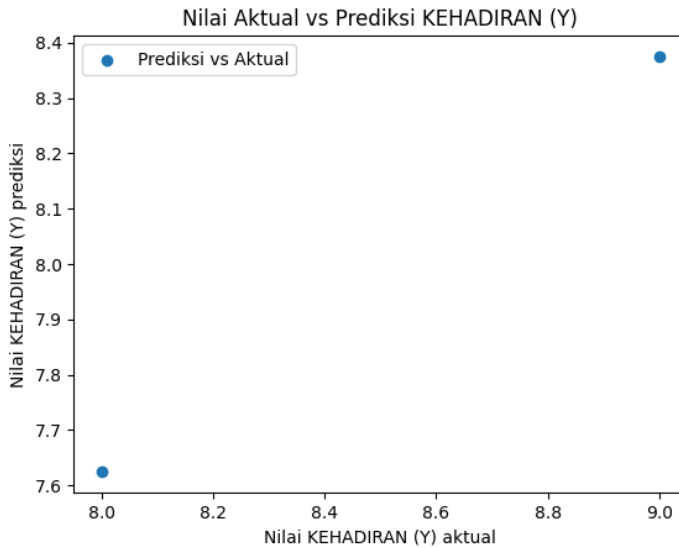
Sumbu horizontal menunjukkan kategori jenis kelamin, yaitu “LAKI-LAKI” dan “PEREMPUAN”, sementara sumbu vertikal menunjukkan frekuensi. Batang biru menunjukkan jumlah observasi untuk setiap kategori jenis kelamin. Berdasarkan plot, dapat dilihat bahwa frekuensi untuk “LAKI-LAKI” dan “PEREMPUAN” sama, yaitu 5. Artinya, dalam dataset tersebut, terdapat 5 orang laki-laki dan 5 orang perempuan. Plot ini memberikan gambaran yang jelas mengenai distribusi jenis kelamin dalam dataset. Informasi ini dapat berguna untuk analisis lebih lanjut, seperti untuk mengetahui proporsi jenis kelamin dalam suatu populasi atau untuk membandingkan distribusi jenis kelamin antara dua kelompok yang berbeda.

```
Error Absolute Rata-rata (MAE): 0.50  
Error Kuadrat Rata-rata (MSE): 0.27  
Error Kuadrat Rata-rata Akar (RMSE): 0.52  
Koefisien Determinasi ( $R^2$ ): -0.06
```

Gambar 27. Output Evaluasi Model

Model yang Anda gunakan memiliki performa yang kurang optimal, hal ini ditunjukkan dari nilai metrik evaluasinya. Error Absolut Rata-rata (MAE) sebesar 0,50 menunjukkan bahwa secara rata-rata, prediksi model menyimpang 0,50 unit dari nilai sebenarnya. Error Kuadrat Rata-rata (MSE) sebesar 0,27 mengindikasikan besarnya kesalahan kuadrat rata-rata model dalam memprediksi data. Error Kuadrat Rata-rata Akar (RMSE) sebesar 0,52 menunjukkan rata-rata selisih absolut antara prediksi model dan nilai sebenarnya. Nilai RMSE yang lebih kecil mengindikasikan model yang lebih akurat. Koefisien Determinasi (R^2) yang negatif (-0,06) menunjukkan bahwa model tidak dapat menjelaskan variasi data dengan baik, malah menunjukkan bahwa model tidak lebih baik dari prediksi dengan nilai konstan. Untuk meningkatkan performa model, beberapa

strategi yang dapat diterapkan adalah memilih model yang lebih kompleks, melakukan tuning hyperparameter, menerapkan teknik regularisasi, menggunakan teknik feature engineering, dan menggunakan teknik ensemble.



Gambar 28. Grafik Prediksi dan Aktual

Grafik tersebut menunjukkan hubungan antara nilai aktual dan prediksi dari model regresi untuk variabel 'KEHADIRAN'. Ada dua titik data yang menunjukkan nilai aktual dan prediksi. Titik pertama memiliki nilai aktual sekitar 8,0 dan nilai prediksi sekitar 7,6. Titik kedua memiliki nilai aktual sekitar 9,0 dan nilai prediksi sekitar 8,4. Dari grafik, terlihat bahwa model memiliki akurasi yang rendah dalam memprediksi nilai 'KEHADIRAN'. Hal ini karena terdapat perbedaan yang signifikan antara nilai prediksi dan nilai aktual untuk kedua titik data. Model memprediksi nilai yang lebih rendah daripada nilai aktual untuk titik pertama, dan memprediksi nilai yang lebih tinggi daripada nilai aktual untuk titik kedua.

Intercept: 14.38
Koefisien slope: -0.75
Persamaan linier: $Y = 14.38 + -0.75X$

Gambar 29. Hasil Persamaan Linear

gambar menunjukkan hasil analisis regresi linier sederhana. Intersep (titik potong dengan sumbu Y) sebesar 14.38 menunjukkan nilai Y ketika X bernilai nol. Koefisien slope (kemiringan) sebesar -0.75 menunjukkan bahwa setiap peningkatan satu unit pada X akan menyebabkan penurunan sebesar 0.75 unit pada Y. Persamaan linier yang terbentuk adalah $Y = 14.38 - 0.75 X$ $Y=14.38-0.75X$. Persamaan ini menggambarkan hubungan linier negatif antara variabel X dan Y. Dengan kata lain, peningkatan variabel X berkorelasi dengan penurunan variabel Y. Misalnya, jika X meningkat sebesar 1, nilai Y akan turun sebesar 0.75. Demikian pula, jika X meningkat sebesar 2, nilai Y akan turun sebesar 1.5. Pemahaman persamaan ini penting untuk memprediksi nilai Y berdasarkan perubahan pada X. Analisis regresi linier ini dapat digunakan dalam berbagai aplikasi seperti peramalan, evaluasi dampak perubahan variabel independen, dan pengambilan keputusan berbasis data yang lebih akurat.

Prediksi Nilai KEHADIRAN (Y) untuk KEDISPLINAN (X) yang diberikan:
Masukkan nilai KEDISPLINAN (X): 8
Prediksi nilai KEHADIRAN (Y) untuk KEDISPLINAN (X) = 8.0: 8.38
Apakah Anda ingin melakukan prediksi lagi? (y/t): y

Prediksi Nilai KEHADIRAN (Y) untuk KEDISPLINAN (X) yang diberikan:
Masukkan nilai KEDISPLINAN (X): 9
Prediksi nilai KEHADIRAN (Y) untuk KEDISPLINAN (X) = 9.0: 7.62
Apakah Anda ingin melakukan prediksi lagi? (y/t): y

Prediksi Nilai KEHADIRAN (Y) untuk KEDISPLINAN (X) yang diberikan:
Masukkan nilai KEDISPLINAN (X): 7
Prediksi nilai KEHADIRAN (Y) untuk KEDISPLINAN (X) = 7.0: 9.12
Apakah Anda ingin melakukan prediksi lagi? (y/t): y

Prediksi Nilai KEHADIRAN (Y) untuk KEDISPLINAN (X) yang diberikan:
Masukkan nilai KEDISPLINAN (X): 6
Prediksi nilai KEHADIRAN (Y) untuk KEDISPLINAN (X) = 6.0: 9.88
Apakah Anda ingin melakukan prediksi lagi? (y/t): t

Gambar 30. Aplikasi Pengujian Variabel X, Y

Gambar tersebut menunjukkan prediksi nilai kehadiran (Y) berdasarkan nilai kedisiplinan (X) menggunakan model regresi linier dengan persamaan $(Y = 14.38 - 0.75X)$. Proses prediksi dimulai dengan memasukkan nilai kedisiplinan (X). Misalnya, ketika nilai kedisiplinan adalah 8, prediksi nilai kehadiran dihitung sebagai $(Y = 14.38 - (0.75 \times 8) = 8.38)$. Selanjutnya, ketika nilai kedisiplinan adalah 9, prediksi nilai kehadiran adalah $(Y = 14.38 - (0.75 \times 9) = 7.63)$. Pada input ketiga, nilai kedisiplinan yang diberikan adalah 7, menghasilkan prediksi kehadiran sebesar $(Y = 14.38 - (0.75 \times 7) = 9.13)$. Untuk input terakhir, nilai kedisiplinan adalah 6, dan prediksi nilai kehadiran menjadi $(Y = 14.38 - (0.75 \times 6) = 9.88)$. Proses ini menunjukkan bagaimana model regresi linier digunakan untuk memprediksi nilai kehadiran berdasarkan nilai kedisiplinan. Setiap kali nilai kedisiplinan dimasukkan, model menghitung nilai kehadiran yang sesuai dengan menggunakan persamaan linier yang telah ditentukan. Prediksi ini membantu memahami dampak dari variasi kedisiplinan terhadap kehadiran. Dengan kata lain, nilai kehadiran dipengaruhi secara negatif oleh nilai

kedisiplinan; semakin tinggi nilai kedisiplinan, semakin rendah nilai kehadiran yang diprediksi. Hal ini ditunjukkan oleh koefisien slope negatif sebesar -0.75 dalam persamaan regresi. Proses ini dapat digunakan untuk melakukan beberapa prediksi dengan memasukkan nilai kedisiplinan yang berbeda dan memperoleh hasil yang sesuai, memberikan gambaran yang lebih jelas tentang hubungan antara kedua variabel tersebut.



VI

PENUTUP

Pentingnya penerapan teknologi machine learning, khususnya algoritma regresi linear, dalam mengevaluasi dan meningkatkan kinerja pegawai. Kajian ini berhasil menunjukkan bahwa dengan menganalisis data kinerja pegawai melalui regresi linear, dapat diperoleh wawasan mendalam mengenai faktor-faktor yang mempengaruhi kinerja, seperti kedisiplinan dan kehadiran.

Machine learning memungkinkan identifikasi pola tersembunyi dalam data yang mungkin tidak terlihat melalui metode analisis tradisional. Dalam kajian ini, algoritma regresi linear dipilih karena kesederhanaannya dan kemampuannya untuk memberikan prediksi yang akurat dan dapat diinterpretasikan. Proses persiapan data yang mencakup pengumpulan, pembersihan, dan normalisasi data merupakan langkah krusial yang memastikan validitas dan reliabilitas hasil analisis.

Evaluasi model menggunakan metrik seperti Mean Squared Error (MSE) dan koefisien regresi memberikan indikasi tentang hubungan linier antara variabel yang dianalisis. Hasilnya menunjukkan bahwa kedisiplinan memiliki pengaruh signifikan terhadap kehadiran, meskipun ada faktor lain yang mungkin mempengaruhi kinerja pegawai. Teknik cross-validation yang digunakan memastikan bahwa model tidak overfitting dan memiliki kemampuan generalisasi yang baik.

Penerapan regresi linear dalam ini tidak hanya memberikan prediksi yang akurat tetapi juga menyediakan dasar yang kuat untuk pengambilan keputusan yang lebih baik dan objektif di Dinas Pertanian. Dengan informasi yang diperoleh, dinas dapat mengembangkan

strategi yang lebih efektif untuk meningkatkan kinerja pegawai, seperti program pelatihan yang disesuaikan, kebijakan kedisiplinan yang lebih baik, dan sistem penghargaan yang tepat.

Secara keseluruhan, kajian ini berkontribusi pada peningkatan efisiensi dan produktivitas sektor pertanian. Dengan memanfaatkan machine learning, Dinas Pertanian dapat membuat keputusan yang lebih terinformasi dan berfokus pada peningkatan kesejahteraan pegawai dan petani, serta mendukung keberlanjutan dan kemajuan sektor pertanian di Kabupaten Labuhanbatu.

DAFTAR PUSTAKA

1. Y. Irawan, "Implementation of Data Mining for Determining Majors Using K-Means Algorithm in Students of Sma Negeri 1 Pangkalan Kerinci," *J. Appl. Eng. Technol. Sci.*, vol. 1, no. 1, pp. 17–29, 2019, doi: 10.37385/jaets.v1i1.18.
2. M. Nabeel, S. Majeed, M. J. Awan, H. Muslih-Ud-din, M. Wasique, and R. Nasir, "Review on effective disease prediction through data mining techniques," *Int. J. Electr. Eng. Informatics*, vol. 13, no. 3, pp. 717–733, 2021, doi: 10.15676/IJEEI.2021.13.3.13.
3. S. Baloch and M. S. Muhammad, "An Intelligent Data Mining-Based Fault Detection and Classification Strategy for Microgrid," *IEEE Access*, vol. 9, pp. 22470–22479, 2021, doi: 10.1109/ACCESS.2021.3056534.
4. H. Niu and E. A. Silva, "Crowdsourced Data Mining for Urban Activity: Review of Data Sources, Applications, and Methods," *J. Urban Plan. Dev.*, vol. 146, no. 2, 2020, doi: 10.1061/(asce)up.1943-5444.0000566.
5. A. Lia Hananto *et al.*, "Analysis of Drug Data Mining with Clustering Technique Using K-Means Algorithm," *J. Phys. Conf. Ser.*, vol. 1908, no. 1, 2021, doi: 10.1088/1742-6596/1908/1/012024.
6. Y. Indah Lestari and S. Defit, "Jurnal Informatika Ekonomi Bisnis Prediksi Tingkat Kepuasan Pelayanan Online Menggunakan Metode Algoritma C.45," vol. 3, pp. 148–154, 2021, doi: 10.37034/infed.v3i3.104.
7. F. Elfaladonna and A. Rahmadani, "Analisa Metode Classification-Decission Tree Dan Algoritma C.45 Untuk Memprediksi Penyakit Diabetes Dengan Menggunakan Aplikasi Rapid Miner,"

- SINTECH (Science Inf. Technol. J.*, vol. 2, no. 1, pp. 10–17, 2019, doi: 10.31598/sintechjournal.v2i1.293.
8. S. Febriani and H. Sulistiani, “Analisis Data Hasil Diagnosa Untuk Klasifikasi Gangguan Kepribadian Menggunakan Algoritma C4.5,” *89Jurnal Teknol. dan Sist. Inf.*, vol. 2, no. 4, pp. 89–95, 2021.
 9. Y. E. Putra and M. Fahrizal, “Rancang Bangun Menggunakan Metode Naive Bayes Dalam Sistem Pakar Penentuan Penyakit Tanaman Nanas Berbasis WEB,” *J. Portal Data*, vol. 1, no. 1, pp. 1–18, 2021.
 10. N. Qorimah, “Implementasi Metode Naïve Bayes Untuk Menentukan Minat Tabungan Pada Perumda BPR Bank Cirebon,” *J. Data Sci. Inform.*, vol. 2, no. 1, pp. 15–20, 2022.
 11. M. Hasnain, M. F. Pasha, I. Ghani, M. Imran, M. Y. Alzahrani, and R. Budiarto, “Evaluating Trust Prediction and Confusion Matrix Measures for Web Services Ranking,” *IEEE Access*, vol. 8, pp. 90847–90861, 2020, doi: 10.1109/ACCESS.2020.2994222.
 12. E. D. Sikumbang, F. Ariani, T. Handayani, and K. Ramanda, “Penerapan Algoritma C4.5 Untuk Menentukan Tingkat Kepuasan Pelanggan Kartu Telkomsel Prabayar,” *J. Sains Komput. Inform.*, vol. 6, no. September, pp. 811–820, 2022.
 13. A. Febriyani, G. K. Prayoga, and O. Nurdiawan, “Index Kepuasan Pelanggan Informa dengan Menggunakan Algoritma C.45,” *JURIKOM (Jurnal Ris. Komputer)*, vol. 8, no. 6, p. 330, 2021, doi: 10.30865/jurikom.v8i6.3686.
 14. P. Beras, R. Oleh, J. P. Gultom, and A. Rikki, “Implementasi Data Mining menggunakan Algoritma C-45 pada Data Masyarakat Kecamatan Garoga untuk Menentukan 11 Pola Implementasi Data Mining menggunakan Algoritma C-45 pada Data Masyarakat Kecamatan Garoga untuk Menentukan Pola Penerima Beras Raskin ARTICL,” vol. 02, no. 01, pp. 11–19, 2020.

15. H. B. Jatmiko, N. Tedi Kurniadi, and D. Maulana, "Optimasi Naïve Bayes Dengan Particle Swarm Optimization Untuk Analisis Sentimen Formula E-Jakarta," *J. Autom. Comput. Inf. Syst.*, vol. 2, no. 1, pp. 22–30, 2022, doi: 10.47134/jacis.v2i1.35.
16. A. Muzaki and A. Witanti, "Sentiment Analysis of the Community in the Twitter To the 2020 Election in Pandemic Covid-19 By Method Naive Bayes Classifier," *J. Tek. Inform.*, vol. 2, no. 2, pp. 101–107, 2021, doi: 10.20884/1.jutif.2021.2.2.51.
17. T. R. S. Hari and S. Sumijan, "Sistem Pakar dengan Menggunakan Metode Naive Bayes dalam Mengidentifikasi Penyakit Karies pada Gigi Manusia," *J. Sistim Inf. dan Teknol.*, vol. 3, pp. 233–238, 2021, doi: 10.37034/jsisfotek.v3i4.71.

TENTANG PENULIS



Rahmadani Ritonga Lahir di Sigambal (Rantauprapat) 27 November 2000. Saat ini penulis tinggal di Sigambal Sumatra Utara. Pendidikan tinggi tempuh pada S-1 di Fakultas Sains dan Teknologi (Lulus 2024) Universitas Labuhanbatu dengan meraih gelar S.Kom. Penulis baru buat buku pertama kalinya dengan judul

“ANALISIS ALGORITMA REGRESI LINEAR UNTUK KINERJA PEGAWAI DINAS PERTANIAN DENGAN MACHINE LEARNING” Penulis berharap awal mula buku ini menjadi berkat untuk mencapai karier penulis di hari selanjutnya, sehingga penulis mampu untuk mengembangkan bakat sebagai seorang penulis sesuai dengan harapan dan cita-cita penulis sejak kecil.



Ibnu Rasyid Munthe, S.T., M.Kom. lahir di Rantau Prapat, Kabupaten Labuhanbatu pada tahun 1987. Ia menempuh pendidikan strata satu (S1) di Universitas Nurtanio Bandung dan melanjutkan pendidikan strata dua (S2) di Universitas Putra Indonesia (YPTK) Padang. sebagai sosok yang produktif, dengan aktif menulis

artikel ilmiah dan buku. Dedikasinya dalam dunia pendidikan tercermin dari perannya sebagai dosen di Universitas Labuhanbatu. Tidak hanya di ruang kelas, Ibnu juga memanfaatkan teknologi digital untuk menyebarkan ilmu pengetahuan. Melalui channel YouTube @manjaddawajada2022, ia berbagi berbagai informasi pembelajaran yang bermanfaat. Channel ini menjadi wadah untuk menyampaikan

pengetahuan kepada khalayak luas, dengan harapan dapat memberikan dampak positif dan inspirasi bagi banyak orang. Terus berinovasi dalam metode pengajaran dan berbagi ilmu, berusaha membawa perubahan positif dalam dunia pendidikan, khususnya di daerah asalnya, Labuhanbatu. Dengan semangat yang tinggi, ia berharap dapat berkontribusi lebih banyak lagi untuk kemajuan pendidikan di Indonesia.



Angga Putra Juledi, S.Kom., M.Kom. Lahir di kota Padang pada tanggal 19 Juli 1994. Dalam menempuh pendidikan dimulai dari sekolah Dasar SDN 19 Padang tamat tahun 2006, SMPN 3 Padang tamat pada tahun 2009, dan di SMA Pertiwi 2 Padang tamat pada tahun 2012. Lalu melanjutkan ke Pendidikan perguruan tinggi swasta yaitu S1 (sarjana universitas putra Indonesia “YPTK” Padang lulus pada tahun 2018 dengan jurusan system informasi, dan melanjutkan program pascasarjana (S2) di universitas putra Indonesia “YPTK” Padang pada tahun 2019 program studi Teknik informatika. Konsentrasi system informasi. Saya mengabdikan diri sebagai salah satu dosen di bimbing ilmu komputer pada fakultas sains dan teknologi dengan program studi system informasi di universitas Labuhanbatu dan menjadi dosen tetap pada tahun 2020 pada kampus tersebut. Pada saat menjadi bagian struktural di universitas Labuhanbatu sebagai kepala bagian sumber daya manusia periode 2023 s/d 2027. Buku pertama terbit pada 31 Desember 2021 dengan judul *Internetworking dan TCP/IP*. Buku kedua terbit pada tanggal 17 Oktober 2023 dengan judul *Panduan Belajar HTML, CSS dan Javascript untuk pemula*. Buku ke-3 terbit pada 09 Mei 2024 dengan judul *dasar-dasar system basis data*.



Masrizal, S.Kom, M.Kom. NIDN 0105039401. Lahir di Manggung 05 Maret 1994. Proses dalam menempuh pendidikan dimulai dari Sekolah Dasar SD.N 13 Manggung, SMP.N 7Pariaman, SMA.N 4 Pariaman dan melanjutkan ke perguruan tinggi swasta dengan jurusan Sistem Informasi S1 (Sarjana) di Universitas Putra Indonesia “YPTK” Padang. Tahun lulusan 2016 dan melanjutkan Magister tahun 2016 Universitas Putra Indonesia “YPTK” Padang. Tahun lulusan 2018 dengan jurusan Sistem Informasi, Fakultas Teknologi Informasi. Pada tahun 2018. Saya mengabdikan diri sebagai salah satu dosen di bidang Ilmu Komputer di Fakultas Sain dan Teknologi (FST) Prodi Sistem Informasi di Universitas Labuhan Batu salah satu menjadi dosen tetap di kampus tersebut. Syukur Alhamdulillah sekarang saya mengabdikan diri di bawah Wakil Rektor III di bidang Pengembangan, Kerjasama dan Alumni sebagai Ka.biro Bagian Pengembangan, Kerjasama dan Alumni serta Devisi Tracer Studi di Universitas Labuhanbatu- Sumatera Utara..

OPTIMALISASI KINERJA PEGAWAI PERTANIAN

Studi Kasus Penggunaan
Algoritma Regresi Linear



Buku ini lahir dari kebutuhan mendesak untuk meningkatkan efisiensi dan efektivitas pegawai Dinas Pertanian dalam menghadapi tantangan globalisasi dan perkembangan teknologi. Algoritma regresi linear, sebagai salah satu teknik machine learning yang paling mendasar dan luas penggunaannya, dipilih dalam penelitian ini untuk mengidentifikasi dan mengoptimalkan faktor-faktor yang mempengaruhi kinerja pegawai. Melalui pendekatan ini, buku ini tidak hanya menyajikan teori dan konsep dasar, tetapi juga memaparkan aplikasi praktis yang dapat diterapkan di berbagai konteks organisasi pertanian.

Kami berharap buku ini dapat menjadi referensi berharga bagi akademisi, praktisi, dan pengambil kebijakan di sektor pertanian. Dengan pemahaman yang lebih baik tentang bagaimana data dapat digunakan untuk mengoptimalkan kinerja, kami percaya bahwa inovasi dan peningkatan berkelanjutan dapat dicapai.

