

BAB II

LANDASAN TEORI

2.1 Tinjauan Pustaka

Pada penelitian terdahulu yang dilakukan oleh Darwis Kurniawan Sumba dengan judul “Sistem Cerdas Prediksi Hasil Tanaman Jagung Di Indonesia Menggunakan *Support Vector Regressi*” pada penelitian ini menggunakan metode *Support Vector Regressi*. Pada penelitian ini dapat diketahui bahwa peneliti dapat mengetahui cara merekayasa aplikasi *data mining* menggunakan algoritma *Support Vector Regression* untuk memprediksi hasil tanaman jagung di Indonesia. Dan juga peneliti dapat mengetahui penerapan algoritma *Support Vector Regression* terhadap memprediksi hasil tanaman jagung di Indonesia dengan mendapatkan hasil akurasi sebesar 87,377% (Sumba, 2018).

Penelitian selanjutnya dilakukan oleh Dewi Puspa Lamondjong dan Mardi Hardjianto, dengan judul “*Data Mining* untuk Memprediksi Jumlah Penjualan Hasil Pertanian Menggunakan *Algoritma Forecasting* (Studi Kasus : Dinas Pertanian Kabupaten Banggai)”, dalam penelitian ini metode yang digunakan adalah metode *Algoritma Forecasting*. Pada penelitian ini dapat diketahui bahwa penulis mencoba untuk memprediksi jumlah penjualan hasil tanaman pangan setiap kali panen dengan menggunakan *data mining* dengan metode *algoritma regressi linear* dan hasil dari prediksi *data mining* dengan menggunakan confusion matrix Akurasinya 96 % , Presisi 90%, Recall 100% (Lamondjong & Hardjianto, 2021).

Penelitian selanjutnya dilakukan oleh Sri Maya Sari Sianturi dan Nelly Astuti Hasibuan, dengan judul “Analisa Data Pertanian Tanaman Pangan Untuk Memprediksi Hasil Panen Dengan Data *Mining Algoritma C4.5* (Studi Kasus : Dinas Tanaman Pangan Dan Holtikultura Provinsi Sumut), dalam penelitian ini menggunakan metode *algoritma C4.5*. pada penelitian ini dapat disimpulkan bahwa dengan *algoritma C4.5* dan pengujian dengan Tanagra menghasilkan pola kombinasi *itemsets* dan *rules* sehingga ilmu pengetahuan dan informasi penting dari data prediksi dapat mempengaruhi di Dinas Tanaman Pangan dan Holtikultura Provinsi Sumut. Proses untuk menghasilkan kombinasi item menggunakan *algoritma C4.5* dilakukan dengan menetapkan nilai *Entropy* dan nilai *Gain*. Untuk pembentukan pohon keputusan. Pengujian yang telah dilakukan dapat disimpulkan bahwa data *mining classification* dengan menggunakan metode pohon keputusan *algoritma C4.5* untuk membentuk pohon keputusan dalam memprediksi hasil panen padi dapat diterapkan menggunakan atung dan mendapatkan hasil yang maksimal (Sianturi, 2019).

Penelitian selanjutnya dilakukan oleh Heru Wahyu Herwanto, Triyanna Widiyaningtyas, dan Poppy Indriana, dengan judul “Penerapan *Algoritma Linear Regression* untuk Prediksi Hasil Panen Tanaman Padi” penelitian ini menggunakan metode *Algoritma Linear Regression*. *Algoritma Linear Regression* dapat diterapkan untuk prediksi hasil panen tanaman padi dengan beberapa variabel yang mempengaruhi, yaitu Luas Lahan, Jumlah Bibit, Pupuk Urea dan Pupuk NPK Phonska. Dari jumlah data sebanyak 300 *instance* dihasilkan tingkat kecocokan model *Multiple Linear Regression* sebesar 94,51%. Artinya sebanyak

94,51% variasi nilai hasil panen bergantung pada variabel bebas yang diukur, yaitu Luas Lahan , Varietas Bibit, Jumlah Bibit, Pupuk Area, dan Pupuk NPK Phonska. Hasil nilai rata-rata akurasi RMSE adalah sebesar 0,432. Hal tersebut menunjukkan bahwa variasi nilai yang dihasilkan oleh suatu model perkiraan yang dihasilkan mendekati akurat (Herwanto *et al.*, 2019).

Penelitian selanjutnya dilakukan oleh Mohammad Nur Fawaiq, Ahmad Jazuli, dan Muhammad Malik Hakim, dengan judul “Prediksi Hasil Pertanian Padi Di Kabupaten Kudus Dengan Metode *Brown’s Double Exponential Smoothing*”. Berdasarkan hasil penelitian yang telah dilakukan, dapat disimpulkan bahwa, *Brown’s Double Exponential Smoothing* bisa diterapkan untuk memprediksi produksi padi di Kabupaten Kudus. Hasil prediksi pada 9 kecamatan di Kabupaten Kudus memperoleh MAPE (*mean absolute percentage error*) sebesar 8.64%, berdasarkan nilai tersebut maka prediksi tergolong dalam kategori tingkat kesalahan rendah karena presentase kesalahan kurang dari 10%. Hasil prediksi padi dapat digunakan sebagai bahan perencanaan yang lebih efektif dan efisien di masa yang akan datang dibandingkan tidak adanya alat prediksi (Fawaiq *et al.*, 2019).

Berikut merupakan tabel perbandingan dari penelitian terdahulu :

Tabel 2.1 Perbandingan Penelitian Terdahulu

No	Nama (Tahun)	Judul	Metode	Hasil
1.	Darwis Kurniawan Sumba (2018)	Sistem Cerdas Prediksi Hasil Tanaman Jagung Di	<i>Support</i> <i>Vector</i> <i>Regressi</i>	Pada penelitian ini dapat diketahui bahwa Peneliti dapat mengetahui cara

		Indonesia Menggunakan <i>Support Vector Regressi</i>		merekayasa aplikasi <i>Data Mining</i> menggunakan <i>Algoritma Support Vector Regression</i> untuk memprediksi Hasil Tanaman Jagung di Indonesia. Dan juga peneliti dapat mengetahui penerapan <i>Algoritma Support Vector Regression</i> terhadap memprediksi Hasil Tanaman Jagung di Indonesia dengan mendapatkan hasil akurasi sebesar 87,377%.
2.	Dewi Puspa Lamondjong dan Mardi Hardjianto (2021)	<i>Data Mining</i> untuk Memprediksi Jumlah Penjualan Hasil Pertanian Menggunakan <i>Algoritma Forecasting</i> (Studi Kasus : Dinas	<i>Algoritma Forecasting</i>	Pada penelitian ini dapat diketahui bahwa penulis mencoba untuk memprediksi jumlah penjualan hasil tanaman pangan setiap kali panen dengan menggunakan data mining dengan metode <i>algoritma regrassi linear</i> dan hasil dari prediksi data

		Pertanian Kabupaten Banggai)		mining dengan menggunakan confusion matrix Akurasinya 96 % , Presisi 90%, Recall 100%.
3.	Sri Maya Sari Sianturi dan Nelly Astuti Hasibuan (2019)	Analisa Data Pertanian Tanaman Pangan Untuk Memprediksi Hasil Panen Dengan Data Mining <i>Algoritma C.45</i> (Studi Kasus : Dinas Tanaman Pangan Dan Holtikutura Provinsi Sumut)	<i>Algoritma C.45</i>	Pada penelitian ini dapat disimpulkan bahwa dengan <i>algoritma C4.5</i> dan pengujian dengan Tanagra menghasilkan pola kombinasi <i>itemsets</i> dan <i>rules</i> sehingga ilmu pengetahuan dan informasi penting dari data prediksi dapat mempengaruhi di Dinas Tanaman Pangan dan Holtikutura Provinsi Sumut. Proses untuk menghasilkan kombinasi item menggunakan <i>algoritma C4.5</i> dilakukan dengan menetapkan nilai <i>Entropy</i> dan nilai <i>Gain</i> .

4.	Heru Wahyu Herwanto, Triyanna Widiyaningtyas, Poppy Indriana (2019)	Penerapan <i>Algoritma Linear Regression</i> untuk Prediksi Hasil Panen Tanaman Padi	<i>Algoritma Linear Regression</i>	Untuk memprediksi hasil tanaman padi dengan beberapa variabel yang mempengaruhi yaitu luas lahan, jumlah bibit, dan pupuk.
5.	Mohammad Nur Fawaiq, Ahmad Jazuli, dan Muhammad Malik Hakim (2019)	Prediksi Hasil Pertanian Padi Di Kabupaten Kudus Dengan Metode <i>Brown's Double Exponential Smoothing</i>	<i>Brown's Double Exponential Smoothing</i>	Hasil prediksi pada 9 kecamatan di Kabupaten Kudus memperoleh MAPE (<i>mean absolute percentage error</i>) sebesar 8.64%, berdasarkan nilai tersebut maka prediksi tergolong dalam kategori tingkat kesalahan rendah karena presentase kesalahan kurang dari 10%. Hasil prediksi padi dapat digunakan sebagai bahan perencanaan yang lebih efektif dan efisien di masa yang akan datang dibandingkan tidak adanya alat prediksi.

Merujuk pada penelitian yang dilakukan oleh Heru Wahyu Herwanto, Triyanna Widiyaningtyas, Poppy Indriana (2019) “Penerapan *Algoritma Linear Regression* untuk Prediksi Hasil Panen Tanaman Padi” maka pada penelitian ini akan dilakukan implementasi metode *Multiple Linear Regression* pada *orange data mining* untuk melihat keakuratan prediksi hasil panen pertanian berikutnya di Kabupaten Malaka.

2.2 Landasan Teori

1. Tanaman Pangan

Tanaman pangan adalah segala jenis tanaman yang dapat menghasilkan karbohidrat, lemak dan protein, oleh karena itu tanaman pangan menjadi sumber utama makanan pokok sebagian besar penduduk Indonesia. Beberapa jenis tanaman pangan yang diusahakan adalah padi, jagung dan kacang hijau (Minarni *et al.*, 2017).

2. Prediksi

Prediksi adalah suatu proses memperkirakan secara sistematis tentang sesuatu yang paling mungkin terjadi di masa depan berdasarkan informasi masa lalu dan sekarang yang dimiliki, agar kesalahannya (selisih antara sesuatu yang terjadi dengan hasil perkiraan) dapat diperkecil. Prediksi tidak harus memberikan jawaban secara pasti kejadian yang akan terjadi, melainkan berusaha untuk mencari jawaban sedekat mungkin yang akan terjadi (Triyanto *et al.*, 2019).

3. *Data Mining*

Data Mining adalah proses yang menggunakan teknik statistik, matematika, kecerdasan buatan, dan pembelajaran mesin untuk mengekstraksi dan mengidentifikasi informasi yang berguna dan pengetahuan terkait dari database besar. Istilah *data mining* merupakan inti dari suatu disiplin ilmu yang tujuan utamanya mencari, menggali, atau menambang pengetahuan dari data atau informasi yang kita miliki (Sumba, 2018).

Data Mining adalah proses untuk mencari nilai tambah berupa informasi yang selama ini tidak diketahui secara manual dari suatu basis data. Informasi yang penting ini diperoleh dengan cara mengekstraksi dan mengenali pola yang penting atau berkesan dari data yang terdapat dalam bisnis data. *Data mining* itu juga dapat diartikan sebagai proses penambangan data dalam jumlah data yang besar dan menghasilkan sebuah *output* (keluaran) berupa pengetahuan (Kandi, 2020).

4. Regresi Linear

Regresi merupakan teknik membangun model yang digunakan untuk prediksi nilai dari data masukan yang diberikan. Regresi adalah ukuran statistik yang digunakan untuk menentukan kekuatan hubungan antara variabel dependen (tak bebas) dengan variabel independen (bebas). Regresi linear adalah teknik analisis data yang memprediksi nilai data yang tidak diketahui dengan menggunakan nilai data lain yang terkait dan diketahui.

Metode utama untuk melakukan prediksi yakni membangun model regresi dengan mencari hubungan antara satu atau lebih variabel independen atau predictor (X) dengan variabel dependen atau respons (Y). *Linear regression* memodelkan hubungan antara variabel skalar dan satu atau lebih variabel penjelas.

Metode *linear regression* tersusun atas dasar pola hubungan data yang relevan di masa lalu. Secara umum, *algoritma linear regression* dibagi menjadi dua jenis, yaitu *simple linear regression* dan *multiple linear regression*. *Simple linear regression* merupakan hubungan antara satu variabel dependen dengan satu variabel independen, sedangkan *multiple linear regression* merupakan hubungan antara satu variabel dependen dengan dua atau lebih variabel independen.

Dalam penelitian ini, digunakan *multiple linear regression* karena terdapat lebih dari satu buah variabel independen. Data penelitian terdiri atas satu variabel dependen (Y) dan empat variabel independen (X). Variabel dependen tersebut adalah Hasil Panen (Herwanto et al., 2019).

Perhitungan yang digunakan untuk *multiple linear regression* dinyatakan dalam

$$Y = a + a_1X_1 + b_2X_2 + \dots + b_nX_n \quad (1)$$

Dengan:

Y = Variabel dependen atau variabel tak bebas (nilai yang diprediksi)

X_1, X_2, \dots, X_n = Variabel independen atau variabel bebas

a = Konstanta

$b_1, b_2, \dots, b_n =$ Koefisien regresi

Dengan mengacu pada (1) untuk menghitung hasil panen tanaman pangan dengan *algoritma linear regression*, digunakan

$$\gamma = a + b_1X_1 + b_2X_2 \quad (2)$$

Dengan:

Y = Hasil panen

a = Konstanta

b_1 dan b_2 = Koefisien variabel bebas

X_1 dan X_2 = Luas tanam dan luas panen

5. *Orange Data Mining Tools*

Data mining tools yang digunakan untuk membantu dalam penelitian ini adalah *orange*. *Orange* merupakan sebuah *tools open source* untuk pengolahan *data mining*. *Orange* memiliki *widget* yang berfungsi sebagai unit komputasi untuk membaca, memproses, melakukan visualisasi, melakukan analisis, mengeksplorasi data, dan lain-lain. *Widget* disusun sedemikian rupa sehingga membentuk *workflow* (alur kerja) dan berkomunikasi satu sama lain di lingkungan *Orange*. *Widget* data memungkinkan *Orange* untuk manipulasi data text atau gambar (Hartono *et al.*, 2020).