Penerapan Data Mining Untuk Memprediksi Penjualan Makanan Pada Kopi Krintji Menggunakan Algoritma Regresi Linear dan C4.5

TUGAS AKHIR

Muhammad Rafi Nauf Akbar
41518010026

PROGRAM STUDI TEKNIK INFORMATIKA
FAKULTAS ILMU KOMPUTER
UNIVERSITAS MERCU BUANA
JAKARTA
2021/2022

https://lib.mercubuana.ac.id/

Penerapan Data Mining Untuk Memprediksi Penjualan Makanan Pada Kopi Krintji Menggunakan Algoritma Regresi Linear dan C4.5

Tugas Akhir

Diajukan Untuk Melengkapi Salah Satu Syarat
Memperoleh Gelar Sarjana Komputer

Oleh:
Muhammad Rafi Nauf Akbar
41518010026

UNIVERSITAS
MERCU BUANA

PROGRAM STUDI TEKNIK INFORMATIKA
FAKULTAS ILMU KOMPUTER
UNIVERSITAS MERCU BUANA
JAKARTA
2021/2022

https://lib.mercubuana.ac.id/

LEMBAR PERNYATAAN ORISINALITAS

Yang bertanda tangan dibawah ini:

NIM	: 41518010026
Nama	: Muhammad Rafi Nauf Akbar
Judul Tugas Akhir	: Penerapan Data Mining Untuk Memprediksi Penjualan Makanan Pada Kopi Krintji Menggunakan Algoritma Regresi Linear dan C4.5

Menyatakan bahwa Laporan Tugas Akhir saya adalah hasil karya sendiri dan bukan plagiat. Apabila ternyata ditemukan didalam laporan Tugas Akhir saya terdapat unsur plagiat, maka saya siap untuk mendapatkan sanksi akademik yang terkait dengan hal tersebut.

Jakarta, 29 Juli 2022

Muhammad Rafi Nauf Akbar

UNIVERSITAS
MERCU BUANA

https://lib.mercubuana.ac.id/

SURAT PERNYATAAN PERSETUJUAN PUBLIKASI TUGAS AKHIR

Sebagai mahasiswa Universitas Mercu Buana, saya yang bertanda tangan di bawah ini :

Nama Mahasiswa	:	Muhammad Rafi Nauf Akbar
NIM	:	41518010026
Judul Tugas Akhir	:	Penerapan Data Mining Untuk Memprediksi Penjualan Makanan Pada Kopi Krintji Menggunakan Algoritma Regresi Linear dan C4.5

Dengan ini memberikan izin dan menyetujui untuk memberikan kepada Universitas Mercu Buana **Hak Bebas Royalti Noneksklusif** (*None-exclusive Royalty Free Right*) atas karya ilmiah saya yang berjudul diatas beserta perangkat yang ada (jika diperlukan).

Dengan Hak Bebas Royalti/Noneksklusif ini Universitas Mercu Buana berhak menyimpan, mengalihmedia/formatkan, mengelola dalam bentuk pangkalan data (*database*), merawat dan mempublikasikan tugas akhir saya.

Selain itu, demi pengembangan ilmu pengetahuan di lingkungan Universitas Mercu Buana, saya memberikan izin kepada Peneliti di Lab Riset Fakultas Ilmu Komputer, Universitas Mercu Buana untuk menggunakan dan mengembangkan hasil riset yang ada dalam tugas akhir untuk kepentingan riset dan publikasi selama tetap mencantumkan nama saya sebagai penulis/pencipta dan sebagai pemilik Hak Cipta.

Demikian pernyataan ini saya buat dengan sebenarnya.

Jakarta, 29 Juli 2022

METERAI TEMPEL

Muhammad Rafi Nauf Akbar

UNIVERSITAS MERCU BUANA

iii

https://lib.mercubuana.ac.id/

SURAT PERNYATAAN LUARAN TUGAS AKHIR

Sebagai mahasiswa Universitas Mercu Buana, saya yang bertanda tangan di bawah ini :

Nama Mahasiswa	:	Muhammad Rafi Nauf Akbar
NIM	:	41518010026
Judul Tugas Akhir	:	Penerapan Data Mining Untuk Memprediksi Penjualan Makanan Pada Kopi Krintji Menggunakan Algoritma Regresi Linear dan C4.5

Menyatakan bahwa :

1. Luaran Tugas Akhir saya adalah sebagai berikut :

No	Luaran	Jenis		Status	
1	Publikasi Ilmiah	Jurnal Nasional Tidak Terakreditasi		Diajukan	
		Jurnal Nasional Terakreditasi	✓		
		Jurnal International Tidak Bereputasi		Diterima	✓
		Jurnal International Bereputasi			
	Disubmit/dipublikasikan di :	Nama Jurnal	: Jurnal Ilmu Teknik dan Komputer (JITKOM)		
		ISSN	: 2548-740X		
		Link Jurnal	: https://publikasi.mercubuana.ac.id/index.php/jitkom		
		Link File Jurnal Jika Sudah di Publish	:		

2. Bersedia untuk menyelesaikan seluruh proses publikasi artikel mulai dari submit, revisi artikel sampai dengan dinyatakan dapat diterbitkan pada jurnal yang dituju.
3. Diminta untuk melampirkan scan KTP dan Surat Pernyataan (Lihat Lampiran Dokumen HKI), untuk kepentingan pendaftaran HKI apabila diperlukan

Demikian pernyataan ini saya buat dengan sebenarnya.

Jakarta, 29 Juli 2022

METERAI TEMPEL 154AJX980362480

Muhammad Rafi Nauf Akbar

UNIVERSITAS MERCU BUANA

https://lib.mercubuana.ac.id/

LEMBAR PERSETUJUAN PENGUJI

NIM	:	41518010026
Nama	:	Muhammad Rafi Nauf Akbar
Judul Tugas Akhir	:	Penerapan Data Mining Untuk Memprediksi Penjualan Makanan Pada Kopi Krintji Menggunakan Algoritma Regresi Linear dan C4.5

Tugas Akhir ini telah diperiksa dan disidangkan sebagai salah satu persyaratan untuk memperoleh gelar Sarjana pada Program Studi Teknik Informatika, FakultasIlmu Komputer, Universitas Mercu Buana.

Jakarta, 29 Juli 2022

(Wawan Gunawan, S.Kom, MT)

UNIVERSITAS
MERCU BUANA

https://lib.mercubuana.ac.id/

LEMBAR PERSETUJUAN PENGUJI

NIM	:	41518010026
Nama	:	Muhammad Rafi Nauf Akbar
Judul Tugas Akhir	:	Penerapan Data Mining Untuk Memprediksi Penjualan Makanan Pada Kopi Krintji Menggunakan Algoritma Regresi Linear dan C4.5

Tugas Akhir ini telah diperiksa dan disidangkan sebagai salah satu persyaratan untuk memperoleh gelar Sarjana pada Program Studi Teknik Informatika, FakultasIlmu Komputer, Universitas Mercu Buana.

Jakarta, 29 Juli 2022

(Harni Kusniyati, M.Kom)

UNIVERSITAS
MERCU BUANA

https://lib.mercubuana.ac.id/

LEMBAR PERSETUJUAN PENGUJI

NIM	:	41518010026
Nama	:	Muhammad Rafi Nauf Akbar
Judul Tugas Akhir	:	Penerapan Data Mining Untuk Memprediksi Penjualan Makanan Pada Kopi Krintji Menggunakan Algoritma Regresi Linear dan C4.5

Tugas Akhir ini telah diperiksa dan disidangkan sebagai salah satu persyaratan untuk memperoleh gelar Sarjana pada Program Studi Teknik Informatika, FakultasIlmu Komputer, Universitas Mercu Buana.

Jakarta, 10 Agustus 2022

Penguji

Umniy Salamah, ST., MMSI

UNIVERSITAS
MERCU BUANA

https://lib.mercubuana.ac.id/

LEMBAR PENGESAHAN

NIM	:	41518010026
Nama	:	Muhammad Rafi Nauf Akbar
Judul Tugas Akhir	:	Penerapan Data Mining Untuk Memprediksi Penjualan Makanan Pada Kopi Krintji Menggunakan Algoritma Regresi Linear dan C4.5

Tugas Akhir ini telah diperiksa dan disidangkan sebagai salah satu persyaratan untuk memperoleh gelar Sarjana pada Program Studi Teknik Informatika, Fakultas Ilmu Komputer, Universitas Mercu Buana.

Jakarta, 29 Juli 2022

Menyetujui,

(Eugenius Kau Suni, ST, MT)
Dosen Pembimbing

Mengetahui,

(Wawan Gunawan, S.Kom, MT)	(Ir. Emil R. Kaburuan, Ph.D., IPM.)
Koord. Tugas Akhir Teknik Informatika	Ka. Prodi Teknik Informatika

UNIVERSITAS MERCU BUANA

https://lib.mercubuana.ac.id/

KATA PENGANTAR

Puji syukur kita panjatkan kehadirat Allah SWT atas segala rahmat dan karunia-Nya, dikarenakan Tugas Akhir yang berjudul "Penerapan Data Mining Untuk Memprediksi Penjualan Makanan dan Minuman Menggunakan Algoritma C 4.5 dan Regresi Linear (Studi Kasus : Kopi Krinjti)" dapat diselesaikan dalam jangka waktu yang ditentukan. Laporan Tugas Akhir ini dibuat sebagai syarat untuk LULUS sebagai sarjana Ilmu Komputer dari Universitas Mercu Buana.

Penulis menyadari bahwa pembuatan Tugas Akhir ini tidak terlepas dari bantuan dan bimbingan berbagai pihak. Oleh karena itu, penulis mengucapkan terima kasih kepada :

1. Kedua orang tua saya yang telah membesarkan saya, selalu mendoakan yang terbaik untuk saya agar dapat menyelesaikan tugas akhir dan juga tak luput terus mendukung saya karena bisa menyelesaikan kuliah dengan baik.
2. Bapak Dr. Harwikarya, MT selaku Dosen Pembimbing Akademik yang telah memantau dan membantu studi saya hingga akhir.
3. Bapak Eugenius Kau Suni, ST, MT selaku Dosen Pembimbing Tugas Akhir yang telah membimbing serta memberikan nasihat dalam penyusunan Tugas Akhir sampai selesai.
4. Sahabat dan teman – teman saya yang telah membantu, mensupport, serta memberikan masukkan dalam menyelesaikan Tugas Akhir.
5. Semua pihak dan personal yang tidak disebutkan satu per satu yang terlibat dalam pembuatan Tugas Akhir ini sehingga dapat selesai dengan baik.

Akhir kata, penulis berharap Tugas Akhir ini dapat menjadi referensi untuk dikembangkan lebih baik lagi kedepannya, karena penulis menyadari hasil Tugas Akhir ini masih jauh dari sempurna. Masih terdapat kekurangan dalam eksperimen, cara penjelasan maupun kekeliruan penulisan. Untuk itu, kritik dan saran pembaca sangat dihargai dan diharapkan. Semoga Tugas Akhir ini dapat memberikan manfaat bagi para pembaca.

https://lib.mercubuana.ac.id/

Jakarta, 29 Juli 2022
Penulis

Muhammad Rafi Nauf Akbar

UNIVERSITAS
MERCU BUANA

https://lib.mercubuana.ac.id/

DAFTAR ISI

HALAMAN SAMPUL i
HALAMAN JUDUL……..i
LEMBAR PERNYATAAN ORISINALITAS ii
SURAT PERNYATAAN PERSETUJUAN PUBLIKASI TUGAS AKHIR... iii
SURAT PERNYATAAN LUARAN TUGAS AKHIR..... iv
LEMBAR PERSETUJUAN PENGUJI..... v
LEMBAR PENGESAHAN vii
ABSTRAK ix
ABSTRACT x
KATA PENGANTAR..... xi
DAFTAR ISI..... xiii
NASKAH JURNAL 1
KERTAS KERJA..... 8
BAB 1. LITERATUR REVIEW 9
BAB 2. ANALISIS DAN PERANCANGAN.....17
BAB 3. SOURCE CODE 20
BAB 4. DATASET..... 34
BAB 5. TAHAPAN EKSPERIMEN..... 37
BAB 6. HASIL SEMUA EKSPERIMEN..... 41
DAFTAR PUSTAKA 48
LAMPIRAN DOKUMEN HAKI..... 51
LAMPIRAN KORESPONDENSI 54

UNIVERSITAS
MERCU BUANA

https://lib.mercubuana.ac.id/

Penerapan Data Mining Untuk Memprediksi Penjualan Makanan Pada Kopi Krintji Menggunakan Algoritma Regresi Linear dan C4.5

Muhammad Rafi Nauf Akbar 1*, Eugenius Kau Suni 2**
*Jurusan Teknik Informatika, Fakultas Ilmu Komputer,
Universitas Mercu Buana
raffinauff@gmail.com
eugenius@mercubuana.ac.id

ABSTRACT

Penelitian yang digagas kali ini menggunakan algoritma C 4.5 dan regresi linear untuk dapat memprediksi penjualan makanan dan minuman pada Kopi Krintji guna memastikan stok selalu tersedia dan kebutuhan customer selalu terpenuhi. Penelitian ini bertujuan untuk menerapkan data mining dalam memprediksi penjualan makanan dan minuman pada Kopi Krintji menggunakan algoritma C 4.5 dan memprediksi ketersediaan stok makanan dan minuman pada Kopi Krintji menggunakan algoritma regresi linear. Metode yang digunakan pada penelitian ini menggunakan metode prediksi. Hasil yang didaptkan menggunakan perhitungan Google Colab. Pada penelitian menggunakan parameter MAE (Mean absolute error), MSE (mean squared error) dan R2 score. Skenario pengujian menggunakan algoritma regresi linear dan C 4.5 dengan parameter MAE, MSE, dan R2 Score mendapatkan hasil yang standar. Pada algoritma regresi linear nilai MAE 59697.34, nilai MSE 9224945948.40, dan R2 Score 0.64. Sedangkan pada algoritma C 4.5 nilai MAE 59936.13, nilai MSE 9128114236.26 dan nilai R2 Score 0.64.

Kata Kunci: prediksi, data mining, regresi linear, C4.5

PENDAHULUAN

A. Latar Belakang

Perkembangan kafe di Indonesia sangat pesat, terlebih lagi anak muda bahkan orang dewasa sering kali berkumpul bersama di waktu senggang. Bahkan selama masa pandemi Covid-19 budaya nongkrong di kafe tetap berjalan dengan protokol kesehatan yang ketat (Kusbiyanto, April Heri, 2021). Salah satu kafe yang menarik perhatian pengunjung dan termasuk sangat ramai adalah Kopi Krintji di Kebayoran Baru, Jakarta Selatan. Kopi Krintji menjadi salah satu kafe yang sedang tren saat ini dimana hampir setiap hari selalu ramai oleh pengunjung. Kafe ini didirikan oleh Anita Sukarman sejak tahun 2017 dan terkenal murah untuk suatu kafe di daerah Jakarta Selatan dengan suasana yang sangat sejuk.

Kopi Krintji menjual berbagai macam menu yang sangat disukai oleh para pengunjung mulai dari pilihan berbagai jenis kopi, pasta, burger, dan lainnya. Tingginya penjualan makanan dan minuman di Kopi Krintji membuat stok persediaan cepat habis karena

https://lib.mercubuana.ac.id/

ramainya pengunjung yang memesan makanan dan minuman tersebut. Beberapa jenis kopi yang sering kali menjadi favorit pengunjung antara lain kopi arabika, robusta, liberika, dan ekselesa (Wanita et al., 2021).

Banyaknya pengunjung yang datang ke Kopi Krintji, sering kali kafe tersebut kehabisan stok makanan dan minuman, sehingga saat pelanggan memesan, pesanan tersebut tidak tersedia. Hal ini mengakibatkan kekecewaan terhadap *customer*. Oleh karena itu penelitian ini membantu Kopi Krintji dalam memprediksi stok makanan dan minuman kedepannya, sehingga tidak terjadi lagi kekurangan atau kehabisan persediaan. Pendekatan teknologi informatika dalam membantu Kopi Krintji ini menjadi sangat penting karena sudah sangat lazim strategi penjualan makanan dan minuman diberbagai tempat sudah menggunakan bantuan teknologi maju antara lain penggunaan aplikasi dalam pemesanan dan juga prediksi *trend* penjualan (Setiawan et al., 2020).

Pada Kopi Krintji, data pemesanana makanan dan minuman selalu berulang dan jumlahnya sangat banyak. Hal ini dapat diproses lebih lanjut menggunakan pendekatan *data mining* untuk mengetahui *trend* data yang timbul pada kurun waktu tertentu. Sebab *data mining* mampu menghitung nilai persediaan makanan dan minuman yang dibutuhkan di waktu yang akan datang (Meizar et al., 2022).

Pada penelitian sebelumnya berjudul "Penerapan Data Mining Untuk Prediksi Penjualan Walpaper Menggunakan Algoritma C4.5" dimana metode ini khususnya algoritma C4.5 sangat bermanfaat dalam memprediksi faktor *customer* dalam pembelian *wallpaper*. Pada analisinya diperoleh jumlah motif *wallpaper* menjadi faktor tertinggi dalam penjualan *wallpaper*. Dalam melakukan analisinya, peneliti menggunakan algoritma C4.5. Hasilnya diketahui bahwa faktor dari segi harga, ukuran, kualitas bahan dan warna tidak begitu mempengaruhi penjualan walpaper, karena masih banyak pelanggan yang meminati walpaper dengan harga yang terbilang mahal, ukuran yang kecil, kualitas wallpaper yang baik, dan warna yang sedikit. Sehingga penjual harus menjaga persediaan barang di gudang secara teratur (Eska, 2018).

Pada penelitian lain yang berjudul "Prediksi Kebutuhan Obat Menggunakan Regresi Linier" disimpulkan bahwa jenis obat dan berapa banyak kebutuhan obat yang dibutuhkan oleh salah satu mitra dapat diprediksi melalui data pesanan dan data penjualan yang ada. Prediksi menggunakan regresi linear yang dilakukan dapat membantu distributor dalam memenuhi kebutuhan obat yang dibutuhkan (Monika, Sinar; Rakhman, 2017).

Penelitian yang digagas kali ini menggunakan algoritma C 4.5 dan regresi linear untuk dapat memprediksi penjualan makanan dan minuman pada Kafe Kopi Krintji guna memastikan stok selalu tersedia dan kebutuhan *customer* selalu terpenuhi. Algoritma C 4.5 dikembangkan oleh Ross Quinlan yang digunakan untuk memproses *data mining* dengan membentuk sebuah pohon keputusan (*decission tree*) (Lukhayu Pritalia, 2018). Algoritma C 4.5 ini dikenal handal untuk digunakan dalam menghitung nilai kecenderungan terhadap sekumpulan data tertentu. Data penjualan makanan dan minuman di Kopi Krintji dapat diproses lebih lanjut menggunakan algoritma C 4.5 untuk mengetahui prediksi stok makanan dan minuman yang ada. Sedangkan Regresi Linear digunakan untuk memberikan nilai prediksi ketersediaan makanan dan minuman yang dijual (Suryanto, 2019). Ini secara khusus melakukan pengolahan data *grossales* dan *item sold* untuk mengetahui kecenderungan makanan dan minuman yang paling banyak terjual dan diminati pembeli yang akan berkaitan langsung dengan stok persediaan makanan dan minuman pada Kafe Kopi Krintji.

B. Rumusan Masalah

Berdasarkan latar belakang di atas, maka rumusan masalah dalam penelitian ini adalah :

1. Bagaimana penerapan *data mining* untuk memprediksi penjualan makanan dan minuman menggunakan algoritma C 4.5 pada Kopi Krintji?
2. Bagaimana penerapan *data mining* untuk memprediksi ketersediaan stok makanan dan minuman menggunakan algoritma regresi linear pada Kopi Krintji?

C. Tujuan

Berdasarkan rumusan masalah di atas maka yang menjadi tujuan dari penelitian ini adalah:

1. Untuk menerapkan *data mining* dalam memprediksi penjualan makanan dan minuman pada Kopi Krintji menggunakan algoritma C4.5
2. Untuk memprediksi ketersediaan stok makanan dan minuman pada Kopi Krintji menggunakan algoritma regresi linear.

STUDI LITERATUR

A. Prediksi

https://lib.mercubuana.ac.id/

Prediksi merupakan ramalan atau perkiraan tentang suatu peristiwa apa yang terjadi pada masa yang akan datang. Prediksi menunjukkan apa yang akan terjadi pada suatu keadaan atau situasi tertentu. Hasil prediksi seringkali menjadi masukan atau input bagi proses perencanaan dan pengambilan keputusan (Kurniawan, 2019). Menurut kamus besar bahasa Indonesia, prediksi adalah hasil dari kegiatan memprediksi atau meramal atau memperkirakan nilai pada masa yang akan datang dengan menggunakan data masa lalu (Kafil, 2019).

B. Data Mining

Data mining adalah sebuah proses percarian secara otomatis informasi yang berguna dalam tempat penyimpanan data berukuran besar (Triyandana et al., 2022). Istilah lain yang sering digunakan adalah *knowledge discovery (mining) in databases (KDD),* kegiatan yang melakukan pemakaian data historis yang bertujuan mencari potensi data yang akan di pakai untuk diteliti, setelah itu akan dianalisis agar mudah dipahami (Wijaya & Dwiasnati, 2020).

C. Algoritma C.45

Ross Quinlan seorang peniliti bidang *data mining* dan teori keputusan merumuskan algoritma C4.5 untuk memproses *data mining* dengan membentuk sebuah pohon keputusan (*Decission Tree*) (Lukhayu Pritalia, 2018). Pada algoritma C 4.5, nilai *entropy*, *gain*, dan *split* yang terdapat di data training yang ada akan menghasilkan *gain ratio*. Atribut yang memiliki *gain ratio* terbesar dipilih untuk membuat simpul akar pohon. Lalu menghitung nilai *entropy*, *gain* dan *split* dari masing-masing atribut dengan menghilangkan atribut yang telah dipilih sebelumnya (Dwiasnati & Devianto, 2019) . Atribut yang memiliki nilai *gain ratio* terbesar dipilih untuk membuat simpul pohon . Ulangi perhitungan tersebut hingga semua atribut memiliki kelas. Jika semua atribut telah memiliki kelas maka tampilkan pohon keputusan awal dan *generate* aturan keputusan awal.

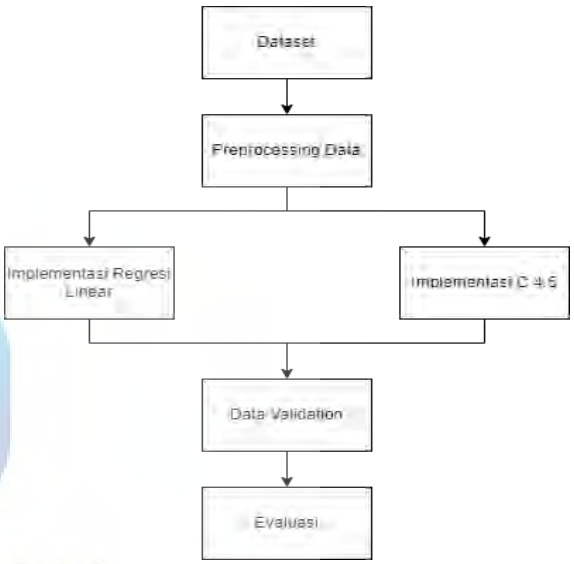
D. Algoritma Regresi Linear

Sebuah metode yang mengukur hubungan korelasi antara dua variabel atau lebih yang berguna untuk data prediksi melalui garis lurus (Suryanto, 2019). Regresi Linear juga merupakan metode statistik yang akan melakukan prediksi dengan menggunakan pengembangan hubungan matematis antara variabel, yaitu variabel independen (X) dan variabel dependen (Y) (Ayuni & Fitrianah, 2019). Analisis ini untuk mengetahui arah hubungan antara variabel independen dengan variabel dependen apakah positif atau negatif dan untuk memprediksi nilai dari variabel dependen apabila nilai variabel independen mengalami kenaikan atau penurunan (Fitriani et al., 2021).

METHODOLOGI

A. Tahapan Penelitian

Pada penelitian ini, terdapat langkah – langkah agar yang ditetapkan tercapai. Berikut adalah langkah – langkah yang akan digunakan.



Gambar 1. Diagram Metodologi

1. *Dataset*

Data yang akan digunakan adalah dataset yang didapatkan dari Kopi Krintji pada periode Bulan Juli – November dengan total dataset yang digunakan sebanyak 2000 data. Penelitian ini yaitu data persediaan stok Kopi Krintji dari Bulan Juli – November 2021 yang akan ditarik ke *database*.

2. *Preprocessing Data*

Pada *preprocessing data* dilakukan langkah awal yaitu mengubah data mentah yang memungkinkan dijadikan *data mining* untuk meningkatkan kualitas data. Pada *cleaning data* dilakukan pembersihan data dari *missing value*, dimana atribut yang memiliki duplikat data dan *data transformasi* mengubah data kedalam bentuk yang akan dianalisis dengan proses perhitungan regresi linear dan C 4.5.

3. Implementasi Algoritma Regresi Linear dan C 4.5

   a. Algoritma Regresi Linear

   Dalam melakukan penelitian pada regresi linear, terdapat langkah - langkah yang akan di lakukan, yang terdiri dari:

https://lib.mercubuana.ac.id/

1) menentukan tujuan dalam melakukan analisis regresi linear dengan cara mempelajari hubungan yang di peroleh dari analisis ini dan di nyatakan dalam persamaan matematika yang dapat menyatakan hubungan antar masing - masing variabel.
2) Mengindentifikasi Variabel Faktor penyebab (X) dan Variabel faktor akibat (Y), pada analisis ini variabel faktor penyebab (X) adalah *item sold* dan variabel faktor akibat (Y) adalah *gross sales*
3) Melakukan pengumpulan data, jika langkah 1 dan 2 sudah di lakukan maka pada langkah ini kita dapat mengumpulkan data dari tiap masing - masing variabel.
4) Melakukan perhitungan menggunakan *Google Colab*

b. Algoritma C 4.5

Pada algoritma C 4.5 akan melatih dengan *data training*. Tahapan ini juga akan mencari nilai terbaik dari beberapa parameter yang akan digunakan.

4. Validasi Model

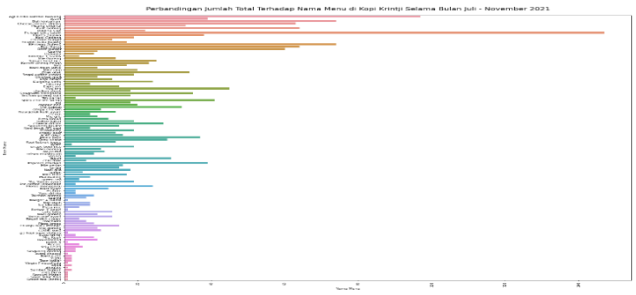
Pada tahapan validasi, Tujuan dari validasi adalah memberikan keyakinan bahwa model benar-benar mempresentasikan sistem sesuai pemodelan yang dilakukan, sehingga semua kesimpulan yang mungkin ditarik dari hasil simulasi mengenai sistem dapat diterima dan dipertanggungjawabkan sampai batas-batas tertentu.

5. Evaluasi

Pada tahapan evaluasi ini diperlukan untuk menganalisa dan mengukur sejauh mana keakuratan hasil yang telah dicapai. Lalu hasil evaluasi akan diberikan dengan menggunakan *classification report* yang terdiri dari MSE, MAE dan R2 *Score*.

HASIL DAN DISKUSI

Berikut ini adalah pembelian pada menu Kopi Krintji pada periode bulan Juli – November 2021.



Gambar 2. Grafik Pembelian Menu Kopi Krintji

Dari grafik batang diatas, peneliti dapat mengetahui data yang paling banyak dan sedikit di dalam kolom "*Item Name*". Kolom ini menunjukkan menu yang sering dipesan oleh *customer* Kopi Krintji dan juga menu yang jarang dipesan oleh customer Kopi Krintji pada periode bulan Juli – November 2021. Berikut adalah *detail* informasi yang didapatkan dari grafik diatas :

1) Menu yang selalu terjual atau dapat dikatakan menu paling laris adalah menu Es Kopi susu gadis yang berhasil dipesan oleh customer Kopi Krintji sebanyak 147 kali selama periode bulan Juli - November 2021.
2) Sedangkan menu yang jarang dipesan atau dapat dikatakan sebagai menu tidak laris karena hanya sedikit dipesan oleh customer Kopi Krintji adalah menu-menu seperti Nasgor + Gadis, Rotbar + gadis, go kopi susu stobery, paket 6, *extra cheese*, *Virgin Pinacollada*, *affogato*, *cold brew*, Sambel Matah, Saus Telur Asin, dan *Green Tea* (Botol) yang kesebelas menu tersebut hanya memiliki total penjualan sebanyak 1 kali saja selama periode bulan Juli - November 2021.

Dibawah ini adalah grafik *heatmap* untuk melihat hubungan korelasi timbal balik dari masing – masing kolom.

Gambar 3. Grafik *Heatmap*

Grafik pada Gambar 3. Grafik *Heatmap* menunjukan nilai korelasi atau nilai keterhubungan yang mempunyai sifat timbal balik antar masing - masing kolom. Dalam mengetahui nilai korelasi dapat melihat pada data yang bersifat numerik, sedangkan data yang bersifat kategorik tidak dapat diketahui atau tidak dapat dilihat nilai korelasinya, oleh karena itu pada grafik di atas cukup melihat data numerik agar dapat melihat nilai korelasi yang terdapat pada tiap kolom.

Penjelasan cara membaca sekaligus menganalisis nilai korelasi antar kolom atau variabel yang ada di sumbu Y dan sumbu X adalah sebagai berikut :

1) Variabel "*Price*" yang ada di sumbu Y

UNIVERSITAS MERCU BUANA

memiliki nilai korelasi dengan variabel "*Gross Sales*" yang ada di sumbu X sebesar 0.57 yang dimana kedua variabel ini memiliki hubungan timbal-balik yang bagus.

2) Variabel "*Item Sold*" yang ada di sumbu Y memiliki nilai korelasi dengan variabel "*Price*" yang ada di sumbu X yakni sebesar 0.018 yang dimana kedua variabel ini tidak terlalu memiliki hubungan timbal balik yang cukup bagus bahkan cenderung tidak memiliki korelasi.

3) Terakhir adalah variabel "*Gross Sales*" yang ada di sumbu Y yang memiliki nilai korelasi atau hubungan timbal-balik dengan variabel "*Item Sold*" yang ada di sumbu X mencapai 0.81 yang dimana kedua variabel ini memiliki hubungan timbal balik yang sangat bagus yang dibuktikan dengan nilai korelasi antar masing-masing variabel.

Pada penelitian ini terdapat eksperimen yang telah dilakukan, yaitu *data training* 65% dan *data testing* 35%. Skenario pengujian yang telah dilakukan menggunakan parameter MAE (*Mean absolute error*), MSE (*mean squared error*) dan R2 *score*. Berikut ini adalah hasil dari *data testing* dan *data training* dari algoritma regresi linear dan C 4.5.

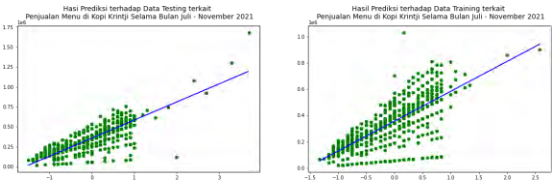
a. Algoritma Regresi Linear

Berikut adalah hasil nilai MAE, MSE, dan R2 *Score* pada algoritma regresi linear.

Tabel 1. Hasil Nilai Akurasi MAE, MSE, dan R2 *Score* Algoritma Regresi Linear

	Data Testing	*Data Training*
MAE	64771.56	59697.34
MSE	9394488337.61	9224945948.40
R2 *Score*	0.65	0.64

Berdasarkan Tabel 1. Hasil Akurasi Algoritma Regresi Linear dapat dilihat hasil dari pengujian dan pelatihan, didapatkan hasil evaluasi MAE pada *data testing* sebesar 64771.56 sedangkan pada *data training* sebesar 59697.34. Hasil evaluasi dari MSE pada *data testing* sebesar 9394488337.61 sedangkan pada *data training* sebesar 9224945948.40. Hasil dari evaluasi dari R2 *Score* 0.65 sedangkan pada *data training* 0.64.



Gambar 4. Grafik Regresi Linear *data testing* dan *data training*

Berdasarkan dari Gambar 4. Grafik Akurasi Regresi Linear *data testing* dan *data training* kedua grafik diatas, baik hasil dari data testing dan data training, keduanya memiliki hubungan yang linear serta juga memiliki hubungan yang positif karena data-data yang ada di dalam sumbu X mengalami peningkatan yang disertai dengan nilai data *point* yang berada di sumbu Y.

Dengan adanya peningkatan nilai yang dapat dilihat dari masing - masing garis grafik yang di atas, kedua grafik tersebut dapat dikatakan grafik yang bersifat linear. Pada dua grafik tersebut memiliki perbedaan yang cukup signifikan. Dapat dilihat pada *data testing*, garis grafik yang mengalami peningkatan meskipun garisnya tidak terlalu lurus ke atas, sedangkan pada *data training* garis grafik mengalami peningkatan dan garisnya lurus ke atas meskipun lurusnya tidak sempurna.

Selain perbedaan pada garis dari kedua data tersebut, terdapat pula perbedaan dalam persebaran titik data pada prediksi *data testing* persebaran data yang diberikan cukup mengumpul pada beberapa titik, walaupun ada juga beberapa titik data yang menyebar tetapi perbandingannya hanya sedikit dengan data yang mengumpul pada beberapa titik, sedangkan pada *data training* titik datanya bersifat lebih menyebar ke beberapa titik, meskipun ada juga beberapa titik data yang berkumpul tetapi perbandingannya sangatlah kecil.

b. Algoritma C 4.5

Berikut adalah hasil nilai MAE, MSE, dan R2 *Score* pada algoritma C 4.5.

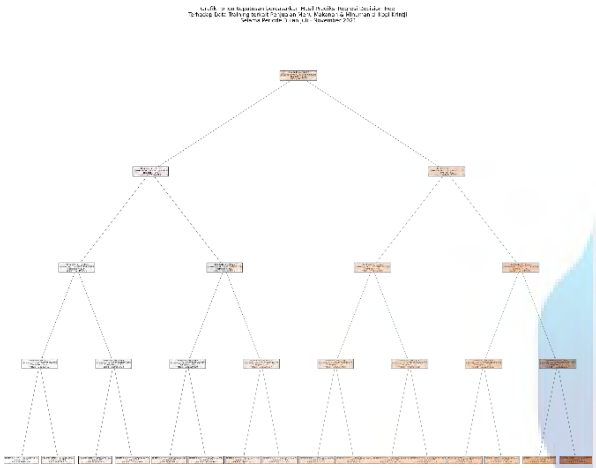
Tabel 2. Hasil Nilai Akurasi MAE, MSE, dan R2 *Score* Algoritma C 4.5

	Data Testing	*Data Training*
MAE	66289.38	59936.13
MSE	10369279663.55	9128114236.26
R2 Score	0.62	0.64

UNIVERSITAS MERCU BUANA

Berdasarkan Tabel 2. Hasil Nilai Akurasi MAE, MSE, dan R2 *Score* Algoritma C 4.5 dapat dilihat hasil dari pengujian dan pelatihan MAE, MSE, dan R2 *Score*. Pada penelitian ini didapatkan hasil evaluasi MAE pada *data testing* sebesar 66289.38 sedangkan pada *data training* sebesar 59936.13. Hasil evaluasi dari MSE pada *data testing* sebesar 10369279663.55 sedangkan pada *data training* sebesar 9128114236.26. Hasil dari evaluasi R2 *Score* sebesar 0.62 sedangkan pada *data training* sebesar 0.64.

Berikut ini adalah hasil dari pohon keputusan *data training* dari algoritma C 4.5.



Gambar 5. Pohon Keputusan *Data Training*

Hasil grafik visualisasi yang terdapat pada Gambar 5. Pohon Keputusan *Data Training* merupakan sebuah grafik pohon keputusan yang khusus hanya dimiliki oleh algoritma *decision tree* baik untuk kebutuhan klasifikasi ataupun prediksi regresi.

Di dalam grafik tersebut terdapat beberapa istilah terminologi dalam pohon keputusan untuk *data training* dan berikut diantaranya adalah :

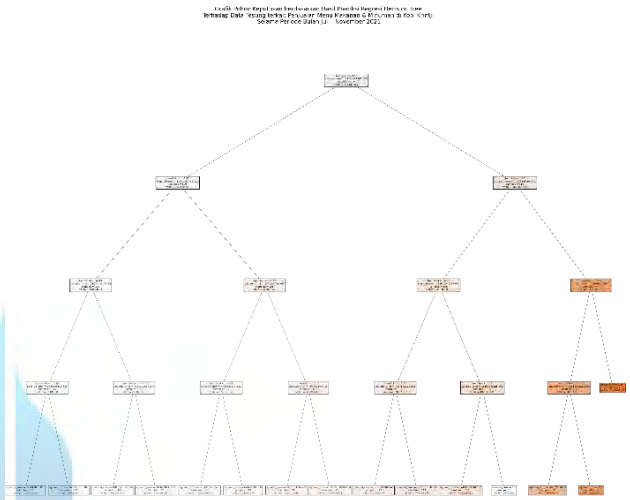
1) *Root node* di dalam grafik pohon keputusan untuk *data training* di atas yakni memiliki nilai *Item Sold* <= -0.042 dengan nilai *squared_error* mencapai 25319338061.538 dari *samples* ke 1300 dengan nilai *value* = 341020.0.

2) *Splitting* juga dapat dikatakan sebagai garis pembagi dalam grafik pohon keputusan.

3) Pada *decision node* ini berisi *node* pilihan yang *output*nya bisa berupa *decision node* kembali atau *leaf node*. Salah satu nilai dari *decision node* adalah nilai *Item Sold* <= -0.792 dengan nilai *metrics squared_error* yakni 3297543835.515 dari *samples data* 328 dengan *value* = 174152.439

4) Pada *leaf node* dari grafik pohon keputusan *data training* adalah ketika nilai *squared_error* mencapai 1949609375.0 berdasarkan *samples* ke 80 dengan nilai *value* = 194375.0.

Berikut ini adalah hasil dari pohon keputusan *data testing* dari algoritma C 4.5.



Gambar 6. Pohon Keputusan *Data Testing*

Grafik visualisasi pada Gambar 4. Pohon Keputusan *Data Testing* merupakan sebuah grafik pohon keputusan yang khusus hanya dimiliki oleh algoritma *decision tree* baik untuk kebutuhan klasifikasi ataupun prediksi regresi. Di dalam grafik tersebut terdapat beberapa istilah terminologi dalam pohon keputusan untuk *data testing* dan berikut diantaranya:

1. *Root node* di dalam grafik pohon keputusan untuk *data testing* diatas yakni memiliki nilai *Item Sold* <= 0.05 dengan nilai *squared*_error mencapai 27104942563.265 dari *samples* ke 700 dengan nilai *value* = 341302.857.

2. *Splitting* juga dapat dikatakan sebagai garis pembagi dalam grafik pohon keputusan.

3. Pada *decision node* adalah nilai *Item Sold* <= -1.25 dengan nilai *metrics squared_error* yakni 2788339310.345 dari *samples data* ke 145 dengan *value* = 164800.0.

4. Pada *leaf node* dari grafik pohon keputusan *data testing* adalah ketika nilai *squared_error* mencapai 975848888.889 berdasarkan *samples* ke 15 dengan nilai *value* = 101533.333.

https://lib.mercubuana.ac.id/

KESIMPULAN

Berdasarkan hasil penelitian yang telah dilakukan untuk memprediksi penjualan makanan dan minuman pada Kopi Kritnji, didapatkan grafik yang menunjukan menu terlaris selama periode penjualan bulan Juli - November 2021 yaitu menu Es kopi susu gadis, menjadi menu yang paling banyak terjual atau dapat dikatakan menu *Best Seller* yang ada di Kopi Krintji, karena itu dari penelitan ini Kopi Krintji harus memaksimalkan stok persediaan Es kopi susu gadis, sehingga Kopi Krintji akan lebih mudah dalam meminimalisir kerugian karena stok persediaan yang di lakukan yaitu dengan melihat jumlah menu yang paling banyak diminati. Penelitian ini juga melakukan pengujian menggunakan algotima regresi linear dan C4.5 dengan parameter MAE, MSE, dan R2 *Score* mendapatkan hasil yang standar. Pada algoritma regresi linear nilai MAE 64771.56, nilai MSE 9394488337.61, dan R2 *Score* 0.65. Sedangkan pada algoritma C 4.5 nilai MAE 66289.38, nilai MSE 10369279663.55 dan nilai R2 *Score* 0.62.

DAFTAR PUSTAKA

Ayuni, G. N., & Fitrianah, D. (2019). Penerapan metode Regresi Linear untuk prediksi penjualan properti pada PT XYZ. *Jurnal Telematika*, *14*(2), 79–86. https://journal.ithb.ac.id/telematika/article/view/321

Dwiasnati, S., & Devianto, Y. (2019). Optimasi Prediksi Keputusan Calon Nasabah Potensial Berbasis Particle Swarm Optimization. *Optimasi Prediksi Keputusan Calon Nasabah Potensial Berbasis Particle Swa*, *6*(2), 286–292.

Eska, J. (2018). *Penerapan Data Mining Untuk Prediksi Penjualan Wallpaper Menggunakan Algoritma C4.5*. *2*. https://doi.org/10.31227/osf.io/x6svc

Fitriani, P., Dani, U., & Prayogi, A. (2021). Jurnal informasi komputer logika. *Jurnal Informasi Komputer Logika*, *2*, 1–3.

Kafil, M. (2019). Penerapan Metode K-Nearest Neighbors Untuk Prediksi Penjualan Berbasis Web Pada Boutiq Dealove Bondowoso. *JATI (Jurnal Mahasiswa Teknik Informatika)*, *3*(2), 59–66. https://doi.org/10.36040/jati.v3i2.860

Kurniawan, R. (2019). *Decision Tree Algorithm for Predicting Return Goods in Online Shopping*. 30–31.

Kusbiyanto, April Heri, I. Q. (2021). Perilaku Nongkrong Generasi Milenial Terhadap Perkembangan Bisnis Kafe Di Masa Pandemi Covid-19 Di Kota Palangka Raya. *Institut Agama Islam Negeri Palangka Raya*, *7*(1), 155–168.

Lukhayu Pritalia, G. (2018). Penerapan Algoritma C4.5 untuk Penentuan Ketersediaan Barang E-commerce. *Indonesian Journal of Information Systems*, *1*(1), 47–56. https://doi.org/10.24002/ijis.v1i1.1727

Meizar, A., Fahrozi, W., Indra, E., & Saputra, M. (2022). Analisis Trend Moment Pada Datamining Forecasting Dalam Memprediksi Jumlah Persediaan Obat Herbal. *Jurnal Sistem Informasi Dan Ilmu Komputer Prima(JUSIKOM PRIMA)*, *5*(2), 103–106. https://doi.org/10.34012/jurnalsisteminformasidanilmukomputer.v5i2.2452

Monika, Sinar; Rakhman, A. (2017). Fakultas Teknik – Universitas Muria Kudus 153. *Prosiding SNATIF Ke-4 Tahun 2017*, *2015*, 153–160.

Setiawan, H., Rahayu, W., & Kurniawan, I. (2020). Perancangan Aplikasi Pemesanan Makanan dan Minuman pada Rumah Makan Cepat Saji D'besto. *Jurnal Riset Dan Aplikasi Mahasiswa Informatika (JRAMI)*, *1*(03), 347–354. https://doi.org/10.30998/jrami.v1i03.356

Suryanto, A. A. (2019). Penerapan Metode Mean Absolute Error (Mea) Dalam Algoritma Regresi Linear Untuk Prediksi Produksi Padi. *Saintekbu*, *11*(1), 78–83. https://doi.org/10.32764/saintekbu.v11i1.298

Triyandana, G., Putri, L. A., & Umaidah, Y. (2022). *Penerapan Data Mining Pengelompokan Menu Makanan dan Minuman Berdasarkan Tingkat Penjualan Menggunakan Metode K-Means*. *6*(1), 40–46.

Wanita, F., Mashud, Angriawan, R., & Elma Pratiwi, C. (2021). Rancang Bangun Sistem Informasi Pengendalian Persediaan (Control Buffer Stock) Untuk Efisiensi Kewirausahaan Penjualan Kopi Pada Soft Coffee. *Jurnal Teknologi Informasi Universitas Lambung Mangkurat (JTIULM)*, *6*(1), 9–18. https://doi.org/10.20527/jtiulm.v6i1.70

Wijaya, H. D., & Dwiasnati, S. (2020). Implementasi Data Mining dengan Algoritma Naïve Bayes pada Penjualan Obat. *Jurnal Informatika*, *7*(1), 1–7. https://doi.org/10.31311/ji.v7i1.6203

UNIVERSITAS MERCU BUANA

KERTAS KERJA

Ringkasan

Kertas kerja ini berisi tentang kelengkapan material dari artikel jurnal dengan judul "Penerapan Data Mining Untuk Memprediksi Penjualan Makanan Pada Kopi Krintji Menggunakan Algoritma Regresi Linear dan C4.5". Seluruh hasil penelitian Tugas Akhir yang tidak dimasukkan kedalam artikel jurnal. Pada kertas kerja ini disajikan terdiri dari literatur review, analisis perancangan, source code, dataset yang digunakan, tahapan eksperimen dan seluruh hasil eksperimen.

Bagian 1 : Literature Review menjabarkan mengenai beberapa jurnal yang terkait dengan penelitian.

Bagian 2 : Analisis dan Perancangan menjabarkan analisis masalah terhadap penelitian

Bagian 3 : *Source Code* menjabarkan kumpulan kode disetiap proses yang dilakukan peneliti terdiri dari membaca data, preprocessing data, data visualisasi, data modelling, implementasi algoritma.

Bagian 4 : Pengambilan dataset yang akan dilakukan.

Bagian 5 : Tahapan Eksperimen berisi pengumpulan data, *preprocessing data*, implementasi algoritma, data validasi, dan evaluasi.

Bagian 6 : Hasil semua eksperimen secara keseluruhan yang mencakup skenario pengujian yang dilakukan oleh peneliti.

https://lib.mercubuana.ac.id/