

BAB I

PENDAHULUAN

Pendahuluan dalam tugas akhir adalah bagian awal yang memberikan penjelasan dan pengantar mengenai topik penelitian yang akan dibahas. Pendahuluan bertujuan untuk memperkenalkan latar belakang penelitian, memberikan justifikasi mengapa topik tersebut penting untuk diteliti, dan menguraikan tujuan serta pertanyaan penelitian yang akan dijawab.

1.1 Latar Belakang

Kesejahteraan penduduk meliputi banyak hal, salah satunya adalah tercukupinya kebutuhan ekonomi. Meningkatkan pemanfaatan potensi wilayah dapat berdampak positif dalam meningkatkan kesejahteraan hidup masyarakat di suatu wilayah. Dalam pengelolaannya, potensi wilayah harus dikelola dengan baik sehingga dapat memberikan manfaat yang signifikan bagi pembangunan ekonomi, sosial, dan lingkungan. Hal ini dilaksanakan dengan menitik beratkan pada upaya memanfaatkan potensi daerah yang dimiliki, baik Sumber Daya Alam maupun Sumber Daya Manusia [1].

Potensi suatu wilayah adalah segala kekayaan yang dimiliki oleh suatu daerah, yang mencakup sumber daya alam maupun sumber daya manusia baik yang sudah dikelola maupun yang belum dikelola [2]. Pemanfaatan sumber daya alam secara bijaksana dapat memberikan kontribusi besar dalam memperkuat sektor industri dan pertanian, serta meningkatkan pendapatan masyarakat melalui peningkatan produksi dan nilai tambah. Sumber daya manusia yang berkualitas juga menjadi faktor kunci dalam pembangunan wilayah, di mana ketersediaan tenaga kerja yang terampil dan profesional dapat meningkatkan daya saing dan produktivitas sektor usaha.

Pemanfaatan Sumber Daya Alam (SDA) memiliki berbagai macam aktivitas. Salah satunya pemanfaatan SDA di Sektor pertanian. Sektor pertanian memiliki beberapa sub sektor, salah satunya adalah sektor tanaman pangan [3]. Sektor tanaman pangan mempunyai peran penting dan strategis dalam menyokong perekonomian suatu

daerah, terlebih dalam menaikan tingkat kesejahteraan dan kemakmuran rakyat [4]. Tanaman pangan sendiri memiliki komoditas utama seperti jagung, padi, kedelai, kacang tanah, ubi jalar, ubi kayu, dan kacang hijau [5].

Berlandaskan data dari Badan Pusat Statistik Produksi sektor tanaman pangan tertinggi saat ini` berada di provinsi Jawa Timur. Produksi tertinggi berada di tanaman padi dengan produksi mencapai 9,908 juta ton. Tingginya produksi tanaman pangan menjadi alasan penetapan lokasi pengambilan data di daerah Jawa Timur dan pengambilan sampel data berdasarkan kabupaten dengan tingkat produksi tertinggi.

Penataan potensi daerah merupakan salah satu model rekognisi segenap sumber daya dan persoalan yang terdapat di suatu daerah [7]. Penelitian ini bertujuan untuk melakukan penataan potensi daerah dengan berfokus pada sektor tanaman pangan dengan menggunakan data iklim. Dalam konteks ini, terdapat fenomena di mana petani cenderung lebih memilih menanam komoditas yang memiliki nilai jual tinggi tanpa memperhatikan kondisi iklim yang sebenarnya [8]. Keadaan tersebut menjadi sebuah permasalahan, karena dapat mengakibatkan ketidakcocokan antara tanaman yang ditanam dengan kondisi iklim setempat, mengurangi produktivitas dan menghadirkan risiko kerugian bagi petani [9]. Disisi lain, pertumbuhan Produk Domestik Bruto (PDB) pada triwulan ke empat tahun 2021, sektor pertanian mengalami penurunan sebesar -1,56% yang mana jika dibandingkan dengan sektor lain, pertanian jauh lebih rendah dibanding sektor lain [10]. Hal ini menunjukkan adanya kebutuhan untuk melakukan klasifikasi yang lebih baik, dengan tujuan menentukan tanaman mana yang cocok untuk ditanam di setiap daerah berdasarkan data iklim dan luas lahan yang tersedia.

Proses identifikasi dan klasifikasi memegang peranan penting dalam memberikan informasi tentang jenis tanaman yang cocok untuk ditanam sesuai dengan kondisi iklim di suatu wilayah. Memberikan gambaran mengenai jenis tanaman yang cocok untuk ditanam dari iklim suatu wilayah dapat membantu petani dalam memaksimalkan nilai produksi suatu komoditas tanaman pangan. Sebagai alternatif untuk mengatasi fenomena tersebut, pembuatan model machine learning seperti

klasifikasi dengan menggunakan SVM dan prediksi produktivitas dengan menggunakan regresi linear dapat dilakukan untuk membantu petani meningkatkan hasil produksi mereka. SVM dipilih karena mampu mengatasi masalah klasifikasi dengan data yang kompleks dan tidak linear, seperti data iklim [11]. Sedangkan regresi linear dipilih karena mampu menghasilkan koefisien untuk setiap variabel bebas, yang menunjukkan seberapa besar pengaruh variabel tersebut terhadap variabel terikat. Hal ini memungkinkan analisis untuk mengidentifikasi variabel yang paling penting dalam memprediksi variabel terikat. [12].

1.2 Rumusan Masalah

Berdasarkan penjelasan latar belakang masalah yang sudah ditulis, dapat diperoleh rumusan permasalahan:

1. Sektor pertanian merupakan sektor penyumbang PDB terendah dibanding sektor lainnya di Indonesia.
2. Banyak petani yang cenderung menanam komoditas yang populer tanpa memperhatikan kondisi iklim.

1.3 Pertanyaan Penelitian

Pertanyaan penelitian merupakan pernyataan atau pertanyaan yang mengarahkan arah penelitian dan menentukan fokus dari studi yang dilakukan.

1. Model seperti apa yang dapat memberikan rekomendasi komoditas tanaman pangan?
2. Model seperti apa yang dapat memperkirakan hasil produksi dari suatu komoditas tanaman pangan?

1.4 Tujuan Penelitian

Tujuan penelitian adalah untuk memberikan pemahaman yang jelas tentang tujuan utama dan tujuan khusus yang ingin dicapai dalam penelitian.

1. Menerapkan SVM sebagai metode untuk mengklasifikasikan jenis tanaman berdasarkan data iklim.

2. Melakukan penerapan Regresi Linear sebagai teknik untuk melakukan prediksi produktivitas suatu jenis tanaman. Menganalisis pengaruh outlier pada performa model dan mengevaluasi performa model dengan dan tanpa outlier.

1.5 Batasan Masalah

Adapun batasan masalah yang berlandaskan pada rumusan masalah adalah sebagai berikut:

1. Penelitian ini berfokus pada pertanian sektor tanaman pangan
2. Dataset yang digunakan dalam penelitian ini berasal dari citra satelit FLDAS dan BPS dengan rentang waktu antara tahun 2004 hingga 2021 di wilayah Jawa Timur.
3. Jumlah dataset sebanyak 252 data dengan indikator tahun, daerah, curah hujan, kelembapan, suhu, suhu tanah, kelembapan tanah, produksi, luas, tanaman, dan produktivitas

1.6 Manfaat Penelitian

Berdasar rumusan masalah dan tujuan, maka bisa diperoleh manfaat penelitian sebagai berikut:

1. Manfaat bagi penulis sendiri dapat menerapkan ilmu kecerdasan buatan yang sudah dipelajari dan diterapkan ke masyarakat.
2. Membantu para pemangku kebijakan agar mendapat wawasan dalam mengelola dan menata potensi suatu daerah, sehingga bisa lebih maksimal dalam pengelolaan potensi.
3. Manfaat bagi petani diharapkan bisa bermanfaat dalam memberikan rekomendasi komoditas tanaman pangan yang akan ditanam dan juga dapat memperkirakan hasil produksi.
4. Memberikan kontribusi dalam menaikkan PDB pada sektor perta

BAB II

TINJAUAN PUSTAKA DAN LANDASAN TEORI

Dalam bagian ini akan dijelaskan mengenai tinjauan pustaka yang akan membahas penelitian sebelumnya dan teori teori yang digunakan dalam penelitian ini

2.1 Tinjauan Pustaka

Tinjauan pustaka adalah proses mengumpulkan dan meninjau sumber-sumber terkait dengan topik penelitian yang sedang dilakukan. Untuk melengkapi penelitian ini, telah dilakukan pengumpulan informasi dari beberapa penelitian sebelumnya sebagai bahan perbandingan terkait kekurangan dan kelebihan yang telah ada.

Penelitian pertama yang dilakukan oleh Lontsi Saadio Cedric memiliki tujuan untuk membuat model berupa alat pendukung keputusan yang bertujuan membantu petani dalam mengambil keputusan dalam memperkirakan hasil pertanian berdasarkan kondisi iklim di negara Afrika, untuk memerangi perubahan iklim dengan lebih baik dan memastikan ketahanan pangan di masa depan lalu membuat model untuk prediksi hasil panen tingkat lanjut berdasarkan model pembelajaran mesin. Pendekatan metode yang digunakan pada penelitian ini adalah regresi logistik multivariat, *k-Nearest Neighbor*, dan model pohon keputusan. Hasil menemukan bahwa model pohon keputusan berkinerja baik dengan koefisien determinasi (R^2) mencapai 95.3% sedangkan model *K-Nearest Neighbor*(KNN) dan regresi logistik masing-masing dengan $R^2 = 93.15\%$ dan $R^2 = 89.78\%$. Hasil prediksi model pohon keputusan dan model *K-Nearest Neighbor* berkorelasi dengan data yang diharapkan dengan nilai akurasi 87%. Penelitian pertama dan penelitian ini memiliki kesamaan dalam hal melakukan prediksi hasil tanam dan melakukan klasifikasi jenis tanaman yang sesuai untuk ditanam di wilayah tertentu. Sedangkan perbedaan dari penelitian ini adalah pada teknik model yang digunakan. Penelitian tersebut menggunakan model algoritma KNN dan juga regresi logistik, sedangkan pada penelitian ini menggunakan model algoritma Regresi Linear dan juga SVM [13]

Penelitian kedua yang dilakukan oleh Fergie Joanda memiliki tujuan untuk membuat model yang dapat membantu petani dalam melakukan penanaman tanaman pangan dengan mempertimbangkan kondisi cuaca, sehingga model bisa memberikan rekomendasi jenis tanaman yang akan ditanam. Variabel data yang digunakan pada penelitian ini meliputi suhu rata-rata pertahun, curah hujan, lama penyinaran, dan hasil produksi. Penelitian ini menggunakan pendekatan algoritma *Decision Tree J48*, dan memberikan akurasi rata rata mencapai 69.48%. Penelitian kedua dan penelitian ini memiliki kesamaan dalam hal pembuatan model untuk memberikan rekomendasi jenis tanaman yang paling sesuai untuk ditanam. Sedangkan perbedaan dari penelitian ini adalah pada algoritma dan pengambilan data. Algoritma yang digunakan untuk menunjang penelitian ini adalah algoritma SVM. Untuk pengambilan data, penelitian tersebut mengambil data yang berasal dari provinsi Sulawesi Utara, sedangkan untuk penelitian ini mengambil data yang berasal dari provinsi Jawa Timur. Penelitian kedua dan penelitian ini memiliki kesamaan dalam hal pembuatan model untuk memberikan rekomendasi jenis tanaman yang paling sesuai untuk ditanam. Sedangkan perbedaan dari penelitian ini adalah pada algoritma dan pengambilan data. Algoritma yang digunakan untuk menunjang penelitian ini adalah algoritma SVM. Untuk pengambilan data, penelitian tersebut mengambil data yang berasal dari provinsi Sulawesi Utara, sedangkan untuk penelitian ini mengambil data yang berasal dari provinsi Jawa Timur [14].

Penelitian ketiga yang dilakukan oleh Dwi Wahyu Triscowati memiliki tujuan untuk membuat model klasifikasi dan pemetaan potensi lahan baku jagung di jawa timur secara otomatis. Data yang digunakan pada penelitian ini menggunakan data satelit *Landsat-8*, *Sentinel-1*, dan *Sentinel-2*, dengan algoritma yang digunakan adalah *Random Forest*. Hasil akurasi tertinggi pada penelitian ini terletak di daerah Banyuwangi dengan nilai mencapai 89%. Penelitian ketiga memiliki kesamaan dengan penelitian ini dalam hal memetakan potensi lahan untuk tanaman pangan menggunakan beberapa citra. Sedangkan perbedaan dari penelitian ini adalah pada objek yang diteliti dan juga teknik algoritma. Objek yang diteliti pada penelitian tersebut hanya berfokus

pada tanaman jagung, sedangkan pada penelitian ini memiliki beberapa objek yaitu berupa jagung, padi, kedelai, kacang tanah, ubi, dan kacang hijau. Untuk teknik algoritma yang digunakan pada penelitian ini adalah algoritma SVM [15].

Pada penelitian keempat yang dilakukan oleh Royyannuur mempunyai tujuan untuk membuat model yang dapat memberikan rekomendasi dari hasil klasifikasi tanaman pangan yang memiliki potensi paling baik untuk ditanam pada suatu daerah berdasarkan kondisi lingkungan. Penelitian tersebut menggunakan data *Famine Early Warning Systems Network* (FEWS NET) dan *Land Data Assimilation System* (FLDAS). Data tersebut diolah dengan menggunakan algoritma *Random Forest* dengan nilai akurasi mencapai 80%. Penelitian ini memiliki persamaan dengan penelitian yang dilakukan oleh Royyannuur Kurniawan Endrayanto dalam hal tujuannya yaitu memberikan rekomendasi tanaman pangan yang cocok untuk ditanam berdasarkan data suatu citra. Sedangkan perbedaan penelitian tersebut dengan penelitian ini adalah metode algoritma yang dipakai. Penelitian tersebut menggunakan metode algoritma *Random Forest*, sedangkan penelitian ini menggunakan metode algoritma SVM [16].

Pada penelitian kelima yang dilakukan oleh Merlinda Wibowo memiliki tujuan untuk membandingkan metode klasifikasi yang mana nantinya akan memberikan rekomendasi Tanaman Pangan. Penelitian ini membandingkan beberapa metode seperti *Naïve Bayes*, *Decision Tree*, *Support Vector Machine (SVM)*, *Neural Network*, *Random Tree*, *Random Forest*, dan *K Nearest Neighbor (KNN)*. Akurasi model tertinggi yang diperoleh pada penelitian tersebut adalah *Random Tree* dengan akurasi mencapai 99%. Penelitian ini memiliki kesamaan dengan penelitian yang disebutkan sebelumnya dalam hal tujuannya untuk mengembangkan metode yang dapat memberikan rekomendasi jenis tanaman pangan yang cocok untuk ditanam di suatu wilayah. Namun, perbedaan utama antara penelitian ini dan penelitian sebelumnya adalah pada sumber data yang digunakan. Pada penelitian sebelumnya, data diambil dari daerah Yogyakarta, sedangkan dalam penelitian ini, data diambil dari wilayah Jawa Timur. Selain itu, penelitian ini menggunakan algoritma SVM sebagai metode utama [17].

Tabel 2. 1 Daftar Penelitian Terkait

No	Nama Peneliti (Tahun)	Objek	Metode	Hasil
1.	Lontsi Saadio Cedric (2022)	Tanaman Pangan	<i>Decision Trees, K-Nearest Neighbor, Regresi Logistik Multivariat</i>	Penelitian tersebut menggunakan model algoritma KNN dengan akurasi 89.78% dan juga regresi logistik dengan MAE 0.117654, sedangkan pada penelitian ini menggunakan model algoritma Regresi Linear dan juga <i>Support Vector Machines (SVM)</i> .
2.	Fergie Joanda Kaunang (2018)	Tanaman Pangan	<i>Decision Tree J48</i>	Penelitian tersebut menggunakan algoritma <i>Decision Tree J48</i> dengan akurasi 69.48% dengan data Sulawesi Utara sedangkan pada penelitian ini menggunakan algoritma SVM dan regresi linear dengan menggunakan data jawa timur.
3.	Dwi Wahyu Triscowati (2021)	Lahan Jagung	<i>Random Forest</i>	Algoritma yang digunakan pada penelitian tersebut adalah Random Forest dengan akurasi mencapai 89%
4	Royyannuur Kurniawan Endrayanto (2021)	Tanaman Pangan	<i>Random Forest</i>	Penelitian tersebut menggunakan algoritma <i>Random Forest</i> dengan akurasi 80% dengan menggunakan citra gambar satelit, sedangkan pada penelitian ini menggunakan algoritma SVM dan regresi dengan menggunakan data iklim dan pertanian.
5	Merlinda Wibowo (2021)	Tanaman Pangan	<i>Naïve Bayes, Decision Tree, Support Vector</i>	Penelitian tersebut membandingkan kinerja algoritma dengan Akurasi model tertinggi yang diperoleh pada penelitian tersebut adalah Random Tree dengan akurasi mencapai 99%. Sedangkan

			<i>Machine (SVM), Neural Network, Random Tree, Random Forest, dan K Nearest Neighbor (KNN)</i>	pada penelitian ini adalah memberikan rekomendasi tanaman pangan dengan menggunakan algoritma SVM
--	--	--	--	---

2.2 Dasar Teori

Dasar teori adalah konsep atau prinsip-prinsip dasar yang membentuk landasan untuk memahami topik penelitian. Pada umumnya, dasar teori berfungsi sebagai dasar bagi peneliti untuk merumuskan hipotesis dan memilih metode penelitian yang sesuai.

2.2.1 Model *Machine learning*

Machine learning adalah bentuk AI yang memungkinkan sistem untuk belajar dari data daripada melalui pemrograman eksplisit. Namun, pembelajaran mesin bukanlah proses yang sederhana. Pembelajaran mesin menggunakan berbagai algoritma yang mana algoritma tersebut memungkinkan untuk meningkatkan, mendeskripsikan, dan memprediksi hasil dari suatu data. Saat suatu algoritma belajar dari suatu data, maka sangat mungkin untuk menghasilkan model yang lebih tepat berdasarkan data tersebut. Model *machine learning* adalah suatu output yang dihasilkan saat melatih suatu data dengan menggunakan suatu algoritma. Setelah dilakukan pelatihan pada suatu data, maka model akan memberikan suatu output [18].

Machine learning mempunyai dua jenis tipe data untuk pelatihan, yaitu *supervised* dan juga *unsupervised*. Perbedaan antara kedua data ini adalah adanya label atau output pada subset data penelitian. Pada *supervised*, atribut output sudah ditentukan dengan atribut input, lalu model dapat melakukan prediksi dan mengklasifikasikan terhadap atribut yang telah ditentukan. Sebaliknya, pada data *unsupervised* melibatkan teknik pengenalan pola tanpa terdapat label atau output pada data. Artinya, semua variabel yang digunakan dalam analisis adalah digunakan sebagai input dan karena pendekatannya, teknik ini sangat cocok untuk *clustering* [19].

2.2.2 *K-Fold Cross Validation*

K-fold Cross Validation adalah suatu teknik validasi yang digunakan untuk menguji dan mengevaluasi akurasi model pada dataset. Dalam teknik ini, K merupakan jumlah subset yang digunakan dalam proses validasi. Misalnya, jika K bernilai 5, maka dataset akan dibagi menjadi 5 subset dengan proporsi yang seimbang. Setiap subset akan bergantian menjadi data uji, sementara subset lainnya digunakan sebagai data

latih. Proses ini diulang sebanyak K kali, di mana setiap subset memiliki kesempatan menjadi data uji. Dari setiap percobaan ini, akan dihasilkan K nilai performa yang mewakili performa model dalam berbagai kondisi pengujian. Selanjutnya, nilai-nilai performa tersebut akan dijumlahkan dan dirata-ratakan untuk mendapatkan estimasi akurasi model secara keseluruhan. Model dengan performa terbaik dapat dipilih berdasarkan nilai rata-rata tertinggi. K -fold Cross Validation memiliki keunggulan dalam efisiensi klasifikasi dataset, karena memanfaatkan seluruh dataset sebagai data uji dan data latih secara bergantian. Namun, perlu diperhatikan bahwa metode ini memerlukan komputasi yang lebih besar karena melibatkan proses sebanyak K kali.

2.2.3 Support Vector Machine

Support Vector Machine (SVM) adalah salah satu algoritma machine learning yang umum digunakan untuk tugas klasifikasi dan regresi. SVM dapat digunakan untuk mengklasifikasikan data dengan menggambar *hyperplane* atau batas keputusan yang memisahkan data ke dalam kelas-kelas yang berbeda. Terdapat dua jenis SVM, yaitu SVM linear dan SVM non-linear. SVM linear adalah ketika data dapat dipisahkan oleh sebuah *hyperplane linear*. Dalam SVM linear, algoritma mencari *hyperplane* yang paling optimal untuk memisahkan data. *Hyperplane* tersebut akan menjadi batas keputusan untuk klasifikasi data [20].

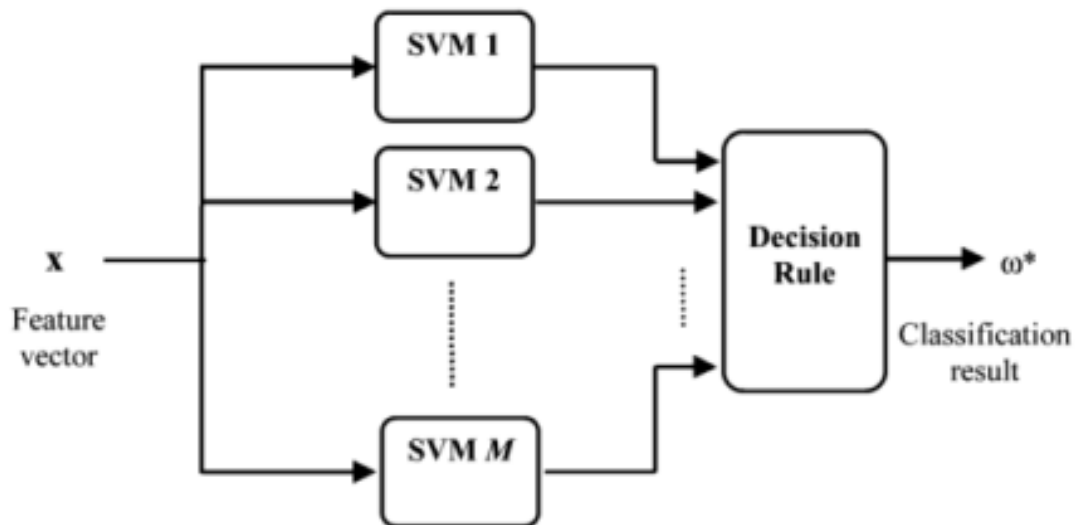
SVM non-linear adalah ketika data tidak dapat dipisahkan oleh sebuah *hyperplane linear*. Dalam SVM non-linear, algoritma memetakan data ke dalam dimensi yang lebih tinggi menggunakan teknik seperti kernel trick, sehingga data dapat dipisahkan oleh *hyperplane*. Kernel adalah fungsi matematika yang menghitung produk titik dalam ruang fitur asli untuk memetakan data ke dalam dimensi yang lebih tinggi, di mana data dapat dipisahkan oleh *hyperplane*. [21].

Salah satu jenis kernel yang paling umum digunakan dalam SVM non-linear adalah kernel RBF (Radial Basis Function). Kernel RBF sangat baik untuk memetakan data ke dalam ruang dimensi yang lebih tinggi, di mana data dapat dipisahkan oleh *hyperplane*. Fungsi RBF dapat dinyatakan sebagai:

$$K(\mathbf{x}, \mathbf{x}') = \exp\left(-\frac{\|\mathbf{x} - \mathbf{x}'\|^2}{2\sigma^2}\right) \quad (2.1)$$

Di mana \mathbf{x} dan \mathbf{x}' adalah vektor input, $\|\mathbf{x} - \mathbf{x}'\|$ adalah jarak antara vektor input \mathbf{x} dan \mathbf{x}' , dan γ adalah sebuah parameter kernel yang mengontrol lebar fungsi kernel. Semakin besar nilai γ , semakin sempit lebar fungsi kernel dan semakin kompleks modelnya. Sebaliknya, semakin kecil nilai γ , semakin lebar fungsi kernel dan semakin sederhana modelnya [21].

Ketika data sudah dimasukkan ke dalam ruang dimensi tinggi, maka *hyperplanes* dapat mengklasifikasikan semua data yang diberikan [22]. Akan tetapi SVM hanya mendukung klasifikasi biner, maka dari itu untuk klasifikasi *multiclass* prinsip yang sama digunakan setelah memecah masalah multiklasifikasi menjadi submasalah yang lebih kecil, yang semuanya merupakan masalah klasifikasi biner. Metode populer yang digunakan untuk melakukan multi-klasifikasi pada SVM salah satunya yaitu *One-Vs-All* atau *One-Vs-Rest*. Melalui metode ini dilakukan pembagian data multiclass sebagai data klasifikasi biner sehingga algoritma klasifikasi biner dapat diterapkan untuk mengkonversi data klasifikasi biner [23]. Hal itu bisa dilihat pada gambar 2.2



Gambar 2.1 *One Vs Rest*

Data yang dikonversi menjadi data klasifikasi biner akan terlihat seperti berikut:

- 1) Padi vs [Jagung, Kedelai, Ubi Jalar, Ubi Kayu, Kacang Hijau, Kacang Tanah]
- 2) Jagung vs [Padi, Kedelai, Ubi Jalar, Ubi Kayu, Kacang Hijau, Kacang Tanah]
- 3) Kedelai vs [Padi, Jagung, Ubi Jalar, Ubi Kayu, Kacang Hijau, Kacang Tanah]
- 4) Ubi Jalar vs [Padi, Jagung, Kedelai, Ubi Kayu, Kacang Hijau, Kacang Tanah]
- 5) Ubi Kayu vs [Padi, Jagung, Kedelai, Ubi Jalar, Kacang Hijau, Kacang Tanah]
- 6) Kacang Hijau vs [Padi, Jagung, Kedelai, Ubi Jalar, Ubi Kayu, Kacang Tanah]
- 7) Kacang Tanah vs [Padi, Jagung, Kedelai, Ubi Jalar, Ubi Kayu, Kacang Hijau]

Dengan menggunakan pendekatan *one-versus-rest* di mana nantinya model akan melatih pengklasifikasi biner C , $fc(x)$, di mana data dari kelas c diperlakukan sebagai positif, dan data dari semua kelas lain diperlakukan sebagai negatif [24].

2.2.4 Multiclass Confusion Matrix

Perhitungan evaluasi kinerja dari model klasifikasi SVM yang sudah dibuat dilakukan dengan menggunakan *Multiclass Confusion Matrix*, yaitu sebuah metode untuk mengevaluasi performa model klasifikasi pada suatu dataset yang memiliki lebih dari dua kelas atau label [25].

Tabel 2. 2 Confusion Matrix Multiclass

CONFUSION MATRIX MULTICLASS							
	KELAS 1	KELAS 2	KELAS 3	KELAS 4	KELAS 5	KELAS 6	KELAS 7
Kelas 1	TP(1)	FP(2)	FP(3)	FP(4)	FP(5)	FP(6)	FP(7)
Kelas 2	FP(1)	TP(2)	FP(3)	FP(4)	FP(5)	FP(6)	FP(7)
Kelas 3	FP(1)	FP(2)	TP(3)	FP(4)	FP(5)	FP(6)	FP(7)
Kelas 4	FP(1)	FP(2)	FP(3)	TP(4)	FP(5)	FP(6)	FP(7)
Kelas 5	FP(1)	FP(2)	FP(3)	FP(4)	TP(5)	FP(6)	FP(7)
Kelas 6	FP(1)	FP(2)	FP(3)	FP(4)	FP(5)	TP(6)	FP(7)
Kelas 7	FP(1)	FP(2)	FP(3)	FP(4)	FP(5)	FP(6)	TP(7)

Dengan keterangan TP(k) merupakan *True Positive* pada kelas k , FP(k) adalah *False Positive* pada kelas k , dan FN(k) merupakan *False Negative* pada kelas k

Perhitungan *confusion matrix* mencakup perhitungan *accuracy*, *precision*, *recall*, dan *F1 Score* [26]. persamaan dari perhitungan *accuracy*, *precision*, dan *recall* bisa dilihat sebagai berikut:

$$Accuracy = \frac{(TP(1) + TP(2) + \dots + TP(k))}{Total\ data\ aktual} \quad (2.2)$$

$$Precision = \frac{TP(k)}{(TP(k) + FP(k))} \quad (2.3)$$

$$Precision = \frac{TP(k)}{(TP(k) + FP(k))} \quad (2.4)$$

$$F1\ Score(k) = 2 \times \frac{(Presisi(k))}{(Presisi(k) + Recall(k))} \quad (2.5)$$

Accuracy digunakan untuk mengukur seberapa akurat model klasifikasi dalam memprediksi kelas yang benar. *Accuracy* didefinisikan sebagai jumlah prediksi yang benar (*true positive*) dibagi dengan total data aktual. TP(k) adalah *true positive* untuk kelas k. Total data aktual adalah jumlah keseluruhan data pada kelas k [25].

2.2.5 Linear Regresion

Linear Regression adalah prosedur statistik untuk menghitung nilai variabel dependen dari variabel independen. Regresi linier mengukur hubungan antara dua atau lebih variabel. Ini adalah teknik pemodelan di mana variabel dependen diprediksi berdasarkan satu atau lebih Variabel independen [27].

Persamaan yang menjelaskan bagaimana hubungan antara variabel X dan Y pada suatu model kesalahan disebut model regresi. Model umum yang digunakan adalah sebagai berikut :

$$Y = a + b_1X_1 + b_2X_2 + \dots + b_nX_n \quad (2.5)$$

Persamaan yang digunakan untuk menghitung nilai prediksi variabel dependen (Y) berdasarkan nilai-nilai variabel independen (x_1, x_2, \dots, x_n) yang telah diukur atau diamati. Dalam persamaan tersebut, b_0 adalah intercept atau nilai konstanta yang menunjukkan nilai y ketika semua variabel independen bernilai 0. Sedangkan $b_1, b_2,$

..., b_n adalah koefisien regresi yang menunjukkan seberapa besar pengaruh variabel independen (x_1, x_2, \dots, x_n) terhadap variabel dependen (y) [28].

2.2.6 Mean Absolute Error

Dalam melakukan regresi linear, penting untuk melakukan evaluasi terhadap model yang dihasilkan untuk menentukan apakah model tersebut cukup baik untuk digunakan sebagai prediksi atau tidak [29]. Pengukuran evaluasi kinerja model regresi menggunakan *Means Absolute Error* (MAE) yang menghitung rata-rata dari nilai absolut selisih antara nilai prediksi dan nilai yang diamati. Oleh karena itu, MAE memberikan bobot yang sama pada semua error, baik itu yang besar maupun kecil. MAE sangat berguna dalam kasus-kasus di mana terdapat banyak outlier yang besar atau ketika kita tidak ingin memperhatikan perbedaan antara kesalahan positif dan negatif [30]. Persamaan untuk menghitung MAE adalah sebagai berikut

$$MAE = \frac{1}{n} \sum_{i=1}^n |Y_i - \hat{Y}_i| \quad (2.6)$$

Dalam analisis MAE, terdapat dua nilai penting yang sering digunakan yaitu nilai observasi (Y_i) dan nilai prediksi (y_i).