



**Analisa Algoritma Regresi Linear dan Decision Tree Dalam
Prediksi Penjualan Produk
(Studi Kasus: Lookma Boutique)**

TUGAS AKHIR

William JayaMangala Putra
41518010013

UNIVERSITAS
MERCU BUANA

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



**Analisa Algoritma Regresi Linear dan Decision Tree Dalam
Prediksi Penjualan Produk
(Studi Kasus: Lookma Boutique)**

Tugas Akhir

Diajukan Untuk Melengkapi Salah Satu Syarat
Memperoleh Gelar Sarjana Komputer

Oleh:

William JayaMangala Putra
41518010013

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

LEMBAR PERNYATAAN ORISINALITAS

Yang bertanda tangan dibawah ini:

NIM : 41518010013

Nama : William JayaMangala Putra

Judul Tugas Akhir : Analisa Algoritma Regresi Linear dan Decission Tree Dalam Prediksi Penjualan Produk (Studi Kasus : Lookma Boutique)

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, 28 Juli 2022



William JayaMangala Putra

UNIVERSITAS
MERCU BUANA

SURAT PERNYATAAN PERSETUJUAN PUBLIKASI TUGAS AKHIR

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

Nama Mahasiswa : William JayaMangala Putra.
NIM : 41518010013
Judul Tugas Akhir : Analisa Algoritma Regresi Linear dan Decision Tree Dalam Prediksi Penjualan Produk (Studi Kasus : Lookma Boutique)

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, 28 Juli 2022

UNIVERSITAS
MERCU BUANA



William JayaMangala Putra

SURAT PERNYATAAN LUARAN TUGAS AKHIR

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

Nama Mahasiswa : William JayaMangala Putra
NIM : 41518010013
Judul Tugas Akhir : Analisa Algoritma Regresi Linear dan Decision Tree Dalam Prediksi Penjualan Produk (Studi Kasus: Lookma Boutique)

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	: JEPIN (Jurnal Edukasi dan Penelitian Informatika)			
	ISSN	: 2548-9364			
	Link Jurnal	: https://jurnal.untan.ac.id/index.php/jepin/index			
	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, 28 Juli 2022



William JayaMangala Putra

LEMBAR PERSETUJUAN PENGUJI

LEMBAR PERSETUJUAN PENGUJI

NIM : 41518010013
Nama : William JayaMangala Putra
Judul Tugas Akhir : Analisa Algoritma Regresi Linear dan Decision Tree Dalam Prediksi Penjualan Produk (Studi Kasus: Lookma Boutique)

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, 10 Agustus 2022

Rengui
Ummiy Salamah, ST, MMSI
MERCU BUANA

LEMBAR PERSETUJUAN PENGUJI

LEMBAR PERSETUJUAN PENGUJI

NIM : 41518010013
Nama : William JayaMangala Putra
Judul Tugas Akhir : Analisa Algoritma Regresi Linear dan Decission Tree Dalam Prediksi Penjualan Produk (Studi Kasus: Lookma Boutique)

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, 28 Juli 2022



(Saruni Dwiastuti, ST, MM, M.Kom)
(Dosen Penguji 2)



UNIVERSITAS
MERCU BUANA

LEMBAR PERSETUJUAN PENGUJI

LEMBAR PERSETUJUAN PENGUJI

NIM : 41518010013
Nama : William JayaMangala Putra
Judul Tugas Akhir : Analisa Algoritma Regresi Linear dan Decision Tree Dalam Prediksi Penjualan Produk (Studi Kasus: Lookma Boutique)

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, 28 Juli 2022


(Muhammad Rifqi, S.Kom, M.Kom)
(Dosen Penguji 3)

UNIVERSITAS
MERCU BUANA

LEMBAR PENGESAHAN

NIM : 41518010013
Nama : William JayaMangala Putra
Judul Tugas Akhir : Analisa Algoritma Regresi Linear dan Decission Tree
Dalam Prediksi Penjualan Produk (Studi Kasus: Lookma
Boutique)

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, 28 Juli 2022

Menyetujui,



Wawan Gunzway, S.Kom, MT
Dosen Pembimbing

Mengetahui,



Wawan Gunzway, S.Kom, MT
Koord. Tugas Akhir Teknik Informatika



Ir. Emil R. Kaburuan, Ph.D., IPM.
Ka. Prodi Teknik Informatika

UNIVERSITAS
MERCU BUANA

KATA PENGANTAR

Segala Puji syukur dipanjatkan kepada Tuhan Yang Maha Esa yang telah memberikan rahmat dan karunia-Nya, sehingga dapat menyelesaikan skripsi dengan judul “*Analisa Algoritma Regresi Linear dan Decision Tree Dalam Prediksi Penjualan Produk (Studi Kasus: Lookma Boutique)*” sebagai syarat untuk menyelesaikan Program Sarjana (S1) pada Program Sarjana Fakultas Ilmu Komputer Jurusan Informatika Universitas Mercu Buana.

Dalam penyusunan skripsi ini banyak hambatan serta rintangan yang penulis hadapi namun pada akhirnya dapat melaluinya berkat adanya bimbingan dan bantuan dari berbagai pihak baik secara moral maupu spiritual. Untuk itu pada kesempatan ini penulis menyampaikan ucapan terimakasih kepada:

1. Kedua orang tua yang telah memberika support serta kasih sayang hingga saya bisa menyelesaikan tugas akhir ini
2. Bapak Wawan Gunawan S.Kom.,MT selaku dosen pembimbing saya yang telah membantu dan memberikan arahan kepada saya dalam menyelesaikan tugas akhir ini.
3. Bapak Emil R. Kaburuan Ph.D selaku Kepala Program Studi Informatika yang telah memberikan arahan dan bimbingannya selama perkuliahan.
4. Bapak Wawan Gunawan S.Kom.,MT selaku Koordinator Tugas Akhir Prodi Informatika yang telah memberikan arahan selama periode tugas akhir.
5. Bapak Sabar Rudiarto,M.Kom selaku dosen pembimbing akademik saya yang telah membimbing selama perkuliahan hingga sampai pada tugas akhir.
6. Alivia Putri Rahma yang berjuang dan selalu memotivasi saya untuk mengerjakan dan menemani skripsi.
7. Teman-teman Informatika 2018, sahabat dan kerabat yang selalu memberikan support dan waktu untuk selalu mengingatkan dalam menyelesaikan tugas akhir ini.

Penulis menyadari bahwa penyusunan Jurnal Tugas Akhir ini masih jauh dari kesempurnaan dan masih terdapat kekurangan dan kelemahan, walaupun demikian, penulis telah berusaha semaksimal mungkin untuk mendapatkan hasil yang terbaik. Akhir kata penulis Akhir kata, penulis berharap jika ada kritik atau saran apapun yang sifatnya membangun bagi saya dengan senang hati akan saya terima.

Jakarta, 28 Juli 2022



William JayaMangala Putra



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
ABSTRAK	ix
ABSTRACT	x
KATA PENGANTAR	xi
DAFTAR ISI.....	xiii
NASKAH JURNAL.....	1
KERTAS KERJA	9
BAB 1. LITERATUR REVIEW	20
BAB 2. ANALISIS DAN PERANCANGAN.....	29
BAB 3. SOURCE CODE	31
BAB 4. DATASET	39
BAB 5. TAHAPAN EKSPERIMEN	40
BAB 6. HASIL SEMUA EKSPERIMEN	43
DAFTAR PUSTAKA	54
LAMPIRAN DOKUMEN HAKI	56
LAMPIRAN KORESPONDENSI	58

NASKAH JURNAL



JEPIN

(Jurnal Edukasi dan Penelitian Informatika)

ISSN(e): 2548-9364 / ISSN(p) : 2460-0741

Vol. X

No. y

mm yy

Analisa Algoritma Regresi Linear dan Decision Tree dalam Prediksi Penjualan Produk (Studi Kasus: Lookma Boutique)

William JayaMangala Putra^{#1}, Wawan Gunawan^{*2}[#]Universitas Mercu Buana, Jakarta, Indonesia

141518010013@student.mercubuana.ac.id

Universitas Mercu Buana, Jakarta, Indonesia

²Wawan.Gunawan@mercubuana.ac.id

Abstrak— Perkembangan bisnis secara online di Indonesia sekarang sangat pesat, dengan proses dilakukan dengan cara memesan barang melalui reseller atau distributor menggunakan salah satu jejaring sosial internet. Belanja online merupakan salah satu cara berbelanja melalui alat komunikasi atau jejaring sosial yang digunakan dalam transaksi jual beli, penjualan di masa yang akan datang menentukan volume penjualan. Salah satu toko online yaitu Lookma Boutique, toko tersebut menghasilkan penjualan yang cukup besar setiap bulan dan tahun, tetapi hasil penjualan yang dihasilkan tidak stabil dan tidak mampu diprediksi kenaikan dan penurunan penjualannya. Oleh karena itu dibutuhkan suatu metode yang bisa prediksi produk penjualan. Penelitian ini berfokus untuk prediksi hasil penjualan dari Lookma Boutique untuk mengetahui efektifitas penjualan berdasarkan popularitas penjualan produk berdasarkan total produk dibayar dan produk pesanan yang dibuat. Dengan perbandingan algoritma yang memiliki keunggulan dalam melakukan prediksi dengan menggunakan algoritma Regresi Linear dan Algoritma Decision Tree. Hasil penelitian ini menunjukkan bahwa dari semua pengujian menggunakan MAE memberikan hasil 0.84, MSE memberikan hasil 1.79 dan R2 Score menghasilkan hasil 1.00.

Kata kunci— Prediksi, Algoritma Regresi Linear, Algoritma Decision Tree, Analisis, Data

I. PENDAHULUAN

Perkembangan bisnis secara online di Indonesia sekarang sangat pesat, dengan proses dilakukan dengan cara memesan barang melalui reseller atau distributor menggunakan salah satu jejaring sosial internet. Belanja online merupakan salah satu cara berbelanja melalui alat komunikasi atau jejaring sosial yang digunakan dalam transaksi jual beli[1], penjualan di masa yang akan datang menentukan volume penjualan. Salah satu toko online yaitu Lookma Boutique, toko tersebut menghasilkan penjualan yang cukup besar setiap bulan dan tahun, tetapi hasil penjualan yang dihasilkan tidak stabil dan tidak mampu diprediksi kenaikan dan penurunan penjualannya. Pembelian item dilakukan berdasarkan informasi merek dagang, size, model dan warna keputusan yang menggunakan model keputusan yang berbentuk seperti pohon. Selain itu, prediksi berguna untuk melihat keadaan produk yang dijual, Oleh karena itu, resiko kesalahan akibat kesalahan perencanaan harus diminimalkan. prediksi biasanya digunakan untuk menemukan informasi dari sejumlah besar data. Kedua algoritma tersebut akan diuji untuk mendapatkan hasil yang terbaik dalam memprediksi penjualan produk dalam satu periode, prediksi ini diharapkan dapat mendukung dan membantu kinerja dalam mengoptimalkan penjualan barang di masa mendatang berdasarkan data yang telah direkam

Korespondensi : Fulan Fulana

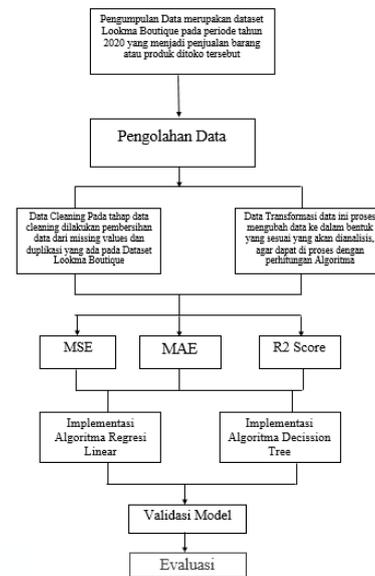
Universitas Mercu Buana

sebelumnya. toko online Lookma Boutique Pada penelitian sebelumnya peramalan dilakukan dengan metode Regresi Linear yang diimplementasikan penelitian tersebut melakukan prediksi untuk mendapatkan jumlah kasus klaim yang akan ditangani. Prediksi Jumlah kasus klaim akan dibagi berdasarkan total keseluruhan klaim, jenis layanan dan jenis manfaat.

Dengan mengetahui jumlah kasus klaim yang ditangani akan mempermudah perencanaan peningkatan dan perbaikan proses penyelesaian penanganan kasus klaim. Jumlah kasus klaim dilihat dari jumlah peserta yang ditanggung dan yang dilayani untuk setiap polis yang terbit[2]. Selanjutnya penelitian ini di dapatkan bahwa metode linear regresi layak dan efektif untuk memprediksi pendaftaran paten untuk tahun selanjutnya berdasarkan data pendaftaran, bahwa metode linear regresi layak dan efektif untuk memprediksi pendaftaran paten untuk tahun selanjutnya berdasarkan data pendaftaran dari tahun 2014 sampai tahun 2018[3]. Pada penelitian terdahulu terakhir Jumlah mahasiswa baru suatu universitas selama 5 tahun terakhir yaitu mulai tahun 2016 hingga tahun 2020 mengalami kenaikan dan penurunan. Data tersebut selanjutnya akan digunakan untuk memprediksi jumlah mahasiswa 5 tahun mendatang. Tujuan dibuat sebuah prediksi adalah untuk mengetahui rasio dosen yang tersedia dengan jumlah mahasiswa baru, mempersiapkan ruang kuliah dan juga fasilitas lainnya[4]. Menjelaskan bahwa kedua metode ini adalah kelebihan dan kekurangan masing-masing saat memprediksi masalah, terutama memprediksi. Apabila toko mengetahui peningkatan atau penurunan dari hasil penjualannya maka toko dapat segera mengambil tindakan untuk mengubah dan meningkatkan strategi yang jauh lebih baik lagi. Selain itu toko harus tetap meningkatkan penjualan untuk keberlangsungan toko[5]. Dari penelitian sebelumnya mengenai penjualan smartphone dengan menggunakan Algoritma Regresi Linear hasil akurasi tertinggi dengan nilai akurasi Hasil pengujian menunjukkan bahwa prediksi penjualan smartphone dengan metode Linear Regresi adalah tergolong dalam kategori sangat baik. Terbukti dengan pengujian nilai MAPE dan MSE untuk smartphone merek Lenovo kurang dari 10. Berdasarkan pemaparan diatas diperlukan perbandingan algoritma algoritma regresi linear dan algoritma Decision Tree agar menjadi acuan terhadap akurasi dari algoritma perbandingan tersebut[6].

Penelitian ini berfokus untuk prediksi hasil penjualan dari Lookma Boutique untuk mengetahui efektifitas penjualan berdasarkan poplitas penjualan produk berdasarkan total produk dibayar dan produk pesanan yang dibuat. Dengan perbandingan algoritma yang memiliki keunggulan dalam melakukan prediksi.

II. METODOLOGI PENELITIAN



Gambar 1. Tahapan Metodologi

A. Pengumpulan Data

Metode pengumpulan data yang digunakan dalam penelitian ini merupakan dataset Lookma Boutique pada periode tahun 2020 yang menjadi penjualan barang atau produk ditoko tersebut. Jumlah data pada dataset ini yaitu 1.400 data dan memiliki 13 atribut.

B. Pengolahan Data

Data yang telah diperoleh perlu dilakukan pre-processing. Pre-processing merupakan langkah penting yang digunakan untuk mengubah data mentah yang memungkinkan menerapkan teknik data mining dan juga untuk meningkatkan kualitas data. Beberapa Tahap di pengolahan data: Pada tahap data cleaning dilakukan pembersihan data dari missing values dan duplikasi yang ada pada Dataset Lookma Boutique. Missing values adalah kondisi dimana suatu atribut memiliki nilai yang tidak valid. Duplikasi data adalah kondisi dimana semua atribut suatu data memiliki kesamaan dengan semua atribut suatu data lain. Pengecekan missing values dilakukan pada keseluruhan dataset, sedangkan pengecekan duplikasi dilakukan pada data per tahun, selanjutnya Data Transformasi data ini proses mentransformasi atau mengubah data ke dalam bentuk yang sesuai yang akan dianalisis, agar dapat di proses dengan perhitungan Algoritma Decision Tree dan Algoritma Regresi Linear.

C. Algoritma Regresi Linear

Regresi Linear adalah Metode regresi merupakan sebuah metode statistik yang melakukan prediksi menggunakan pengembangan hubungan matematis antara variabel, yaitu variabel dependen (Y) dengan variabel independen (X)[7]. Variabel dependen merupakan variabel akibat atau variabel yang dipengaruhi, sedangkan variabel

independen merupakan variabel sebab atau variabel yang mempengaruhi. Prediksi terhadap nilai variabel dependen dapat dilakukan jika variabel independennya diketahui[8]. Umumnya penjualan atau permintaan suatu produk dinyatakan sebagai variabel dependen yang besar atau nilainya dipengaruhi oleh variabel independen. 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[9]. Berikut metode regresi linear dapat dilihat dari persamaan 1[10]. Berikut untuk menentukan nilai konstanta dapat dilihat dari persamaan 2. Berikut untuk menentukan nilai koefisien regresi dapat dilihat dari persamaan 3.

$$Y = a + bx \quad (1)$$

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

$$b = \frac{n(\sum XY) - (\sum X)(\sum Y)}{n\sum X^2 - (\sum X)^2} \quad (3)$$

D. Algoritma Decision Tree

Mempersiapkan data training, bisa diambil asal data histori yang pernah terjadi sebelumnya serta telah dikelompokkan pada kelas-kelas tertentu[11] dan menentukan akar berasal pohon dengan menghitung nilai gain yang tertinggi dari masing-masing atribut atau berdasarkan nilai index entropy terendah. Sebelumnya dihitung terlebih dahulu nilai index entropy[12].

E. Validasi Model

Pada tahapan validasi model merupakan pengambilan keputusan dapat mempercayai model yang digunakan sebagai bagian dari proses pengambilan keputusan[13]. Membandingkan output ukuran kinerja model simulasi dengan ukuran kinerja yang sesuai dari sistem nyata adalah metode yang paling sesuai untuk melakukan validasi model simulasi. Uji validasi model dapat dengan melihat pada nilai-nilai simpangan[14].

F. Evaluasi

Pada Tahap Evaluasi dan Peninjauan model ini akan dibahas mengenai hasil evaluasi yang telah dikerjakan. Phase ini akan dilakukan perbandingan kuantitatif dengan mempertimbangkan nilai komparasi MAE, MSE, R2 Score[15]. Untuk menentukan MAE (*Mean Absolute Error*) dapat dilihat dari persamaan 4. MSE (*Mean Squared Error*) dapat dilihat dari persamaan 5. Metode *R2 Score* dapat dilihat dari persamaan 6.

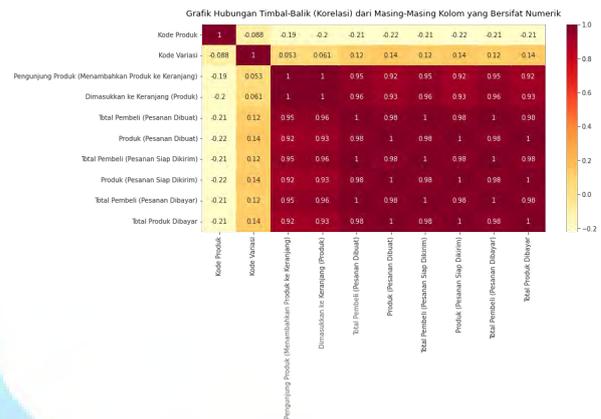
$$MAE : \sum \frac{y - \hat{y}}{n} \quad (4)$$

$$MSE : \frac{1}{2} \sum_{i=1}^n (y_i - \hat{y}_i)^2 \quad (5)$$

$$R2 \text{ Score} : R^2 = 1 - \frac{RSS}{TSS} \quad (6)$$

III. HASIL DAN PEMBAHASAN

A. Melihat korelasi atau nilai keterkaitan masing-masing kolom.



Gambar 2. Peta Heatmap

Pada gambar 2 grafik diatas dinamakan sebagai grafik heatmap dengan menunjukkan nilai korelasi/nilai keterhubungan yang bersifat hubungan timbal balik antar masing-masing kolom yang bersifat numerik sedangkan kolom data yang bersifat kategorik tidak memiliki hubungan timbal balik. Membaca dari SUMBU Y ke SUMBU X atau sebaliknya dengan memperhatikan nilai yang ada di box sebagai penanda nilai hubungannya. Berikut adalah penjelasan terkait nilai korelasi antar kolom/variabel yang ada di SUMBU Y dan SUMBU X:

Pada Variabel "Produk (Pesanan Siap Dikirim)" yang ada di SUMBU Y memiliki nilai korelasi dengan variabel "Dimasukkan ke Keranjang (Produk)" yang ada di SUMBU X yaitu sebesar 0.93 yang dimana kedua variabel ini memiliki hubungan timbal balik yang sangat bagus. Selain itu variabel ini juga memiliki nilai hubungan timbal balik dengan variabel "Kode Produk" yang ada di SUMBU X adalah -0.22 atau dengan kata lain kedua variabel ini memiliki hubungan timbal balik yang tidak bagus.

Variabel "Produk (Pesanan Dibuat)" yang ada di SUMBU Y memiliki nilai korelasi dengan variabel "Total Pembeli (Pesanan Dibuat)" yang ada di SUMBU X sebesar 0.98 yang dimana kedua variabel ini memiliki hubungan timbal balik yang sangat bagus, dst. Variabel "Total Pembeli (Pesanan Dibuat)" yang ada di SUMBU Y memiliki nilai korelasi dengan variabel "Pengunjung Produk (Menambahkan Produk ke Keranjang)" yang ada di SUMBU X adalah mencapai 0.95 yang dimana kedua variabel ini memiliki hubungan timbal balik yang sangat bagus, dst.

Variabel "Kode Variasi" yang ada di SUMBU Y memiliki nilai korelasi dengan variabel "Total Produk Dibayar" yang ada di SUMBU X sebesar 0.14 yang dimana kedua variabel ini cenderung memiliki hubungan timbal balik yang cukup bagus dibandingkan dengan variabel "Kode Produk" yang ada di SUMBU X yang memiliki nilai hubungan timbal balik -0.088

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.

MAE digunakan untuk mengukur keakuratan suatu model statistik dalam melakukan prediksi atau peramalan.

Mean Squared Error (MSE) merupakan nilai rata-rata kesalahan kuadrat diantara nilai aktual dan nilai peramalan. Metode Mean Squared Error secara umum digunakan untuk mengecek estimasi berapa nilai kesalahan pada peramalan.

R-Squared (R2 Score) merupakan nilai metric yang bersifat presentase dengan menunjukkan sifat kecocokan dari hasil seberapa banyak nilai variasi dari variabel dependen (variabel y) yang dapat dijelaskan oleh variabel independen (variabel x) dalam pengujian regresi.

Pada bagian ini akan ditunjukkan hasil dari data testing dan data training dari kedua algoritma yaitu algoritma regresi linear dan algoritma decision tree.

B. Algoritma Regresi Linear

Setelah melakukan proses pemanggilan salah satu algoritma machine learning yang akan digunakan yaitu regresi linear sekaligus melatihnya dengan data training, telah mendapatkan the best value/nilai terbaik dari masing-masing parameter dan juga telah melihat nilai prediksi dari data testing dan data training, selanjutnya adalah akan melihat hasil prediksi yang didapatkan dari ketiga metrics yang saya gunakan yaitu mean absolute error, mean squared error, dan r2 score.

$$x: 2749 \quad y: 2378 \quad x^2: 5498 \quad y^2: 4756 \quad xy: 155214$$

Menghitung konstanta a

$$a = \frac{(\Sigma Y)(\Sigma X^2) - (\Sigma X)(\Sigma XY)}{n \Sigma X^2 - (\Sigma X)^2}$$

$$a = \frac{(2378)(5498) - (2749)(155214)}{1233(5498) - (2749)^2}$$

$$a = \frac{13.074.244 - 426.683.286}{6.779.034 - 5498^2}$$

$$a = \frac{-29.609.042}{6.773.536}$$

Korespondensi : Fulan Fulana

Menghitung konstanta b

$$a = -4,37128289$$

$$b = \frac{n(\Sigma XY) - (\Sigma X)(\Sigma Y)}{n \Sigma X^2 - (\Sigma X)^2}$$

$$b = \frac{1233(155214) - (2749)(2378)}{1233(5498) - (2749)^2}$$

$$b = \frac{191.378.862 - 6.537.122}{6.779.034 - 7.557.001}$$

$$b = \frac{-777.967}{-184.841.740}$$

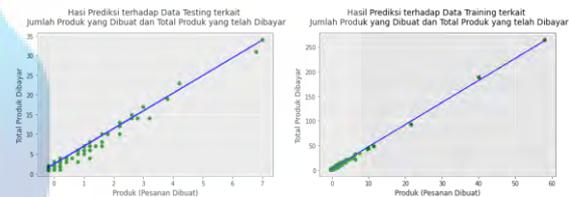
$$b = -237,595862$$

Maka persamaan regresinya adalah:

$$y = a + bX$$

$$y = -4,37128289 - 237,595862X$$

Dari pengujian didapatkan hasil pengujian data tesing MAE 0.76, sedangkan pengujian data testing MSE 0.99, pengujian data testing R2 Score 0.97 dan pengujian data training parameter MAE 0.84, MSE 1.79 dan R2 Score 1.00.

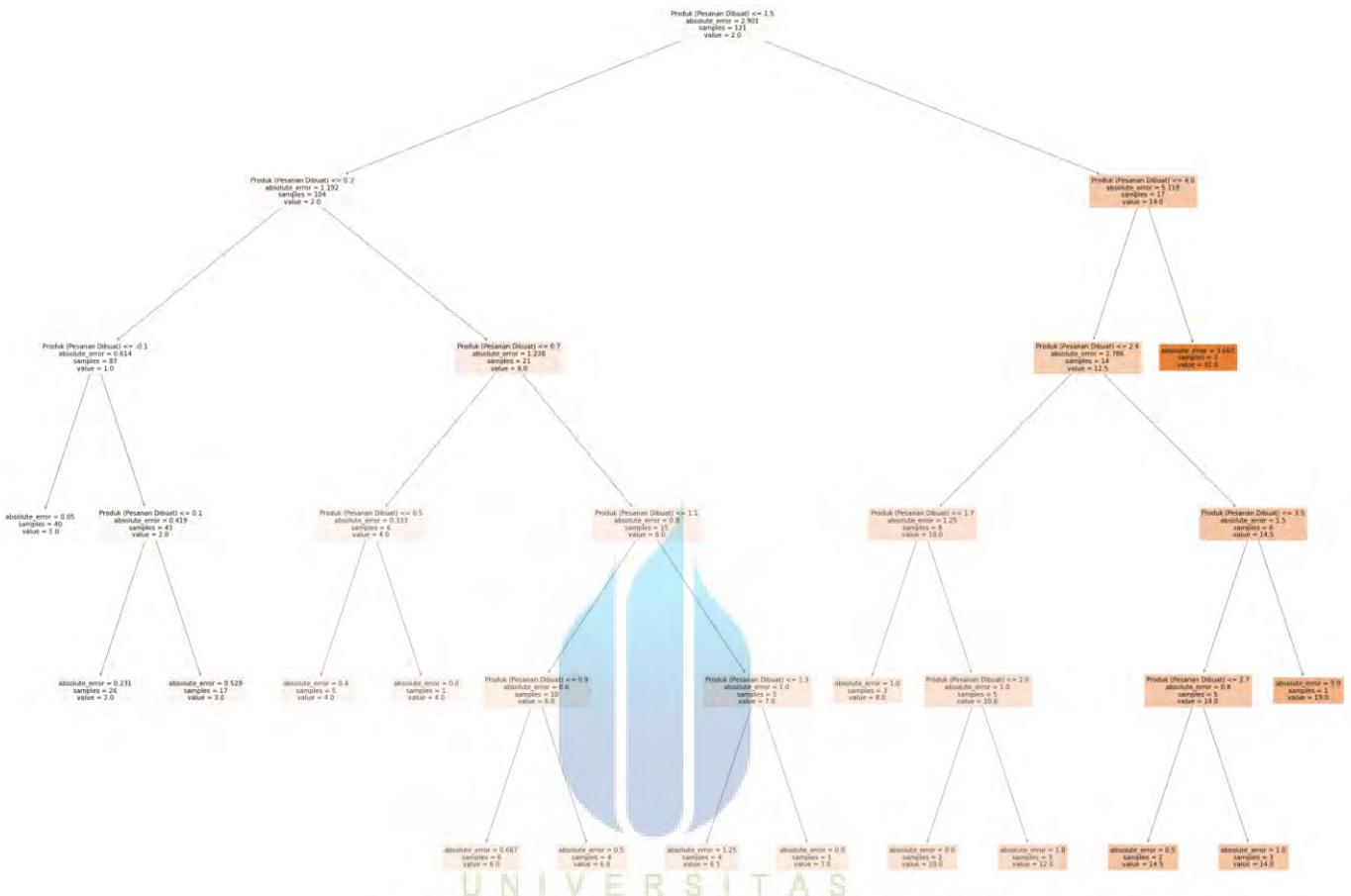


Gambar 3. Grafik Regresi Linear

Berdasarkan dari kedua grafik diatas, baik hasil dari data testing & data training, keduanya memiliki hubungan yang linear serta juga memiliki hubungan yang positif karena data-data yang ada di dalam variabel "Produk (Pesanan Dibuat)" selaku variabel dependen mengalami peningkatan yang disertai dengan nilai variabel "Total Produk Dibayar" selaku variabel independen. Tetapi juga dapat perbedaan yang dimana pada grafik bagian data testing terlihat titik datanya lebih tersebar dari pada titik data pada grafik linear untuk data training tetapi walaupun begitu bentuk garis keduanya membentuk pola linear sehingga dapat dikatakan bahwasanya algoritma regresi linear dalam memprediksi penjualan produk di Lookma Boutique sangat tepat dan akurat.

Universitas Mercu Buana

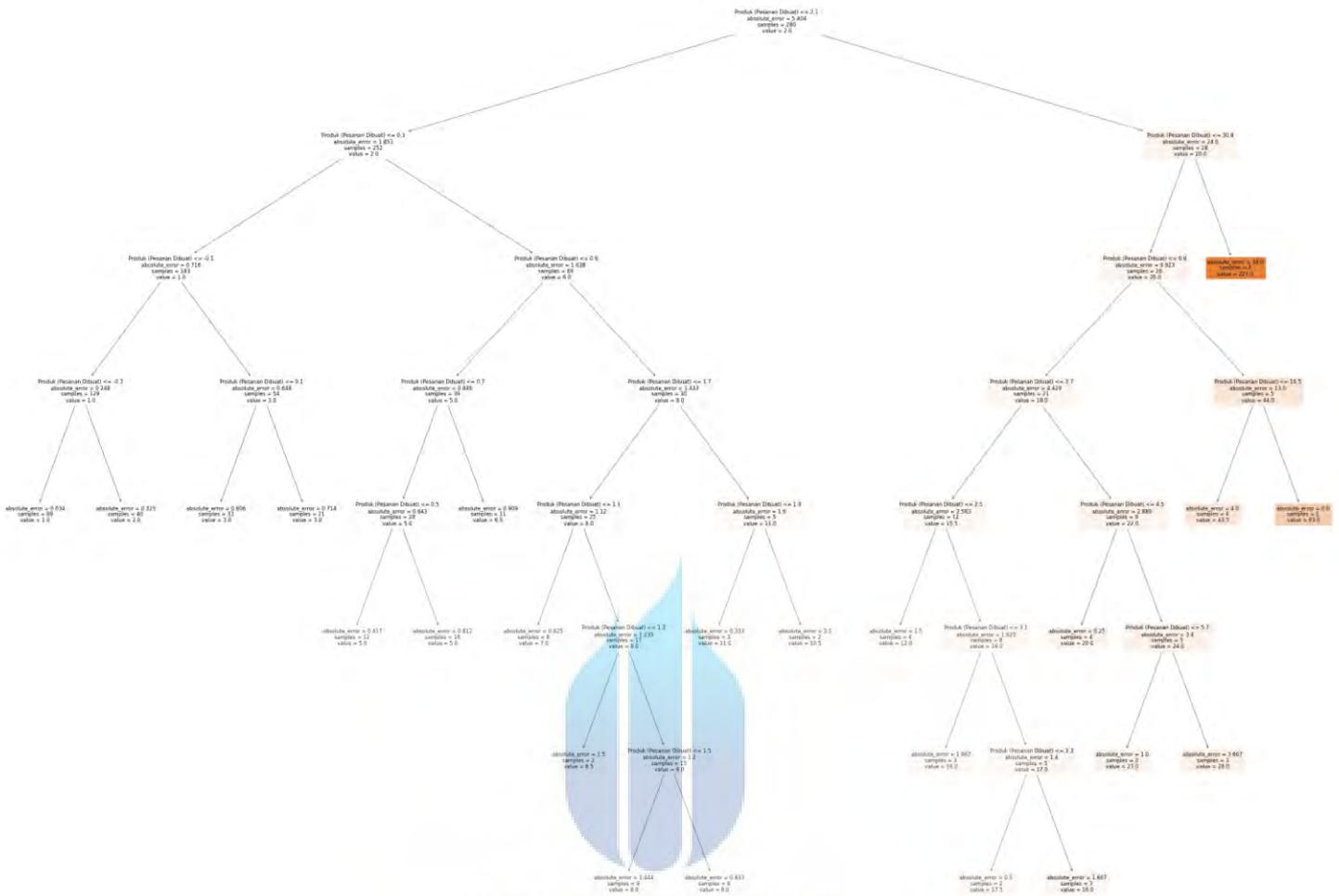
C. Algoritma Decision Tree



Gambar 4. Decission Data Testing

Hasil grafik visualisasi pada gambar 4 merupakan sebuah grafik pohon keputusan yang khusus hanya dimiliki oleh algoritma decision tree baik untuk kebutuhan klasifikasi ataupun prediksi. Di dalam grafik tersebut terdapat beberapa istilah terminologi dalam pohon keputusan untuk data testing dan berikut diantaranya Root node juga dapat dikatakan sebagai node tertinggi (yang paling teratas) di dalam grafik struktur pohon keputusan. Root node di dalam grafik pohon keputusan untuk data testing diatas yakni memiliki nilai yaitu Produk (Pesanan Dibuat) <= 1.5 dengan nilai absolute_error yakni 2.901 dari samples ke 121 dengan nilai value = 2.0. Splitting juga dapat dikatakan sebagai garis pembagi dalam grafik pohon

keputusan. Decision node: merupakan istilah dalam grafik pohon keputusan untuk memecah sub-node menjadi sub-node yang lebih lanjut. Salah satu contoh dari decision node di dalam grafik pohon keputusan terhadap data testing adalah ketika nilai pada Produk (Pesanan Dibuat) <= 0.9 dengan nilai metrics absolute_error adalah 0.6 dari samples data 10 dan juga value = 6.0. Terminal node (leaf node) bagian di dalam grafik pohon keputusan yang tidak memiliki cabang. Sebagai contoh dari leaf node dalam grafik pohon keputusan terhadap data testing adalah ketika nilai absolute_error = 3.667 berdasarkan samples = 3 dengan nilai value = 31.0.



UNIVERSITAS
MERCU BUANA

Gambar 5. Decision Data Training

Sama seperti tahap sebelumnya (membuat grafik pohon keputusan untuk data testing), saat ini akan membuat sebuah penjelasan terhadap grafik pohon keputusan berdasarkan data training pada gambar 5. Di dalam grafik tersebut terdapat beberapa istilah terminologi dalam pohon keputusan untuk data training dan berikut diantaranya Root node juga dapat dikatakan sebagai node tertinggi (yang paling teratas) di dalam grafik struktur pohon keputusan. Root node di dalam grafik pohon keputusan untuk data training diatas yakni memiliki nilai yaitu Produk (Pesanan Dibuat) ≤ 2.1 dengan nilai absolute_error yakni 5.404 dari samples ke 280 dengan nilai value = 2.0. Splitting juga dapat dikatakan sebagai garis pembagi dalam grafik pohon keputusan. Decision node grafik pohon keputusan untuk memecah sub-node menjadi sub-node yang lebih lanjut. Salah satu contoh dari decision node di dalam grafik pohon keputusan terhadap data training adalah ketika nilai pada Produk (Pesanan Dibuat) ≤ 3.7 dengan nilai metrics absolute_error adalah 4.429 dari

samples data ke 21 dan juga value = 18.0. Terminal node (leaf node) bagian di dalam grafik pohon keputusan yang tidak memiliki cabang. Sebagai contoh dari leaf node dalam grafi pohon keputusan terhadap data training adalah ketika nilai absolute_error = 3.667 berdasarkan samples = 3 dengan nilai value = 28.0. Berdasarkan dari pengujian diatas didapatkan hasil pengujian data testing MAE 0.92, sedangkan pengujian data testing MSE 1.36, pengujian data testing R2 Score 0.96 dan pengujian data training MAE 0.92, kemudian MSE 17.43, terakhir R2 Score 0.96.

III. KESIMPULAN

Berdasarkan hasil pengujian dari semua percobaan pada penelitian ini yang dilakukan prediksi penjualan dalam. Dimana skenario pengujian ini yang dilakukan yaitu dengan parameter MSE, MAE dan R2 Score mendapatkan hasil yang sangat baik. Berdasarkan hasil ini membantu meramalkan penjualan dalam kurun waktu kedepan dimana hasil pengujian akurasi pada

algoritma regresi linear dengan parameter MAE mendapatkan 0.76, MSE 0.99 dan R2 Score 0.97, untuk pengujian pada algoritma decision tree dengan parameter MAE mendapatkan hasil 0.92, MSE 1.36 dan R2 Score 0.96. Hal ini mendapatkan kemungkinan besar dikarenakan penggunaan variabel yang digunakan dalam perhitungan regresi linear, sehingga semakin memperkuat nilai prediksi penjualan produk di Lookma Boutique sangat akurat dan produk penjualan yang laris dalam kurun waktu 1 tahun yaitu READY LMB KOREAN TENNIS SKIRT TALI PITA PUTIH DENGAN LUBANG" dan "KOREA CELANA PENDEK WANITA HW short hot pants hits murah berkualitas. Dari kedua algoritma yaitu Algoritma Regresi Linear dan Decision Tree yang paling unggul yaitu regresi linear karena algoritma ini adalah yang terbaik untuk digunakan yang paling kompleks dibandingkan dengan algoritma lain yang juga mencoba menemukan hubungan antara variabel independen dan dependen bisa dapat dikatakan dapat digunakan untuk memprediksi nilai yang ada pada masa depan.

IV. REFERENSI

- [1] M. J. Paruntu *et al.*, "KEPUASAN KONSUMEN BUKALAPAK DI KOTA MANADO ANALYSIS OF PURCHASE DECISIONS AND USE OF SOCIAL MEDIA ON BUKALAPAK CUSTOMER SATISFACTION IN MANADO CITY Oleh : E-mail : Jurnal EMBA Vol . 9 No . 3 Juli 2021 , Hal . 97 - 104," vol. 9, no. 3, pp. 97–104, 2021.
- [2] F. Nurzaman, "Penerapan Algoritma Regresi Linier Untuk Prediksi Jumlah Klaim Pada Asuransi Kesehatan," *Semin. Nas. Teknol. Inf.*, pp. 105–114, 2017.
- [3] D. Novianty, N. D. Palasara, and M. Qomaruddin, "Algoritma Regresi Linear pada Prediksi Permohonan Paten yang Terdaftar di Indonesia," *J. Sist. dan Teknol. Inf.*, vol. 9, no. 2, p. 81, 2021, doi: 10.26418/justin.v9i2.43664.
- [4] N. Almumtazah, N. Azizah, Y. L. Putri, and D. C. R. Novitasari, "Prediksi Jumlah Mahasiswa Baru Menggunakan Metode Regresi Linier Sederhana," *J. Ilm. Mat. Dan Terap.*, vol. 18, no. 1, pp. 31–40, 2021, doi: 10.22487/2540766x.2021.v18.i1.15465.
- [5] W. P. Hidayanti, "Penerapan Algoritma K-Nearest Neighbor Untuk Klasifikasi Efektivitas Penjualan Vape (Rokok El ektrik) pada ' Lombok Vape On ' Pendahuluan dihasilkan tidak stabil dan tidak mampu diprediksi Dari penelitian yang dilakukan , berusaha untuk mengklasifikasik," vol. 3, no. 2, 2020.
- [6] T. Indarwati, T. Irawati, and E. Rimawati, "Penggunaan Metode Linear Regression Untuk Prediksi Penjualan Smartphone," *J. Teknol. Inf. dan Komun.*, vol. 6, no. 2, pp. 2–7, 2019, doi: 10.30646/tikomsin.v6i2.369.
- [7] F. Ginting, E. Buulolo, and E. R. Siagian, "Implementasi Algoritma Regresi Linear Sederhana Dalam Memprediksi Besaran Pendapatan Daerah (Studi Kasus: Dinas Pendapatan Kab. Deli Serdang)," *KOMIK (Konferensi Nas. Teknol. Inf. dan Komputer)*, vol. 3, no. 1, pp. 274–279, 2019, doi: 10.30865/komik.v3i1.1602.
- [8] T. Khotimah and R. Nindiyasari, "Forecasting Dengan Metode Regresi Linier Pada Sistem Penunjang Keputusan Untuk Memprediksi Jumlah Penjualan Batik (Studi Kasus Kub Sarwo Endah Batik Tulis Lasem)," *J. Mantik Penusa*, vol. 1, no. 1, pp. 71–92, 2017.
- [9] G. N. Ayuni and D. Fitriana, "Penerapan metode Regresi Linear untuk prediksi penjualan properti pada PT XYZ," *J. Telemat.*, vol. 14, no. 2, pp. 79–86, 2019, [Online]. Available: <https://journal.ithb.ac.id/telematika/article/view/321>.
- [10] A. A. Suryanto, "Penerapan Metode Mean Absolute Error (Mea) Dalam Algoritma Regresi Linear Untuk Prediksi Produksi Padi," *Saintekbu*, vol. 11, no. 1, pp. 78–83, 2019, doi: 10.32764/saintekbu.v11i1.298.
- [11] M. F. Arifin and D. Fitriana, "Penerapan Algoritma Klasifikasi C4.5 Dalam Rekomendasi Penerimaan Mitra Penjualan Studi Kasus : PT Atria Artha Persada," *InComTech*, vol. 8, no. 2, pp. 87–102, 2018, doi: 10.22441/incomtech.v8i1.2198.
- [12] Alwafi Ridho Subarkah, "PREDIKSI PENJUALAN PRODUK ROTI MENGGUNAKAN ALGORITMA C4.5 PADA PT. PRIMA TOP BOGA," *Progr. Stud. Tek. Inform. Sekol. Tinggi Teknol. Pelita Bangsa*, vol. 9, no. 2, pp. 10–17, 2018.
- [13] R. D. Shaputra and S. Hidayat, "Implementasi regresi linear untuk prediksi penjualan pada aplikasi point of sales restoran," *Automata*, 2021, [Online]. Available: <https://103.220.113.195/AUTOMATA/article/view/17355>.
- [14] G. W. N. Sandrawira Anggraini, Sarjon Defit, "Analisis Data Mining Penjualan Ban Menggunakan," *J. Ilmu Tek. Elektro Komput. dan Inform.*, vol. 4, no. 2, pp. 136–143, 2018.
- [15] W. M. Baihaqi, M. Dianingrum, and K. A. N. Ramadhan, "Regresi Linier Sederhana Untuk Memprediksi Kunjungan Pasien Di Rumah Sakit Berdasarkan Jenis Layanan Dan Umur

Pasien,” *J. Simetris*, vol. 10, no. 2, pp. 671–680, 2019, [Online]. Available: <https://www.jurnal.umk.ac.id/index.php/simet/article/view/3484>.



Universitas Mercu Buana

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

KERTAS KERJA

Ringkasan

Kertas kerja ini berisi tentang kelengkapan material dari artikel jurnal dengan judul “Analisa Algoritma Regresi Linear dan Decission Tree Dalam Prediksi Penjualan Produk (Studi Kasus: Lookma Boutique)”. 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.

PENDAHULUAN

Perkembangan bisnis secara online di Indonesia sekarang sangat pesat, dengan proses dilakukan dengan cara memesan barang melalui reseller atau distributor menggunakan salah satu jejaring sosial internet. Belanja online merupakan salah satu cara berbelanja melalui alat komunikasi atau jejaring sosial yang digunakan dalam transaksi jual beli (Paruntu et al., 2021), penjualan di masa yang akan datang menentukan volume penjualan. Salah satu toko online yaitu Lookma Boutique, toko tersebut menghasilkan penjualan yang cukup besar setiap bulan dan tahun, tetapi hasil penjualan yang dihasilkan tidak stabil dan tidak mampu diprediksi kenaikan dan penurunan penjualannya. Pembelian item dilakukan berdasarkan informasi merek dagang, size, model dan warna keputusan yang menggunakan model keputusan yang berbentuk seperti pohon. Selain itu, prediksi berguna untuk melihat keadaan produk yang dijual, Oleh karena itu, resiko kesalahan akibat kesalahan perencanaan harus diminimalkan. prediksi biasanya digunakan untuk menemukan informasi dari sejumlah besar data. Kedua algoritma tersebut akan diuji untuk mendapatkan hasil yang terbaik dalam memprediksi penjualan produk dalam satu periode, prediksi ini diharapkan dapat mendukung dan membantu kinerja dalam mengoptimalkan penjualan barang di masa mendatang berdasarkan data yang telah direkam sebelumnya. toko online Lookma Boutique

Pada penelitian sebelumnya peramalan dilakukan dengan metode Regresi Linear yang diimplementasikan penelitian tersebut melakukan prediksi untuk mendapatkan jumlah kasus klaim yang akan ditangani. Prediksi Jumlah kasus klaim akan dibagi berdasarkan total keseluruhan klaim, jenis layanan dan jenis manfaat. Dengan mengetahui jumlah kasus klaim yang ditangani akan mempermudah perencanaan peningkatan dan perbaikan proses penyelesaian penanganan kasus klaim. Jumlah kasus klaim dilihat dari jumlah peserta yang ditanggung dan yang dilayani untuk setiap polis yang terbit (Nurzaman, 2017). Selanjutnya penelitian ini di dapatkan bahwa metode linear regresi layak dan efektif untuk memprediksi pendaftaran paten untuk tahun selanjutnya berdasarkan data pendaftaran, bahwa metode linear regresi layak dan efektif untuk memprediksi pendaftaran paten untuk tahun selanjutnya berdasarkan data pendaftaran dari tahun 2014 sampai tahun

Universitas Mercu Buana

2018(Novianty et al., 2021). Pada penelitian terdahulu terakhir Jumlah mahasiswa baru suatu universitas selama 5 tahun terakhir yaitu mulai tahun 2016 hingga tahun 2020 mengalami kenaikan dan penurunan. Data tersebut selanjutnya akan digunakan untuk memprediksi jumlah mahasiswa 5 tahun mendatang. Tujuan dibuat sebuah prediksi adalah untuk mengetahui rasio dosen yang tersedia dengan jumlah mahasiswa baru, mempersiapkan ruang kuliah dan juga fasilitas lainnya(Ginting et al., 2019).Menjelaskan bahwa kedua metode ini adalah kelebihan dan kekurangan masing-masing saat memprediksi masalah, terutama memprediksi. Apabila toko mengetahui peningkatan atau penurunan dari hasil penjualannya maka toko dapat segera mengambil tindakan untuk mengubah dan meningkatkan strategi yang jauh lebih baik lagi. Selain itu toko harus tetap meningkatkan penjualan untuk keberlangsungan toko.

Dari penelitian sebelumnya mengenai penjualan smartphone dengan menggunakan Algoritma Regresi Linear hasil akurasi tertinggi dengan nilai akurasi Hasil pengujian menunjukkan bahwa prediksi penjualan smartphone dengan metode Linear Regresi adalah tergolong dalam kategori sangat baik. Terbukti dengan pengujian nilai MAPE dan MSE untuk smartphone merek Lenovo kurang dari 10. Berdasarkan pemaparan diatas diperlukan perbandingan algoritma algoritma regresi linear dan algoritma Decision Tree agar menjadi acuan terhadap akurasi dari algoritma perbandingan tersebut(Indarwati et al., 2019).

Penelitian ini berfokus untuk prediksi hasil penjualan dari Lookma Boutique untuk mengetahui efektifitas penjualan berdasarkan popularitas penjualan produk berdasarkan total produk dibayar dan produk pesanan yang dibuat. Dengan perbandingan algoritma yang memiliki keunggulan dalam melakukan prediksi.

RUMUSAN MASALAH

Salah satu toko online lookma boutique menghasilkan penjualan cukup besar setiap bulan dan tahun, tetapi hasil penjualan yang dihasilkan tidak stabil atau mampu diprediksi kenaikan dan penurunan penjualan oleh karena itu butuh kedua algoritma untuk memprediksi penjualan dalam kurun waktu 1 tahun, oleh karena itu disimpulkan bagaimana melakukan prediksi penjualan secara efektif.

Universitas Mercu Buana

TUJUAN DAN MANFAAT

a. Tujuan

Membandingkan kedua algoritma tersebut akan diuji untuk mendapatkan hasil yang terbaik dalam memprediksi penjualan produk dalam satu periode 1 tahun, prediksi ini dapat mendukung dan membantu kinerja dalam mengoptimalkan penjualan.

b. Manfaat

Dengan adanya penelitian dalam bidang peramalan dapat membantu dalam menentukan jumlah produk yang akan diproduksi berdasarkan hasil prediksi sehingga dapat membuat keputusan yang tepat dan prediksi dengan tingkat kesalahan yang minim.

BATASAN MASALAH

Adapun batasan masalah dalam penelitian ini adalah sebagai berikut:

- Dalam memprediksi variable yang digunakan adalah total produk dibayar dan produk pesanan yang dibuat.
- Output dari hasil penelitian merupakan hasil prediksi satu waktu didepan.

STUDI LITERATUR

1. Algoritma Regresi Linear

Metode Regresi, yaitu sebuah metode statistik juga yang melakukan peramalan dengan menggunakan pengembangan hubungan matematis antara variable yaitu variable dependen (Y) dengan variable Independen (X). Dalam metode regresi suatu model perlu dispesifikasikan sebelum dilakukan pengumpulan data (Khotimah & Nindiyasari, 2017). Model matematis dalam regresi ini terdapat pada persamaan 1.

$$Y = a + bx \quad (1)$$

Dimana,

Y = Variable Response atau Variable Akibat (Dependent)

Universitas Mercu Buana

X = Variable Predictor atau Variable faktor penyebab (Independent)

a = konstanta

b = koefisien regresi.

Nilai a dan b dapat dihitung dengan menggunakan persamaan 2 dan 3

$$a = \frac{(\sum y)(\sum x^2) - (\sum x)(\sum xy)}{n(\sum x^2) - (\sum x)^2} \quad (2)$$

$$b = \frac{n(\sum xy) - (\sum x)(\sum y)}{n(\sum x^2) - (\sum x)^2} \quad (3)$$

2. Algoritma Decision Tree

Algoritma C4.5 artinya perluasan asal algoritma ID3 dan memakai prinsip Decision Tree yang seperti. algoritma C4.5 memiliki kelebihan bisa mengolah data numerik, dapat menangani nilai atribut yang hilang, menghasilkan aturan-aturan yang praktis diinterpretasikan serta performanya adalah salah satu yang tercepat dibandingkan menggunakan algoritma lain (Moh. Irfan Rusdi, Triawan Adi Cahyanto, n.d.).

Menentukan pohon keputusan pada algoritma C4.5 terdapat 4 rumus sebagai berikut :

Menghitung nilai *entropy* dengan melakukan nilai persamaan :

$$Entropy(S) = -\sum_{i=1}^n p_i * \log_2 p_i \quad (1)$$

Keterangan:

S = Jumlah kasus

n = Total himpunan S

p_i = Proporsi ke i

Menghitung nilai *Gain* dengan melakukan persamaan :

$$Gain(S, A) = (S) - \sum_{|S_i|} \frac{|S_i|}{|S|} * Entropy(S_i) \quad n \ i = 1 \quad (2)$$

Keterangan :

S = Jumlah kasus

A = Karakter

n = Total A

$|S_i|$ = Jumlah kasus pada partisi ke i

$|S|$ = Jumlah kasus dalam S

Selanjutnya nilai *Split Info* dihitung dengan persamaan :

$$SplitInfo(S, A) = - \sum \frac{S_i}{S} \log \log \frac{S_i}{S} \quad n \ i = 1 \quad (3)$$

Keterangan :

S = Jumlah kasus

A = Karakter

S_i = Jumlah sample karakter i

Yang menentukan sebuah atribut dapat dijadikan akar maupun cabang suatu pohon keputusan didapat dari nilai *Gain Ratio* dengan persamaan :

$$Gain\ Ratio(S, A) = \frac{Gain(S, A)}{SplitInfo(S, A)} \quad (4)$$

Keterangan :

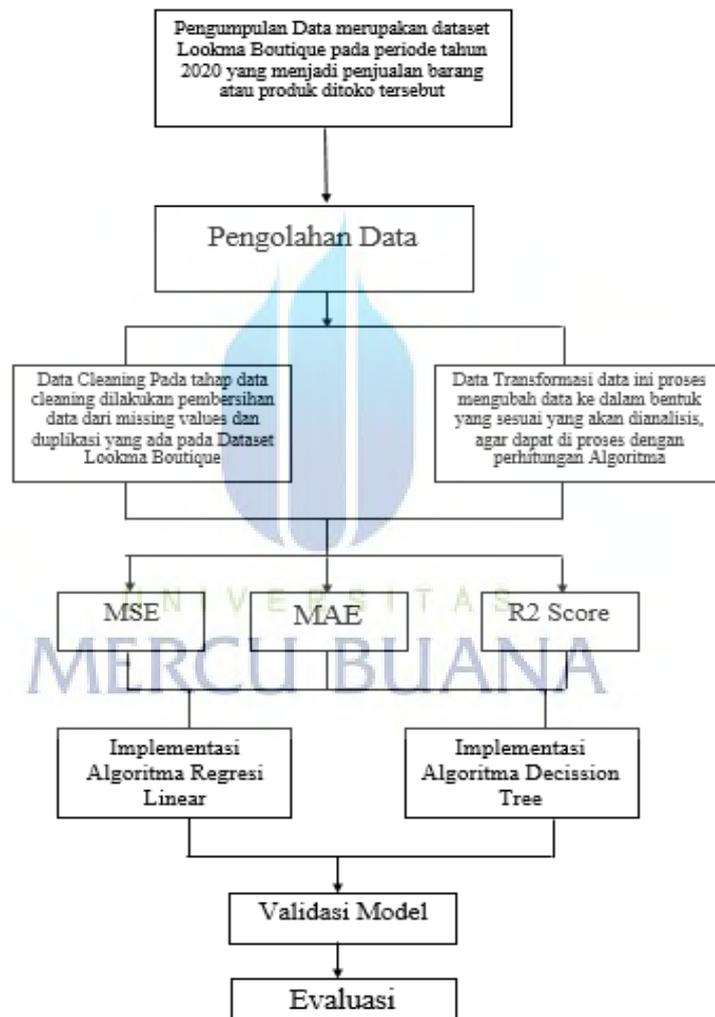
$Gain(S, A)$ = Info gain pada karakter A

$SplitInfo(S, A)$ = Split info pada karakter A

3. Prediksi

Suatu aktivitas memperkirakan peristiwa dimasa yang akan datang tentunya dengan bantuan penyusunan planning terlebih dahulu, dimana planning ini dirancang berdasarkan kapasitas serta kemampuan permintaan atau produksi(Lusiana & Yuliarty, 2020).

METODELOGI



Gambar 6. Diagram Metodologi

Penjelasan terkait metodologi pada Gambar 1, dapat dijelaskan sebagai berikut:

1. Pengumpulan Data

Data yang dikumpulkan berupa data Penjualan Lookma Boutique untuk melakukan penelitian dan mencari informasi-informasi yang dibutuhkan dalam menyelesaikan penelitian ini.

2. Pengolahan Data Awal (Pre-processing)

Data yang telah diperoleh perlu dilakukan pre-processing. Pre-processing merupakan langkah penting yang digunakan untuk mengubah data mentah menjadi format yang memungkinkan untuk menerapkan teknik data mining dan juga untuk meningkatkan kualitas data.

Terdapat beberapa jenis pre-processing yang akan digunakan dalam penelitian ini, antara lain:

2.1 Data Cleaning

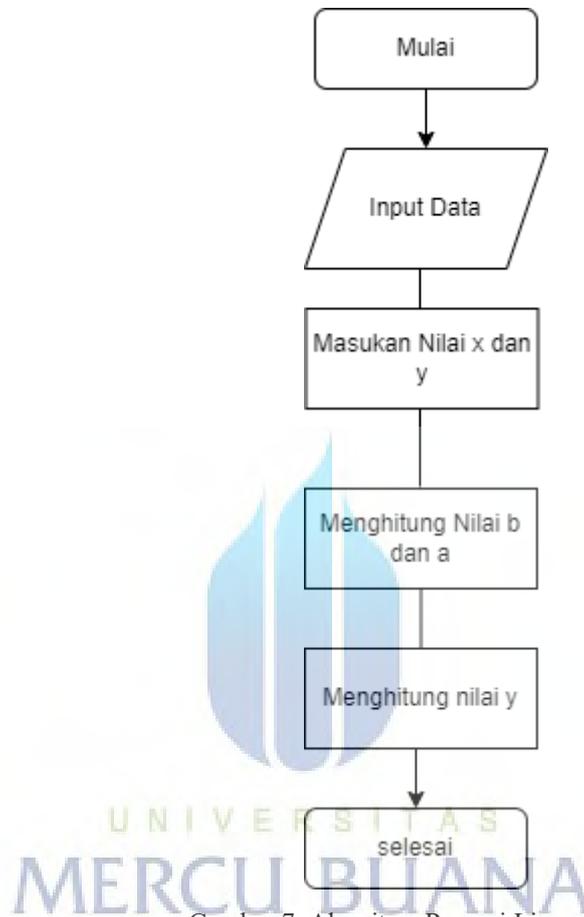
Pada tahap data cleaning dilakukan pembersihan data dari missing values dan duplikasi yang ada pada Lookma Boutique. Missing values adalah kondisi dimana suatu atribut memiliki nilai yang tidak valid. Duplikasi data adalah kondisi dimana semua atribut suatu data memiliki kesamaan dengan semua atribut suatu data lain. Pengecekan missing values dilakukan pada keseluruhan dataset, sedangkan pengecekan duplikasi dilakukan pada data per tahun.

2.2 Data Transformation

Transformasi data adalah proses mentransformasi atau mengubah data ke dalam bentuk yang sesuai, agar dapat di proses dengan perhitungan Algoritma Decision Tree dan Algoritma Regresi Linear.

3. Implementasi Regresi Linear dan Algoritma Decision Tree

3.1. Regresi Linear

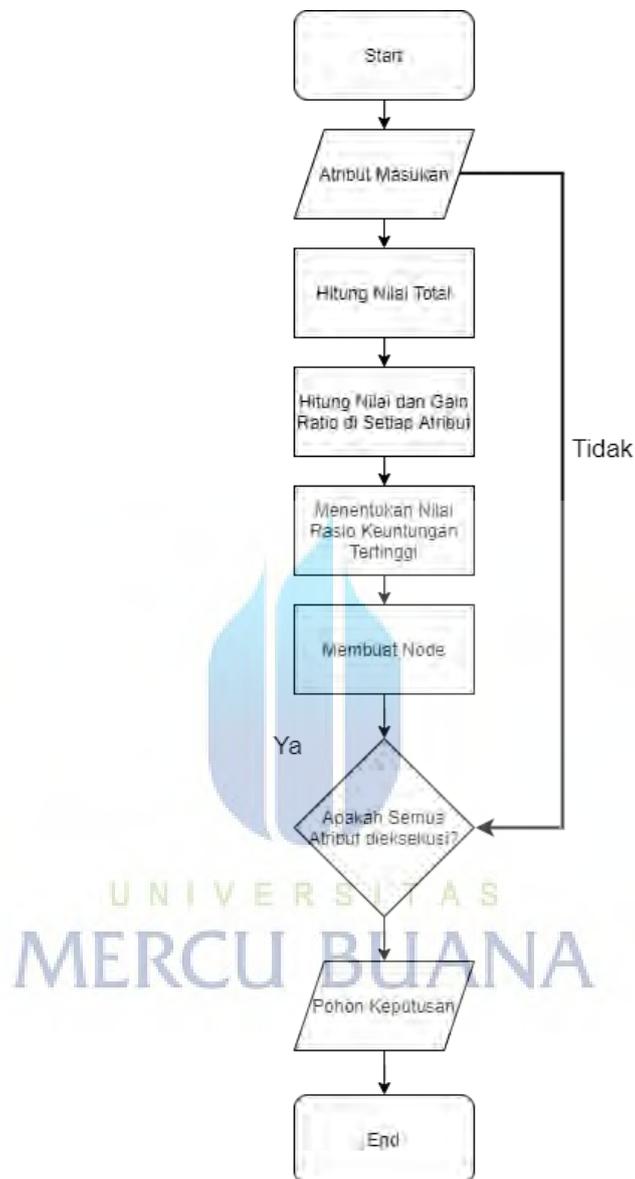


Gambar 7. Algoritma Regresi Linear

Gambar 2 dapat dijelaskan sebagai berikut yaitu, Regresi Linear adalah Metode regresi merupakan sebuah metode statistik yang melakukan prediksi menggunakan pengembangan hubungan matematis antara variabel, yaitu variabel dependen (Y) dengan variabel independen (X). Variabel dependen merupakan variabel akibat atau variabel yang dipengaruhi, sedangkan variabel independen merupakan variabel sebab atau variabel yang mempengaruhi. Prediksi terhadap nilai variabel dependen dapat dilakukan jika variabel independennya diketahui. Umumnya penjualan atau permintaan suatu produk dinyatakan sebagai variabel dependen yang besar atau nilainya dipengaruhi oleh variabel independen.

Universitas Mercu Buana

3.2 Algoritma Decision Tree



Gambar 8. Algoritma Decision Tree

Dari diagram 2 dapat dijelaskan mempersiapkan data training, bisa diambil asal data histori yang pernah terjadi sebelumnya serta telah dikelompokkan pada kelas-kelas tertentu dan menentukan akar berasal pohon dengan menghitung nilai gain yang tertinggi dari masing-masing atribut atau berdasarkan nilai index entropy terendah. Sebelumnya dihitung terlebih dahulu nilai index entropy (Hasan et al., 2018).

4. Validasi Model

Pada tahapan validasi model merupakan pengambil keputusan dapat mempercayai model yang digunakan sebagai bagian dari proses pengambilan keputusan. Membandingkan output ukuran kinerja model simulasi dengan ukuran kinerja yang sesuai dari sistem nyata adalah metode yang paling sesuai untuk melakukan validasi model simulasi.

5. Evaluasi

Pada Tahap Evaluasi dan Peninjauan model ini akan dibahas mengenai hasil evaluasi yang telah dikerjakan. Phase ini akan dilakukan perbandingan kuantitatif dengan mempertimbangkan nilai komparasi *MAE*, *MSE*, *R2 Score*.

