

## IMPLEMENTASI ALGORITMA NAIVE BAYES UNTUK MEMPREDIKSI HASIL PANEN PADI

Abdul Basit<sup>1</sup>

<sup>1</sup> Universitas Muhammadiyah Sukabumi,  
Jl. R. Syamsudin, SH. No. 50 Kota Sukabumi, Jawa Barat  
Email : irhamagung2016@gmail.com,

### ABSTRACT

*Rice (Oryza Sativa) is one of the important plants in Indonesia. Rice is produced from rice plants, which is a staple food for the people of Indonesia. One of them is rice producer, Pelabuhan Ratu Distric, the result harvest in Pelabuhan ratu every year experiencing ups and downs that are not stable. As an effort to prevent extreme price increases the need for accurate and accurate predictions about the result harvest in Kecamatan Pelabuhan ratu. lack of predictive advice the result rice harvest resulting in lack of information needed to increase rice production in Kecamatan Pelabuhan ratu. Data mining is one of implementation which is applied to find a model and patterns that are able to make predictions on a data based on previous data in a certain period of time. One algorithm used in data mining techniques for classification is naive bayes. Naive Bayes algorithm is Prediction techniques based on simple possibilities on applying Bayes rules with strong independence.*

**Keywords:** Prediction, Harvest Result, Naive Bayes

### ABSTRAK

Padi (Oryza Sativa) merupakan salah satu tanaman yang penting di Indonesia. Dari tanaman padi dihasilkan beras, yang merupakan bahan makanan pokok bagi rakyat Indonesia. Salah satunya penghasil padi yaitu Kecamatan Pelabuhan ratu, Hasil panen di Kecamatan Pelabuhan ratu setiap tahunnya mengalami naik turun yang tidak stabil, sebagai upaya mengatasi agar tidak terjadi kenaikan harga yang ekstrim perlu adanya ramalan yang tepat dan akurat tentang hasil panen di Kecamatan Pelabuhan ratu. Kurangnya saran prediksi hasil panen padi mengakibatkan kurangnya informasi yang dibutuhkan untuk meningkatkan produksi padi di Kecamatan Pelabuhan ratu. Data mining adalah salah satu bentuk implementasi yang diterapkan untuk mencari sebuah model dan pola yang mampu melakukan prediksi pada suatu data berdasarkan data sebelumnya di periode waktu tertentu. Salah satu algoritma yang digunakan dalam teknik data mining untuk klasifikasi adalah Naive Bayes. Teorema Bayes merupakan teknik prediksi berdasarkan kemungkinan sederhana pada penerapan aturan Bayes dengan ketidaktergantungan yang kuat.

**Kata Kunci:** Prediksi, Hasil Panen, Naive Bayes

### 1. PENDAHULUAN

Padi (Oryza Sativa) merupakan salah satu tanaman yang penting di Indonesia. Dari tanaman padi dihasilkan beras, yang merupakan bahan makanan pokok bagi rakyat Indonesia. Dimana padi merupakan komoditas utama dibagian pertanian di

setip desa. Padi dapat tumbuh dengan baik didaerah panas dengan curah hujan yang tinggi. Produksi padi di Indonesia mengambil pangsa sekitar (9%) dari total produksi dunia. Indonesia negara penghasil beras ke tiga terbesar di dunia, setelah China dan India.

Produksi padi di Indonesiatahun 2014 sebanyak 70,85 juta ton gabah kering giling (GKG)atau mengalami penurunan sebanyak 0,43 juta ton (0,61 persen) dibandingkan tahun 2013.

Hasil panen di Kecamatan Pelabuhan ratu setiap tahunnya mengalami naik turun yang tidak stabil, sebagai upaya mengatasi agar tidak terjadi kenaikan harga yang ekstrim perlu adanya ramalan yang tepat dan akurat tentang hasil panen di Kecamatan Pelabuhan ratu. Kurangnya saran prediksi hasil panen padi mengakibatkan kurangnya informasi yang dibutuhkan untuk meningkatkan produksi padi di Kecamatan Pelabuhan ratu.

Pada penelitian ini dibuat sebuah aplikasi prediksi, yaitu sebuah aplikasi yang didesain untuk memprediksi hasil panen padi di tahun yang akan datang di Kecamatan Pelabuhan ratu. Ada bermacam-macam metode yang dapat digunakan untuk menghitung prediksi seperti metode Trend Least Square, GARCH, Moving Average, Regresi Linier dan Nonlinier.

Data mining adalah salah satu bentuk implementasi yang diterapkan untuk mencari sebuah model dan pola yang mampu melakukan prediksi pada suatu data berdasarkan data sebelumnya di periode waktu tertentu. Data mining adalah bentuk penggalian data yang digunakan untuk menggali pengetahuan dari jumlah data yang besar.

Salah satu algoritma yang digunakan dalam teknik data mining yang memakai teori Bayes untuk klasifikasi adalah Naïve Bayes. Teorema Bayes merupakan teknik prediksi berdasarkan kemungkinan sederhana pada penerapan aturan Bayes dengan ketidaktergantungan yang kuat. Naïve Bayes banyak digunakan untuk proses klasifikasi karena Naïve Bayes lebih disukai disebabkan kecepatan dan kesederhanaannya.

## 2. METODOLOGI PENITIAN

### 2.2 Pengertian 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 (Herdianto, 2016)

Prediksi atau permalan adalah proses perkiraan (pengukuran) besarnya atau jumlah sesuatu pada waktu ayang akan datang berdasarkan data pada masa lampau yang dianalisis secara ilmiah khususnya menggunakan metode statistika(Sudjana, 2017).

Prediksi adalah perhitungan yang objektif dan dengan menggunakan data-data masa lalu, untuk menentukan sesuatu di masa yang akan datang(Sumayang, 2017).

### 2.2 Tanaman Padi

Padi tergolong dalam famili *Gramineae* (rumput-rumputa). Padi dapat beradaptasi pada lingkungan aerob dan nonaerob. Batang padi berbulu dan berongga, dari buku batang inilah tumbuh anakan atau daun. Akar padi adalah akar serabut yang sangat sensitif dalam penyerapan har, tetapi peka terhadap kekeringan. Biji padi mengandung butiran pati *amilosa* dan *amilopektin* yang mempengaruhi mutu dan rasa nasi (Septiana, 2016)

Padi merupakan komoditas tanaman pangan penghasil beras yang memegang peranan penting dalam kehidupan ekonomi Indonesia. Yaitu beras sebagai makanan pokok sangat sulit digantikan oleh bahan pokok lainnya. Diantaranya jagung, umbi-umbian, sagu dan sumber karbohidrat lainnya. Sehingga keberadaan beras menjadi prioritas utama masyarakat dalam memenuhi kebutuhan asupan karbohidrat yang dapat mengenyangkan dan merupakan sumber

karbohidrat utama yang mudah diubah menjadi energi. Padi sebagai tanaman pangan dikonsumsi kurang lebih 90% dari keseluruhan penduduk Indonesia untuk makanan pokok sehari-hari (Usman, 2017).

Tanaman padi (*Oryza sativa* L.) merupakan tanaman pangan yang memiliki arti penting bagi hampir seluruh penduduk Indonesia karena beras mampu memenuhi kebutuhan kalori sebagian besar penduduk Indonesia. Dewasa ini dengan bertambahnya penduduk Indonesia produktivitas tanaman padi di Indonesia masih belum seperti yang diharapkan. Berdasarkan data yang ada, rata-rata produktivitas padi gogo baru sebesar 2,95 ton/ha sementara untuk padi sawah sebesar 5,08 ton/ha (Amrullah, 2017).

### 2.3 Algoritma Naive Bayes

*Naive Bayes Classifier* merupakan sebuah metode klasifikasi yang berakar pada teorema Bayes. Ciri utama dari *Naive Bayes Classifier* ini adalah asumsi yang sangat kuat (naïf) akan independensi dari masing-masing kondisi / kejadian (Wibawa, 2018)

Kelebihan dari metode ini adalah :

1. Menangani kuantitatif dan data diskrit.
2. Kokoh untuk titik noise yang diisolasi, misalkan titik yang dirata-ratakan ketika mengestimasi peluang bersyarat data.
3. Hanya memerlukan sejumlah kecil data pelatihan untuk mengestimasi parameter (rata-rata, variasi, dan variabel) yang dibutuhkan untuk klasifikasi.
4. Menangani nilai yang hilang dengan mengabaikan instansi selama perhitungan estimasi peluang.
5. Cepat dan efisien ruang.
6. Kokoh terhadap atribut yang tidak relevan

Persamaan dari teorema Bayes adalah :

$$P(H|X) = \frac{P(X|H).P(H)}{P(X)} \quad (1)$$

Di mana :

X : Data dengan class yang belum diketahui

H : Hipotesis data merupakan suatu class spesifik

P(H|X) : Probabilitas hipotesis H berdasar kondisi X (posteriori probabilitas)

P(H) : Probabilitas hipotesis H (prior probabilitas)

P(X|H) : Probabilitas X berdasarkan kondisi pada hipotesis H

P(X) : Probabilitas X

Untuk menjelaskan metode *Naive Bayes*, perlu diketahui bahwa proses klasifikasi memerlukan sejumlah petunjuk untuk menentukan kelas apa yang cocok bagi sampel yang dianalisis tersebut. Karena itu, metode *Naive Bayes* di atas disesuaikan sebagai berikut:

$$P(C|F_1 \dots F_n) = \frac{P(C)P(F_1 \dots F_n|C)}{P(F_1 \dots F_n)} \quad (2)$$

Di mana Variabel C merepresentasikan kelas, sementara variabel  $F_1 \dots F_n$  merepresentasikan karakteristik petunjuk yang dibutuhkan untuk melakukan klasifikasi. Maka rumus tersebut menjelaskan bahwa peluang masuknya sampel karakteristik tertentu dalam kelas C (Posterior) adalah peluang munculnya kelas C (sebelum masuknya sampel tersebut, seringkali disebut prior), dikali dengan peluang kemunculan karakteristik-karakteristik sampel pada kelas C (disebut juga likelihood), dibagi dengan peluang kemunculan karakteristik-karakteristik sampel secara global (disebut juga evidence). Karena itu, rumus di atas dapat pula ditulis secara sederhana sebagai berikut:

$$Posterior = \frac{Prior \times likelihood}{evidence} \quad (3)$$

Nilai Evidence selalu tetap untuk setiap kelas pada satu sampel. Nilai dari posterior tersebut nantinya akan dibandingkan dengan nilai-nilai posterior

kelas lainnya untuk menentukan ke kelas apa suatu sampel akan diklasifikasikan.

### 3. HASIL DAN PEMBAHASAN

#### 3.1 Dataset

Dataset yang digunakan berasal dari data sekunder yang didapatkan di Kec. Palabuhan ratu dengan beberapa parameter yang digunakana. Adapun dataset yang digunakan dalam penelitian ini adalah sebagai berikut:

**Tabel 3.1** Dataset

No	Tahun	Musim			Luas Sawah (Ha)	Total (Ton)	Hasil
		Pertama	Kedua	Ketiga			
1	1	1289	1290	1293	1.391	3872	Meningkat
2	2	1285	1287	1285	1385	3857	Meningkat
3	3	1280	1281	1282	1380	3843	Meningkat
4	4	1278	1279	1280	1375	3837	Menurun
5	5	1274	1275	1277	1372	3826	Meningkat
6	6	1270	1272	1273	1370	3815	Menurun
7	7	1268	1269	1271	1365	3808	Meningkat
8	8	1264	1265	1267	1363	3796	Meningkat
9	9	1260	1263	1262	1360	3785	Menurun
10	10	1257	1258	1254	1358	3769	Menurun

#### 3.2 Cleaning data

Pada proses ini dilakukan perbersihan, pengkoreksian atau menghapus data yang salah atau tidak lengkap. Dalam kasus ini dataset yang diolah bebas dari proses *cleaning*.

**Tabel 3.2** Cleaning data

No	Tahun	Musim			Luas Sawah (Ha)	Total (Ton)	Status
		Pertama	Kedua	Ketiga			

			a	a			
1	1	1289	1290	1293	1.391	3872	Meningkat
2	2	1285	1287	1285	1385	3857	Meningkat
3	3	1280	1281	1282	1380	3843	Meningkat
4	4	1278	1279	1280	1375	3837	Menurun
5	5	1274	1275	1277	1372	3826	Meningkat
6	6	1270	1272	1273	1370	3815	Menurun
7	7	1268	1269	1271	1365	3808	Meningkat
8	8	1264	1265	1267	1363	3796	Meningkat
9	9	1260	1263	1262	1360	3785	Menurun
10	10	1257	1258	1254	1358	3769	Menurun

#### 3.3 Transformasi Data

Pada proses ini dilakukan pengubahan data untuk mengelompokan data yang beragam menjadi beberapa kategori. Adapun transformasi data yang digunakan pada penelitian ini adalah sebagai berikut:

**Tabel 3.3** Transformasi Data

Atribut/Variabel		Jumlah Data	Hasil	
Total		25	Meningkat	Menurun
Musim	Pertama	25	20	5
	Kedua	25	22	3
	Ketiga	25	23	2
Luas Sawah	Besar	17	10	6
	Sedang	18	12	6
	Kecil	20	17	3

#### 3.4 Data Testing

Berdasarkan tabel 4.1, terdapat data yang kemudian nantinya hanya akan digunakan 25 data untuk *data training* dan akan dilakukan proses klasifikasi terhadap data baru yang akan di *testing*.

#### 3.5 Modelling Naive Bayes

Adapun beberapa tahapan proses perhitungan yang dilakukan dalam penelitian ini:

1. Menghitung jumlah atribut/variable Setelah mendapatkan jumlah *data*

training dan data testing, maka selanjutnya adalah menghitung probabilitas prior dan probabilitas posterior, dalam bentuk persamaan di bawah ini:

$$\begin{aligned} \text{Total data training} &= 25 \\ \text{Data Meningkat} &= 17 \\ \text{Data Menurun} &= 8 \\ P(\text{Meningkat}) &= 17/25 = 0,68 \\ P(\text{Menurun}) &= 8/25 = 0,32 \end{aligned}$$

- Menghitung jumlah kasus dan atribut/variabel yang sama Selanjutnya untuk mengetahui hasil perhitungan dari probabilitas prior, maka dilakukan perhitungan dengan cara merinci jumlah kasus dari tiap-tiap atribut variabel data, adapun hasil perhitungan probabilitas prior dengan menggunakan algoritma naive bayes dapat dilihat pada tabel berikut:

**Tabel 3.4** Perhitungan Probabilitas Prior

Atribut/Variabel	Jumlah Data	Hasil		P(X/Ci)		
		Meningkat	Menurun	Meningkat	Menurun	
Total	25	17	8	0,68	0,32	
Musim	Pertama	25	20	5	0,8	0,2
	Kedua	25	23	2	0,92	0,08
	Ketiga	25	22	3	0,88	0,12
Luas Sawah	Besar	15	10	5	0,667	0,333
	Sedang	18	12	6	0,667	0,333
	Kecil	20	17	3	0,85	0,15

- Menghitung semua hasil variabel dari setiap klasifikasi Setelah mengetahui nilai probabilitas dari setiap atribut/variable terhadap probabilitas tiap kelas atau yang dirumuskan dalam bentuk persamaan  $P(X/Ci)$ , maka langkah berikutnya adalah melakukan perhitungan terhadap total keseluruhan probabilitas setiap kelas. Berikut

persamaan untuk menghitung probabilitas setiap kelas:

$$P(X|\text{Hasil} = \text{Meningkat}) = 0,8 * 0,92 * 0,88 * 0,667 * 0,667 * 0,85 = 0,2449238514$$

$$P(X|\text{Hasil} = \text{Menurun}) = 0,2 * 0,08 * 0,12 * 0,333 * 0,333 * 0,15 = 0,000031936$$

4. Menghitung  $P(X|H) * P(H)$ , maka:

$$P(X|\text{Hasil} = \text{Meningkat}) = 0,2449238514 * 0,68 = 1,66548219$$

$$P(X|\text{Hasil} = \text{Menurun}) = 0,000031936 * 0,32 = 0,0000102195$$

5. Hasil Akurasi

Berdasarkan hasil perhitungan pemodelan naive bayes diatas, diketahui bahwa nilai  $P(X|\text{Meningkat})$  lebih besar daripada nilai  $P(X|\text{Menurun})$ , sehingga dapat diambil kesimpulan tersebut akan masuk kedalam klasifikasi Meningkat dan tidak termasuk kedalam klasifikasi Menurun.

#### 4. KESIMPULAN

Pembuatan aplikasi ini untuk membantu mengetahui gambaran prediksi pada hasil panen padi pada tahun yang akan datang di Kecamatan Pelabuhan ratu. Sistem yang telah mengacu pada rumusan masalah yang ada yaitu suatu sistem yang dapat memprediksi hasil panen padi menggunakan algoritma naive bayes pada tahun yang akan datang meningkat atau menurun. Kesimpulan yang dapat diambil yaitu :

- Penggunaan algoritma naive bayes untuk prediksi hasil panen padi berdasarkan data-data hasil panen padi sebelumnya dan luas lahan sawah. Adapun beberapa kelas-kelas yang digunakan antara lain data-data hasil panen padi permusim, luas lahan sawah serta total dari jumlah permusim.
- Bisa mengetahui gambaran hasil panen padi pada tahun yang akan datang dan agar tidak terjadi kenaikan harga yang ekstrem.

3. Sistem ini hanya sebagai alat bantu untuk memprediksi hasil panen padi di Kecamatan Pelabuhan ratu.

of Rice (*Oryza sativa* L.) under Different Jajar Legowo System and Planting Space. *J. Agroland*, 24(1), 27–35.

#### DAFTAR PUSTAKA

- [1] Amrullah. (2017). Peningkatan produktivitas tanaman padi ( *Oryza sativa* L.) melalui pemberian nano silika. *Jurnal Pangan*, 23(1), 17–32.
- [2] Herdianto. (2016). *Prediksi*. Eprint. [http://eprints.umg.ac.id/1566/3/BAB II.pdf](http://eprints.umg.ac.id/1566/3/BAB%20II.pdf)
- [3] Jannah, Q. W. (2015). *MAKALAH BOTANI TUMBUHAN BEPEMBULUH INTEGRASI AL-QUR'AN DENGAN TUMBUHAN PADI (Oriza sativa)*. Tangan-Tangan Kreatif. <http://qonitawarda.blogspot.com/2015/06/integrasi-ayat-al-quran-dengan-tumbuhan.html>
- [4] Septiana, A. F. (2016). *Kajian Adopsi Inovasi*.
- [5] Sudjana. (2017). *prediksi*. Seputarpengertian. <https://www.seputarpengertian.co.id/2017/11/pengertian-peramalan-forecasting-menurut-para-ahli.html>
- [6] Sugeng Priyadi Abu Abdilbarr. (2013). *Hadits-hadits Anjuran Bercocok Tanam (Bagian I)*. Petanirumahan. <https://petanirumahan.com/2013/07/11/hadits-hadits-anjuran-bercocok-tanam-bagian-i/>
- [7] Sumayang. (2017). *prediksi*. Seputarpengertian. <https://www.seputarpengertian.co.id/2017/11/pengertian-peramalan-forecasting-menurut-para-ahli.html>
- [8] Usman. (2017). PERTUMBUHAN DAN HASIL TANAMAN PADI (*Oryza sativa* L) PADA BERBAGAI POLA JAJAR LEGOWO DAN JARAK TANAM Growth and Yield
- [9] Wibawa. (2018). Perbandingan Kinerja Metode Naive Bayes dan K-Nearest Neighbor untuk Klasifikasi Artikel Berbahasa Indonesia. *Jurnal Teknologi Informasi Dan Ilmu Komputer*, 5(4), 427. <https://doi.org/10.25126/jtiik.201854773>