

Perbandingan Teknik Data Mining Untuk Prediksi Penjualan Pada UMKM Gerabah

Nosiel¹, Sriyanto², Ferdy Maylani³

^{1,2,3}Magister Teknik Informatika, Institut Informatika dan Bisnis Darmajaya Lampung
Email: ¹Nosiel.nosiel.2021211026@mail.darmajaya.ac.id, ²Sriyanto@darmajaya.ac.id,
³Ferdy.maylani.2021211015@mail.darmajaya.ac.id

Abstract

Basically, maximum sales are those that can meet existing demands, determining the amount of production planned to meet the production level to meet the planned sales level or market demand level, the amount of demand and supply is an uncertainty. The amount of demand and supply is uncertainty, Pottery MSMEs in producing pottery of course adjust to the wishes and needs of their customers, the needs and interests of customers in the production of Pottery MSMEs always change from time to time. In addition to efforts to meet the desires and needs of customers, MSME Pottery also needs to pay attention to the production capacity it has, minimize waste products and can determine the amount of pottery production at this time and so on. The purpose of this study is to compare 3 data mining methods such as Decision Tree, Naive Bayes, K-NN to determine the amount of production to meet consumer demand. Applying the comparison of 3 methods in data mining in determining the amount of pottery production, it is hoped that MSMEs can overcome fluctuations in consumer demand with minimal production costs. In this study, testing was carried out using 3 methods of Decision Tree, Naive Bayes, and KNN, using data that was processed based on the Knowledge Discovery in Database (KDD) stage. The results of this mining activity are expected to provide a predictive conclusion on the sales of pottery products from the methods we use, based on the results of the study showing that the best method of predicting sales of pottery is the Naive Bayes method.

Keywords: Sales, Pottery SMEs, Decision Tree, Naive Bayes, K-NN.

Abstrak

Pada dasarnya penjualan yang maksimal adalah yang dapat memenuhi permintaan-permintaan yang ada, penentuan jumlah produksi yang di rencanakan untuk memenuhi tingkat produksi guna memenuhi tingkat penjualan yang di rencanakan atau tingkat permintaan pasar, jumlah permintaan dan persediaan merupakan suatu ketidakpastian. Jumlah permintaan dan persediaan adalah ketidakpastian, UMKM Gerabah dalam memproduksi gerabah tentu nya menyesuaikan dengan keinginan serta kebutuhan pelanggannya, kebutuhan dan minat pelanggan terhadap hasil produksi UMKM Gerabah selalu berubah ubah setiap waktunya. Selain upaya memenuhi keinginan dan kebutuhan para pelanggan, UMKM Gerabah juga perlu memperhatikan kapasitas produksi yang di miliknya tersebut, minimalisasi produk sisa dan dapat menentukan jumlah produksi gerabah pada saat ini dan seterusnya. Tujuan dari penelitian ini adalah membandingkan 3 metode data mining seperti *Decision Tree*, *Naive Bayes*, *K-NN* untuk menentukan jumlah produksi dalam memenuhi permintaan konsumen. Penerapan perbandingan 3 metode dalam data mining dalam menentukan jumlah produksi gerabah, diharapkan UMKM dapat mengatasi fluktuasi permintaan konsumen dengan biaya produksi yang minimal. Pada penelitian ini dilakukan pengujian menggunakan 3 metode Decision Tree, Naive Bayes, dan KNN, dengan menggunakan data yang diolah berdasarkan tahapan *Knowledge Discovery in Database* (KDD). Hasil dari kegiatan mining ini diharapkan dapat memberikan kesimpulan prediksi penjualan produk gerabah dari metode-metode yang kita gunakan, berdasarkan hasil penelitian menunjukkan bahwa metode yang terbaik dari penelitian prediksi penjualan gerabah adalah metode *Naive Bayes*.

Keywords : Penjualan, UMKM Gerabah, *Decision Tree*, *Naive Bayes*, *K-NN*.

1. PENDAHULUAN

Usaha mikro kecil dan menengah (UMKM) memiliki peranan yang sangat vital di dalam pembangunan dan pertumbuhan ekonomi, tidak hanya di negara-negara berkembang seperti Indonesia tetapi juga di negara-negara maju. Di Indonesia peranan UMKM selain berperan dalam pertumbuhan pembangunan dan ekonomi, UMKM juga memiliki peranan yang sangat penting dalam mengatasi masalah pengangguran. Tumbuhnya usaha mikro menjadikannya sebagai sumber pertumbuhan kesempatan kerja dan pendapatan. Dengan banyak menyerap tenaga kerja berarti UMKM juga punya peran strategis dalam upaya pemerintah dalam memerangi kemiskinan dan pengangguran. Kontribusi sektor usaha mikro kecil dan menengah terhadap produk domestik bruto meningkat dari 57,84% menjadi 60,34% dalam 5 tahun terakhir. Serapan tenaga kerja pada sektor ini juga meningkat dari 96,99% menjadi 97,22% pada periode yang sama, Sedangkan perkembangan UMKM menurut sektor ekonomi di Provinsi Lampung tahun 2017 berdasarkan jenis usaha antara lain:



Gambar 1. Data UMKM di Provinsi Lampung

Sumber : (Dinas Koperasi dan Ukm Provinsi Lampung, 2017)

Daerah Desa Negara Ratu Dusun Sidoharjo 1 merupakan desa yang terletak di Kecamatan Natar Kabupaten Lampung Selatan Provinsi Lampung. Desa ini memiliki potensi-potensi yang dapat dikembangkan dan beberapa potensi tersebut dalam bentuk UMKM. Potensi tersebut antara lain UMKM Gerabah, UMKM ini merupakan industri rumahan yang dimiliki oleh kelompok usaha.

UMKM Gerabah adalah UMKM yang membuat perkakas dari tanah liat yang di bentuk, dikeringkan, kemudian dibakar sehingga dapat digunakan sebagai alat-alat yang berguna untuk membantu kehidupan manusia. Contoh jenis-jenis produk Gerabah antara lain : Guji, Pot Bunga, Meja, Kursi, Bak Mandi Celengan seperti Ayam, Katak dan jenis hewan lainnya. Proses penjualan produk Gerabah yaitu di pasarkan dengan cara survei ke tempat-tempat tertentu yang belum pernah di masuki, selain itu pemasaran Gerabah sendiri di lakukan pada acara event-event tertentu seperti menyambut Bulan Ramdhan, Pameran, Festival, Pasar Malam dan Sekatenan. Ada juga pembeli yang datang ke rumah-rumah, baik pembeli tersebut bertindak sebagai distributor mau pun sebagai konsumen langsung.

Pada dasarnya penjualan yang maksimal yaitu yang dapat memenuhi permintaan-permintaan yang ada, penentuan jumlah produksi yang direncanakan untuk memenuhi tingkat produksi guna memenuhi tingkat penjualan yang di rencanakan atau tingkat permintaan pasar, Jumlah permintaan dan persediaan merupakan suatu ketidakpastian. Jumlah permintaan dan persediaan adalah ketidakpastian, UMKM Gerabah dalam memproduksi gerabah tentu nya menyesuaikan dengan keinginan serta kebutuhan pelanggannya, kebutuhan dan minat pelanggan terhadap hasil produksi UMKM Gerabah selalu berubah ubah setiap waktunya. Selain upaya memenuhi keinginan dan kebutuhan para pelanggan, UMKM Gerabah juga perlu memperhatikan kapasitas produksi yang di milikinya tersebut, minimalisasi produk sisa dan dapat menentukan jumlah produksi gerabah pada saat ini dan seterusnya.

Berdasarkan hasil wawancara dan penelitian, maka penulis akan menggunakan perbandingan 3 metode data mining seperti Decision Tree, Naive Bayes, K-NN untuk menentukan jumlah produksi dalam memenuhi permintaan konsumen. Penerapan perbandingan 3 metode dalam data mining dalam menentukan jumlah produksi gerabah, diharapkan UMKM dapat mengatasi fluktuasi permintaan konsumen dengan biaya produksi yang minimal.

2. KERANGKA TEORI

2.1 Prediksi/Forecasting

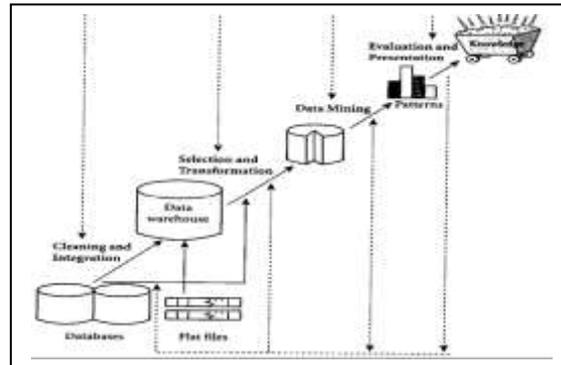
Prediksi/forecasting adalah menentukan jumlah kebutuhan bulan mendatang terkait dengan dukungan data historis (historical data) atau serangkaian waktu/periode yang dianalisis sehingga dapat diperhitungkan untuk memprediksi jumlah kebutuhan pada bulan mendatang. Prediksi juga dapat digunakan dalam pengklasifikasian, tidak hanya untuk memprediksi time series, karena sifatnya yang bisa menghasilkan class berdasarkan atribut yang ada (Sutardi:2007,44).

2.2 Data Mining

Data mining mengacu pada proses pencarian informasi yang tidak diketahui sebelumnya dari sekumpulan data besar [4]. Pada prosesnya data mining akan mengekstrak informasi yang berharga dengan cara menganalisis adanya pola-pola ataupun hubungan keterkaitan tertentu dari data-data yang berukuran besar [5]. Definisi lain data mining adalah serangkaian proses yang memperkerjakan satu atau lebih teknik pembelajaran komputer untuk

menganalisis dan mengekstrak pengetahuan secara otomatis atau serangkaian proses untuk menggali nilai tambah dari suatu kumpulan data berupa pengetahuan yang selama ini tidak diketahui secara manual [6].

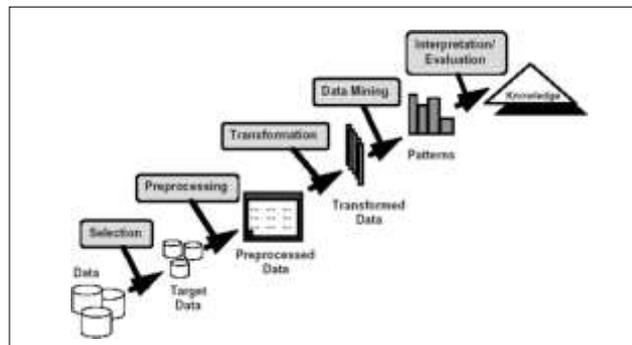
Karena data mining adalah suatu rangkaian proses, data mining dapat dibagi menjadi beberapa tahap. Tahap-tahap tersebut bersifat interaktif di mana pemakai terlibat langsung atau dengan perantara knowledge base. Tahap-tahap ini diilustrasikan pada Gambar 2.



Gambar 2. Tahap-tahap *data mining*

2.3 Knowledge Discovery in Database

Knowledge Discovery in Database (KDD) adalah proses menentukan informasi yang berguna serta pola-pola yang ada dalam data. Informasi ini terkandung dalam basis data yang berukuran besar yang sebelumnya tidak diketahui dan potensial bermanfaat. Data Mining merupakan salah satu langkah dari serangkaian proses iterative KDD (Kusri, dkk 2009:7). Berikut tahapan proses KDD dapat dilihat pada gambar 3.



Gambar 3. Tahapan Dalam KDD

Tahapan proses KDD terdiri dari:

1. Data Selection. Pada proses ini dilakukan pemilihan himpunan data, menciptakan himpunan data target, atau memfokuskan pada subset variable (sampel data) dimana penemuan (discovery) akan dilakukan. Hasil seleksi disimpan dalam suatu berkas yang terpisah dari basis data operasional.
2. Pre-Processing dan Cleaning Data. Pre-Processing dan Cleaning Data dilakukan membuang data yang tidak konsisten dan noise, duplikasi data, memperbaiki kesalahan data, dan bisa diperkaya dengan data eksternal yang relevan.
3. Transformation. Proses ini mentransformasikan atau menggabungkan data ke dalam yang lebih tepat untuk melakukan proses mining dengan cara melakukan peringkasan (agregasi).
4. Data Mining. Proses Data Mining yaitu proses mencari pola atau informasi menarik dalam data terpilih dengan menggunakan teknik, metode atau algoritma tertentu sesuai dengan tujuan dari proses KDD secara keseluruhan.
5. Interpretation / Evaluasi. Proses untuk menerjemahkan pola-pola yang dihasilkan dari Data Mining. Mengevaluasi (menguji) apakah pola atau informasi yang ditemukan bersesuaian atau bertentangan dengan fakta atau hipotesa sebelumnya. Pengetahuan yang diperoleh dari pola-pola yang terbentuk dipresentasikan dalam bentuk visualisasi.

2.3 Naive Bayes

Naive Bayes adalah teknik prediksi berbasis probabilistic sederhana yang berdasar pada penerapan Teorema Bayes (aturan Bayes) dengan asumsi independensi (ketidakketergantungan) yang kuat. (Prasetyo, 2012).

2.4 Decision Tree

Decision Tree adalah metode untuk menemukan fungsi pendekatan yang bernilai diskrit dan tahan terhadap data-data yang memiliki kesalahan (noisy data) serta mampu mempelajari ekspresi-ekspresi disjunctive seperti OR. (Lesmana, 2012).

2.5 K-Nearest Neighbor (k-NN)

K-NN adalah salah satu metode yang menggunakan algoritma terbimbing dimana hasil dari query instance yang baru diklasifikasikan berdasarkan mayoritas dari kategori pada k-NN [13]. Tujuan dari algoritma ini yaitu atribut dan training sample digunakan untuk mengklasifikasikan obyek yang baru .

Langkah penyelesaian masalah menggunakan Metode k-NN adalah sebagai berikut :

1. Menentukan nilai parameter k (jumlah tetangga terdekat)
2. Menghitung jarak setiap sample data dengan data yang diuji.
3. Mengurutkan data berdasarkan jarak dari yang terkecil hingga yang terbesar.
4. Mengamati jumlah keputusan yang terbanyak untuk k data yang diambil.
5. Jika terdapat dua atau lebih kelas ω_i yang merupakan tetangga terdekat dari data uji x , maka terjadilah kondisi
6. seimbang (konflik) dan digunakan strategi pemecahan konflik.

3. METODOLOGI

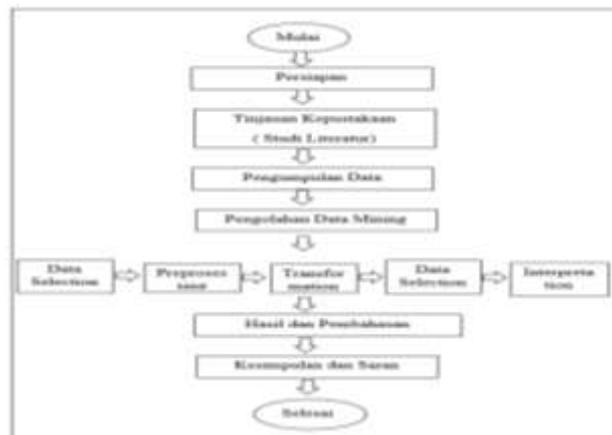
Berdasarkan penelitian penulis akan menggunakan perbandingan 3 metode dalam data mining untuk menentukan jumlah produksi dalam memenuhi permintaan konsumen. Penerapan perbandingan 3 metode dalam data mining untuk menentukan jumlah produksi gerabah, diharapkan UMKM dapat mengatasi fluktuasi permintaan konsumen dengan biaya produksi yang minimal.

3.1 Metode pengumpulan

Metode pengumpulan data yang digunakan untuk memperoleh data-data yang dibutuhkan dalam penelitian ini adalah :

- a. Wawancara
Metode wawancara dilakukan dengan cara menyampaikan sejumlah pertanyaan dari pewawancara untuk di jawab oleh narasumber pemilik UMKM untuk mendapatkan suatu informasi.
 - b. Studi Pustaka
Merupakan metode pengumpulan data yang diperoleh dari hasil olahan orang lain berupa dokumen, buku pustaka, jurnal, dengan membaca berbagai bahan penulisan, mengenai permasalahan yang berhubungan dengan penulisan dan khususnya penelitian yang berkaitan karya ilmiah.
 - c. Observasi/survei
Metode ini digunakan dengan cara terjun langsung dan mengamati apa saja yang di perlukan di UMKM Gerabah untuk menemukan informasi dan pengetahuan yang diperlukan untuk bahan penelitian.
-

3.2 Tahapan Penelitian Knowledge Discovery in Database (KDD)



Gambar 4. Tahapan Penelitian

Tahapan penelitian yang dituangkan dalam diagram alir ini. Menggambarkan proses penelitian yang akan ditempuh sekaligus menggambarkan penelitian secara keseluruhan. Tahapan yang akan ditempuh yaitu:

1. Persiapan. Tahap ini merupakan tahapan subjek (populasi) UMKM Gerabah. Objek yang diambil yaitu penjualan produk gerabah. Batasan dan menyusun rencana penelitian.
2. Tinjauan kepustakaan. Dalam tinjauan kepustakaan dilakukan telaah dan studi literatur mengenai prediksi penjualan dan yang berhubungan.
3. Pengumpulan data dilakukan dengan cara wawancara kepada pemilik UMKM Gerabah, observasi dan dokumen.
4. Pengolahan data mining. Data yang dikumpulkan diolah sesuai dengan tahapan Knowledge Discovery in Database (KDD).
5. Hasil dan Pembahasan. Pembahasan pada tahapan ini menjelaskan hasil dari proses data mining yang dilakukan dengan menggunakan metode Decision Tree, Naive Bayes dan K-Nearest Neighbor.
6. Kesimpulan. Membuat kesimpulan dari hasil penelitian untuk pihak UMKM Gerabah agar dapat menjadi lebih baik lagi.

3.3 Pengolahan Data Mining

Adapun pengolahan data mining yang dilakukan pada penelitian ini yaitu, mengikuti tahapan dalam Knowledge Discovery in Database (KDD), untuk menghasilkan informasi sesuai dengan urutan yang sudah ditentukan, berikut tahapannya :

3.3.1 Data Selection

Pemilihan (seleksi) data dari sekumpulan data operasional yang perlu dilakukan sebelum tahap penggalian informasi dalam KDD dimulai. Data hasil seleksi akan di gunakan untuk proses data mining, disimpan dalam suatu berkas, terpisah dari basis data operasional. Data yang digunakan dalam penelitian ini berasal dari UMKM Gerabah, yaitu data penjualan produk gerabah 2020.

Semua atribut yang ada pada data penjualan diantaranya field, Nama Gerabah, Jenis Gerabah, Lokasi, Bulan, Produksi, Penjualan, Persediaan, Status Produksi. Dari semua data yang ada akan diseleksi dan digunakan hanya 3 field yang digunakan untuk proses knowledge discovery in database (KDD). Field tersebut yaitu:

1. Nama produk gerabah merupakan atribut yang terdapat pada tabel data penjualan yang berisi informasi tentang nama produk gerabah
2. Jenis Gerabah merupakan atribut yang terdapat dalam tabel penjualan yang berisi informasi jenis gerabah
3. Lokasi merupakan atribut yang terdapat dalam tabel penjualan yang berisi informasi lokasi
4. Bulan merupakan atribut yang terdapat pada tabel data penjualan yang berisi informasi bulan transaksi penjualan
5. Produksi merupakan atribut yang terdapat pada tabel data penjualan yang berisi informasi produksi gerabah
6. Penjualan merupakan atribut yang terdapat pada tabel data penjualan yang berisi informasi transaksi penjualan
7. Persediaan merupakan atribut yang terdapat pada tabel data penjualan yang berisi informasi transaksi penjualan
8. Status produksi merupakan atribut yang terdapat pada tabel data penjualan yang berisi informasi Status produksi gerabah

3.3.2 Preprocessing

Sebelum proses data mining dapat dilaksanakan perlu dilakukan tahap preprocessing, pada tahap ini akan dilakukan proses integrasi data untuk penggabungan data dari database yang berbeda, selanjutnya dilakukan data cleaning untuk menghasilkan dataset yang bersih sehingga dapat digunakan dalam tahap berikutnya yaitu mining. Berikut merupakan penjelasan dari kedua proses :

1. Integrasi Data

Tahap ini adalah proses penggabungan data dari berbagai database yang berbeda, sehingga data tersebut saling berintegrasi. Data integrasi dilakukan pada atribut-atribut yang mengidentifikasi entitas-entitas yang unik. Pada tahapan ini tidak ada penggabungan data dikarenakan data yang diambil berasal dari satu database.

2. Data Cleaning

Tahap ini adalah tahap awal dari proses KDD. Pada tahapan ini data yang tidak relevan, missing value, dan radudant harus dibersihkan. Hal ini dikarenakan data yang relevan, tidak missing value, dan tidak radudant merupakan syarat awal dalam melakukan data mining. Suatu data dikatakan missing value jika terdapat atribut dalam dataset yang tidak berisi nilai atau kosong, sedangkan data dikatakan radudant jika dalam satu dataset lebih dari satu record yang berisi nilai yang sama, setelah melakukan cleaning terhadap data yang lebih memenuhi syarat berdasarkan data penjualan.

3.3.3 Transformation

Tahapan Transformation merupakan tahap merubah data yang telah dipilih, sehingga data tersebut sesuai untuk proses data mining. Proses transformasi dalam KDD merupakan proses kreatif dan sangat tergantung pada jenis atau pola informasi yang akan dicari dalam basis data.

Pada tahapan ini dari seluruh data operasional didapatkan data pengelompokan atribut yang digunakan untuk proses transformasi data mining, yaitu atribut bulan dan klasifikasi sebagai kriteria data yang menjadi target dalam proses mining.

3.3.4 Data Mining

Tahap ini merupakan proses mencari pola atau informasi menarik dalam data terpilih dengan menggunakan teknik atau metode tertentu berdasarkan proses KDD secara keseluruhan.

3.3.4.1 Naive Bayes

Metode yang digunakan pada penelitian ini adalah metode naive bayes, klaifikasi Bayesian adalah klasifikasi statistik yang bisa memprediksi probabilitas sebuah class. Klasifikasi ini dihitung berdasarkan Teoema Bayes. (Widiastuti, 2010) Persamaan dari teorema Bayes dirumuskan seperti Persamaan 1 berikut ini :

$$P(H|X) = \frac{P(X|H) \cdot P(H)}{P(X)} \dots\dots\dots (1)$$

3.3.4.2 Decision Tree

Metode yang digunakan pada penelitian ini adalah metode decision tree, hal yang harus dilakukan dalam metode decision tree adalah menghitung entropy dan information gain. (Ranny dkk, 2012)

Persamaan 2 Rumus *entropy* :

$$Entropi (y) = -p_1 \log_2 p_1 - p_2 \log_2 p_2 - \dots - p_n \log_2 p_n \dots\dots\dots (2)$$

Persamaan 3 Rumus *Information Gain*

$$gain(y, A) = entropi(y) - \sum_{c \in \text{nilai}(A)} \frac{y_c}{y} entropi(y_c) \dots\dots\dots (3)$$

3.3.4.3 K-NN

Metode yang digunakan pada penelitian ini adalah metode K-Nearest Neighbor (KNN) dimana metode ini memiliki atribut yang diinisialisasikan sebagai k, yaitu jumlah tetangga yang dijadikan acuan pada KNN, nilai k adalah bilangan bulat positif, berjumlah kecil dan ganjil.

Adapun tahapan pengerjaan metode KNN dalam penelitian ini adalah sebagai berikut :

1. Penentuan nilai k. Penentuan nilai k yang digunakan dalam klasifikasi tidak memiliki aturan yang baku, namun pada penelitian ini nilai k yang digunakan adalah 3, 5, 7, 9, 11, 13, 15.

2. Perhitungan jarak antar data *training* dan data uji(test). Teknik perhitungan jarak yang digunakan dalam metode KNN ini adalah Jarak *Euclidean Distance*. Dengan rumus sebagai berikut :
Contoh perhitungan jarak dalam KNN menggunakan *Euclidean Distance* akan dijelaskan pada bab selanjutnya dengan menggunakan data *real*.

$$\sqrt{\sum_{i=1}^K (X_i - Y_i)^2}$$

3. Pengurutan data hasil perhitungan. Jarak yang telah didapatkan kemudian diurutkan dari yang paling dekat jaraknya sampai yang paling jauh (ascending).
4. Menentukan kelompok data hasil uji berdasarkan label mayoritas dari K tetangga terdekat.

3.3.5 Interpretation/Evaluasi

Pola informasi yang dihasilkan dari proses data mining perlu ditampilkan dalam bentuk yang mudah dimengerti oleh pihak yang berkepentingan. Tahap ini mencakup pemeriksaan apakah pola atau informasi yang ditemukan bertentangan dengan fakta atau hipotesis yang ada sebelumnya. Pada tahap ini didapatkan pola penjualan elektronik dari proses data mining dengan metode K-Nearest Neighbor, pola atau informasi yang dihasilkan dari proses data mining adalah berupa rules yang didapat dari perhitungan.K-Nearest Neighbor.

4. HASIL DAN PEMBAHASAN

4.1 Hasil Wawancara

Dari hasil survei dan wawancara dengan pemilik UMKM Gerabah, di dapatlah data penjualan dan produksi pada umkm gerabah

Tabel 1. Dataset penjualan produk UMKM Gerabah

No	Lokasi	Bulan	Nama Gerabah	Jenis Gerabah	Penjualan	Persediaan	Produk	Status Produksi
01	Indralaya	Januari	Guji, Meja Kursi, Bak mandi, Pot Bunga,Celengan	standar,jumbo set,minimalis, kaca	360	90	450	Bertambah/berkurang
02	Muara 2	Feb	Guji, Meja Kursi, Pot Bunga	standar,jumbo set,minimalis, kaca	540	90	450	Bertambah/berkurang
03	Menggal	Maret	Guji, Meja Kursi, Bak mandi, Pot Bunga,Celengan	standar,jumbo set,minimalis, kaca, Celengan seperti Ayam, Gajah, Macan, Katak	360	120	480	Bertambah/berkurang
04	Tugu Mulyo	April	Guji, Meja Kursi, Pot Bunga,Celengan	standar,jumbo set,minimalis, kaca, Celengan seperti Ayam, Gajah, Macan, Katak	540	60	480	Bertambah/berkurang
05	Jambi	Mei	Guji, Meja Kursi, Pot Bunga,Celengan	standar,jumbo set,minimalis, kaca, Celengan seperti Ayam, Gajah, Macan, Katak	360	150	510	Bertambah/berkurang
06	Bengkulu	Juni	Guji, Meja Kursi, Pot Bunga,Celengan	standar,jumbo set,minimalis, kaca, Celengan seperti Ayam, Gajah, Macan, Katak	540	30	510	Bertambah/berkurang
07	Padang	Juli	Guji, Meja Kursi, Pot Bunga	standar,jumbo set,minimalis, kaca	360	180	540	Bertambah/berkurang
08	Medan	Agust	Guji, Meja Kursi, Pot Bunga	standar,jumbo set,minimalis, kaca	540	0	540	Bertambah/berkurang

09	Unit 2	Sept	Guji, Meja Kursi, Bak mandi, Pot Bunga	standar.jumbo set,minimalis, kaca	360	210	570	Bertambah/berkurang
10	Simpang Pematang	Okto	Guji, Meja Kursi, Bak mandi, Pot Bunga	standar.jumbo set,minimalis, kaca	540	30	570	Bertambah/berkurang
11	Kayu agung	Nov	Guji, Meja Kursi, Bak mandi, Pot Bunga	standar.jumbo set,minimalis, kaca	360	240	600	Bertambah/berkurang
12	Kenten	Des	Guji, Meja Kursi, Bak mandi, Pot Bunga	standar.jumbo set,minimalis, kaca	540	60	600	Bertambah/berkurang

4.2 Implementasi RapidMiner

RapidMiner merupakan salah satu software data mining pengolahan data set untuk mencari pola data sesuai dengan tujuan dari pengolahan data tersebut, tidak semua algoritma yang ada dapat sesuai atau dapat mengolah data set yang ada, harus dilakukan penyesuaian pola data dan sesuai dengan tujuan dari pengolahan data tersebut.



Gambar 5. Model Sub Proses *Clasification Decision Tree*

Adapun hasil Akurasi dari performance vector sebesar 75,00 % untuk proses yang di laksanakan pada model diatas dapat dilihat pada Gambar 5 :

	Isu Bertambah	Isu Berkurang	class presisi
pred Bertambah	5	1	83.33%
pred Berkurang	2	4	66.67%
class recall	71.43%	80.00%	

Gambar 6. Hasil Perhitungan Nilai Akurasi

Berdasarkan Gambar 6. hasil pemodelan yang telah di proses oleh tools RapidMiner selain menghasilkan dalam bentuk pola pemodelan ini juga dapat mengetahui keakuratan data. Keakuratan data yang didapatkan sebesar 75.00%.



Gambar 7. Model Klasifikasi Decision Tree

Gambar 7 menggambarkan model pengelompokan penjualan berdasarkan pengklasifikasian pada metode Decision Tree menggunakan tool RapidMiner.



Gambar 8. Model Sub Proses Clasification Naive Bayes

Adapun hasil Accuracy dari performance vector sebesar 100,00 % untuk proses yang di laksanakan pada model diatas dapat dilihat pada Gambar 9 :

accuracy: 100.00%

	Ya Beruntung	Ya Beruntung	class precision
pred Beruntung	5	0	100.00%
pred Beruntung	0	5	100.00%
class recall	100.00%	100.00%	

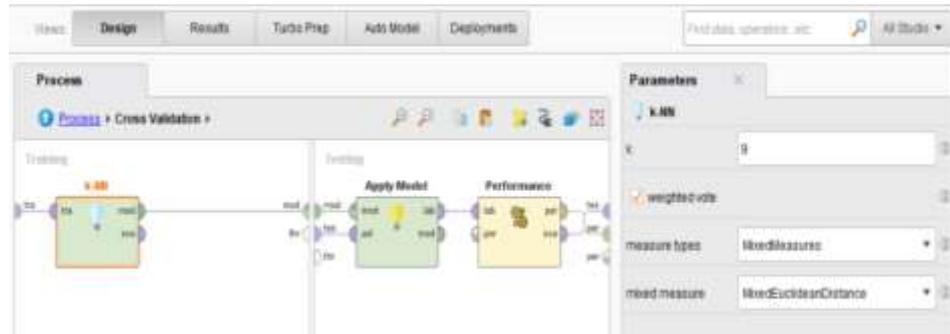
Gambar 9. Hasil Perhitungan Nilai Akurasi

Berdasarkan Gambar 9 hasil pemodelan yang telah di proses oleh tools RapidMiner didapatkan keakuratan data sebesar 100.00%.



Gambar 10. Model Klasifikasi Naive Bayes

Gambar 11. menggambarkan Simple Distribution penjualan berdasarkan pengklasifikasian pada metode Naive Bayes menggunakan tool RapidMiner.



Gambar 11. Model Sub Proses Clasification KNN

Adapun hasil Accuracy dari performance vector sebesar 90,00 % untuk proses yang di laksanakan pada model diatas dapat dilihat pada Gambar 12 :

	True Benarung	True Benarabah	class precision
pred. Benarung	6	2	75.00%
pred. Benarabah	0	4	100.00%
class recall	100.00%	66.67%	

Gambar 12. Hasil Perhitungan Nilai Akurasi

Berdasarkan Gambar 12 hasil pemodelan yang telah di proses oleh tools RapidMiner didapatkan keakuratan data sebesar 90.00%.

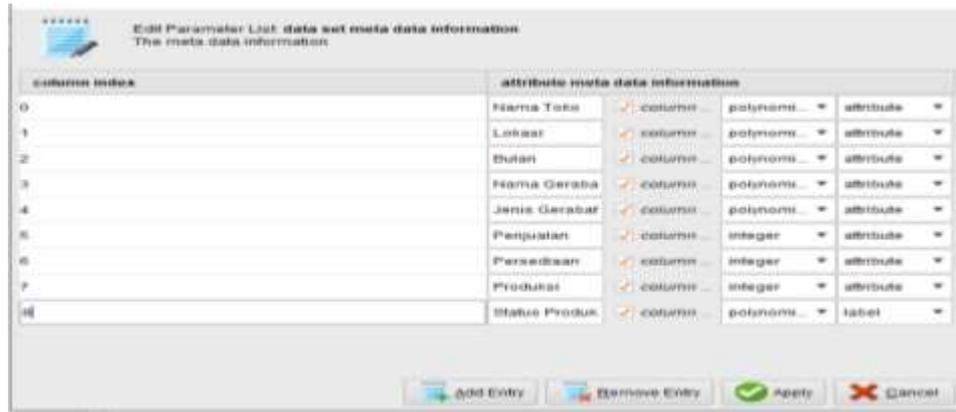


Gambar 13. Model Klasifikasi KNN

Gambar 13 menggambarkan KNN Classification penjualan berdasarkan pengklasifikasian pada metode *K-Nearest Neighbor* menggunakan tool RapidMiner.

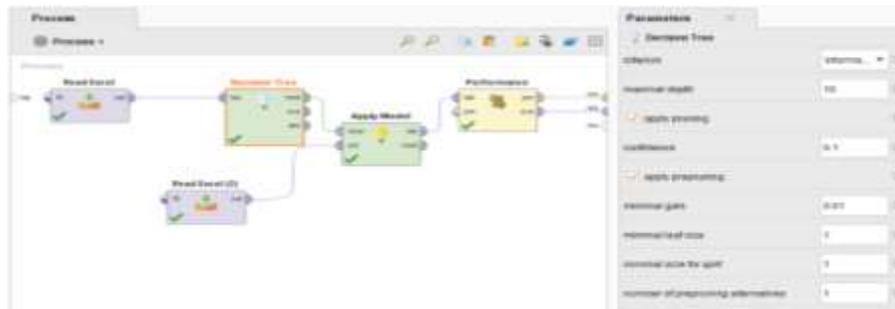
4.3 Susunan Operetor Algoritma

Setelah melakukan klasifikasi, tahap selanjutnya yang dilakukan adalah drag and drop operator read excel kemudian lakukan Importing tabel Microsoft Excel kedalam proses untuk membaca data training dalam format excel, kemudian ganti atribut target menjadi label. Dapat dilihat pada Gambar 14.

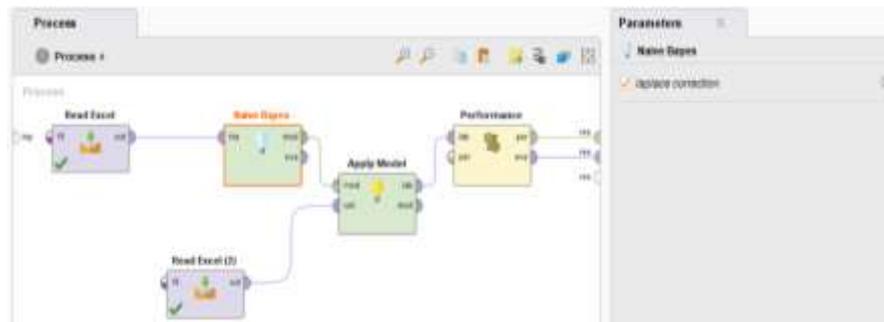


Gambar 14. Importing data Training Pada Operator Read Excel

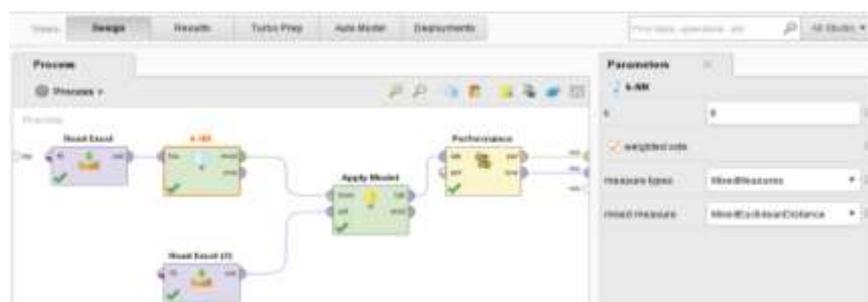
Selanjutnya drag and drop operator Decision Tree dengan mengganti criterion menjadi information_gain, minimal leaf size menjadi 1, minimal leaf size for split menjadi 1, number of prepruning alternatives menjadi 1. Kemudian hubungkan operator read excel dengan operator Decision Tree. Setelah itu hubungkan Decision Tree dengan operator apply model, kemudian ulangi drag operator read excel (2) dan importing tabel data testing, sambungkan pada operator apply model setelah itu hubungkan pada result. Dapat dilihat pada Gambar 15.



Gambar 15. Susunan Operator algoritma Decision Tree



Gambar 16. Susunan Operator algoritma Naive Bayes



Gambar 17. Susunan Operator algoritma KNN

4.4 Hasil Metode

4.4.1 Hasil Algoritma Decision Tree

Tahap ini adalah tahap terakhir dari proses data mining rapidminer dimana setelah semua operator terhubung kemudian klik icon play tombol F11, maka akan muncul sebuah tab result, yang isinya sebuah prediksi dari seluruh data yang memenuhi nilai criterion menjadi information_gain, minimal leaf size menjadi 1, minimal leaf size for split menjadi 1, number of prepruning alternatives menjadi 1, didapatkan hasil prediksi penjualan seperti gambar berikut :

Row No.	Status Produk	predicted/status...	confidence...	confidence...	Lokasi	Bulan	Nama Gerab...	Jenis Gerab...	Pengujian	Persediaan	Produksi
1	Bertambah/bekurang	Bekurang	1	0	Indragiri	Januari	Gup. Meja Ku...	standar.jumb...	300	90	450
2	Bertambah/bekurang	Bertambah	0	1	Siak 2	Februari	Gup. Meja Ku...	standar.jumb...	540	90	450
3	Bertambah/bekurang	Bekurang	1	0	Winggala	Maret	Gup. Meja Ku...	standar.jumb...	300	120	480
4	Bertambah/bekurang	Bertambah	0	1	Tequi Malya	April	Gup. Meja Ku...	standar.jumb...	540	80	480
5	Bertambah/bekurang	Bekurang	1	0	Jambi	Mei	Gup. Meja Ku...	standar.jumb...	300	150	510
6	Bertambah/bekurang	Bertambah	0	1	Bengkulu	Juni	Gup. Meja Ku...	standar.jumb...	540	30	510
7	Bertambah/bekurang	Bekurang	1	0	Padang	Juli	Gup. Meja Ku...	standar.jumb...	300	180	540
8	Bertambah/bekurang	Bertambah	0	1	Medan	Agustus	Gup. Meja Ku...	standar.jumb...	540	0	540
9	Bertambah/bekurang	Bekurang	1	0	Unit 2	September	Gup. Meja Ku...	standar.jumb...	300	210	570
10	Bertambah/bekurang	Bertambah	0	1	Simpang Pe...	Oktober	Gup. Meja Ku...	standar.jumb...	540	30	570
11	Bertambah/bekurang	Bekurang	1	0	Kayu Agung	November	Gup. Meja Ku...	standar.jumb...	300	240	600
12	Bertambah/bekurang	Bertambah	0	1	Kartapati	Desember	Gup. Meja Ku...	standar.jumb...	540	60	600

Gambar 18. Hasil Prediksi penjualan gerabah

Gambar 18 merupakan hasil prediksi penjualan gerabah dimana prediksi penjualan pada tahun 2020

4.4.2 Hasil Algoritma Naive Bayes

Tahap ini adalah tahap terakhir dari proses data mining rapidminer dimana setelah semua operator terhubung kemudian klik icon play tombol F11, maka akan muncul sebuah tab result, yang isinya sebuah prediksi dari seluruh data, didapatkan hasil prediksi penjualan seperti gambar berikut :

R.	Status Produk	predicted/status Prod...	confidence...	confidence...	Lokasi	Bulan	Nama Gerab...	Jenis Gerab...	Pengujian	Persediaan	Produksi
1	Bertambah/bekurang	Bekurang	1	0	Indragiri	Januari	Gup. Meja Ku...	standar.jumb...	300	90	450
2	Bertambah/bekurang	Bertambah	0	1	Siak 2	Februari	Gup. Meja Ku...	standar.jumb...	540	90	450
3	Bertambah/bekurang	Bekurang	1	0	Winggala	Maret	Gup. Meja Ku...	standar.jumb...	300	120	480
4	Bertambah/bekurang	Bertambah	0	1	Tequi Malya	April	Gup. Meja Ku...	standar.jumb...	540	80	480
5	Bertambah/bekurang	Bekurang	1	0	Jambi	Mei	Gup. Meja Ku...	standar.jumb...	300	150	510
6	Bertambah/bekurang	Bertambah	0	1	Bengkulu	Juni	Gup. Meja Ku...	standar.jumb...	540	30	510
7	Bertambah/bekurang	Bekurang	1	0	Padang	Juli	Gup. Meja Ku...	standar.jumb...	300	180	540
8	Bertambah/bekurang	Bertambah	0	1	Medan	Agustus	Gup. Meja Ku...	standar.jumb...	540	0	540
9	Bertambah/bekurang	Bekurang	1	0	Unit 2	September	Gup. Meja Ku...	standar.jumb...	300	210	570
10	Bertambah/bekurang	Bertambah	0	1	Simpang Pe...	Oktober	Gup. Meja Ku...	standar.jumb...	540	30	570
11	Bertambah/bekurang	Bekurang	1	0	Kayu Agung	November	Gup. Meja Ku...	standar.jumb...	300	240	600
12	Bertambah/bekurang	Bertambah	0	1	Kartapati	Desember	Gup. Meja Ku...	standar.jumb...	540	60	600

Gambar 19. Hasil Prediksi penjualan gerabah

Gambar 19 merupakan hasil prediksi penjualan gerabah dimana prediksi penjualan pada tahun 2020

4.4.3 Hasil Algoritma k-NN

Tahap ini adalah tahap terakhir dari proses data mining rapidminer dimana setelah semua operator terhubung kemudian klik icon play tombol F11, maka akan muncul sebuah tab result, yang isinya sebuah prediksi dari seluruh data yang memenuhi nilai $k=9$, didapatkan hasil prediksi penjualan seperti gambar berikut :

R.	Status Produksi	predicted(Status P..	confidence...	confidence...	Lokasi	Bulan	Nama Gerabah	Jenis Gerabah...	Penjualan	Persediaan	Prediksi
1	Bertambah/berkurang	Berurang	1	0	Indrataya	Januari	Gaj. Meja Kursi. Bat...	standar jumb...	350	90	450
2	Bertambah/berkurang	Bertambah	0	1	Muara 2	Februari	Gaj. Meja Kursi. Pot...	standar jumb...	540	90	450
3	Bertambah/berkurang	Berurang	1	0	Wengata	Maret	Gaj. Meja Kursi. Bat...	standar jumb...	350	120	480
4	Bertambah/berkurang	Bertambah	0	1000	Tugu Majo	April	Gaj. Meja Kursi. Pot...	standar jumb...	540	60	480
5	Bertambah/berkurang	Berurang	1	0	Jambi	Mei	Gaj. Meja Kursi. Pot...	standar jumb...	350	150	510
6	Bertambah/berkurang	Bertambah	0	1	Bengkulu	Juni	Gaj. Meja Kursi. Pot...	standar jumb...	540	30	510
7	Bertambah/berkurang	Berurang	1	0	Padang	Juli	Gaj. Meja Kursi. Pot...	standar jumb...	350	180	540
8	Bertambah/berkurang	Bertambah	0	1	Medan	Agustus	Gaj. Meja Kursi. Pot...	standar jumb...	540	0	540
9	Bertambah/berkurang	Berurang	1	0	Ulu 2	September	Gaj. Meja Kursi. Bat...	standar jumb...	350	210	570
10	Bertambah/berkurang	Bertambah	0	1	Simpang Ps...	Oktober	Gaj. Meja Kursi. Bat...	standar jumb...	540	30	570
11	Bertambah/berkurang	Berurang	1	0	Kayo agang	November	Gaj. Meja Kursi. Bat...	standar jumb...	350	240	600
12	Bertambah/berkurang	Bertambah	0	1	Kertapati	Desember	Gaj. Meja Kursi. Bat...	standar jumb...	540	60	600

Gambar 20. Hasil Prediksi penjualan gerabah

Gambar 20 merupakan hasil prediksi penjualan gerabah dimana prediksi penjualan pada tahun 2020

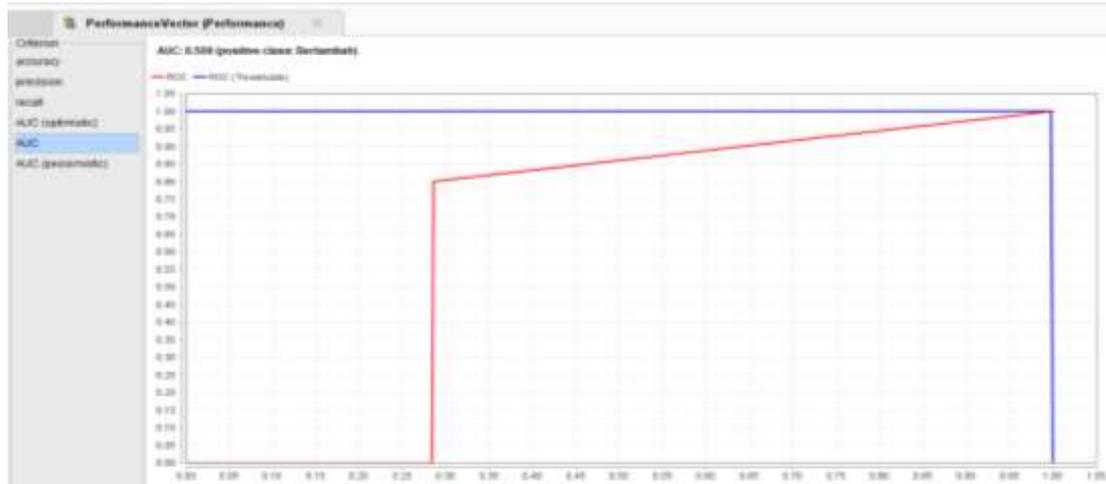
Hasil dari ketiga metode yang kita uji

Tabel 2. Hasil pengujian dari 3 metode data mining

Nama Metode	Accuracy
Decision Tree	75.00 %
Naive Bayes	100.00 %
<i>K-Nearest Neighbor</i>	90.00 %

Dari ketiga metode Decision Tree, Naive bayes dan KNN metode yang paling tinggi akurasi adalah metode Naive bayes dikarenakan metode naive bayes hanya mencentang parameters *laplace correction* saja dan didapatkan hasil prediksi penjualan 100.00%, untuk metode Decision tree mengubah nilai *criterion* menjadi *information_gain*, *minimal leaf size* menjadi 1, *minimal leaf size for split* menjadi 1, *number of prepruning alternatives* menjadi 1, didapatkan hasil prediksi penjualan 75.00%, dan untuk metode KNN mengubah nilai K menjadi 9, didapatkan hasil prediksi penjualan 90.00 %.

Tahap ini adalah tahap menampilkan grafik AUC dari 3 metode yang kita uji seperti gambar berikut
Tampilan grafik metode Decision Tree



Gambar 21. Grafik metode Decision Tree

Tampilan grafik metode Naive Bayes



Gambar 22. Grafik metode Naive Bayes

Tampilan grafik metode KNN



Gambar 23. Grafik metode KNN

5. KESIMPULAN

Pada penelitian ini dilakukan pengujian menggunakan 3 metode Decision Tree, Naive Bayes, dan KNN, dengan menggunakan data yang diolah berdasarkan tahapan Knowledge Discovery in Database (KDD). Berdasarkan hasil perhitungan data mining menggunakan teknik klasifikasi dan algoritma untuk metode Decision Tree mendapatkan akurasi 75,00 %, untuk metode Naive Bayes mendapatkan akurasi 100,00 %, dan untuk metode KNN mendapatkan

akurasi 90,00 %, berdasarkan hasil penelitian menunjukkan bahwa metode yang terbaik dari penelitian prediksi penjualan gerabah adalah metode *Naive Bayes*.

UCAPAN TERIMA KASIH

Alhamdulillah puji syukur kepada Allah swt, karena kehendak dan ridha-Nya peneliti dapat menyelesaikan penelitian ini. Peneliti sadari penelitian ini tidak akan selesai tanpa doa, dukungan dan dorongan dari berbagai pihak.

DAFTAR PUSTAKA

- ¹Azwanti, Nurul, 'Analisa Algoritma C4 . 5 Untuk Memprediksi Penjualan Motor Pada Pt . Capella Dinamik Nusantara', 13.1 (2018)
- Sma, D I, And Negeri Surakarta, 'Perbandingan 3 Metode Dalam Data Mining Untuk Prediksi Penerima Beasiswa Berdasarkan Prestasi Di Sma Negeri 6 Surakarta', I (2015), 15
- Yulia R.A, 2018. Penerapan Data Mining Untuk Prediksi Penjualan Produk Elektronik Terlaris Menggunakan Metode K-Nearest Neighbor (Studi Kasus : Pt. Bintang Multi Sarana Palembang), Program Studi Sistem Informasi Fakultas Sains Dan Teknologi Universitas Islam Negeri Raden Fatah Palembang
- Sabita, H., Fitria, F., & Herwanto, R. (2021). Analisa Dan Prediksi Iklan Lowongan Kerja Palsu Dengan Metode Natural Language Progaming Dan Machine Learning. *Jurnal Informatika*, 21(1), 14-22.
- Basriati, Sri, Elfira Safitri, Rahmawati, and Wahyu Wulandari, 'Penerapan Metode Fuzzy Sugeno Untuk Menentukan Jumlah Produksi Roti Optimum', *Seminar Nasional Teknologi Informasi, Komunikasi Dan Industri (SNTIKI) 11 Fakultas Sains Dan Teknologi, UIN Sultan Syarif Kasim Riau Pekanbaru*, November, 2019, 471-77
- Setiawan, M. (2017, October). Metode K-Means Untuk Sistem Informasi Pengelompokan Mahasiswa Baru Pada Perguruan Tinggi. In *Prosiding Seminar Nasional Darmajaya* (Vol. 1, No. 1, pp. 130-145).
- Juliana, Eki, and Ragil Kurniawan, 'Implementasi Metode Fuzzy Tsukamoto Dalam Memprediksi Jumlah Produksi Tmg', *Jurnal Ilmiah ILKOMINFO - Ilmu Komputer & Informatika*, 4.1 (2021), 9-15 <<https://doi.org/10.47324/ilkominfo.v4i1.107>>
- Kusumadewi, Sri dan Purnomo, Hari. "Aplikasi Logika Fuzzy". Graha Ilmu. Yogyakarta2010.
- Lodewyik, Dorteus R. "Penerapan Logika Fuzzy Metode Sugeno untuk Menentukan Jumlah Produksi Roti Berdasarkan Data Persediaan danJumlah Permintaan". Ilmu Matematika dan Terapan, Vol.9, No.2, hal. 121-134. 2015.
- Sutojo & dkk, (2011). Kecerdasan Buatan. Yogyakarta: Andi.
- Hidayat Nurul, Suwandi Galih Putra, Sistem Diagnosis Penyakit Mata Menggunakan Metode Fuzzy Tsukamoto. *Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer - Vol. 3, No. 4, April 2019, ISSN: 3531-3537 (Print) 2548-964X (Online)*
- Maryaningsih, Siswanto & Mesterjon, Metode Logika Fuzzy Tsukamoto Dalam Sistem Pengambilan Keputusan Penerimaan Beasiswa. *JurnalMedia Infotama*, 9(1 february 2013), pp. 1-13. 2013.
- Reflan, Nuari, Apriyuni Aflahah, Kusri, and Juwari, 'Implementasi Metode K-Nearest Neighbor (Knn) Untuk Memprediksi Varietas Padi Yang Cocok Untuk Lahan Pertanian', *Jurnal Informa Politeknik Indonusa Surakarta*, 4 (2018), 2-8
- Sma, D I, and Negeri Surakarta, 'Perbandingan 3 Metode Dalam Data Mining Untuk Prediksi Penerima Beasiswa Berdasarkan Prestasi Di Sma Negeri 6 Surakarta', I (2015), 15
- Waliyansyah, Rahmat Robi, and Citra Fitriyah, 'Perbandingan Akurasi Klasifikasi Citra Kayu Jati Menggunakan Metode Naive Bayes Dan K-Nearest Neighbor (k-NN)', *Jurnal Edukasi Dan Penelitian Informatika (JEPIN)*, 5.2 (2019), 157 <<https://doi.org/10.26418/jp.v5i2.32473>>
- Wibowo, H., & Indriyani, F. (2018, October). K-Nearest Neighbor Method For Monitoring Of Production And Preservation Information (Treatment) Of Rubber Tree Plant. In *International Conference on Information Technology and Business (ICITB)* (pp. 29-44).