

Penerapan Data Mining Untuk Memprediksi Harga Bawang Merah Di Yogyakarta Menggunakan Metode K-Nearest Neighbor

Implementation Of Data Mining To Predict The Price Of Shallots In Yogyakarta Using The K-Nearest Neighbor Method

Dapit Virdaus¹, Putri Taqwa Prasetyaningrum²

¹Program Studi Sistem Informasi, Fakultas Teknologi Informasi, Universitas Mercu Buana Yogyakarta, Jl.
Wates Km. 10 Yogyakarta 55753, Indonesia
Email: dapitvirdauss@gmail.com¹, putri@mercubuana-yogya.ac.id.com²

ABSTRAK

Bawang merah merupakan salah satu sayuran musiman dengan hasil panen terbesar. Namun, harganya terkadang tidak stabil, bisa sangat murah atau sangat mahal. Kondisi ini menyulitkan mereka yang membutuhkan bawang merah, dan juga membuat petani merugi karena harga yang tidak stabil. Keadaan yang membuat petani bawang merah merugi merupakan masalah serius, mengingat bawang merah merupakan tanaman atau bumbu yang hampir wajib ada di setiap hidangan. Oleh karena itu, diperlukan suatu penelitian untuk mengatasi masalah tersebut, maka penelitian yang mengimplementasikan data mining dengan menggunakan metode *algoritma K-Nearest Neighbor* akan diterapkan untuk memprediksi harga bawang merah menggunakan data yang telah tersedia sebelumnya. Pada penelitian ini dilakukan empat skenario atau kondisi yang dijadikan sasaran data bawang merah dalam pengolahan datanya agar diperoleh hasil yang lebih variatif dan lebih presisi dengan tingkat akurasi prediksi tertinggi. Setelah melakukan berbagai macam percobaan pada data yang diadakan dalam penelitian ini, atau lebih tepatnya penerapan empat kondisi pada data, diperoleh hasil akurasi tertinggi dengan nilai 91,67%. Variabel yang paling berpengaruh terhadap panen adalah faktor cuaca atau curah hujan, namun untuk faktor harga pada suatu waktu atau pada kondisi dimana faktor produksi yang besar tidak menjamin harga akan lebih murah atau stabil. Dalam hal ini terdapat campur tangan dalam urusan politik dan ekonomi yang tidak dapat diakses oleh penulis sebagai peneliti.

Kata kunci: *data mining, k-nearest neighbor, bawang merah, prediksi.*

ABSTRACT

Shallots are one of the seasonal vegetables with the largest yield. However, the price is sometimes unstable, it can be very cheap or extremely expensive. This condition causes difficulty for those who need shallots, and it also makes farmers lose money because of unstable prices. The situation that makes shallot farmers lose profits is a serious problem, considering that onion is a plant or spice that is almost mandatory in every dish. Therefore, a research is needed in order to solve this problem, a research which implements data mining using the K-Nearest Neighbor algorithm method will be applied in order to predict the price of shallots using previously available data. In this research, four scenarios or conditions were subjected to onion data in the data processing in order to obtain more varied and more precise results with the highest level of prediction accuracy. After carrying out various kinds of experiments on the data held in this study, or more precisely the application of four conditions to the data, the highest accuracy results were obtained with a value of 91.67%. The variable that has the most influence on harvest is weather or rainfall factor, but for the price factor at some time or in conditions where large production factors do not guarantee prices will be cheaper or stable. In this case there is interference in political and economic affairs that the writer as the researcher is unable to access.

Keywords: *data mining, k-nearest neighbor, shallots, prediction*

1. PENDAHULUAN

PENERAPAN DATA MINING UNTUK MEMPREDIKSI HARGA BAWANG MERAH DI YOGYAKARTA MENGGUNAKAN METODE K-NEAREST NEIGHBOR

Indonesia merupakan negara dengan kekayaan akan sumber daya alamnya, dimana sektor pertanian merupakan salah satu daya dorong ekonomi Indonesia. Pada survei pertanian hortikultura tanaman sayuran dan buah-buahan semusim mencakup dua puluh dua jenis tanaman sayuran semusim dan empat jenis tanaman buah-buahan semusim (Hortikultura 2018).

Hasil panen bawang merah juga terpengaruh oleh musim berdasarkan curah hujan yang terjadi, sehingga apabila curah hujan yang tinggi dapat merusak bawang merah. Bahkan harga bawang merah di sejumlah pasar tradisional Yogyakarta mengalami kenaikan yang cukup signifikan pada bulan Juli 2020 saja mencapai Rp 40.000 hingga Rp 50.000 perkilogram (Putri 2020).

Pada penelitian ini pengimplementasikan algoritma yang digunakan pada *data mining* menggunakan algoritma *K-Nearest-Neighbor*, yang mana alasan pemilihan algoritma *K-Nearest-Neighbor* adalah karena algoritma ini memiliki cara hitung yang sederhana dan simpel namun memiliki tingkat akurasi yang tinggi dan termasuk salah satu dari sepuluh algoritma atau metode paling populer pada perhitungan data mining, serta juga tangguh terhadap data *training sample* yang *noisy*.

Berdasarkan permasalahan tersebut, penulis tertarik untuk mengambil judul penelitian dengan judul “Penerapan *Data Mining* Untuk Memprediksi Harga Bawang Merah di Yogyakarta Menggunakan Metode *K-Nearest-Neighbor*”. Dimana penerapan *data mining* diharapkan dapat mengatasi permasalahan yang terjadi dengan memanfaatkan data-data terdahulu guna dapat mencari informasi dan rakangkaian pola sehingga dapat memecahkan masalah yang ada.

2. TINJAUAN PUSTAKA

Dalam penelitian Fatah (2016) yang berjudul “Prediksi Harga *Cryptocurrency* Dengan Metode *K-Nearest Neighbours*”, penelitian ini membahas tentang uang elektronik menjadi pilihan yang mulai ramai digunakan oleh banyak orang. Hasil dari penelitian ini

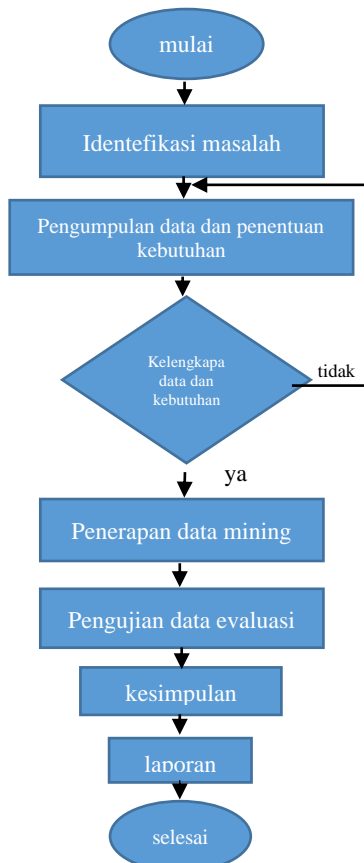
diketahui bahwa model KNN yang paling baik dalam memprediksi harga *Cryptocurrency* adalah KNN dengan parameter nilai $K=3$ dan *Nearest Neighbour Search Algorithm* : *Linear NN Search*. Dengan nilai *Mean Absolute Error* (MAE) sebesar 0.0018 dan *Root Mean Squared Error* (RMSEP) sebesar 0.0089.

Pada penelitian Sinta (2015) yang berjudul “Metode Ensemble *K-Nearest-Neighbor* Untuk Prediksi Harga Beras di Indonesia”, yang mana membahas tentang kenaikan harga beras di Indonesia. Hasil prediksi harga beras di Indonesia menunjukkan bahwa metode ensemble KNN memiliki kinerja yang lebih baik dibandingkan dengan metode KNN tunggal. Nilai-nilai tersebut semakin kecil jika nilai K yang dicobakan semakin besar, namun jika nilai K yang dicobakan sangat besar atau mendekati ukuran data *training* maka ketiga nilai tersebut memberikan hasil yang besar. Kisaran nilai prediksi harga beras hampir sama dengan harga beras sebenarnya.

Pada penelitian Muliono (2019) yang berjudul “Analisis Algoritma *K-Nearest-Neighbors* dalam Prediksi Waktu Kelulusan Mahasiswa”. Penelitian ini bertujuan untuk melihat pola atau kategori mahasiswa seperti apa yang akan lulus tepat waktu sehingga dapat membuat akreditasi kampus tetap baik. Hasil penelitian ini akan membandingkan data real dan hasil tingkat akurasi algoritma yang akan disajikan dalam *ROC Curve*. Pada penelitian ini terdapat 5 tingkat akurasi yang digunakan untuk melakukan evaluasi dan validasi terhadap data *training*, dimulai dari akurasi $0.90 - 1.00 = \text{Excellent classification}$ sampai akurasi $0.50 - 0.60 = \text{Failure}$.

3. METODOLOGI PENELITIAN

Pada penelitian ini data yang digunakan merupakan data bawang merah dari tahun 2014-2018 yang meliputi luas panen, curah hujan, jumlah produksi dan harga jual. Serta metode yang digunakan pada penelitian ini adalah algoritma *K-Nearest-Neighbors*. Penelitian akan melalui beberapa tahapan atau proses, berikut merupakan flowchart jalan penelitian.



Gambar 3.1 Jalan Penelitian

Pengumpulan Data

Proses pengumpulan data atau informasi yang akan digunakan sebagai data uji dalam data mining sehingga menghasilkan informasi baru dari olah data menggunakan berbagai macam algoritma data mining. Pengumpulan data dapat menggunakan teknik wawancara atau permohonan permintaan data penelitian.

Data Cleaning

Melakukan pembersihan data yang hilang atau tidak lengkap sehingga tidak ada lagi *missing value* agar nantinya dapat diproses dengan benar serta tidak membuat tingkat akurasi prediksi menurun.

Data Integration dan Transformation

Pada tahap ini data yang banyak dari berbagai tabel akan diintegrasikan atau digabungkan menjadi satu *dataset* yang utuh yang kemudian ditransformasikan atau diubah menjadi data label atau target. Sesuai dengan judul penelitian ini, maka data yang akan diubah menjadi label adalah data harga.

4. PEMBAHASAN

Data yang digunakan pada penelitian ini merupakan data tanaman sayuran semusim atau lebih tepatnya data bawang merah, yang mana

data terdiri dari empat atribut. Pertama Luas panen yang dihitung dalam satuan hektar, kemudian jumlah produksi dalam satuan ton, dan Curah hujan dalam satuan milimeter serta harga jual dalam satuan rupiah. Dimana masing-masing data tersaji dalam tiap bulan dari tahun 2014 smpa 2018 dapat dilihat seperti tabel 4.1 ini.

Tabel 4.1 Sampel Data

Luas Panen (Ha)	Produksi (Ton)	Curah Hujan (Mm)	Harga (Rp)
3	142	296,7	24428
25	1969	298,1	20986
101	8464	157,4	22417
38	2809	176,6	22344
262	21099	96,9	20468
65	5652	66	22212
14	1683	51	22221

Data Selection

Data yang digunakan dalam penelitian ini yaitu data bawang merah berdasarkan lima tahun berturut-turut dimana data dari tahun 2014, 2015, 2016, 2017 dan 2018 yang berasal dari data bps.go.id khusus untuk daerah Yogyakarta. Kemudian data tersebut diseleksi dan akan digunakan untuk diolah dalam memprediksi harga bawang merah. Adapun atribut yang digunakan dalam penentuan prediksi luas panen, produksi, curah hujan dan harga yang akan dijadikan labelnya.

Tabel 4.2 Data Selection

Id	Luas Panen (Ha)	Produksi (Ton)	Curah Hujan (Mm)	Harga (Rp)
Jan-14	3	142	296,7	24428
Feb-14	25	1969	298,1	20986
Mar-14	101	8464	157,4	22417
Apr-14	38	2809	176,6	22344
Mei-14	262	21099	96,9	20468
Jun-14	65	5652	66	22212
Jul-14	14	1683	51	22221
Agu-14	68	4654	0	19852

Pre-Processing

Tahap *preprocessing* dilakukan untuk menjadikan data yang akan digunakan sebagai data pengujian agar data tidak memiliki permasalahan lagi di dalamnya, yang artinya data akan dicek dan dibersihkan hingga tidak memiliki *missing value*.

Name	Type	Missing	Filter (5/5 attributes)	Search for
id	Polynomial	0	Label Sep-18 (1)	Max Agu-14
Label Harga	Polynomial	0	Label Sangat Mahal (1)	Max Murah
Luas Panen	Integer	0	Min 3	Max 600
Produksi	Integer	0	Min 142	Max 62194
Curah Hujan	Real	0	Min 0	Max 693

Gambar 4.1 Data tanpa Missing Value

Tabel 4.3 Pembersihan Data

Luas Panen (Ha)	Produksi (Ton)	Curah Hujan (Mm)	Harga (Rp)
3	142	296,7	24428
25	1969	298,1	20986
101	8464	157,4	22417
38	2809	176,6	22344
262	21099	96,9	20468
65	5652	66	22212
14	1683	51	22221

Transformation

Pada tahap *transformation* ini adalah hasil dari pengelompokan data *preprocessing* yang mana salah satu atributnya akan diubah menjadi label. Dimana label itu artinya atribut target yang akan dijadikan sebagai atribut yang menentukan nilai class suatu data. Dalam hal ini atribut yang akan digunakan atau dijadikan sebagai atribut label adalah harga. Karena data harga masih berbentuk nilai *integer* maka akan diubah menjadi *nominal* atau *polynomial*, yang mana akan menjadi 3 kategori.

Dalam hal ini tidak ada aturan baku yang penulis temui terkait cara atau langkah pembagian kelas menjadi beberapa kategori, namun penulis belajar dari rumus statistika distribusi frekuensi, kondisi data yang sebenarnya dan langkah perhitungan yang telah dilakukan oleh penelitian terdahulu, pada penelitian yang membahas tentang prediksi harga komoditas cabai rawit di Yogyakarta (Reczy 2020).

Pada tahap ini, data atribut harga yang sebelumnya *Integer* akan diubah menjadi *polynomial* untuk dijadikan sebagai label. Berikut ini langkah-langkah mengubah data harga menjadi *polynomial*.

1. Mencari nilai maksimal dan minimal pada data harga.
Maksimal = 33583, Minimal = 13537
2. Mengurangi nilai maksimal dan minimal untuk mencari *range* data.
 $33583 - 13537 = 20046$

3. Hasil *range* data dibagi 3 untuk membagi menjadi 3 kategori.

$$\frac{20046}{3} = 6682$$

4. Hasil data minimal ditambah dengan hasil *range* dan hasilnya menjadi batas *range* untuk kategori 1 yaitu *range* harga murah Rp. 6682 – Rp. 20219
 $13537 + 6682 = 20219$
5. Hasil *range* ditambah dengan hasil batas *range* kategori 1 dan hasilnya untuk kategori 2 yaitu *range* harga mahal Rp. 20219 – Rp. 26901
 $6682 + 20219 = 26901$
6. Hasil *range* ditambah dengan hasil batas *range* kategori 2 dan hasilnya menjadi batas untuk kategori 3 yaitu *range* harga sangat mahal Rp. 26901 – Rp. 33583
 $6682 + 26901 = 33583$

Tabel 4.4 Kelas Harga

Harga (Rp)	Kategori
>= Rp. 6682 – Rp. <= 20219	Murah
>Rp. 20219 – Rp. <= 26901	Mahal
>Rp. 26901 – Rp. <= 33583	Sangat Mahal

Penerapan K-Nearest-Neighbors

Berikut ini merupakan tahapan dimana penggunaan algoritma *K-Nearest Neighbor* pada penerapan *data mining*, adapun langkah-langkah dari *K-Nearest-Neighbor* :

1. Penentuan nilai *k*. Penentuan nilai *k* yang digunakan tidak memiliki aturan yang baku, namun pada penelitian ini nilai *k* yang digunakan adalah 3.
2. Hitung jarak antar *data training* dan data uji (*test*) yang ada pada tahap *transformation* dengan menggunakan perhitungan *Euclidean Distance* sebagai berikut :

Hitung jarak data uji dengan data tes pertama

$$\sqrt{(141 - 286)^2 + (17252 - 23536)^2 + (0 - 324)^2} = 6294$$

Hitung jarak data uji dengan data tes kedua

$$\sqrt{(141 - 23)^2 + (17252 - 1606)^2 + (0 - 508)^2} = 43912$$

Hitung jarak data uji dengan data tes ketiga

$$\sqrt{(141 - 3)^2 + (17252 - 259)^2 + (0 - 268)^2} = 28013$$

Hitung jarak data uji dengan data tes keempat

$$\sqrt{(141 - 5)^2 + (17252 - 317)^2 + (0 - 292)^2} = 615$$

Hitung jarak data uji dengan data tes kelima

$$\sqrt{(141 - 15)^2 + (17252 - 565)^2 + (0 - 349)^2}$$

= 667

Hitung jarak data uji dengan data tes keenam

$$\sqrt{(141 - 60)^2 + (17252 - 5259)^2 + (0 - 403)^2}$$

= 4328

Tabel 4.5 Hasil Pengurutan Jarak

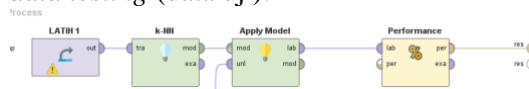
Luas (Ha)	Produksi (Ton)	Curah Hujan (Mm)	Euclidean Distance	Rangking	Harga
5	317	292	615	1	Mahal
15	565	349	667	2	Mahal
60	5259	403	4328	3	Sangat Mahal
286	23536	324	6294	4	Murah
23	1606	508	43912	5	Murah
3	259	268	28013	6	Murah

Pada tabel 4.5, dapat dilihat hasil dari nilai $k=3$ adalah dua harga mahal dan satu harga murah. Maka akan diambil suara mayoritas dua dari tiga sehingga harga yang keluar adalah mahal berdasarkan jarak terkecil antar tetangga dengan nilai $k=3$.

Implementasi Pada RapidMiner

Pada pengimplementasian dengan Rapidminer akan dilakukan beberapa perilaku atau kondisi terhadap data bawang merah agar diharapkan dapat memberikan hasil yang lebih bervariasi serta pemahaman lebih mendalam terhadap data dan metode yang digunakan. Karena seperti yang diketahui salah satu tujuan data mining adalah untuk menemukan pola dan informasi yang baru, agar bisa jadi model pembelajaran. Pada penerapannya ini nilai k yang akan digunakan adalah 3, 5, dan 11. Karena seperti yang diketahui dalam metode dalam *K-Nearest-Neighbor* nilai k tersebut tidak memiliki aturan baku. Berikut pembagian kondisinya :

1. *Data training* dan *data testing* akan dibagi secara manual, dimana data dari tahun 2014 sampai dengan tahun 2017 akan dijadikan *data training*. Sementara data tahun 2018 akan dijadikan *data testing*, yang apa bila dikalkulasikan pembagian ini berjumlah 80% *data training* (data latihan) dan 20% *data testing* (data uji).



Gambar 4.2 Susunan Operator Kondisi 1

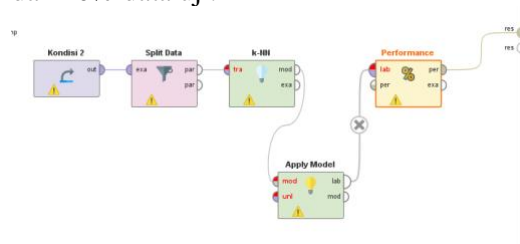
accuracy: 75.00%

	true Murah	true Mahal	true Sangat Mahal	class precision
pred. Murah	9	3	0	75.00%
pred. Mahal	0	0	0	0.00%
pred. Sangat Mahal	0	0	0	0.00%
class recall	100.00%	0.00%	0.00%	

Gambar 4.3 Tingkat Akurasi

Tingkat akurasi yang didapatkan pada kondisi pertama $k=3$ adalah 75,00%, $k=5$ adalah 75,00% dan $k=11$ 75,00%. Nilai k tidak mempengaruhi tingkat akurasi.

2. Pembagian data secara otomatis dengan menggunakan operator split data pada RapidMiner dengan rasio 0,8 data latihan dan 0,2 data uji yang apa bila jika dihitung dalam skala persen maka 80% data latihan dan 20% data uji.



Gambar 4.4 Susunan Operator Kondisi 2

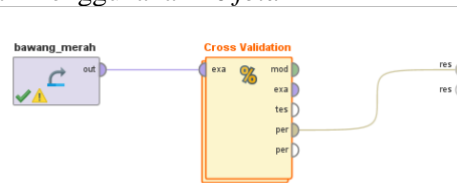
accuracy: 91.67%

	true Murah	true Mahal	true Sangat Mahal	class precision
pred. Murah	11	1	0	91.67%
pred. Mahal	0	0	0	0.00%
pred. Sangat Mahal	0	0	0	0.00%
class recall	100.00%	0.00%	0.00%	

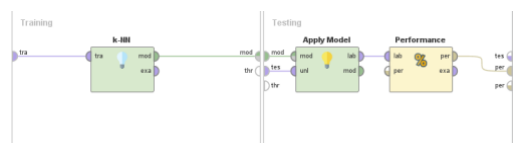
Gambar 4.5 Tingkat akurasi

Tingkat akurasi yang didapatkan pada kondisi pertama $k=3$ adalah 83,33%, $k=5$ adalah 91,67% dan $k=11$ 91,67%. Nilai k mempengaruhi tingkat akurasi, dimana nilai $k \geq 5$ memiliki tingkat akurasi tertinggi.

3. Data akan proses dengan operator *Cross Validation*, dimana pada pemrosesan ini akan menggunakan 10 fold



Gambar 4.6 Penempatan Cross Validation



Gambar 4.7 Proses di Cross Validation

accuracy: 88.33% +/- 8.05% (micro average: 88.33%)

	true Murah	true Mahal	true Sangat Mahal	class precision
pred Murah	53	6	1	88.33%
pred Mahal	0	0	0	0.00%
pred Sangat Mahal	0	0	0	0.00%
class recall	100.00%	0.00%	0.00%	

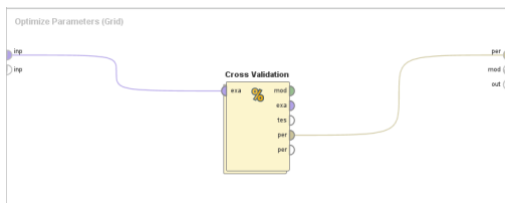
Gambar 4.8 Tingkat Akurasi

Tingkat akurasi yang didapatkan pada kondisi pertama $k = 3$ adalah 83,33%, $k = 5$ adalah 88,33% dan $k = 11$ 88,33%. Nilai k mempengaruhi tingkat akurasi, dimana nilai $k \geq 5$ memiliki tingkat akurasi tertinggi.

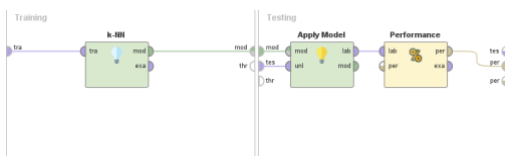
4. Data akan diproses dengan *Cross Validation* tetapi pada awalnya pemrosesannya akan ditambahkan operator *Optimize Parameters (Grid)*. Dimana operator *Optimize Parameters (Grid)* ini merupakan operator yang akan menemukan nilai optimal dari parameter yang dipilih dalam prosesnya, sehingga bisa menentukan nilai k dengan hasil yang paling maksimal, tanpa harus mencoba menginput nilai k secara berulang-uang.



Gambar 4.9 Letak *Optimize Parameters (Grid)*



Gambar 4.10 *Cross Validation*



Gambar 4.11 Proses di *Cross Validation*

accuracy: 88.33% +/- 8.05% (micro average: 88.33%)

	true Murah	true Mahal	true Sangat Mahal	class precision
pred Murah	53	6	1	88.33%
pred Mahal	0	0	0	0.00%
pred Sangat Mahal	0	0	0	0.00%
class recall	100.00%	0.00%	0.00%	

Gambar 4.12 Tingkat Akurasi

ParameterSet

Parameter set:

```
Performance:
PerformanceVector [
----accuracy: 88.33% +/- 8.05% (micro average: 88.33%)
ConfusionMatrix:
True:  Murah  Mahal  Sangat Mahal
Murah:  53    6    1
Mahal:   0    0    0
Sangat Mahal:  0    0    0
]
k-NN.k = 11
```

Gambar 4.13 Output Nilai k yang Disarankan

Berdasarkan empat kondisi yang dilakukan terhadap data bawang merah, yang mana kondisi itu terbagi sebagai berikut :

1. Data bawang merah dibagi menjadi dua berdasarkan tahun, yang mana tahun terakhir dijadikan data uji dengan rasion 80% data latih dan 20% data uji. Didapatkan hasil akurasi masing dari nilai $k = 3$ akurasinya adalah 75%, nilai $k = 5$ akurasinya 75% dan nilai $k = 11$ akuasrinya 75%. Sehingga dapat dinyatakan pada kondisi ini nilai k tidak mempengaruhi tingkat akursi.
2. Kondisi kedua data secara otomatis dengan menggunakan operator *split data* pada *RapidMiner* dengan rasio 0,8 data latih dan 0,2 data uji yang apa bila jika dihitung dalam skala persen maka 80% data latih dan 20% data uji. Didapatkan hasil akurasi masing-masing dari nilai $k = 3$ akurasinya adalah 83,33%, nilai $k = 5$ akurasinya 91,67% dan nilai $k = 11$ akuasrinya 91,67%. Sehingga dapat dinyatakan pada kondisi ini nilai k dengan akurasi tertinggi didapatkan pada angka 5, yang mana apabila nilai k di atas 5 maka hasilnya tetap sama, terbukti dengan nilai $k = 11$ hasilnya juga 91,67%.
3. Kondisi ketiga data akan proses dengan operator *Cross Validation*, dimana pada pemrosesannya menggunakan 10 *fold*. Pada uji coba ketiga ini didapatkan hasil akurasi dari nilai $k = 3$ akurasinya adalah 83,33%, nilai $k = 5$ akurasinya 88,33% dan nilai $k = 11$ akuasrinya 88,33%. Sehingga dapat dinyatakan pada kondisi ini nilai

k dengan akurasi tertinggi didapatkan pada angka 5, yang mana apabila nilai k di atas 5 maka hasilnya tetap sama, terbukti dengan nilai $k = 11$ hasilnya juga 88,33%.

4. kondisi keempat ini data diproses dengan *Cross Validation* tetapi pada awalnya pemrosesannya akan ditambahkan operator *Optimize Parameters (Grid)*. Pada uji coba terakhir ini didapatkan tingkat akurasi 88,33% dengan rekomendasi nilai $k = 11$.

Maka dari rangkaian panjang uji coba terhadap empat kondisi yang diberlakukan pada data bawang merah dimana masing-masing kondisi memberikan hasil yang berbeda-beda. Sehingga didapatkan hasil akurasi tertinggi pada empat kondisi pengujian ini adalah 91,67%, yang mana ini merupakan hasil uji coba pada kondisi kedua, dimana data dibagi menggunakan operator *split data*.

4. KESIMPULAN

Berdasarkan proses perhitungan dengan menggunakan algoritma *K-Nearest-Neighbor* pada data bawang merah yang mana data tersebut dibuat dalam empat kondisi maka dapat disimpulkan :

1. Algoritma *K-Nearest-Neighbor* telah berhasil diimplementasikan dengan baik.
2. Hasil akurasi tertinggi yang didapatkan dengan menggunakan *K-Nearest-Neighbor* adalah 91,67% pada pemodel data di kondisi uji coba ke 2 dari empat kondisi yang diterapkan.
3. Nilai k atau tetangga terdekat yang menghasilkan tingkat akurasi tertinggi yaitu $k = 5$, yang mana ini juga merupakan penerapan pemodelan data pada kondisi ke 2 dari empat kondisi yang diuji cobakan.
4. Variabel curah hujan dan luas area tidak mempengaruhi harga jual bawang merah baik menurun atau meningkat, yang mana ini terbukti dari analisa dan temuan selama uji coba berulang yang dilakukan di beberapa kondisi yang berbeda serta fakta data yang sebenarnya. Sehingga hipotesa peneliti yang menggira variabel curah hujan dan luas area mempengaruhi harga bawang merah tidak terbukti, dimana ini menjelaskan pula ada campur tangan

manusia dalam bidang politik dan ekonomi yang mempengaruhi nilai jual bawang merah, karena terbukti tidak peduli bagaimana pun jumlah produksi bawang merah harga tetap tidak menentu atau terprediksi. Pada hal jika barang sedikit permintaan meningkat harga pasti mahal atau sebaliknya, namun pada kondisi bawang merah ini, itu tidak berlaku.

5. Dari hasil analisis dan percobaan data yang telah dilakukan, maka bagi petani yang hendak bercocok tanam bawang merah pada musim cukup kering lebih baik guna menghindari pembusukan, yang mana pada bulan Juli sampai Agustus biasanya waktu paling pas. Terbukti dari data 2018 sampai 2014 dimana tiga bulan atau 90 hari dihitung dari mulai penanaman, panen bawang merah selalu mencapai angka tertinggi dalam tiap tahunnya.

5. UCAPAN TERIMA KASIH

Penulis selaku peneliti mengucapkan terima kasih sebesar-besarnya kepada kedua orang tua yang senantiasa mendukung penulis dalam melakukan penelitian ini serta memberikan semangat dan tentu saja bantuan dana dalam memperlancar penelitian ini, serta kepada dosen pembimbing skripsi

DAFTAR PUSTAKA

- Fatah, Haerul, dan Agus Subekti. (2018). "Jurnal PILAR Nusa Mandiri." *PREDIKSI HARGA CRYPTOCURRENCY DENGAN METODE 14*: 137.
- Hortikultura, Subdirektorat Statistik. (2018). "STATISTIK TANAMAN SAYURAN DAN BUAH-BUAHAN SEMUSIM, 2018." Dalam *STATISTIK TANAMAN SAYURAN DAN BUAH-BUAHAN SEMUSIM, 2018*, disunting oleh Subdirektorat Statistik Hortikultura, 11. BPS-Statistics Indonesia.
- Kusrini. (2009). *Algoritma Data Mining*. Yogyakarta.
- Larose, Daniel T. (2005). *DISCOVERING. New Jersey: John Wiley & Sons, Inc. 2005. Discovering Knowledge in Data: An Introduction to Data Mining*. New Jersey & Sons, Inc.: jhon wiley.
- Lerdorf, Rasmus, Kevin Tatroe, Bob Kaehms, dan Ric McGredy. (2002).

- Programming PHP*. First Edition. Sebastopol: O'Reilly Media, Inc.
- Muliono, Rizki, Juanda Hakim Lubis, dan Nurul Khairina. (2019). "Analisis Algoritma K-Nearest Neighbors." *semantika* 12: 12.
- Putri, Sri Cahyani. 2020. *Harga Bawang Merah di Kota Yogya Tembus Rp 50 Ribu Per KG*. Diakses 06 10, 2020. <https://jogja.tribunnews.com/2020/06/06/harga-bawang-merah-di-kota-yogya-tembus-rp-50-ribu>.
- RapidMiner. (2020). *About RapidMiner*. Diakses 06 22, 2020. <https://rapidminer.com/us/>.
- Reczy, Sebastianus. (2020). *Penerapan Algoritma K-Nearest Neighbor untuk prediksi harga cabai rawit di Yogyakarta*. Yogyakarta: Sanata Dharma.
- Sinta, Dewi.(2015). "METODE ENSEMBLE K-NEAREST NEIGHBOR UNTUK." *METODE ENSEMBLE K-NEAREST NEIGHBOR UNTUK*.
- Sumarni, Nani, dan Achmad Hidayat. (2005). *Budidaya Bawang Merah*. Panduan Teknis PTT Bawang Merah. Bandung: BALAI PENELITIAN TANAMAN SAYURAN .
- Yustanti, Wiyli. (2012). "Algoritma K-Nearest Neighbour untuk Memprediksi." *jurnal matematika, statistika dan komputasi* 9: 57.