

## **LAPORAN PENELITIAN**

### **Prediksi Jumlah Kendaraan Bermotor Menggunakan Algoritma Regresi Linear**

Diajukan sebagai salah satu syarat untuk menempuh Tugas Akhir  
pada Mata Kuliah Data Science



Disusun Oleh :

Anggun Saraswati Putri Iryani (312010269)

**PROGRAM STUDI TEKNIK INFORMATIKA  
FAKULTAS TEKNIK  
UNIVERSITAS PELITA BANGSA  
BEKASI  
2023**

# **BAB I**

## **PENDAHULUAN**

Kota-kota cenderung menjadi pusat pertumbuhan populasi dan aktivitas ekonomi. Seiring dengan urbanisasi, kebutuhan akan transportasi meningkat. Kebutuhan akan alat transportasi telah menjadi kebutuhan dasar di kalangan masyarakat saat ini. Mayoritas orang lebih suka memiliki alat transportasi pribadi daripada menggunakan transportasi umum dan ini dapat tercermin dalam peningkatan jumlah kendaraan bermotor.

Setiap tahun, jumlah kendaraan bermotor di Indonesia terus mengalami peningkatan. Berdasarkan data BPS jumlah kendaraan bermotor di Indonesia pada tahun 2021 sebanyak 143.797.227. Jumlah tersebut meningkat 5,7% dibanding tahun sebelumnya menciptakan tantangan terkait kemacetan di berbagai kota besar di negara ini. Masalah lalu lintas yang disebabkan oleh penumpukan kendaraan setiap hari menjadi kenyataan di beberapa area jalan.

Informasi peningkatan jumlah kendaraan bermotor perlu diketahui karena kendaraan bermotor membutuhkan sarana jalan. Peningkatan jumlah kendaraan tanpa dibarengi oleh penambahan sarana jalan akan mengakibatkan timbulnya kemacetan yang parah. Dalam menambah sarana jalan perlu dilakukan perencanaan matang. Diantaranya dengan memprediksi jumlah kendaraan bermotor dimasa mendatang. Dengan mengetahui jumlah kendaraan bermotor di masa mendatang maka dapat diketahui seberapa besar sarana jalan yang dibutuhkan.

Salah satu metode yang dapat digunakan untuk meramalkan perkembangan jumlah pengendara kendaraan bermotor di Jawa Tengah dalam beberapa tahun mendatang adalah dengan melakukan prediksi melalui metode algoritma regresi linier. Regresi linear adalah sebuah konsep statistika regresi yang memodelkan hubungan antara dua variabel dan termasuk dalam konsep Supervised Machine Learning.

Machine learning adalah istilah yang merujuk pada berbagai algoritma yang melakukan prediksi cerdas berdasarkan kumpulan data. Sedangkan Supervised Learning dinyatakan oleh (Cord dan Cunningham, 2008) sebagai cara untuk

memberikan mapping terhadap input dan output dari data yang ada dan menghasilkan prediksi dari mapping yang telah dibuat.

## **BAB II**

### **Tinjauan Pustaka**

#### **2.1 Tinjauan Pustaka**

Bab ini memuat penelitian terdahulu yang berguna sebagai bahan acuan penulis untuk melakukan penelitian ini.

##### **2.1.1 Kajian jurnal pertama**

Sebuah Jurnal dengan judul “ Prediksi Jumlah Kendaraan Bermotor Menggunakan Machine Learning”. Penelitian ini bertujuan untuk memprediksi jumlah kendaraan bermotor dimasa yang akan datang. Dengan mengetahui hal tersebut maka akan diketahui juga seberapa besar sarana jalan yang dibutuhkan.

#### **2.2 Landasan Teori**

##### **2.2.1 Machine learning**

Machine learning merupakan bidang lain dari ilmu komputer yang merancang sebuah algoritma agar memungkinkan sebuah komputer untuk belajar melalui data sehingga sering dikatakan sebagai learn from data. Jadi machine learning adalah pemrograman komputer yang menggunakan data masa lalu yang digunakan untuk pembelajaran model sehingga mendapatkan performa yang optimal dalam menggali informasi dari suatu kumpulan data[1].Inti machine learning adalah untuk membuat model yang merefleksikan pola-pola data. Sedangkan menurut Tom M. Mitchell definisi Machine Learning adalah sebuah program komputer yang belajar dari pengalaman dari tugas yang dibebankan dengan kinerjanya yang terukur. Secara garis besar machine learning dibagi menjadi 3 yaitu Supervised Learning , Unsupervised Learning dan Reinforcement Learning. a. Supervised Learning adalah algoritma machine learning yang proses pembelajarannya dibawah pengawasan. Yang termasuk dalam supervised learning diantaranya adalah classification dan regression. b. Unsupervised Learning adalah algoritma machine learning yang proses pembelajarannya tanpa

pengawasan. Yang termasuk dalam unsupervised learning diantaranya adalah clustering dan dimensionality reduction. c. Reinforcement Learning adalah algoritma machine learning yang bisa membuat agent software mesin bekerja dan bekerja secara otomatis untuk menentukan perilaku ideal sehingga dapat memaksimalkan kinerja algoritma. Yang termasuk dalam Reinforcement Learning diantaranya adalah real-time decision, robot navigation, learning tasks, skills acquisition dan game AI.

#### 2.2.1 Apache spark

*Apache Spark* merupakan *tools Big Data* yang sangat berguna untuk membangun jalur pemrosesan data dengan mudah, didukung oleh beberapa jenis bahasa pemrograman dan menyediakan berbagai *library* yang dapat memenuhi kebutuhan pemrosesan data. Kita dapat mengakses hingga *petabyte* data dari berbagai sumber penyimpanan berbeda dan memprosesnya secara cepat dengan menyiapkan beberapa *node server* yang terinstall *framework Apache Spark*. *Apache Spark* juga dilengkapi dengan *library* untuk memenuhi kebutuhan analisis data seperti *GraphX* untuk komputasi grafik, dan *MLlib* untuk memenuhi kebutuhan pengolahan data menggunakan *machine learning*.

#### 2.2.2. Regresi

Regresi adalah metode statistik yang dipakai untuk memperkirakan hubungan antara sebuah variabel terkait dengan satu variabel independen atau lebih. Metode ini juga bisa digunakan untuk menilai kekuatan hubungan antara variabel dengan perkiraan masa depan. Regresi linear sederhana merupakan suatu model persamaan yang menggambarkan hubungan satu variabel bebas (X) dengan satu variabel tak bebas (Y). Persamaan regresi linear sederhana secara matematik diekspresikan oleh

$$Y = a + bX$$

$\hat{Y}$  = garis regresi

a = konstanta (intersep)

b = konstanta regresi (slope)

X = variabel bebas (predictor)

Besarnya konstanta a dan b dapat ditentukan menggunakan persamaan :

$$a = \frac{(\sum y_i) (\sum x_i^2) - (\sum x_i)(\sum x_i y_i)}{n \sum x_i^2 - (\sum x_i)^2}$$

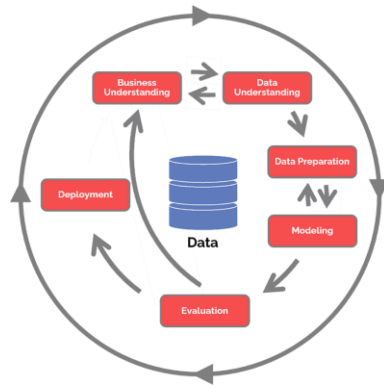
$$b = \frac{n(\sum x_i y_i) - (\sum x_i) (\sum y_i)}{n \sum x_i^2 - (\sum x_i)^2}$$

Dimana n=jumlah data

### **BAB III**

## **METODE PENELITIAN**

METODOLOGI Dalam penelitian ini metode yang digunakan adalah CRISP DM (Cross Industry Standard Processing for Data Mining).



CRISP-DM merupakan standar yang dikembangkan di Eropa yang terdiri atas 6 tahap yaitu

- a. Business Understanding Merupakan tahap dimana permasalahan didefinisikan dengan tepat
- b. Data Understanding Proses dimana mempertemukan antara data yang kita miliki dan data apa yang seharusnya diperlukan
- c. Data Preparation Proses persiapan data untuk diolah menggunakan model yang tepat.
- d. Modeling Pembuatan model sesuai dengan karakteristik data, tujuan teknis dan parameter model.
- e. Evaluasi Menguji model yang dibuat berdasarkan data dan parameter yang relevan.
- f. Deployment Mengembangkan model yang sudah dibuat dengan mempertimbangkan diantaranya kemudahan penggunaan model dan pemeliharannya di masa mendatang

## BAB IV

### Hasil dan Pembahasan

Metode penelitian dimulai dengan tahap pemahaman bisnis, di mana masalah diidentifikasi secara tepat. Yang mana penelitian ini digunakan untuk memprediksi jumlah kendaraan bermotor di Jawa Tengah melalui data dari BPS (Badan Pusat Statistik) Jawa Tengah. Selanjutnya, tahap pemahaman data dilakukan untuk mencocokkan data yang telah diperoleh dengan data yang diperlukan. Proses persiapan data kemudian dilakukan, melibatkan langkah-langkah pra-pemrosesan seperti pembersihan dan reduksi data sebelum data siap untuk diproses menggunakan model yang sesuai. Adapun data yang didapatkan adalah seperti berikut :

Tabel 1 Data Jumlah Kendaraan Bermotor Berdasarkan Kota/Kab Pada Rentang Tahun 2017-2021.

Kabupaten	Tahun	Jumlah
Kabupaten Cilacap	2017	333292
Kabupaten Cilacap	2018	685658
Kabupaten Cilacap	2019	725053
Kabupaten Cilacap	2020	752008
Kabupaten Cilacap	2021	806833
Kabupaten Banyumas	2017	604427
Kabupaten Banyumas	2018	637506
Kabupaten Banyumas	2019	673506
Kabupaten Banyumas	2020	693638
Kabupaten Banyumas	2021	739001
Kabupaten Purbalingga	2017	20932
Kabupaten Purbalingga	2018	329288
Kabupaten Purbalingga	2019	349584
Kabupaten Purbalingga	2020	361588
Kabupaten Purbalingga	2021	382763
Kabupaten Banjarnegara	2017	14698
Kabupaten Banjarnegara	2018	294371
Kabupaten Banjarnegara	2019	312287
Kabupaten Banjarnegara	2020	320372
Kabupaten Banjarnegara	2021	328457
Kabupaten Kebumen	2017	29959
Kabupaten Kebumen	2018	449431
Kabupaten Kebumen	2019	473039
Kabupaten Kebumen	2020	494111
Kabupaten Kebumen	2021	520274
Kabupaten Purworejo	2017	198791
Kabupaten Purworejo	2018	273930
Kabupaten Purworejo	2019	287428
Kabupaten Purworejo	2020	297285
Kabupaten Purworejo	2021	307142

Langkah selanjutnya adalah tahap pemodelan, di mana model dibuat dengan mempertimbangkan karakteristik data yang ada, tujuan teknis, dan parameter-model yang relevan dengan menggunakan regresi linear pada bahasa pemrograman python dengan library Sklearn, Numpy dan Matplotlib. Dalam tahap pemodelan ini dilakukan visualisasi dan diterapkan regresi linear terhadap data yang telah disiapkan.

Persamaan regresi linear sederhana secara matematik diekspresikan oleh

$$Y = a + bX$$

$\hat{Y}$  = garis regresi

a = konstanta (intersep)

b = konstanta regresi (slope)

X = variabel bebas (predictor)

Besarnya konstanta a dan b dapat ditentukan menggunakan persamaan :

$$a = \frac{(\sum y_i) (\sum x_i^2) - (\sum x_i)(\sum x_i y_i)}{n \sum x_i^2 - (\sum x_i)^2}$$

$$b = \frac{n(\sum x_i y_i) - (\sum x_i) (\sum y_i)}{n \sum x_i^2 - (\sum x_i)^2}$$

Dimana n=jumlah data

Terakhir, evaluasi dilakukan untuk menguji keberlanjutan model yang telah dibuat, berdasarkan pada data dan parameter yang sesuai dengan metode evaluasi R<sup>2</sup>. R<sup>2</sup> mencerminkan sejauh mana variabel bebas mempengaruhi variabel terikat. Semakin mendekati nilai satu, maka hasil dari model regresi akan semakin baik.

$$R = 1 - \frac{SS\ Error}{SS\ Total} = 1 - \frac{\sum (y_i - \hat{y}_i)^2}{\sum (y_i - \bar{y})^2}$$

*Keterangan :*

$y_i$  = observasi respon ke - i

$\bar{y}$  = rata - rata

$\hat{y}_i$  = prediksi respon ke - i

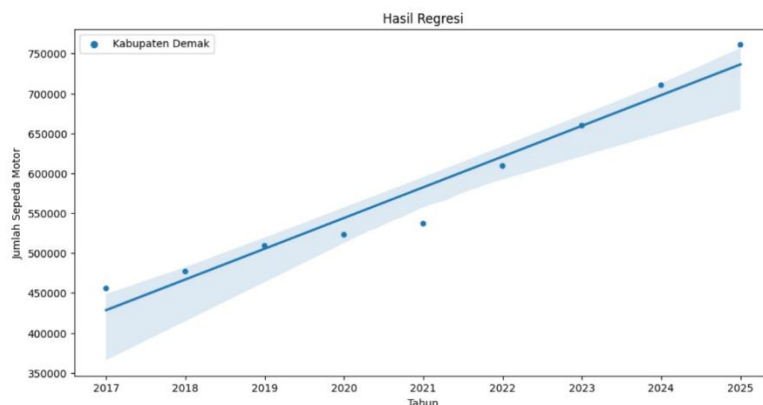
Model regresi merupakan hasil proses identifikasi parameter regresi yang dapat dengan akurat memprediksi variabel terikat berdasarkan variabel bebas. Dalam penelitian ini variabel terikatnya adalah jumlah kendaraan bermotor dan variabel



bebasnya adalah tahun. Model yang digunakan adalah model Regresi Linear. Model ini mempunyai kelebihan dalam penggunaannya yang sederhana, mudah dipahami dan diinterpretasikan.

Table hasil Prediksi jumlah kendaraan bermotor

Tahun	Kabupaten	prediction
2022	Kabupaten	584886
2022	Kabupaten	606123
2022	Kabupaten	545446
2022	Kabupaten	597022
2022	Kabupaten	581852
2022	Kabupaten	590954
2022	Kabupaten	618259
2022	Kabupaten	566683
2022	Kabupaten	569717
2022	Kabupaten	630395
2022	Kabupaten	621293
2022	Kabupaten	557581
2022	Kabupaten	627361
2022	Kabupaten	600056
2022	Kabupaten	548479
2022	Kabupaten	612191
2022	Kabupaten	560615
2022	Kabupaten	572751
2022	Kabupaten	587920



Visualisasi hasil regresi linear

Dalam model Regresi Linear yang digunakan dalam penelitian ini, digunakan metode evaluasi R<sup>2</sup>. R<sup>2</sup> mencerminkan sejauh mana variabel bebas mempengaruhi variabel terikat. Semakin mendekati nilai satu, maka hasil dari model regresi akan semakin baik. Dengan menggunakan metode evaluasi R<sup>2</sup>, didapatkan nilai R<sup>2</sup> dari model yang telah dibuat sebesar 0.86. Hasil evaluasi ini menunjukkan bahwa model regresi ini memiliki tingkat akurasi yang cukup baik.

## **BAB IV**

### **Kesimpulan dan Saran**

Penelitian ini menunjukkan bahwa model regresi linear dengan Bahasa pemrograman Python dapat memberikan prediksi jumlah kendaraan bermotor di Surabaya dengan akurasi cukup baik ( $R^2 = 0.86$ ), dengan hasil bahwa pada tahun 2023 Surabaya akan memiliki 3.192.527 Kendaraan bermotor.

Meskipun memiliki kelebihan dalam memberikan prediksi yang akurat dan mudah diinterpretasikan, model ini memiliki keterbatasan, seperti asumsi linearitas yang dapat mempengaruhi representasi hubungan yang sebenarnya.

Sehingga diperlukan kebijakan adaptif dan pemantauan terus-menerus terhadap perkembangan jumlah kendaraan bermotor untuk mengatasi dampak peningkatan tersebut di masa mendatang.

## DAFTAR PUSTAKA

- Heizer, Jay dan Render, Barry. 2015. *Manajemen Operasi*. Jakarta: Salemba Empat.
- Montgomery, Douglas C., et al. *Introduction to Linear Regression Analysis*. Wiley, 2021.
- Cord, Matthieu, and Pádraig Cunningham, editors. *Machine Learning Techniques for Multimedia: Case Studies on Organization and Retrieval*. Springer, 2008.
- Nichols JA, Herbert Chan HW, Baker MAB. Machine learning: applications of artificial intelligence to imaging and diagnosis. *Biophys Rev*. 2019 Feb;11(1):111-118. doi:10.1007/s12551-018-0449-9. Epub 2018 Sep 4. PMID: 30182201; PMCID: PMC6381354.
- BPS Provinsi Jawa tengah. "Badan Pusat Statistik." Badan Pusat Statistik 23 juli 2020 <https://jateng.bps.go.id/indicator/17/1006/1/jumlah-kendaraan-bermotor-menurut-kabupaten-kota-dan-jenis-kendaraan-di-provinsi-jawa-tengah.html>