

Prediksi Hasil Panen Pertanian Salak di Daerah Tapanuli Selatan Menggunakan Algoritma SVM (Support Vector Machine)

Rakhmat Kurniawan, Aidil Halim, Henni Melisa*

¹Fakultas Sains dan Teknologi, Program Studi Ilmu Komputer, Universitas Islam Negeri Sumatera Utara, Medan, Indonesia

Email: ¹rakhmat.kr@uinsu.ac.id, ²aidilhalimlubis@uinsu.ac.id, ³hennimelisahutabarat@gmail.com

Email Penulis Korespondensi: hennimelisahutabarat@gmail.com

Abstrak—Tapanuli Selatan merupakan salah satu daerah yang dikenal dengan pertanian salak yang luas. Pertanian salak yang terdapat di daerah Tapanuli Selatan ini merupakan salah satu sumber perekonomian dari masyarakat yang berada di daerah Tapanuli Selatan. Dinas Pertanian Daerah Tapanuli Selatan adalah sebuah instansi yang bergerak di berbagai bidang pertanian, dalam kegiatan tahunan melakukan pencatatan hasil produksi pertanian salak. Hasil produksi pertanian salak yang di dapatkan sering mengalami perubahan, maka diperlukan sebuah sistem untuk melakukan prediksi, tujuannya yaitu untuk mengetahui hasil panen pertanian salak. Dengan penerapan algoritma SVM (Support Vector Machine) telah berhasil dilakukan dengan jumlah data sebanyak 28 data hasil panen pertanian salak di daerah Tapanuli Selatan pada tahun 2019-2020. Tingkat error yang di dapatkan dari RMSE (Root Mean Square Error) pada pembuatan model prediksi yaitu sebesar 1,49 sedangkan pada tingkat akurasinya di dapatkan sebesar 44%.

Kata Kunci: Tapanuli Selatan; Pertanian Salak; Prediksi; Support Vector Machine; Akurasi

Abstract—South Tapanuli is an area known for its extensive salak farming. Salak farming in the South Tapanuli area is one of the economic sources of the people in the South Tapanuli area. The South Tapanuli Regional Agriculture Office is an agency engaged in various fields of agriculture, in its annual activity it records the results of salak agricultural production. Salak agricultural production results that are obtained often experience changes, so we need a system to make predictions, the goal is to find out the yield of salak farming. With the application of the SVM (Support Vector Machine) algorithm, it has been successfully carried out with a total of 28 data on salak agricultural yields in the South Tapanuli Region year 2019 - 2020. The error rate obtained from the RMSE (Root Mean Square Error) in making the prediction model is 1,49 while the accuracy level is 44%.

Keywords: South Tapanuli; Salak Agriculture; Prediction; Support Vector Machine; Accuracy

1. PENDAHULUAN

Sains dan teknologi adalah sebagai alat bagi manusia untuk mengurangi peran dan beban manusia dalam melakukan sesuatu. Dengan hadirnya sains dan teknologi manusia sebenarnya dapat mengurangi keterlibatannya secara menyeluruh dalam setiap aktivitas kehidupan. Sains juga dapat mensejahterakan manusia apabila digunakan sesuai aturan yang digariskan oleh agama. Demikian juga penggunaan teknologi yang hampir mendominasi seluruh kehidupan manusia tanpa mengenal batas, tempat dan waktu [1].

Tapanuli Selatan merupakan salah satu Kabupaten yang ada di Provinsi Sumatera Utara. Kabupaten Tapanuli Selatan merupakan salah satu Kabupaten yang dikenal dengan pertanian salak yang luas. Pertanian salak yang terdapat di Kabupaten Tapanuli Selatan merupakan salah satu sumber perekonomian dari Masyarakat yang berada di daerah Kabupaten Tapanuli Selatan. Kabupaten Tapanuli Selatan ini berada di daerah pegunungan memiliki banyak potensi yaitu dalam bidang pertanian, peternakan, objek wisata dan juga dalam bidang perdagangan.

Tanaman salak merupakan tanaman buah asli dari Indonesia yang dapat dijumpai hampir di semua provinsi wilayah Nusantara. Nama ilmiah tanaman salak adalah *salacca edulis reinw*. Dalam satu tahun tanaman salak dapat dipanen tiga kali, jadi ada tiga musim panen dalam satu tahun yaitu panen besar pada bulan November -Februari, panen sedang pada bulan Mei-Agustus, dan panen kecil pada bulan Maret-Oktober. Masaknya buah salak dalam satu tandan tidak seragam, yang masak terlebih dahulu adalah buah dibagian ujung. Karena itu pemetikan buah salak dilakukan secara berkala, dipilih dahulu buah yang sudah masak. Sedangkan sisa buah dalam tandan akan dipanen pada hari berikutnya setelah buah masak [2].

Hasil produksi pertanian salak di daerah ini setiap tahunnya naik turun yang tidak stabil. Kurangnya saran atau belum ada yang memberikan saran bahwa prediksi ini perlu dilakukan untuk meningkatkan produksi pertanian salak. Prediksi sangat dibutuhkan mengetahui gambaran hasil panen kedepannya meningkat atau menurun. Dalam memprediksi hasil panen pertanian salak, diperlukan keahlian khusus untuk menganalisa hasil prediksi pertanian salak menjelang masa panen dengan mempertimbangkan banyak faktor seperti: luas lahan (luas lahan yang digunakan pada pertanian salak sangat berpengaruh terhadap hasil panen yang akan diperoleh, jika lahan yang digunakan luas maka potensi hasil panen juga akan semakin besar), jumlah pokok (jumlah pokok yang ditanam pada pertanian salak tergantung luas lahannya, apabila luas lahan semakin besar, maka jumlah pokok yang akan digunakan juga banyak) dan lain sebagainya.

Prediksi adalah suatu proses memperkirakan secara sistematis tentang sesuatu yang mungkin terjadi di 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 [3].

Dalam penelitian ini akan dilakukan prediksi hasil panen pertanian salak yang diperoleh dari data yang di dapat dari Dinas Pertanian Kabupaten Tapanuli Selatan. Data yang digunakan yakni luas lahan pertanian salak, jumlah tanaman salak, hasil panen/setiap tahunnya. Hasil yang di dapatkan dari pembuatan sistem ini adalah model prediksi yang dapat memprediksi hasil panen pertanian salak di daerah Tapanuli Selatan.

SVM (*Support Vector Machine*) telah berhasil diterapkan pada banyak kasus dengan akurasi tingkat tinggi. Keberhasilan tersebut dapat dilihat pada penelitian Suhardjono yang menghasilkan akurasi 85,81% pada penelitian prediksi waktu kelulusan mahasiswa menggunakan SVM berbasis Pso[4]. Selanjutnya dalam penelitian Pratama[5] dari penelitian tersebut di dapatkan rata-rata akurasi terbaik sebesar 80,55%. Dengan demikian algoritma dinyatakan berhasil jika menghasilkan nilai keakuratan kisaran 70% - 100%.

Pada penelitian ini digunakan algoritma SVM (*Support Vector Machine*) untuk memprediksi hasil panen pertanian salak. Algoritma SVM (*Support Vector Machine*) dapat diartikan sebagai upaya untuk menemukan *hyperlane* terbaik, yang bertindak sebagai pemisah antara dua kategori dalam ruang input [6].

Tujuan penelitian ini adalah untuk mengetahui bagaimana menerapkan algoritma SVM (*Support Vector Machine*) dalam memprediksi hasil panen pertanian salak di Daerah Tapanuli Selatan dan untuk mengetahui bagaimana prediksi yang dihasilkan dari penerapan algoritma SVM (*Support Vector Machine*).

2. METODOLOGI PENELITIAN

2.1 Landasan Teori

2.1.1 Prediksi

Prediksi merupakan salah satu cara atau proses untuk memprediksi atau memperkirakan secara urut dan sistematis mengenai sesuatu yang mungkin dapat terjadi pada masa depan berdasarkan tentang informasi pada masa lalu dan sekarang yang dimiliki, agar tingkat error dan kesalahannya dapat diperkecil. Dalam prediksi tidak harus memberikan suatu jawaban secara pasti tentang kejadian yang nanti akan terjadi pada masa yang akan datang, melainkan berusaha untuk mencari jawaban yang akurat mungkin nanti akan terjadi [7].

2.1.2 Algoritma *Machine Learning*

Kata algoritma atau algoritme muncul dari kata “algoritmi”, yaitu bentuk Latin dari al-Khwarizmi (Muhammad ibn Musa al-Khawarizmi), seorang matematikawan, ahli astronomi, dan ahli geografi dari Persia yang hidup pada tahun 780-850. Dalam matematika dan ilmu komputer, algoritma adalah prosedur langkah demi langkah untuk perhitungan. Algoritma digunakan untuk penghitungan, pemrosesan data, dan penalaran otomatis. Menurut Harold Stuart Stone, algoritma adalah “*Precise instructions (In language understood by the computer) for a fast, efficient, good process that specifies the movies of the computer*” [8].

Saat ini sudah tersedia ratusan bahkan ribuan algoritma *machine learning*. Secara garis besar, algoritma *machine learning* dapat dikelompokkan menjadi tiga kategori, yaitu :

- a. *Supervised learning*
- b. *Unsupervised learning*
- c. *Reinforcement learning*

2.1.3 Algoritma SVM (*Support Vector Machine*)

SVM (*Support Vector Machine*) adalah metode *machine learning* yang bekerja atas prinsip *Structural Risk Minimization* (SRM) dengan tujuan menemukan *hyperlane* terbaik memisahkan dua buah kelas pada *input space* [9]. Secara sederhana SVM (*Support Vector Machine*) memiliki konsep mencari *hyperlane* terbaik, yang berfungsi sebagai batas dari dua buah *class* [10]. Dalam algoritma SVM ada *trik kernel* dimana ada SVM *linear* dan SVM *nonlinear*. Dimana SVM adalah *hyperlane linear* yang bekerja hanya pada data yang hanya dapat dipisahkan dengan cara *linear*. SVM *nonlinear* yaitu data yang berdistribusi pada kelas yang tidak *linear* sering digunakan pendekatan kernel pada fitur awal data set. Dimana kernel dapat diartikan sebagai suatu fungsi yang memetakan fitur data yang memiliki dimensi awal rendah *linear* [11].

Persamaan SVM:

$$f(x) = w^t \phi(x) + b \quad (1)$$

Dimana:

b	= bias
$x = (x_1, x_2, \dots, x_D)^T$	= Variabel Input
$w = (w_0, w_1, \dots, w_D)^T$	= Parameter Bobot
$\phi(x)$	= Fungsi Transformasi Fitur

2.1.4 *Preprocessing Data*

Preprocessing merupakan langkah awal yang dilakukan dalam pengolahan data untuk membantu metode yang digunakan agar dapat berjalan dengan baik dan nantinya menghasilkan faktor kesalahan *Root Mean Square* (RMSE) yang rendah. Pada tahap *preprocessing*, dilakukan proses *cleaning* data, transformasi data dan normalisasi data [12]. Normalisasi data set dilakukan agar data set yang digunakan memiliki nilai yang konsisten dan rentang yang tidak terlalu jauh. Berikut merupakan persamaan atau rumus dari normalisasi data. Pada penelitian ini normalisasi data set menggunakan metode *simple feature scalling* [13].

2.1.5 Root Mean Square Error (RMSE)

Root mean square error (RMSE) digunakan dalam analisis perbandingan data lapangan dengan data hasil proses yang diolah untuk mengetahui tingkat kesalahan yang terjadi [11]. *Root mean square error* (RMSE) merupakan besarnya tingkat kesalahan hasil prediksi, dimana semakin kecil (mendekati 0) nilai RMSE maka hasil prediksi akan semakin akurat [14]. Berikut persamaan RMSE :

$$RMSE = \sqrt{\frac{\sum y_t - y_t^2}{n}} \quad (2)$$

Dimana :

RMSE = *Root Mean Square Error*
n = Jumlah Sampel
y_t = Nilai Aktual Indeks
y_t² = Nilai Prediksi Indeks

2.1.6 Flowchart

Flowchart merupakan diagram yang menyatakan aliran proses dengan menggunakan anotasi bidang-bidang geometri, seperti lingkaran, persegi empat, wajik, oval, dan sebagainya untuk mempresentasikan langkah-langkah kegiatan beserta urutannya dengan menghubungkan masing-masing langkah tersebut menggunakan tanda panah [15].

2.1.7 Python

Python adalah bahasa pemrograman interpretative yang di anggap mudah dipelajari serta berfokus pada keterbacaan kode. Dengan kata lain, *python* diklaim sebagai bahasa pemrograman yang memiliki kode-kode pemrograman yang sangat jelas, lengkap dan mudah untuk dipahami [16]. Terdapat beberapa *library* yang digunakan dalam bahasa pemrograman *python* ini antara lain, *library pandas*, *library numpy* dan *library sklearn*.

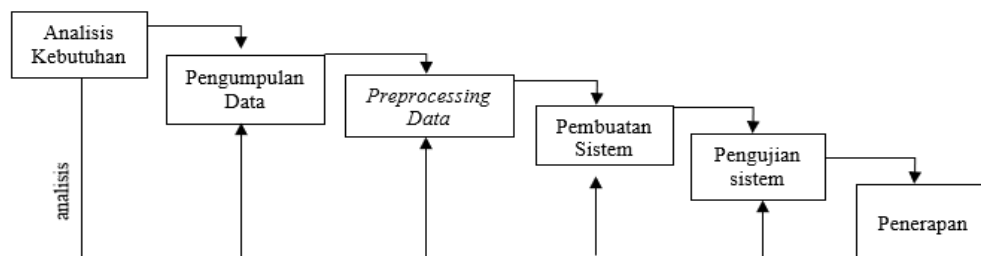
Dimana *library pandas* ini bersifat *open source* serta menyediakan berbagai *tools* berfungsi untuk melakukan analisis data yang cenderung mudah diaplikasikan dan mudah digunakan. *Pandas* merupakan *library* yang populer digunakan dalam bidang data *science*. *Pandas* memiliki banyak fungsi, diantaranya dapat digunakan dalam proses pembuatan tabel, mengubah dimensi dari sebuah *dataframe*, proses pengecekan data, serta fungsi-fungsi lainnya. *Numpy* memiliki banyak sekali kelebihan, diantaranya adalah memudahkan pengguna untuk melakukan operasi komputasi pada data, dan cocok digunakan untuk melakukan akses pada data acak. *Scikit-learn* atau yang biasa disebut *sklearn* merupakan sebuah *library machine learning* yang menyediakan berbagai macam algoritma pembelajaran, baik *supervised* maupun *unsupervised learning* yang dapat langsung digunakan dan dimanfaatkan untuk berbagai keperluan. Penggunaan *sklearn* berfokus pada pemodelan data [17].

2.1.8 Jupyter Notebook

Jupyter notebook adalah aplikasi yang digunakan untuk merancang, mengembangkan, melaksanakan dan berkomunikasi dengan kode khusus [18]. *Anaconda* merupakan sebuah platform bahasa pemrograman *python* yang bersifat *open – source*. Bertujuan untuk menyederhanakan manajemen paket dan penyebaran [19].

2.2 Tahapan Penelitian

Adapun tahapan-tahapannya adalah sebagai berikut:



Gambar 1. Tahapan Penelitian

2.2.1 Analisis Kebutuhan

Perangkat keras yang digunakan pada pembuatan sistem ini adalah sebagai berikut:

- Processor : Intel(R) Core(TM) i3-6006U CPU @ 2.00GHz 1.99 GHz
- Memory : Random Access Memory (RAM) 4.00GB
- Penyimpanan : 64-bit Operating System

Perangkat lunak yang digunakan pada pembuatan sistem ini adalah sebagai berikut:

- Operating system windows 10 pro 64 bit*
- Python 3.7.0*

c. *Jupyter notebook*

2.2.2 Pengumpulan Data

- Penelitian kepustakaan merupakan penelitian yang dilakukan dengan mencari referensi dari buku perpustakaan dan juga jurnal dari internet sebagai sumber untuk landasan teori dengan tujuan untuk memahami dan mencari referensi yang berhubungan dengan judul yang diangkat pada tugas akhir ini.
- Pengumpulan data dari Dinas Pertanian Daerah Tapanuli Selatan, pada penelitian ini menggunakan dataset yang diperoleh dari Dinas Pertanian Daerah Tapanuli Selatan. Data yang digunakan yakni data tahun 2019-2020, data tersebut mempunyai beberapa atribut yang dapat di proses dengan algoritma *support vector machine* (SVM) antara lain, kecamatan, luas lahan, dan hasil panen.

Tabel 1. Atribut dan Skala Pengukuran Dataset

No.	Atribut	Skala Pengukuran
1.	Kecamatan	Batang Angkola Sayur Matinggi Tantom Angkola Angkola Muaratais Angkola Timur Angkola Selatan Angkola Barat Angkola Sangkununur Batang Toru Marancar Muara Batang Toru Sipirok Saipar Dolok Hole Aek Bilah
2.	Luas lahan	Hektar
3.	Hasil panen	Ton

Pada tabel 1 di atas atribut yang digunakan pada penelitian ini adalah kecamatan, luas lahan dan hasil panen. Kemudian untuk skala pengukuran pada atribut kecamatan merupakan nama-nama desa dari daerah Tapanuli Selatan tersebut, skala pengukuran atribut luas lahan merupakan hektar dan untuk skala pengukuran pada atribut hasil panen merupakan ton.

Adapun dataset yang di peroleh dari Dinas Pertanian Daerah Tapanuli Selatan dapat dilihat pada tabel di bawah ini:

Tabel 2. Dataset Dinas Pertanian Daerah Tapanuli Selatan

Tahun	Kecamatan	Luas Lahan (Ha)	Hasil Panen (Ton)
2019	Batang Angkola	140,45	1.404,05
	Sayurmatinggi	1,47	19,11
	Tantom Angkola	0,00	0,00
	Angkola Timur	0,50	8,00
	Angkola Selatan	45,35	453,05
	Angkola Barat	15.883,20	131.126,97
	Angkola Sangkununur	0,00	0,00
	Batang Toru	565,00	5.650,00
	Marancar	2.125,00	17.625,00
	Muara Batang Toru	0,45	4,50
	Sipirok	10,80	108,00
	Arse	1,00	15,00
	Saipar Dolok Hole	0,78	7,80
	Aek Bilah	0,00	0,00
2020	Batang Angkola	0,00	0,00
	Sayurmatinggi	0,00	0,00
	Tantom Angkola	0,00	0,00
	Angkola Muaratais	0,00	0,00
	Angkola Timur	0,50	35,80
	Angkola Selatan	6,00	60,00
	Angkola Barat	5.730,75	145.168,90
	Batang Sangkununur	0,18	0,80
Batang Toru	1.130,00	10.502,25	

Marancar	120,00	600,00
Muara Batang Toru	0,29	2,40
Sipirok	9,30	49,60
Arse	0,00	0,00
Saipar Dolok Hole	0,18	1,73
Aek Bilah	0,00	0,00

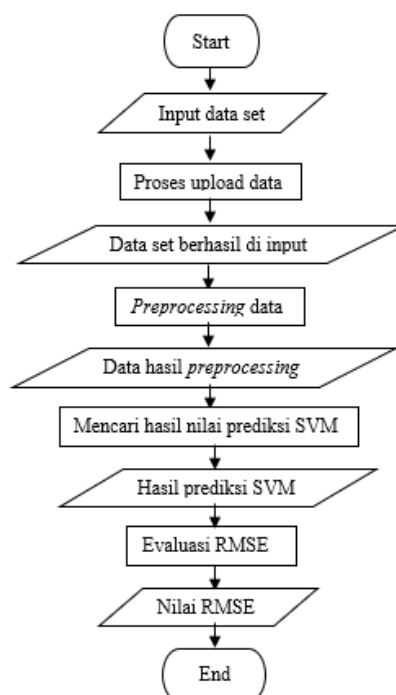
Pada tabel 2 di atas merupakan dataset yang digunakan pada penelitian ini, data tersebut diperoleh dari Dinas Pertanian Daerah Tapanuli Selatan.

2.2.3 Preprocessing Data

Preprocessing data digunakan untuk mendapatkan data yang baik dan siap untuk digunakan. Terdapat beberapa data yang memiliki *missing value*, *noisei*, dan tidak konsisten. Tahapan *preprocessing* data ini dipakai untuk masukan apakah sudah sesuai dengan apa yang dibutuhkan. Pada tahap *preprocessing* data ini dilakukan proses *cleaning*, dimana data yang *missing value* akan dihapus, sehingga data yang dimiliki memiliki atribut yang lengkap. Kemudian dilakukan transformasi data, pada proses ini data berupa huruf pada setiap atribut akan diubah menjadi angka, hal ini dilakukan untuk mempermudah dalam memproses datanya dan dilakukan proses normalisasi data menggunakan metode *simple feature scaling*.

2.2.4 Pembuatan Sistem

Berikut merupakan *flowchart* algoritma SVM yang akan digunakan pada prediksi hasil panen pertanian salak.



Gambar 2. Flowchart Algoritma SVM

Pada gambar 2 di atas merupakan *flowchart* dari algoritma SVM dalam melakukan prediksi hasil panen pertanian salak. Data input dilakukan *preprocessing* data terlebih dahulu, kemudian selanjutnya data dari hasil *preprocessing* data dilakukan perhitungan dengan menggunakan metode algoritma SVM. Dari hasil perhitungan dengan menggunakan algoritma SVM akan di evaluasi menggunakan RMSE (*Root Mean Square Erroe*). Nilai RMSE merupakan hasil akhir dari *flowchart* ini.

2.2.4 Pengujian Sistem

Proses pengujian sistem ini dilakukan dengan menggunakan data hasil panen pertanian salak pada tahun 2019-2020. Jumlah data terdiri dari 28 data dengan 3 variabel yaitu nama kecamatan, luas lahan, dan hasil panen/setiap tahunnya. Data tersebut akan dilakukan *preprocessing* dengan beberapa tahapan yaitu *cleaning* data, transformasi data dan normalisasi data. Kemudian dilakukan perhitungan dengan menggunakan algoritma SVM, kemudian lakukan evaluasi RMSE.

2.2.4 Penerapan

Penerapan sistem ini adalah untuk memprediksi hasil panen pertanian salak di Daerah Tapanuli Selatan, berdasarkan kecamatannya, luas lahan, dan hasil panen/setiap tahunnya.

3. HASIL DAN PEMBAHASAN

3.1 Pembahasan

3.1.1 Preprocessing Dataset

a. *Clenaing* Dataset

Cleaning dataset dilakukan untuk menghapus data yang terindikasi sebagai *missing value*, data yang tidak lengkap nilai fiturnya dan data yang duplikat.

Tabel 3. *Cleaning* Dataset

No.	Kecamatan	Luas Lahan (Ha)	Hasil Panen (Ton)
1.	Batang Angkola	140,45	1.404,05
2.	Sayurimatinggi	1,47	19,11
3.	Angkola Timur	0,50	8,00
4.	Angkola Selatan	45,35	453,05
5.	Angkola Barat	15.883,20	131.126,97
6.	Batang Toru	565,00	5.650,00
7.	Marancar	2.125,00	17.625,00
8.	Muara Batang Toru	0,45	4,50
9.	Sipirok	10,80	108,00
10.	Arse	1,00	15,00
11.	Saipar Dolok Hole	0,78	7,80
12.	Angkola Timur	0,50	35,80
13.	Angkola Selatan	6,00	60,00
14.	Angkola Barat	5.730,75	145.168,90
15.	Batang Sangkunur	0,18	0,80
16.	Batang Toru	1.130,00	10.502,25
17.	Marancar	120,00	600,00
18.	Muara Batang Toru	0,29	2,40
19.	Sipirok	9,30	49,60
20.	Saipar Dolok Hole	0,18	1,73

Pada tabel 3 di atas merupakan hasil *cleaning* dataset, yaitu penghapusan kecamatan yang tidak memiliki luas lahan dan hasil panen pertanian salak di daerah Tapanuli Selatan.

b. Transformasi Dataset

Proses transformasi dilakukan untuk merubah data dalam bentuk kata atau kalimat kedalam bentuk angka agar dapat di proses oleh model prediksi yang di buat. Proses transformasi dilakukan pada kolom kecamatan, proses transformasi yang dilakukan menggunakan metode *LabelEncoder*.

Tabel 4. Hasil Transformasi Data

No.	Kecamatan	Keterangan	Luas Lahan (Ha)	Hasil Panen (Ton)
1.	0	Angkola Barat	15.883,20	131.126,97
2.	0	Angkola Barat	5.730,75	145.168,90
3.	1	Angkola Sangkunur	0,18	0,80
4.	2	Angkola Selatan	45,35	453,05
5.	2	Angkola Selatan	6,00	60,00
6.	3	Angkola Timur	0,50	8
No.	Kecamatan	Keterangan	Luas Lahan (Ha)	Hasil Panen (Ton)
7.	3	Angkola Timur	0,50	35,80
8.	4	Arse	1,00	15
9.	5	Batang Angkola	140,45	1.404,05
10.	6	Batang Toru	565	5.650
11.	6	Batang Toru	1.130,00	10.502,25
12.	7	Marancar	2.125	17.625
13.	7	Marancar	120,00	600,00
14.	8	Muara Batang Toru	0,45	4,5
15.	8	Muara Batang Toru	0,29	2,40
16.	9	Saipar Dolok Hole	0,78	7,8
17.	9	Saipar Dolok Hole	0,18	1,73
18.	10	Sayur Matinggi	1,47	19,11
19.	11	Sipirok	10,80	108
20.	11	Sipirok	9,30	49,60

Pada tabel 4 di atas merupakan hasil dari transformasi data yang dilakukan pada kolom kecamatan, yakni dilakukan perubahan kata menjadi angka.

c. Normalisasi Dataset

Normalisasi dataset dilakukan agar dataset yang digunakan memiliki nilai yang konsisten dan rentang yang tidak terlalu jauh. Untuk proses normalisasi dilakukan dengan menggunakan metode *simple feature scaling*. Untuk proses normalisasi dataset menggunakan rumus $normalisasi = \frac{X}{x_{max}}$. Nilai X yang digunakan pada proses ini terdiri dari variabel luas lahan dan hasil panen, untuk nilai x_{max} diperoleh dari nilai tertinggi (maksimal) dari setiap variabel.

Tabel 5. Normalisasi Dataset

No	Kecamatan	Keterangan	Luas Lahan (Ha)	Hasil Panen (Ton)
1.	5	Batang Angkola	0,0088426	0,0096718
2.	10	Sayur Matinggi	0,0000925	0,0001316
3.	3	Angkola Timur	0,0000314	0,0000551
4.	2	Angkola Selatan	0,0028552	0,0031208
5.	0	Angkola Barat	1	0,9032717
6.	6	Batang Toru	0,0355721	0,0389201
7.	7	Marancar	0,1337891	0,1214103
8.	8	Muara Batang Toru	0,0000283	0,0000309
9.	11	Sipirok	0,0006799	0,0007439
10.	4	Arse	0,0000629	0,0001033
11.	9	Saipar Dolok Hole	0,0000491	0,0000537

Pada tabel 5 di atas merupakan hasil normalisasi dataset yang dilakukan pada kolom luas lahan dan hasil panen pertanian salak di daerah Tapanuli Selatan.

Untuk variabel hasil panen yang juga merupakan variabel y dalam penelitian ini akan dilakukan klasifikasi menjadi dua kelas. Dengan mencari nilai median dari hasil panen.

Tabel 6. Klasifikasi Dataset

No	Kecamatan	Keterangan	Luas Lahan (Ha)	Hasil Panen (Ton)
1.	5	Batang Angkola	0,0088426	1
2.	10	Sayur Matinggi	0,0000925	-1
3.	3	Angkola Timur	0,0000314	-1
4.	2	Angkola Selatan	0,0028552	1
5.	0	Angkola Barat	1	1
6.	6	Batang Toru	0,0355721	1
7.	7	Marancar	0,1337891	1
8.	8	Muara Batang Toru	0,0000283	-1
9.	11	Sipirok	0,0006799	1
10.	4	Arse	0,0000629	-1
11.	9	Saipar Dolok Hole	0,0000491	-1

Pada tabel 6 di atas merupakan hasil klasifikasi dataset yang dilakukan pada kolom hasil panen. Nilai median pada penelitian ini di dapatkan dari perhitungan yang dilakukan di Microsoft Excel, nilai median yang di dapat yaitu sebesar 0,0006799 (54,80 ton). Berarti jika hasil panen lebih besar dari 0,0006799 (54,80 ton) akan diberi label “1” dan untuk hasil panen yang lebih kecil dari 0,0006799 (54,80 ton) akan diberi label “-1”.

3.1.2 Hasil Analisis Data

Dari hasil uji model prediksi menggunakan algoritma SVM (*Support Vector Machine*) dengan menggunakan 11 data training dan 9 data testing di dapatkan hasil panen yang berbeda sesuai dengan data masukkan yang diberikan. Berikut merupakan data hasil prediksi yang di dapatkan.

Tabel 7. Hasil Analisis Data

Kecamatan	Keterangan	Data Real	Hasil Prediksi
3	Angkola Timur	-1	1
2	Angkola Selatan	-1	1
0	Angkola Barat	1	1
1	Angkola Sangkunar	-1	1
4	Batang Toru	1	1
5	Marancar	1	1
6	Muara Batang Toru	-1	1
8	Sipirok	-1	1

7	Saipar Dolok Hole	1	1
---	-------------------	---	---

Hasil yang terdapat pada tabel 7 di atas merupakan hasil dari perhitungan data proses testing data tahun 2020 dimana data real pada tabel tersebut merupakan data asli tahun 2020 sedangkan yang hasil prediksi merupakan hasil yang diperoleh dari proses testing menggunakan algoritma SVM (*Support Vector Machine*). Dari proses perhitungan manual yang menggunakan 9 data *testing* dan 11 data *training* dapat dihitung nilai RMSE dan tingkat akurasi berdasarkan hasil dari perhitungan manual. Adapun nilai RMSE dan akurasi berdasarkan hasil perhitungan manual sebagai berikut :

a. RMSE

Proses perhitungan nilai RMSE menggunakan rumus yang terdapat pada bab 2, adapun hasilnya sebagai berikut :

$$RMSE = \sqrt{\frac{\sum (y_i - \hat{y}_i)^2}{n}}$$

$$RMSE = \sqrt{\frac{\sum (-1-1)^2 + (-1-1)^2 + (1-1)^2 + (-1-1)^2 + (1-1)^2 + (1-1)^2 + (-1-1)^2 + (-1-1)^2 + (1-1)^2}{9}}$$

$$RMSE = \sqrt{\frac{20}{9}}$$

$$RMSE = 1,49071$$

Berdasarkan perhitungan diatas, diperoleh nilai RMSE sebesar 1,49071 untuk hasil prediksi menggunakan perhitungan manual.

b. Akurasi

Proses perhitungan nilai akurasi menggunakan rumus atau persamaan yang terdapat pada *confusion matrix*, adapun hasilnya sebagai berikut :

$$Akurasi = \frac{TP + TN}{TP + FP + TN + FN} \times 100\%$$

$$Akurasi = \frac{4 + 0}{4 + 0 + 0 + 5} \times 100\%$$

$$Akurasi = \frac{4}{5} \times 100\%$$

$$Akurasi = 44\%$$

Berdasarkan perhitungan diatas, diperoleh nilai Akurasi sebesar 44% untuk hasil prediksi menggunakan perhitungan manual.

3.2 Hasil

3.2.1 Implementasi Sistem Prediksi

Model prediksi yang dibuat pada penelitian ini menggunakan algoritma SVM (*Support Vector Machine*) dengan menggunakan bahasa pemrograman *python* dengan *library pandas*, *library numpy*, dan *library sklearn*, dan menggunakan *text editor jupyter notebook*. Adapun tahapan yang dilakukan dalam representasi data akan di implmentasikan ke dalam model prediksi yang dibuat.

Berikut ini merupakan hasil *preprocessing* dataset pada model prediksi hasil panen pertanian salak di Daerah Tapanuli Selatan.

	Kecamatan	Luas_Lahan	Hasil_Panen
0	Angkola Timur	0,0000872	-1
1	Angkola Selatan	0,0010469	-1
2	Angkola Barat	1	1
3	Angkola Sangkununur	0,0000314	-1
4	Batang Toru	0,1971818	1
5	Marancar	0,0209396	1
6	Muara Batang Toru	0,0000506	-1
7	Sipirok	0,0016228	-1
8	Saipar Dolok Hole	0,0000314	-1

Gambar 3. Hasil *preprocessing* pada Model

Pada gambar 3 di atas merupakan hasil *preprocessing* pada model prediksi yang dibuat menggunakan bahasa pemrograman *python*.

3.2.2 Pengujian

Pada tahapan ini didapatkan hasil prediksi pada model prediksi yang dibuat, berikut merupakan hasil prediksi yang didapatkan.

	Prediksi	Asli
0	1	-1
1	1	-1
2	1	1
3	-1	-1
4	1	1
5	1	1
6	1	-1
7	1	-1
8	1	-1

Gambar 4. Hasil Prediksi pada Model

Pada gambar 4 di atas merupakan hasil *preprocessing* pada model prediksi yang dibuat menggunakan bahasa pemrograman *python*. Pada tabel tersebut dapat dilihat perbandingan hasil data asli yang digunakan dengan hasil prediksi yang di dapatkan.

Kemudian dilakukan proses pencarian nilai RMSE (*Root Mean Square Error*) dari keseluruhan dataset yang berjumlah 20 dataset yang di bagi menjadi data *training* dan data *testing*, dengan dataset tahun 2019 sebagai data *training* dan dataset tahun 2020 sebagai data *testing*.

RMSE : 1.49
ACCURACY DATA TESTING : 0.44

Gambar 5. Hasil RMSE (*Root Mean Square Error*)

Gambar 5 di atas merupakan hasil yang di dapatkan dari model prediksi yang dibuat, yakni di dapatkan nilai RMSE (*Root Mean Square Error*) nya sebesar 1,49 dan mendapatkan akurasi 0,44%.

4. KESIMPULAN

Dari hasil penelitian yang dilakukan dengan menggunakan algoritma SVM (*Support Vector Machine*) di dapatkan nilai RMSE (*Root Mean Square Error*) sebesar 1,49071 dari perhitungan manual dengan 11 data *training* dengan 9 data *testing* dan tingkat akurasi dengan akurasi 44%. Dan pada model prediksi yang dibuat di dapatkan hasil RMSE (*Root Mean Square Error*) sebesar 1,49 dengan akurasi 44%. Dari hasil yang di dapatkan dalam perhitungan manual maupun dalam model prediksi yang dibuat keduanya sama-sama mendapatkan hasil RMSE (*Root Mean Square Error*) sebesar 1,49 dengan tingkat akurasi 44%.

REFERENCES

- [1] M. AR, "Sains, Teknologi, Dan Nilai-Nilai Moral," *Elkawnie*, vol. 2, no. 2, p. 109, 2016, doi: 10.22373/ekw.v2i2.2657.
- [2] H. Priyatno, *Potensi Buah Salak Sebagai Suplemen Obat dan Pangan*, 1st ed. Surakarta: Muhammadiyah University Press, 2018.
- [3] E. P. K. Orpa, E. F. Ripanti, and T. Tursina, "Model Prediksi Awal Masa Studi Mahasiswa Menggunakan Algoritma Decision Tree C4.5," *J. Sist. dan Teknol. Inf.*, vol. 7, no. 4, p. 272, 2019, doi: 10.26418/justin.v7i4.33163.
- [4] Suhardjono, W. Ganda, and H. Abdul, "Prediksi Waktu Kelulusan Mahasiswa Menggunakan Svm Berbasis Pso," *Bianglala Inform.*, vol. 7, no. 2, pp. 97–101, 2019.
- [5] A. Pratama, R. C. Wihandika, and D. E. Ratnawati, "Implementasi algoritme support vector machine (SVM) untuk prediksi ketepatan waktu kelulusan mahasiswa," *J. Pengemb. Teknol. Inf. dan Ilmu Komput.*, vol. 2, no. April, pp. 1704–1708, 2018.
- [6] I. Colanus, R. Drajana, and A. Bode, "Support Vector Machine Untuk Prediksi Produksi Tanaman Pangan di Provinsi Gorontalo," *J. NOE*, vol. 4, no. 2, pp. 3–10, 2021.
- [7] R. H. Kusumodestoni and S. Sarwido, "Komparasi Model Support Vector Machines (Svm) Dan Neural Network Untuk Mengetahui Tingkat Akurasi Prediksi Tertinggi Harga Saham," *J. Inform. Upgris*, vol. 3, no. 1, 2017, doi: 10.26877/jiu.v3i1.1536.
- [8] P. Rifkie, *Algoritma Machine Learning*. Bandung: Informatika Bandung, 2021.
- [9] M. Fitri, M. Anastasia, and A. T. Mohammad, *Data Mining Konsep dan Penerapannya*. Yogyakarta: Deepublish, 2021.
- [10] F. S. Jumeilah, "Penerapan Support Vector Machine (SVM) untuk Pengkategorian Penelitian," *J. RESTI (Rekayasa Sist. dan Teknol. Informasi)*, vol. 1, no. 1, pp. 19–25, 2017, doi: 10.29207/resti.v1i1.11.

- [11] I. C. R. Drajana, "Metode Support Vector Machine Dan Forward Selection Prediksi Pembayaran Pembelian Bahan Baku Kopra," *Ilk. J. Ilm.*, vol. 9, no. 2, pp. 116–123, 2017, doi: 10.33096/ilkom.v9i2.134.116-123.
- [12] H. W. Herwanto, T. Widiyaningtyas, and P. Indriana, "Penerapan Algoritme Linear Regression untuk Prediksi Hasil Panen Tanaman Padi," *J. Nas. Tek. Elektro dan Teknol. Inf.*, vol. 8, no. 4, p. 364, 2019, doi: 10.22146/jnteti.v8i4.537.
- [13] N. Suryani, A. S. Fitriani, D. Normalization, and M. M. Normalization, "Prediction Of Election Participant With Malang City Demographic Data Using The K - Nn Algorithm," vol. 6, no. 36, pp. 2369–2376, 2022.
- [14] I. Suprayogi, Trimajon, and Mahyudin, "Model Prediksi Liku Kalibrasi Menggunakan Pendekatan Jaringan Saraf Tiruan (ZST) (Studi Kasus : Sub DAS Siak Hulu)," *J. Online Mhs. Fak. Tek. Univ. Riau*, vol. 1, no. 1, pp. 1–18, 2014.
- [15] R. I. Akhsanu, *Panduan Pembuatan Flowchart*. Surabaya: Fakultas Kesehatan masyarakat Departemen Administrasi Dan Kebijakan Kesehatan, 2017.
- [16] E. Jubile, *Python Untuk Programmer Pemula*. Jakarta: PT Elex Media Komputindo, 2019.
- [17] Nurhayati, *Teknik Ensemble Learning Untuk Peningkatan Performa Akurasi Model Prediksi (Seleksi Mahasiswa Penerima Beasiswa)*. Tangerang: Pascal Books, 2022.
- [18] N. F. Aldi, H. N. Hanum, and H. Roni, *Analisis Sentimen Terhadap Pembatasan Sosial Menggunakan Deep Learning*. Bandung: Kreatif Industri Nusantara, 2020.
- [19] A. M. F. Alwan, A. Roni, and A. R. Maulana, *Tutorial Optimasi Single Exponential Smoothing Menggunakan Algoritma Genetika*. Bandung: Kreatif Industri Nusantara, 2020.