

## BAB II STUDI LITERATUR

### 2.1 Tinjauan Pustaka

Penelitian ini dikerjakan tidak luput dari hasil penelitian yang ada sebelumnya yang akan dijadikan sebagai pengetahuan yang mendukung ataupun perbandingan untuk penelitian ini. Hasil-hasil penelitian tersebut berkaitan dengan topik regresi linear yang dilakukan penelitian ini. Pada penelitian terkait prediksi jumlah kunjungan wisatawan mancanegara pernah dilakukan oleh Vivi Nur Wijyaningrum dan Novi Nur Putrawijaya tahun 2019 [9]. Pada penelitian tersebut menggunakan metode *Support Vector Regression* (SVR). Penelitian ini melakukan pengujian enam parameter SVR yang terdiri dari pengujian banyaknya iterasi, nilai lambda, nilai epsilon, nilai cLR, nilai konstanta C, dan nilai sigma. Hasil pengujian menunjukkan bahwa besar kecilnya nilai parameter SVR yang digunakan akan mempengaruhi nilai MAPE yang dihasilkan dari prediksi. Parameter-parameter optimal yang dapat memberikan nilai error terkecil adalah iterasi = 200, lambda = 3, epsilon = 0.01, cLR = 0.00002, konstanta C = 0.00007, dan sigma = 0.25, yaitu dengan menghasilkan nilai MAPE sebesar 14.32644% dan RMSE sebesar 0.19010.

Penelitian lainnya yang dilakukan oleh Fransiskus Ginting, dkk adalah menghasilkan suatu nilai dari prediksi besaran pendapatan daerah di Kabupaten Deli Serdang menggunakan metode regresi linear sederhana pada tahun 2019 [6]. Dengan memisahkan data tersebut menjadi 2 bulan, 3 bulan, 4 bulan, 6 bulan, dan 1 tahun sehingga diperoleh nilai  $Y = 703.249,01$ . Pendapatan pokok wilayah di Deli Serdang pada tahun 2019 sekitar adalah Rp. 703.249.000.000.01.

Pada penelitian yang dilakukan oleh Nur Nafi'iyah pada tahun pada tahun 2016 [10]. Penelitian tersebut membandingkan metode regresi linear, backpropagation, dan *fuzzy* mamdani untuk memprediksi harga emas. Dalam proses pencarian persamaan regresi linier, peneliti mengubah nilai pada rentang 0 sampai dengan 1. Nilai minimum dan maksimum dari variabel awal masing-masing adalah 466000 dan 574000. Nilai minimum dan maksimum variabel akhir secara berurutan

454000 dan 574000. Nilai minimum dan maksimum variabel harga jual secara berurutan 460000 dan 574000. Di mana hasil persamaan tersebut seperti berikut:  $y' = -0,008 + 0,1102x_1 + 0,879x_2$ . Prediksi dibuat dari bobot yang dihitung, menghasilkan nilai korelasi 0,95. Hal ini menunjukkan bahwa nilai kesalahan *backpropagation* sekitar 0,05 ketika memprediksi harga emas. Harga emas yang diprediksi menggunakan algoritma fuzzy mamdani memiliki nilai error yang sangat tinggi. Nilai korelasinya hanya 0,00000141. Artinya nilai-nilai tersebut tidak mewakili hubungan antar variabel. Penelitian tersebut menyimpulkan bahwa algoritma *backpropagation* mempunyai akurasi 95% dan regresi linear sebesar 93% sedang kan *fuzzy mamdani* mempunyai nilai nilai akurasi sistem tidak sampai 1%.

Penelitian menggunakan metode regresi linear untuk prediksi kebutuhan energi listrik jangka panjang telah dilakukan oleh M. Syafruddin, dkk pada tahun 2014 [11]. Berbagai model regresi linier telah dibangun dengan memvariasikan parameter dalam memprediksi kebutuhan listrik jangka panjang di Provinsi Lampung. Setelah bereksperimen dengan model yang berbeda, kami menyimpulkan bahwa untuk model daya terhubung didapat 6 parameter karena kesalahan standar sangat kecil dan R2 lebih besar dari model 4 dan 5 parameter lainnya. Faktor-faktor yang mempengaruhi permintaan listrik di wilayah Lampung dapat dibedakan menjadi dua parameter yaitu parameter ekonomi dan parameter kelistrikan. Menurut perkiraan daya listrik terhubung untuk perumahan tahun 2014, adalah 1.399,06 MVA, tetapi pada tahun 2028 akan mencapai 2.034,35 MVA, meningkat 31,23% (peningkatan rata-rata 2,46%).

Pada tahun 2017 sebuah penelitian yang dilakukan oleh Petrus Katemba dan Rosita Koro Djoh yang memprediksi tingkat produksi kopi menggunakan metode regresi linear [12]. Dalam penelitian ini, data produksi kopi diolah terlebih dahulu (Preprocessing) menggunakan metode missing value dan data didistribusikan dengan menggunakan Leave One Out cross-validation. kemudian menggunakan regresi linier untuk membuat prediksi dan menggunakan mean squared error (MSE) untuk menilai errornya. Berdasarkan analisis regresi linier sederhana yang meliputi lima periode dari tahun 2011 sampai dengan tahun 2015, didapatkan nilai MSE sebesar 431,12% dan nilai MAPE sebesar 20,001%. Dengan memprediksi 3

tahun ke depan dan menghitung nilai MSE pada iterasi ke-5 sampai iterasi ke-8, nilai MSE yang diperoleh adalah 43,111% hingga 26,944% kemudian nilai MAPE adalah 20.001% hingga 12.500%. Tingkat akurasi yang tinggi didapatkan dengan menggunakan MAPE setelah dilakukan pengujian antara MSE dan MAPE. Data yang digunakan untuk memprediksi kopi adalah data rentang waktu, kemudian Setelah menghitung prediksi untuk periode 2011-2015 nilai paling tinggi yaitu 1.537,38 ton pada 2015 dan terendah 1.109.944 ton pada tahun 2011.

Tabel 2.1 Penelitian Terkait

No	Nama Peneliti	Judul Penelitian	Tahun	Hasil Penelitian
1	Vivi Nur Wijyaningrum dan Novi Nur Putrawijaya	<i>Support Vector Regression</i> Untuk Memprediksi Jumlah Kunjungan Wisatawan Mancanegara Di Pulau Bali	2019	Pada penelitian tersebut parameter optimalnya iterasi = 200, lambda = 3, epsilon = 0.01, cLR = 0.00002, konstanta C = 0.00007, dan sigma = 0.25, yaitu dengan menghasilkan nilai MAPE sebesar 14.32644% dan RMSE sebesar 0.19010.
2	Fransikus Ginting, Efori Buulolo dan Edward R Siagian	Implementasi Algoritma Regresi Linear Sederhana Dalam Memprediksi Besaran Pendapatan Daerah (Studi Kasus: Dinas Pendapatan Kab. Deli Serdang)	2019	Penelitian tersebut membagi data menjadi 2 bulan, 3 bulan, 4 bulan, 6 bulan, 1 Tahun, sehingga diperoleh nilai $Y = 703.249,01$ . Pendapatan pokok wilayah di Deli Serdang pada tahun 2019 sekitar adalah Rp. 703.249.000.000.01.
3	Nur Nafi'iyah	Perbandingan Regresi Linear, <i>Backpropagation</i> , Dan <i>Fuzzy Mamdani</i> Dalam Prediksi Harga Emas	2016	Penelitian tersebut menyimpulkan bahwa algoritma <i>backpropagation</i> mempunyai akurasi 95% dan regresi linear sebesar 93% sedang kan <i>fuzzy mamdani</i> mempunyai nilai nilai

				akurasi sistem tidak sampai 1%.
4	M. syafruddin, Lukmanul Hakim dan Dikpride Despa	Metode Regresi Linier Untuk Prediksi Kebutuhan Energi Listrik Jangka Panjang (Studi Kasus: Provinsi Lampung)	2014	Pada penelitian tersebut Faktor-faktor yang mempengaruhi permintaan listrik di Provinsi Lampung dapat dibedakan menjadi dua parameter yaitu parameter ekonomi dan parameter kelistrikan. Menurut hasil prakiraan sambungan listrik rumah tangga tahun 2014, kapasitas sambungan sebesar 1.399,06 MVA, namun pada tahun 2028 mengalami peningkatan sebesar 2.034,35 MVA atau meningkat 31,23% (rata-rata laju pertumbuhan 2,46 %).
5	Petrus Kutemba dan Rosita Koro Djoh	Prediksi Tingkat Produksi Kopi Menggunakan Regresi Linear	2017	Dengan menggunakan data time series dari penelitian ini, mean square error (MSE) dari tahun 2011 hingga 2015 adalah 431,12% dan nilai MAPE adalah 20,001%. Dengan memprediksi 3 tahun ke depan dan menghitung nilai MSE pada iterasi ke-5 sampai iterasi ke-8, nilai MSE yang diperoleh adalah 43,111% hingga 26,944% kemudian nilai MAPE adalah 20.001% hingga 12.500%. nilai hasil prediksi tertinggi pada penelitian ini adalah pada tahun 2015 sebesar 1.537,38 ton dan nilai terendah pada tahun

				2011 sebesar 1.109,944 ton.
--	--	--	--	-----------------------------

Berdasarkan Tabel 2.1 penulis menggunakan beberapa penelitian terdahulu sebagai acuan pada penelitian ini. Penulis mendapati satu penelitian dengan topik yang sama yaitu memprediksi jumlah kunjungan wisatawan oleh Vivi Nur Wijayaningrum dan Novi Nur Putrawijaya tetapi penelitian tersebut memakai metode atau algoritma *Support Vector Regression* (SVR). Pada penelitian oleh Nur Nafi'iyah penelitian tersebut membandingkan beberapa metode yaitu *backpropagation*, regresi linear, dan fuzzy mamdani. Sedangkan Penelitian yang lainnya semua menggunakan metode regresi linear untuk memprediksi. Yaitu memprediksi besaran pendapatan daerah oleh Fransiskus Ginting, dkk, memprediksi kebutuhan energi listrik jangka panjang oleh M. Syafruddin, dkk, dan memprediksi tingkat produksi kopi oleh Petrus Kutemba dan Rosita Koro Djoh.

Disini Peneliti ingin memprediksi jumlah kunjungan wisatawan menggunakan metode regresi linear sederhana dimana dataset yang diteliti merupakan data runtun waktu (*Time Series*) kemudian diuji keakurasiannya dengan menggunakan *Mean Absolute Percentage Error* (MAPE). Berdasarkan uraian sebelumnya penelitian-peneelitian terdahulu tidak ada yang melakukan hal tersebut. Maka dari itu penelitian ini merupakan penelitian yang berbeda dengan penelitian- penelitian terdahulu yang dipakai sebagai acuan.

## 2.2 Wisatawan

Menurut Spilane, pariwisata adalah perjalanan dari satu tempat ke tempat lain yang sifatnya sementara, baik Anda adalah bagian dari individu atau kelompok yang mencari keseimbangan dan keselarasan dengan lingkungan hidup, masyarakat, budaya, alam, dan ilmu pengetahuan, dengan kata lain pariwisata hanya untuk bersenang-senang. Pariwisata dapat dibagi menjadi dua kelompok, pariwisata nasional dan pariwisata internasional. Kedua kelompok wisata tersebut memiliki dua jenis pelaku wisata yang dikenal sebagai wisatawan domestik dan wisatawan asing atau yang dikenal sebagai wisatawan mancanegara. Pariwisata merupakan industri yang erat kaitannya dengan industri lainnya. Dengan kata lain,

pariwisata adalah kumpulan fenomena dan interaksi yang dihasilkan dari interaksi dengan wisatawan, pengusaha, politisi, hak pariwisata dan masyarakat [13]. Pariwisata merupakan kontributor penting bagi pertumbuhan ekonomi di banyak negara maju dan berkembang. Pariwisata berkontribusi terhadap pertumbuhan ekonomi melalui berbagai bidang seperti memperoleh devisa melalui pariwisata, menarik investasi internasional, meningkatkan penerimaan pajak, dan menciptakan lapangan pekerjaan [2].

### **2.3 Prediksi Atau Peramalan**

Peramalan atau prediksi adalah kegiatan memprediksi apa yang akan terjadi di masa yang akan datang [14]. Sedangkan Metode peramalan adalah teknik yang memperhatikan data dan informasi masa lalu kemudian memperkirakan atau memprediksi nilai masa depan secara matematis ataupun statistik [15].

Dalam merancang suatu metode peramalan, ada tiga langkah yang dilakukan, secara ringkas tahapan-tahapan ini adalah sebagai berikut : [16]:

1. Analisis pada data atau informasi masa lalu atau sebelumnya. Tahap ini diperlukan untuk mengetahui sampel dari data yang dipakai.
2. Metode yang cocok untuk prediksi yang bisa dipakai. Ada metode lain untuk dipilih, tergantung pada kebutuhan Anda. Pilihan metode prediksi dapat menentukan hasil ramalan. Hasil prediksi diuji dengan perhitungan nilai error atau error yang terendah. Hal ini membuktikan, tidak ada metode prediksi yang universal.
3. Melakuakn penyesuaian dari data masa lau dengan menggunakan metode yang cocok.

Suatu prediksi bisa memiliki sifat kualitatif (bukan numerik) ataupun bisa bersifat kuantitatif (numerik). Prediksi dengan sifat yang berbentuk bukan numerik lebih sulit untuk dilakukan prediksi karena variabelnya bersifat relatif [11]. Di sisi lain, dalam peramalan kuantitatif, itu tergantung pada bagaimana peramalan itu digunakan. Metode yang berbeda menyebabkan hasil prediksi yang berbeda. [15].

## 2.4 Regresi Linear Sederhana

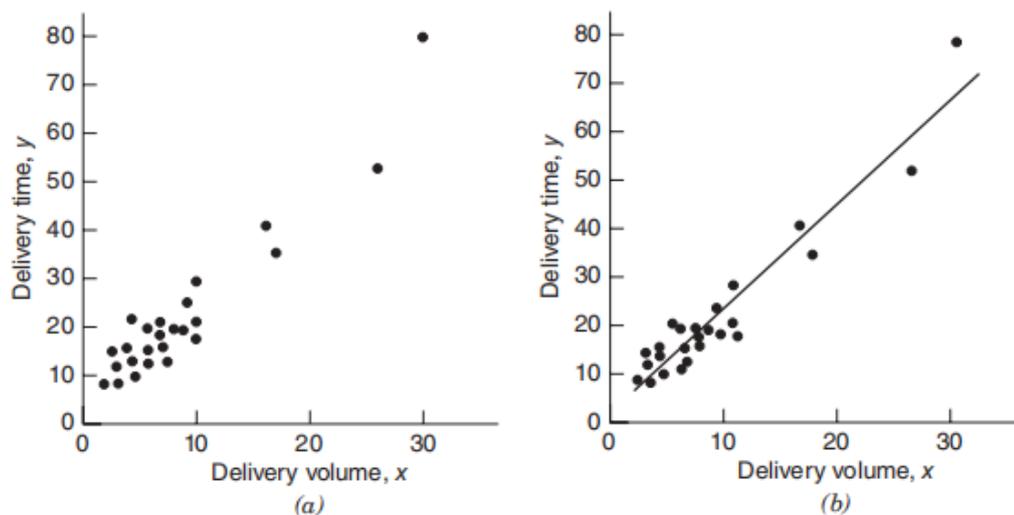
Analisis regresi adalah metode statistik untuk mempelajari dan memodelkan hubungan antar variabel. Ada banyak pengaplikasian metode regresi yang terjadi di hampir setiap bidang, termasuk teknik, fisika dan kimia, ekonomi, manajemen, kehidupan dan biologi, serta ilmu sosial. Faktanya, analisis regresi mungkin merupakan metode statistik yang paling banyak digunakan [20].

Metode Regresi Linear merupakan sebuah perhitungan statistika yang diperlukan untuk mengetahui sejauh mana korelasi dari satu atau banyak variabel terhadap satu variabel [12]. Bila hanya ada satu variabel bebas disebut regresi linier sederhana, dan bila ada banyak variabel bebas disebut regresi linier berganda. Regresi linier sederhana menggunakan metode statistik yang menguji sejauh mana korelasi antara variabel penyebab dan variabel akibat. Faktor penyebab biasanya disebut sebagai X atau prediktor dan variabel akibat biasanya juga disebut sebagai Y atau respon. [6].

Model persamaan untuk regresi linear adalah sebagai berikut [7]:

$$Y = a + bX \quad (2.1)$$

Y adalah variabel dependen yang diharapkan, X adalah variabel independen, dan a adalah titik potong. Misalnya, nilai Y ketika X = 0, b adalah perubahan rata-rata dan kemiringan Y ketika unit X berubah.



Gambar 2.1 Garis regresi dalam persebaran data [20].

Pada suatu metode regresi dapat kita temukan koefisien. Koefisien-koefisien ini merupakan nilai duga parameter pada contoh regresi untuk kondisi yang sebenarnya (true condition), sama halnya menggunakan statistik mean (rata-rata) pada konsep statistika dasar. Namun, koefisien model regresi adalah rata-rata dari nilai yang ditampilkan untuk variabel Y ketika menentukan nilai X.

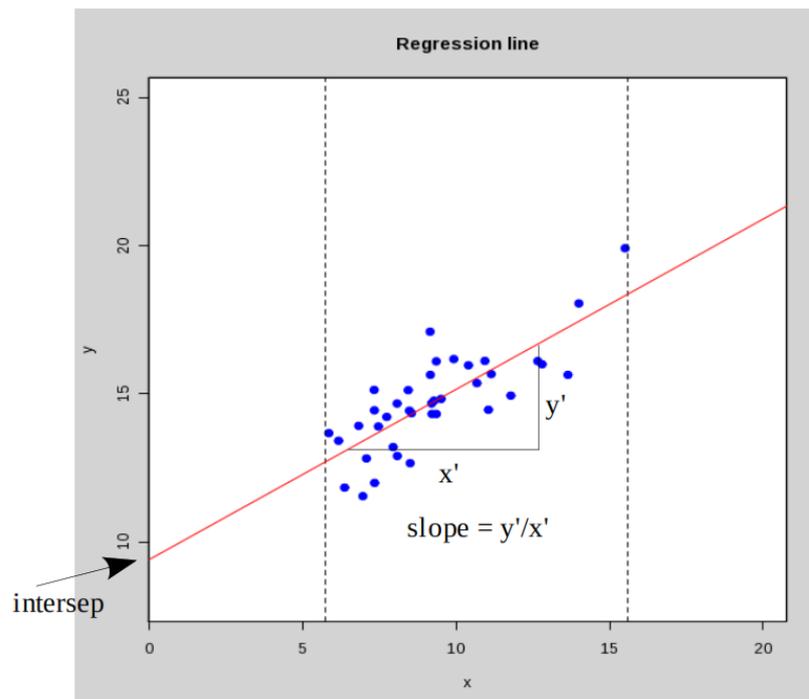
Nilai-nilai a dan b dapat dihitung menggunakan rumus dibawah ini [7]:

$$a = \frac{(\sum y)(\sum x^2) - (\sum x)(\sum xy)}{n(\sum x^2) - (\sum x)^2} \quad (2.2)$$

$$b = \frac{n(\sum xy) - (\sum x)(\sum y)}{n(\sum x^2) - (\sum x)^2} \quad (2.3)$$

Nilai a adalah slope, b adalah intercep dan n adalah banyaknya data yang digunakan dalam perhitungan.

Berikut adalah contoh garis regresi di dalam sebuah grafik [21]:



Gambar 2.2 Intersept dan slope pada grafik regresi.

Pada Gambar 2.2 dari gambar diatas sumbu X berada pada kisaran angka 5 hingga angka 15. Hal ini menyatakan bahwa kita akan melakukan prediksi nilai Y untuk nilai X yang berada dalam rentang tersebut.

## 2.5 Pengujian Akurasi

Keakuratan prediksi ditentukan oleh besarnya penyimpangan atau kesalahan yang terjadi antara data yang diprediksi dengan data yang sebenarnya atau nyata. Kesalahan dalam mempersiapkan prediksi muncul tidak hanya dari kesalahan atau error saja, tetapi juga karena model prediksi tidak dapat mengidentifikasi faktor-faktor lain dalam kumpulan data yang mempengaruhi tingkat bias dari prediksi. Besarnya kesalahan atau defleksi tersebut dapat disebabkan oleh outliers yang besar dengan tidak adanya metode prediksi yang dapat membuat prediksi akurat, atau metode prediksi yang digunakan mungkin tidak dapat memprediksi tren, musiman, atau komponen melingkar yang mungkin secara akurat dalam seri data.

Di antara berbagai cara untuk menghitung besarnya kesalahan tersebut salah satu diantaranya adalah *Mean Absolute Percentage Error* (MAPE).

### 2.5.1 MAPE (*Mean Absolute Percentage Error*)

*Mean Absolute Percent Error* (MAPE) adalah rata-rata perbedaan mutlak yang ada antara nilai yang diprediksi dan nilai awal yang kemudian dinyatakan sebagai persentase dari nilai sebenarnya. Mean Absolute Percentage (MAPE) dapat digunakan untuk mengevaluasi hasil ramalan untuk menentukan seberapa akurat dan tercapainya angka ramalan. Nilai *Mean Absolute Percentage Error* (MAPE) dapat dihitung dengan menggunakan persamaan berikut [22]:

$$MAPE = \frac{\sum \frac{|y - y'|}{y} \cdot 100\%}{n}$$