

## **Penerapan Metode Support Vector Machine (SVM) Dalam Klasifikasi Produktivitas Padi**

**Hamim Tohari\*, Sri Harini, M Ainul Yaqin, Irwan Budi Santoso, Cahyo Crysdiان**

Fakultas Saintek, Program Studi Magister Informatika, UIN Maulana Malik Ibrahim, Malang, Indonesia

Email: <sup>1,\*</sup>hamimtohari77@gmail.com, <sup>2</sup>sriharini@mat.uin-malang.ac.id, <sup>3</sup>yaqinov@ti.uin-malang.ac.id,

<sup>4</sup>irwan@ti.uin-malang.ac.id, <sup>5</sup>cahyo@ti.uin-malang.ac.id

Email Penulis Korespondensi: hamimtohari77@gmail.com

Submitted: 06/11/2023; Accepted: 30/11/2023; Published: 30/11/2023

**Abstrak**—Indonesia merupakan salah satu negara penghasil beras terbesar di dunia. Penduduk Indonesia hampir 95% mengonsumsi beras sebagai bahan pangan pokok wajib, sehingga setiap tahunnya permintaan beras semakin meningkat seiring dengan bertambahnya jumlah penduduk. Jawa Timur dikenal dengan provinsi penghasil padi terbesar di Indonesia. Untuk mengoptimalkan produksi beras, provinsi Jawa Tengah dapat mengelompokkan kota atau kabupaten penghasil padi. Hal ini bertujuan untuk melihat dan mengetahui kota atau kabupaten yang memiliki potensi dalam memproduksi padi sekaligus mengetahui daerah yang memiliki produksi padi kurang maksimal. Untuk melihat terpenuhinya kebutuhan beras di provinsi Jawa Timur maka diperlukan adanya prediksi produktivitas padi di wilayah Jawa Timur sehingga bisa dijadikan dasar dalam upaya peningkatan hasil panen padi untuk periode berikutnya. Dalam penelitian ini digunakan metode Support Vector Machine (SVM) dalam klasifikasi data untuk prediksi hasil panen di provinsi Jawa Timur. Keunggulan dari algoritma SVM yaitu dapat digunakan untuk masalah klasifikasi dan regresi dengan kernel linier ataupun non-linier kernel. Data yang digunakan adalah data statistik pertanian yang didapat dari website [jatim.bps.go.id](http://jatim.bps.go.id). data-data tersebut kemudian dianalisis dengan proses data mining. Hasil dari penelitian ini berupa pola prediksi dengan pohon keputusan yang dapat dijadikan dasar prediksi dalam memperkirakan hasil panen pada periode selanjutnya. Dari pembagian data training sebanyak 80% dan data testing 20% didapatkan hasil dengan akurasi 80% pada saat memprediksi kategori '0' (tidak sesuai target), dan mendapat akurasi 100% pada saat memprediksi kategori '1' (sesuai target). Dan secara keseluruhan, classification model memiliki akurasi 88%. Kontribusi yang ingin dicapai dalam penelitian ini adalah memberikan gagasan dalam pengolahan data di bidang pertanian.

**Kata Kunci:** Padi; Produktivitas; Jawa Timur; Klasifikasi; SVM

**Abstract**—Indonesia is one of the largest rice producing countries in the world. Almost 95% of the Indonesian population consumes rice as a mandatory staple food, so that every year the demand for rice increases along with the increase in population. East Java is known as the largest rice producing province in Indonesia. To optimize rice production, Central Java province can group rice producing cities or districts. This aims to see and find out cities or districts that have the potential to produce rice as well as find out areas that have less than optimal rice production. To see whether the need for rice in East Java province is met, it is necessary to predict rice productivity in the East Java region so that it can be used as a basis for efforts to increase rice yields for the next period. In this research, the Support Vector Machine (SVM) method was used to classify data for predicting crop yields in East Java province. The advantage of the SVM algorithm is that it can be used for classification and regression problems with linear kernels or non-linear kernels. The data used is agricultural statistical data obtained from the website [jatim.bps.go.id](http://jatim.bps.go.id). The data is then analyzed using a data mining process. The results of this research are in the form of a prediction pattern with a decision tree which can be used as a basis for predictions in estimating harvest results in the next period. From dividing 80% training data and 20% testing data, results were obtained with 80% accuracy when predicting category '0' (not on target), and 100% accuracy when predicting category '1' (on target). And overall, the classification model has an accuracy of 88%. The contribution to be achieved in this research is to provide ideas

for data processing in the agricultural sector.

**Keywords:** Rice; Productivity; East Java; Classification; SVM

### **1. PENDAHULUAN**

Pangan merupakan salah satu kebutuhan pokok dan mendasar bagi manusia yang harus dipenuhi. Pemenuhan kebutuhan pangan dilakukan sebagai upaya agar pangan selalu tersedia setiap saat dan juga dapat terjangkau oleh seluruh lapisan masyarakat[1]. Untuk mewujudkan terlaksananya ketahanan pangan bagi masyarakat Indonesia, sektor pertanian menjadi sektor utama dalam upaya penyediaan kebutuhan pangan. Pertanian selalu menjadi sasaran utama dalam upaya pembangunan sektor ekonomi di Indonesia[2].

Kajian tentang sektor pertanian seolah tak bisa lepas dengan diskusi tentang ketahanan pangan. Berdasarkan undang-undang, konsep ketahanan pangan adalah kondisi terpenuhinya pangan bagi rumah tangga yang tercermin dari tersedianya pangan yang cukup, baik jumlah maupun mutunya, aman, merata dan terjangkau[3]. Tujuan utama dari kebijakan pangan adalah untuk menjamin tercapainya ketahanan pangan yang meliputi pasokan, keamanan, diversifikasi, kelembagaan dan organisasi pangan. Karena itu, kebijakan ini dibutuhkan untuk meningkatkan kemandirian pangan. Pembangunan yang mengabaikan kemandirian dalam memenuhi kebutuhan dasar bagi penduduk akan berakibat fatal sehingga bisa menimbulkan ketergantungan pada negara lain. Dengan kata lain, kedaulatan pangan di negara akan terancam[4].

Kesuksesan pemerintah dalam membangun sektor pertanian suatu daerah terlihat dari kemampuan daerah tersebut mewujudkan swasembada pangan. Pengembangan sektor pertanian berdampak pada jangkauan luas

terhadap program pengentasan kemiskinan dan peningkatan sumber daya manusia. Pangan di Indonesia selalu identik dengan beras, dikarenakan selama ini beras merupakan makan pokok bagi sebagian besar masyarakat di negeri ini[5].

Jawa menjadi penyumbang terbesar di sektor pangan nasional melebihi wilayah lain di Indonesia. Oleh karena itu, Jawa memiliki peran penting dalam perubahan status Indonesia dari pengimpor beras terbesar menjadi pengeksport terbesar di tahun 1984 yang ditandai dengan keberhasilan swasembada beras sehingga memperoleh penghargaan dari FAO Organisasi Pangan dan Pertanian Dunia) pada tahun 1985[6].

Dapat dikatakan bahwa sejak tahun 2016 hingga 2019 peranan kategori pertanian mengalami penurunan sebesar 2,09 persen, namun mengalami penguatan pada tahun 2020 sebesar 0,55 persen dibanding tahun 2019[7]. Di samping itu, peran strategis sektor pertanian juga ditunjukkan dari kontribusinya terhadap penyerapan tenaga kerja yang terbesar dibandingkan dengan sektor lainnya, yakni sekitar 33,01 persen berdasarkan hasil Survei Angkatan Kerja Nasional pada Agustus 2020[8]. Peningkatan ini membuktikan sektor pertanian adalah sektor yang paling bertahan di kala pandemi COVID-19 sehingga perlunya perhatian yang lebih dari pemerintah agar dapat mendongkrak gairah ekonomi saat ini di samping peranan penting sektor pertanian terhadap keberlangsungan hidup masyarakat khususnya di Provinsi Jawa Timur[9]. Untuk mengoptimalkan produksi beras, provinsi Jawa Timur dapat mengelompokkan kota atau kabupaten penghasil padi. Hal ini bertujuan untuk melihat dan mengetahui kota atau kabupaten yang memiliki potensi dalam memproduksi padi sekaligus mengetahui daerah yang memiliki produksi padi kurang maksimal[10]. Metode dalam data mining dapat menemukan pola-pola yang tidak terlihat dan menarik dari kumpulan data[11].

Wijayanto dan Yoka melakukan pengelompokan kota/kabupaten penghasil padi di Jawa Tengah berdasarkan produktivitas padinya dengan metode clustering. Clustering yang digunakan adalah dengan menggunakan algoritma K-Means. Hasil yang didapatkan dari metode ini yaitu 3 cluster antara lain C0 menggambarkan daerah yang hasil produktivitas padinya sedang terdiri dari 12 daerah. C1 menggambarkan daerah yang hasil produktivitas padinya rendah terdiri dari 18 daerah. Serta C2 menggambarkan daerah yang hasil produktivitas padinya tinggi terdiri dari 12 daerah[4].

Siti Rokhmah melakukan penelitian menggunakan algoritma C.45 dalam klasifikasi data untuk prediksi hasil panen di wilayah Kabupaten Sukoharjo. Data yang digunakan adalah data statistik pertanian yang didapat dari website sukoharjo.bps.go.id. data-data tersebut kemudian dianalisis dengan teknik data mining, sehingga menghasilkan data training yang siap untuk di olah dengan algoritma C.45, pengujian dilakukan dengan split validation untuk menghasilkan validitas klasifikasi data. Hasil model prediksi dengan algoritma C.45 diuji dengan split validation dengan tingkat performance 100 % terhadap data yang sama dengan tahun yang berbeda[3].

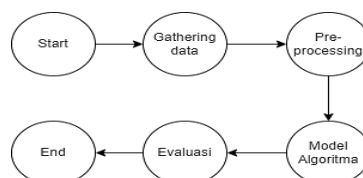
Selain metode di atas, metode data mining yang dapat digunakan untuk klasifikasi adalah Support Vector Machine (SVM). SVM merupakan algoritma dengan metode supervised learning yang menganalisa data dan mengenali pola[12]. Adapun keunggulan dari algoritma SVM, yaitu dapat digunakan untuk masalah klasifikasi dan regresi dengan kernel linier ataupun non-linier kernel yang dapat menjadi satu kemampuan algoritma pembelajaran untuk klasifikasi serta regresi dengan akurasi yang tinggi[13]. SVM dikembangkan oleh Vapnik, Guyon & Boser[14]. Pertama kali SVM ditampilkan sekitar tahun 1992 pada Annual Workshop on Computational Learning Theory[15]. Metode SVM merupakan metode machine learning dengan tujuan mencari hyperplane terbaik yang membagi dua buah kelas di input space[16]. Metode SVM merupakan salah satu teknik yang baru bila dibandingkan dengan teknik lain[17]. Pemilihan fungsi kernel yang tepat dan sesuai merupakan hal yang sangat penting dan diperlukan, sebab fungsi dari kernel tersebut yang akan menentukan feature space dimana fungsi dari klasifier akan dicari[18].

Maka dari itu, sebagai perbandingan, dalam penelitian ini metode SVM akan digunakan untuk klasifikasi produktivitas padi di Jawa Timur. Tujuan penelitian ini adalah berkontribusi dalam memberikan gagasan dalam pengolahan data di bidang pertanian dengan mengaplikasikan metode SVM untuk klasifikasi produktivitas padi.

## 2. METODOLOGI PENELITIAN

### 2.1 Tahapan Penelitian

Dalam penelitian ini akan dilakukan klasifikasi menggunakan metode *Support Vector Machine* (SVM) terhadap dataset produksi padi. Penelitian ini akan bertahap melalui beberapa proses. Adapun tahapan penelitian bisa dilihat pada **gambar 1** di bawah ini.



**Gambar 1.** Tahapan Penelitian Klasifikasi dengan Metode SVM.

Berikut penjelasan untuk uraian diagram di atas:

a. Gathering data

Tahapan awal dalam penelitian ini adalah memilih dataset yang akan digunakan. Pada penelitian ini dataset yang digunakan adalah merupakan dataset sebanyak 190 data yang terdiri dari data lahan, jumlah panen padi, produktivitas padi dan target produktivitas padi dari seluruh kota/kabupaten di Jawa Timur. Dengan data ini diharapkan dapat memperoleh hasil yang akurat dan terpercaya.

b. Pre-processing

Sebelum dilakukan proses pengklasifikasian data, dilakukan metode pemrosesan data awal terlebih dahulu sesuai dengan proses metode Knowledge Discovery in Database (KDD) yaitu preprocessing data. Data atau record dan atribut harus melalui beberapa tahap pengolahan awal data seperti pembersihan data (cleaning), filtering, transformasi data, dan lain-lain sehingga mendapatkan data yang berkualitas.

c. Model Algoritma

Pada penelitian ini algoritma yang digunakan adalah Support Vector Machine (SVM). Untuk tahap pemodelan SVM dilakukan dengan tujuan untuk menemukan *hyperplane* terbaik yang memisahkan dua buah kelas pada *input space*. Tingkat akurasi pada model yang akan dihasilkan dalam proses peralihan dengan SVM sangat bergantung terhadap fungsi kernel dan parameter yang digunakan.

d. Evaluasi

Evaluasi dilakukan dengan cara menganalisa hasil dari algoritma yang digunakan untuk memastikan bahwa hasil perhitungan dan pengujian sesuai dengan tujuan penelitian. Evaluasi dilakukan untuk mengukur hasil klasifikasi guna mengetahui tingkat accuracy, precision, dan recall. Proses evaluasi menggunakan *Confusion Matrix*. Berikut tabel 1 dari *Confusion Matrix*:

**Tabel 1.** Tabel *Confusion Matrix*

Correct Classification	Classified as	
	Actual “+”	Actual “-“
Predicted “+”	TP (True Positive)	FP (False Postive)
Predicted “-“	FN (False Negative)	TN (True Negative)

**2.2 Tahap Pengujian**

Adapun penjelasan mengenai langkah-langkah dalam pengujian sebagai berikut:

a. Insert Dataset

Langkah pertama ialah memasukkan dataset yang akan diujikan.

b. Gathering Information

Pada tahapan ini mengumpulkan sejumlah informasi pada dataset mulai dari jumlah baris, jumlah atribut, korelasi antar atribut, jumlah variabel NULL/NaN, dan lain-lain.

c. Data Preprocessing

Langkah selanjutnya adalah melakukan preprocessing seperti data cleansing, selection feature, transformation data, dan feature engineering. Ini berguna untuk membersihkan nilai value yang kosong dan juga pemilihan atribut karena tidak semua atribut akan dipakai.

d. Split data Validation

Pada tahap ini data akan membagi jumlah keseluruhan dataset menjadi 2 yaitu data training dan data testing dengan nilai yaitu 80:20, dengan data training 80% dan data testing 20%. Setelah membagi dataset langkah selanjutnya adalah memasukkan data training kedalam model algoritma. Setelah melakukan penerapan algoritma, kemudian hasil tersebut akan diujikan dengan data testing sehingga akan mendapatkan bentuk confusion matrix sebagai nilai akhir.

**2.3 Data Penelitian**

Pada penelitian ini, data yang digunakan adalah data pertanian di Provinsi Jawa Timur tahun 2022 yang diperoleh dari website Badan Pusat Statistik (BPS) Jawa Timur. Totalnya ada 190 data. Data tersebut akan diolah menggunakan metode data mining *sSupport Vector Machine* (SVM). Adapun data penelitian sebagaimana terdapat di tabel 2.

**Tabel 2.** Dataset Penelitian

Kabupaten/Kota	Luas Lahan (ha)	Produksi Padi (ton)	Produktivitas Padi (ku/ha)
Kab. Pacitan	17.988,72	90.955,25	50,56
Kab. Ponorogo	64.543,87	370.435,11	57,39
Kab. Trenggalek	22.293,31	117.346,67	52,64
Kab. Tulungagung	40.611,23	221.637,05	54,58
Kab. Blitar	34.255,13	217.566,97	63,51
Kab. Kediri	29.851,89	172.474,88	57,78

Kabupaten/Kota	Luas Lahan (ha)	Produksi Padi (ton)	Produktivitas Padi (ku/ha)
Kab. Malang	47.821,21	283.895,29	59,37
Kab. Lumajang	56.064,69	303.468,63	54,13
Kab. Jember	119.808,13	613.237,38	51,18
Kab. Banyuwangi	77.952,96	462.584,81	59,34
Kab. Bondowoso	49.262,33	246.388,27	50,02
Kab. Situbondo	27.172,07	141.914,27	52,23
Kab. Probolinggo	34.231,53	187.277,08	54,71
Kab. Pasuruan	47.999,47	254.578,42	53,04
Kab. Sidoarjo	31.431,32	196.839,63	62,63
Kab. Mojokerto	49.281,71	287.251,32	58,29
Kab. Jombang	55.229,58	343.427,84	62,18
Kab. Nganjuk	71.223,24	387.897,28	54,46
Kab. Madiun	74.141,77	419.977,93	56,65
Kab. Magetan	42.606,10	263.822,71	61,92
Kab. Ngawi	129.474,04	785.037,99	60,63
Kab. Bojonegoro	133.739,17	715.198,84	53,48
Kab. Tuban	85.194,97	502.136,24	58,94
Kab. Lamongan	151.263,60	920.935,59	60,88
Kab. Gresik	63.241,32	410.323,14	64,88
Kab. Bangkalan	39.691,95	193.329,37	48,71
Kab. Sampang	34.880,84	172.558,93	49,47
Kab. Pamekasan	20.497,42	108.020,20	52,70
Kab. Sumenep	40.558,48	223.000,46	54,98
Kota Kediri	1.882,04	10.449,52	55,52
Kota Blitar	734,37	5.259,99	71,63
Kota Malang	1.753,52	11.524,42	65,72
Kota Probolinggo	1.309,07	8.067,16	61,63
Kota Pasuruan	1.432,61	7.939,57	55,42
Kota Mojokerto	742,57	4.324,18	58,23
Kota Madiun	2.174,03	11.585,31	53,29
Kota Surabaya	1.455,52	8.180,52	56,20
Kota Batu	963,70	5.912,16	61,35

## 2.4 Pre-Processing

Sebelum proses data mining dimulai, data diproses terlebih dahulu untuk diberi label dalam 2 kelas. Dalam hal ini dibagi dalam 2 kelas, yaitu : tidak sesuai target (0) dan sesuai target (1). Adapun nilai produktivitas padi akan sesuai target jika  $> 60$  dan tidak sesuai target jika  $\leq 60$ . Berikut tabel 3 yang merupakan hasil dari pre-processing data:

**Tabel 3.** Hasil Pre-Pocessing

Kabupaten/Kota	Luas Lahan	Produksi Padi	Produktivitas Padi	Target
Kab. Pacitan	17.988,72	90.955,25	50,56	0
Kab. Ponorogo	64.543,87	370.435,11	57,39	0
Kab. Trenggalek	22.293,31	117.346,67	52,64	0
Kab. Tulungagung	40.611,23	221.637,05	54,58	0
Kab. Blitar	34.255,13	217.566,97	63,51	1
Kab. Kediri	29.851,89	172.474,88	57,78	0
Kab. Malang	47.821,21	283.895,29	59,37	0
Kab. Lumajang	56.064,69	303.468,63	54,13	0
Kab. Jember	119.808,13	613.237,38	51,18	0
Kab. Banyuwangi	77.952,96	462.584,81	59,34	0
Kab. Bondowoso	49.262,33	246.388,27	50,02	0
Kab. Situbondo	27.172,07	141.914,27	52,23	0
Kab. Probolinggo	34.231,53	187.277,08	54,71	0
Kab. Pasuruan	47.999,47	254.578,42	53,04	0
Kab. Sidoarjo	31.431,32	196.839,63	62,63	1
Kab. Mojokerto	49.281,71	287.251,32	58,29	0
Kab. Jombang	55.229,58	343.427,84	62,18	1
Kab. Nganjuk	71.223,24	387.897,28	54,46	0
Kab. Madiun	74.141,77	419.977,93	56,65	0

Kabupaten/Kota	Luas Lahan	Produksi Padi	Produktivitas Padi	Target
Kab. Magetan	42.606,10	263.822,71	61,92	1
Kab. Ngawi	129.474,04	785.037,99	60,63	1
Kab. Pamekasan	20.497,42	108.020,20	52,70	0
Kab. Sumenep	40.558,48	223.000,46	54,98	0
....	...	...	...	...
Kota Blitar	734,37	5.259,99	71,63	1
Kota Malang	1.753,52	11.524,42	65,72	1
Kota Probolinggo	1.309,07	8.067,16	61,63	1
Kota Pasuruan	1.432,61	7.939,57	55,42	0
Kota Mojokerto	742,57	4.324,18	58,23	0
Kota Madiun	2.174,03	11.585,31	53,29	0
Kota Surabaya	1.455,52	8.180,52	56,20	0
Kota Batu	963,70	5.912,16	61,35	1

Berikut atribut penelitian sebagaimana dijelaskan di tabel 4:

**Tabel 4.** Atribut Penelitian

Atribut	Keterangan	Skala
Y	Target (Label)	Nominal
X <sub>1</sub>	Kota/Kabupaten	Nominal
X <sub>2</sub>	Luas Lahan	Numerik
X <sub>3</sub>	Produksi Padi	Numerik
X <sub>4</sub>	Produktivitas Padi	Numerik

Berikut deskripsi atribut penelitian sebagaimana dijelaskan di tabel 5:

**Tabel 5.** Deskripsi Atribut Penelitian

Atribut	Deskripsi
Y	Target Produktivitas Padi yang dicapai 0 = tidak sesuai target 1 = sesuai target
X <sub>1</sub>	Nama Kota/Kabupaten
X <sub>2</sub>	Total luas lahan yang digunakan menanam padi
X <sub>3</sub>	Jumlah Produksi Padi dalam satuan ton
X <sub>4</sub>	Nilai Produktivitas Padi

## 2.5. Tahapan Pemodelan

*Support Vector Machine* (SVM) ialah model *classification supervised learning*. Tujuan SVM ialah menemukan batas pemisah (*hyperplane*) untuk memaksimalkan margin sehingga didapat performa yang maksimal. SVM dapat memprediksi secara optimal, baik terhadap regresi ataupun klasifikasi. Di masa kini, permasalahan tidak bisa dipecahkan secara linier. Hal itu menyebabkan perlu diterapkan pemecahan masalah secara non-linier[19]. Untuk mengubah data dari non-linier ke dalam bentuk linier diperlukan suatu fungsi yang bernama kernelisasi. Kernel didefinisikan menjadi fungsi yang menggambarkan fitur data mulai dari perspektif terendah menjadi fitur baru dengan perspektif yang lebih besar. 3 variasi kernel yang sering digunakan untuk pemecahan masalah secara non-linier dengan persamaan sebagai berikut :

a) Kernel Polynominal

$$K(x, y) = (x \cdot y + c)^d \tag{1}$$

b) Kernal Gaussian RBF

$$K(x, y) = \exp\left(\frac{-||x-y||^2}{2\sigma^2}\right) \tag{2}$$

c) Kernel Linear

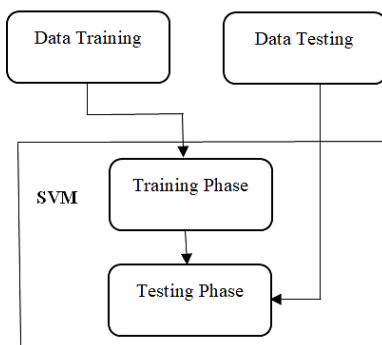
$$K(x, y) = x \cdot y \tag{3}$$

Dengan  $x$  dan  $y$  adalah pasangan dua data pada semua bagian data training. Parameter  $\sigma$  adalah sigma,  $c$  adalah *complexity*,  $d$  adalah *degree* dan  $-||x - y||^2$  adalah kuadrat dari jarak vektor  $x$  dan  $y$ .

Menurut teorema Vapnik, pemilihan margin maksimum akan diberikan kemampuan generalisasi yang baik. Untuk menentukan persamaan *hyperplane*  $\langle w, x \rangle + b = 0$  harus diketahui lebih dahulu nilai  $w$  dan  $b$ . Disini nilai  $w$  merupakan nilai yang berkaitan dengan margin. Untuk memperolehnya digunakan bantuan *hyperplane* kanonik. Dalam hal ini,  $x^+$  adalah data yang terletak pada kelas  $y = + 1$  dan terdekat ke *hyperplane*

dan  $x^-$  adalah data yang terletak pada kelas  $y = -1$  dan terdekat ke *hyperplane*. Lebar margin  $y$  adalah jarak  $x^+$  ke *hyperplane* atau jarak  $x^-$  ke *hyperplane*[20].

Sebelum menerapkan data-data dalam algoritma SVM, data training dan data testing ditentukan terlebih dahulu dari data yang merupakan hasil transformasi. Data training dimasukkan ke dalam fase training pada algoritma SVM. Setelah didapatkan modelnya, data testing dimasukkan ke dalam fase testing pada algoritma SVM. Tahapan permodelan dapat dilihat pada gambar 2.



Gambar 2. Tahap Permodelan

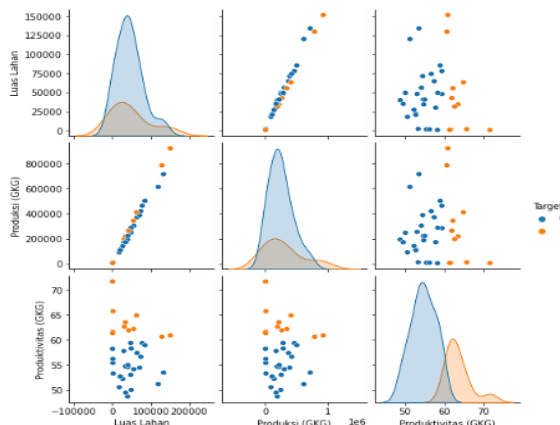
Langkah selanjutnya adalah Penerapan metode SVM dan membagi dataset menjadi 2 yaitu data training dan data testing. Data dibagi 2 dengan bobot nilai sebesar 80:20 dimana 80% data akan digunakan sebagai data training dan 20% data akan digunakan sebagai data testing. Adapun proses pemodelan metode SVM dalam penelitian ini menggunakan platform Jupyter Notebook untuk membuat sebuah program dengan menggunakan bahasa pemrograman *python*.

### 2.6 Tahap Evaluasi

Langkah selanjutnya adalah melakukan evaluasi terhadap hasil kinerja Metode SVM dengan menggunakan *Confusion Matrix*. *Confusion Matrix* adalah tabel dengan 4 kombinasi berbeda dari nilai prediksi dan nilai aktual. Ada empat istilah yang merupakan representasi hasil proses klasifikasi pada *confusion matrix* yaitu *True Positive (TP)*, *True Negative (TN)*, *False Positive (FP)*, dan *False Negative (FN)*[12]. Pada tahapan ini akan dilakukan pembuatan *confusion matrix* dengan menggunakan library pendukung dari ScikitLearn dan memvisualisasikannya menggunakan seaborn. Langkah terakhir dalam penelitian ini adalah melakukan validasi dan evaluasi terhadap hasil pengujian. Pada tahapan ini akan didapatkan hasil akhir dari kemampuan algoritma dalam klasifikasi produktivitas padi di Jawa Timur. Hasil yang akan didapatkan adalah berupa *accuracy score*, *recall*, dan *precision*.

## 3. HASIL DAN PEMBAHASAN

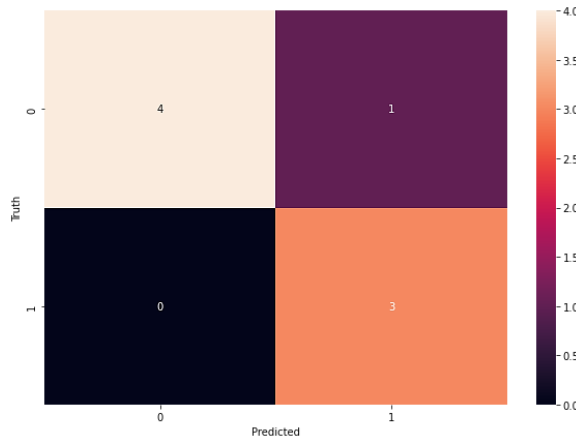
Pada penelitian ini, data yang digunakan adalah data produksi padi di Jawa Timur tahun 2022 yang diperoleh dari website BPS Jawa Timur. Total ada 190 data yang terdiri dari beberapa atribut yaitu : nama kota/kabupaten, luas lahan, jumlah produksi padi dan nilai produktivitas. Sebelum proses data mining dimulai, data diproses terlebih dahulu untuk diberi label dalam 2 kelas, yaitu : tidak sesuai target (0) dan sesuai target (1). Data diproses menggunakan salah satu algoritma data mining yaitu *Support Vector Machine (SVM)*. Proses permodelan SVM menggunakan platform Jupyter Notebook. Berikut visualisasi bentuk persebaran data sebagaimana di gambar 3.



Gambar 3. Visualisasi Bentuk Persebaran Data

Dalam proses penerapan algoritma SVM, langkah pertama adalah membagi dataset menjadi 2, yaitu data training dan data testing. Data dibagi 2 dengan bobot nilai sebesar 80:20 dimana 80% data akan digunakan sebagai data training dan 20% data akan digunakan sebagai data testing. Setelah pembagian data dilakukan, data pun siap untuk diproses menggunakan algoritma SVM sehingga mendapatkan hasil klasifikasi.

Pada tahap evaluasi, hasil kinerja Metode SVM dievaluasi dengan menggunakan *Confusion Matrix*. Pada tahapan ini akan dilakukan pembuatan *confusion matrix* dengan menggunakan library pendukung dari ScikitLearn. Berikut hasil visualisasi confusion matrix sebagaimana di gambar 4.



Gambar 4. Visualisasi Confusion Matrix

Pada gambar 4 diketahui, nilai True Positive (TP) = 4 dan False Positive (FP) = 1. Sedangkan False Negative (FN) = 0 dan True Negative (TN) = 3. Class 0 mewakili “tidak sesuai target” dan class 1 mewakili “sesuai target” dimana sumbu x mewakili “actual values” dan sumbu y mewakili “predicted values”.

**TP** : Dimana status aslinya adalah “tidak sesuai target” (class 0), dalam model ini dapat dinyatakan bahwa status tersebut memang benar — benar “tidak sesuai target”(class 0). berhasil memprediksi sebanyak 4.

**TF** : Dimana status aslinya adalah “sesuai target” (class 1), dalam model ini dapat dinyatakan bahwa status tersebut memang benar-benar “sesuai target”(class 1). berhasil memprediksi sebanyak 1.

**FP** : Status aslinya adalah “sesuai target” (Class 1), dalam model ini dapat dinyatakan bahwa status tersebut adalah “tidak sesuai target” (class 0). dengan nilai prediksi 0.

**FN** : Status aslinya adalah “tidak sesuai target” (class 0), dalam model ini dapat dinyatakan bahwa status tersebut adalah “sesuai target” (class 1) dengan nilai prediksi 3.

Tahap terakhir adalah melakukan validasi dan evaluasi terhadap hasil pengujian. Pada tahapan ini akan didapatkan hasil akhir dari kemampuan algoritma dalam klasifikasi dataset produksi padi di Jawa Timur. Hasil yang akan didapatkan adalah berupa *accuracy score*, *recall*, dan *precision*. Berikut hasil dari validasi menggunakan platform Jupyter Notebook sebagaimana di gambar 5.

	precision	recall	f1-score	support
0	1.00	0.80	0.89	5
1	0.75	1.00	0.86	3
accuracy			0.88	8
macro avg	0.88	0.90	0.87	8
weighted avg	0.91	0.88	0.88	8

Gambar 5. Hasil Validasi

Precision adalah akurasi prediksi positif. Recall adalah fraksi positif yang diidentifikasi dengan benar. Skor F1 adalah rata-rata presisi dan recall harmonik tertimbang sehingga skor terbaik adalah 1.0 dan yang terburuk adalah 0.0. Support adalah jumlah kemunculan aktual kelas dalam kumpulan data yang ditentukan. Sehingga hasil validasi sebagai berikut di tabel 6.

Tabel 6. Hasil Validasi

	Akurasi	Presisi	Recall
Tidak Sesuai Target (0)	0,89	1,00	0,80
Sesuai Target (1)	0,86	0,75	1,00

Dari data produktivitas padi seluruh kota/kabupaten yang ada di Jawa Timur, dibuat pembagian data training sebanyak 80% dan data testing 20% mendapatkan akurasi 80% pada saat memprediksi kategori ‘0’

(tidak sesuai target), dan mendapat akurasi 86% pada saat memprediksi kategori '1' (sesuai target). Dan secara keseluruhan, classification model memiliki akurasi 88%.

#### 4. KESIMPULAN

Berdasarkan dari hasil penelitian maka dapat diambil kesimpulan bahwa penerapan metode *Support Vector Machine* untuk klasifikasi produktivitas padi di Jawa Timur bisa bekerja secara optimal dan baik. Ini dibuktikan dengan pembagian data training sebanyak 80% dan data testing 20% mendapatkan akurasi 89% pada saat memprediksi kategori '0' (tidak sesuai target), dan mendapat akurasi 86% pada saat memprediksi kategori '1' (sesuai target). Dan secara keseluruhan, classification model memiliki akurasi 88%.

#### REFERENCES

- [1] I. Febriani, M. Safii and O. Alfina, "Implementasi Data Mining Peningkatan Produksi Beras Menggunakan Metode K-Means Clustering," *Majalah Ilmiah Methoda*, vol. 12, no. 3, 2022.
- [2] A. Setia Budi, P. Hadi Susilo, and N. Nafi'iyah, "SVM ALGORITHM FOR PREDICTING RICE YIELDS," *Jurnal Teknologi Informasi dan Pendidikan*, vol. 13, no. 2, 2020.
- [3] S. Rokhmah, A. Susilowati, and M. Intan, "Klasifikasi Data untuk Prediksi Hasil Panen Tanaman Padi di Wilayah Kabupaten Sukoharjo Menggunakan Algoritma C4.5," *JURTI*, vol. 6, no. 2, 2022.
- [4] S. Wijayanto and M. Y. Fathoni, "Pengelompokan Produktivitas Tanaman Padi di Jawa Tengah Menggunakan Metode Clustering K-Means," *Jurnal Penelitian Ilmu dan Teknologi Komputer*, vol. 13, no. 2, 2021.
- [5] B. Satria, E. M. Harahap and Jamilah, "Peningkatan Produktivitas Padi Sawah (*Oryza sativa* L.) Melalui Penerapan Beberapa Jarak Tanam dan Sistem Tanam," *Jurnal Agroekoteknologi FP USU*, vol. 5, no. 3, 2017.
- [6] S. M. S. Sianturi and N. A. Hasibuan, "Analisa Data Pertanian Tanaman Pangan untuk Memprediksi Hasil Panen dengan Data Mining Algoritma C.45 (Studi Kasus : Dinas Tanaman Pangan dan Holtikultura Provinsi Sumut)," *Jurnal Pelita Informatika*, vol. 7, no. 4, 2019.
- [7] A. Desiani, Irmeilyana, H. Hanum, Y. Andriani, S. I. Maiyanti, C. M. Uteh and I. Rayani, "Penerapan Metode Support Vector Machine Dalam Klasifikasi Bunga Iris," *Indonesian Journal of Applied Informatics*, vol. 7, no. 1, 2022.
- [8] A. Zalvadila, L. Syafie, and H. Darwis, "Klasifikasi Penyakit Tanaman Bawang Merah Menggunakan Metode SVM dan CNN," *Jurnal Pengembangan IT*, vol. 8, no. 3, 2023.
- [9] E. Rasmikayati and D. A. Faisal, "DINAMIKA PRODUKTIVITAS PADI DITINJAU DARI FLUKTUASI SUSUT HASIL SERTA FAKTOR SOSIAL, EKONOMI DAN BUDAYA YANG MEMPENGARUHINYA," *Agribisnis dan Sosial Ekonomi Pertanian*, vol. 1, no. 2, 2016.
- [10] E. Haryatmi and S. P. Hervianti, "Penerapan Algoritma Support Vector Machine untuk Model Prediksi Kelulusan Mahasiswa Tepat Waktu," *JURNAL RESTI (Rekayasa Sistem dan Teknologi Informasi)*, vol. 5, no. 2, 2021.
- [11] R. R. Fiska, "Penerapan Teknik Data Mining dengan Metode Support Vector Machine (SVM) untuk Memprediksi Siswa yang Berpeluang Drop Out (Studi Kasus di SMKN 1 Sutera)," *SATIN-Sains dan Teknologi Informasi*, vol. 3, no. 1, 2017.
- [12] M. F. Akbarollah, W. Wiyanto, D. Ardiatma, and A. T. Zy, "Penerapan Algoritma K-Nearest Neighbor Dalam Klasifikasi Penyakit Jantung," *Journal of Computer System and Informatics (JoSYC)*, vol. 4, no. 4, 2023, doi: 10.47065/josyc.v4i4.4071.
- [13] D. E. Safitri and A. S. Fitriani, "IMPLEMENTASI METODE KLASIFIKASI DENGAN ALGORITMA SUPPORT VECTOR MACHINE KERNEL GAUSSIAN RBF UNTUK PREDIKSI PARTISIPASI PEMILU TERHADAP DEMOGRAFI KOTA SURABAYA," *Indonesian Journal of Business Intelligence (IJUBI)*, vol. 5, no. 1, 2022, doi: 10.21927/ijubi.v5i1.2259.
- [14] D. Vernanda, N. N. Purnawan, T. H. Apandi, and Haryati, "ANALISIS DATA UNTUK KLASIFIKASI TINGKAT KEMATANGAN BUAH NANAS MENGGUNAKAN SVM," *Jurnal Ilmiah Ilmu dan Teknologi Rekayasa*, vol. 4, 2022, doi: 10.31962/jiitr.vvii.67.
- [15] Y. Amrozi, D. Yuliati, A. Susilo, N. Novianto, and R. Ramadhan, "Klasifikasi Jenis Buah Pisang Berdasarkan Citra Warna dengan Metode SVM," *Jurnal Sisfokom (Sistem Informasi dan Komputer)*, vol. 11, no. 3, 2022, doi: 10.32736/sisfokom.v11i3.1502.
- [16] F. R. Lumbanraja, F. Lufiana, Y. Heningtyas and K. Muludi, "Implementasi Support Vector Machine (SVM) untuk Klasifikasi Penderita Diabetes Mellitus," *Jurnal Komputasi*, vol. 10, no. 1, 2022.
- [17] I. M. Parapat, M. T. Furqon and Sutrisno, "Penerapan Metode Support Vector Machine (SVM) Pada Klasifikasi Penyimpangan Tumbuh Kembang Anak," *Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer*, vol. 2, no. 10, 2018.

- [18] A. Handayanto, K. Latifa, N. D. Saputro, and R. R. Waliyansyah, “Analisis dan Penerapan Algoritma Support Vector Machine (SVM) dalam Data Mining untuk Menunjang Strategi Promosi,” *JUITA : Jurnal Informatika*, vol. 7, no. 2, 2019.
- [19] E. Haryatmi and S. Pramita Hervianti, “Penerapan Algoritma Support Vector Machine Untuk Model Prediksi Kelulusan Mahasiswa Tepat Waktu,” *Jurnal RESTI (Rekayasa Sistem dan Teknologi Informasi)*, vol. 5, no. 2, 2021, doi: 10.29207/resti.v5i2.3007.
- [20] A. Darmawan, N. Kustian dan W. Rahayu, “Implementasi Data Mining Menggunakan Model SVM untuk Prediksi Kepuasan Pengunjung Taman Tabebuya,” *Jurnal String*, vol. 2, no. 3, 2018.