

Analisis Pemberian Bantuan UMKM Menggunakan Algoritma K-NN dan C4.5

Akrim Teguh Suseno^{*}, M Al Amin, Fajar Mahardika

Fakultas Sains dan Teknologi, Program Studi Teknologi Komputer, Institut Teknologi dan Sains Nahdlatul Ulama
Pekalongan, Pekalongan, Indonesia

email: ^{1*}akrim.teguh@gmail.com, ²aimoh@gmail.com, ³fajarmahardika@itsnupekalongan.ac.id

Email Penulis Korespondensi: akrim.teguh@gmail.com

Submitted: 21/09/2023; Accepted: 30/11/2023; Published: 30/11/2023

Abstrak—Pandemi Covid-19 memberikan dampak terhadap ekonomi negara di dunia termasuk Indonesia. Rendahnya neraca perdagangan, konsumsi masyarakat dan diberlakukannya pembatasan sosial berskala besar (PSBB) membuat para pelaku usaha yang merugi hingga gulung tikar khususnya Usaha Mikro, Kecil dan Menengah (UMKM). Bantuan terhadap UMKM telah diberikan dari pemerintah untuk mempertahankan berjalannya ekonomi di lingkungan mikro. Namun pemberian bantuan tersebut perlu dilakukan analisis lebih lanjut karena terdapat laporan bahwa bantuan tidak tepat sasaran sehingga menyebabkan anggaran yang dikeluarkan menjadi tidak efektif. Pada penelitian ini akan dilakukan analisis terhadap pemberian bantuan dari pemerintah terhadap UMKM khususnya di kabupaten Pekalongan dengan menggunakan teknik data mining khususnya klasifikasi. Algoritma yang digunakan yaitu *K-Nearest Neighbor* (K-NN) dan C.45 yang kemudian dibandingkan untuk menentukan tingkat efektifitas paling tinggi dalam rekomendasi pemberian bantuan UMKM di Kab. Pekalongan. Data yang digunakan berjumlah 312 UMKM dan setelah melewati proses data preprocessing sehingga mendapatkan data yang akurat yaitu sebesar 279. Data tersebut dibagi menjadi 2 yaitu data training sebesar 200 data dan data testing sebesar 79 data. Hasil dari penelitian ini, pada algoritma K-NN mendapatkan tingkat accuracy sebesar 94,94%, precision 94,73 % dan recall 94,73% sedangkan untuk algoritma C.45 mendapatkan tingkat accuracy sebesar 86,08%, precision 87,21% dan recall 88,3%. Berdasarkan hasil penelitian tersebut maka dapat disimpulkan bahwa penggunaan teknik data mining dengan algoritma K-NN dan C.4.5 memiliki tingkat akurasi yang tinggi untuk rekomendasi bantuan kepada UMKM namun untuk hasil terbaik dapat menggunakan algoritma K-NN yang memiliki tingkat akurasi sebesar 94,94%.

Kata Kunci: Data Mining; K-NN; C4.5; UMKM; Klasifikasi

Abstract—The Covid-19 pandemic has had an impact on the economies of countries in the world, including Indonesia. The low trade balance, public consumption and the implementation of large-scale social restrictions (PSBB) have caused business actors to suffer losses and even go out of business, especially Micro, Small and Medium Enterprises (MSMEs). Assistance to MSMEs has been provided by the government to maintain the functioning of the economy in the micro environment. However, the provision of this assistance needs to be carried out further analysis because there are reports that the assistance is not on target, causing the budget spent to be ineffective. In this research, an analysis will be carried out on the provision of assistance from the government to MSMEs, especially in Pekalongan district, using data mining techniques, especially classification. The algorithms used are K-Nearest Neighbor (K-NN) and C.45 which are then compared to determine the highest level of effectiveness in recommendations for providing MSME assistance in the District. Pekalongan. The data used was 312 MSMEs and after going through the data preprocessing process we got accurate data, namely 279. The data was divided into 2, namely training data of 200 data and testing data of 79 data. The results of this research, the K-NN algorithm obtained an accuracy level of 94.94%, precision 94.73% and recall 94.73%, while the C.45 algorithm obtained an accuracy level of 86.08%, precision 87.21% and recall 88.3%. Based on the results of this research, it can be concluded that the use of data mining techniques with the K-NN and C.4.5 algorithms has a high level of accuracy for recommendations for assistance to MSMEs, however for the best results you can use the K-NN algorithm which has an accuracy level of 94.94% .

Keywords: Data Mining; K-NN; C4.5; MSMEs; Classification

1. PENDAHULUAN

Pada tahun 2020 banyak negara didunia terkena pandemi covid-19 yang menyebabkan banyak kerugian dari berbagai sektor salah satunya adalah ekonomi. Sebagian besar negara mengalami dampak inflasi yang luar biasa termasuk Indonesia. Kondisi pertumbuhan ekonomi di Indonesia pada triwulan II tahun 2020 turun hingga - 5,32%. Sektor ekonomi yang sangat terdapat adalah Usaha Mikro, Kecil dan Menengah (UMKM) karena pemerintah memberlakukan Pembatasan Sosial Berskala Besar (PSBB) untuk membatasi persebaran covid-19 sehingga penjualan menurun drastis dan tidak sedikit yang gulung tikar. Pada tahun 2022 perkembangan ekonomi Indonesia membaik karena berbagai kebijakan pemerintah dalam peningkatan ekonomi negara. Salah satu kebijakan untuk pemulihan ekonomi di Indonesia yang telah dilakukan pemerintah dengan memberikan bantuan kepada UMKM yang terdampak pandemi covid-19 dengan total bantuan 258 triliun rupiah yang disalurkan kepada 3,59 juta UMKM di Indonesia. Namun pada tahun 2023 terdapat ancaman penurunan terhadap kondisi ekonomi yang dapat menimbulkan inflasi negara. Hal tersebut disebabkan oleh banyaknya negara yang menaikkan suku bunga untuk mempertahankan kondisi keuangan negara akibat pandemi covid-19. Suku bunga yang paling berdampak di banyak negara adalah kenaikan suku bunga The Fed Amerika Serikat yang Mei 2023 telah menaikkan hingga 5%. Hal tersebut dapat memicu kenaikan harga di berbagai sektor khususnya pada bahan import dari material hingga kebutuhan pokok. Oleh sebab itu, dalam kondisi tersebut kemungkinan pemerintah akan memberikan solusi untuk menjaga kestabilan ekonomi di Indonesia salah satunya memberikan

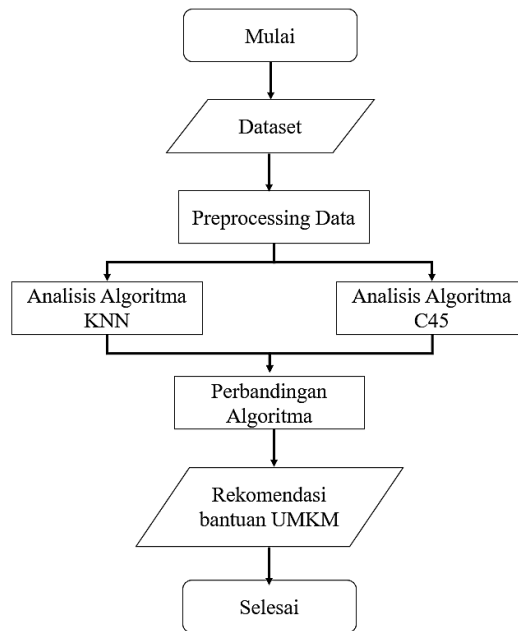
bantuan kepada para pelaku usaha yang terdampak khususnya UMKM. Namun dalam pemberian bantuan masih terdapat permasalahan yaitu berdasarkan bantuan UMKM di pandemi covid-19 masih banyaknya laporan bahwa bantuan UMKM kurang tepat sasaran dan pemerintah pun kesulitan untuk menganalisis kesesuaian bantuan kepada para pelaku UMKM [1], [2]. Hal tersebut didukung adanya sentimen masyarakat yang terhadap bantuan covid-19 yang cenderung negatif[3]. Padahal implementasi bantuan dari pemerintah berpengaruh terhadap perkembangan dan kemajuan UMKM di berbagai daerah di Indonesia [4]–[6]. Oleh sebab itu diperlukan analisis untuk mengukur tingkat akurasi dalam pemberian bantuan oleh pemerintah sehingga nantinya bantuan yang diberikan lebih tepat sasaran. Teknik analisis tersebut dapat dilakukan dengan disiplin ilmu komputer dengan menemukan pola-pola baru dari suatu data yang besar melalui metode *statistics*, *machine learning*, *artificial intelligence* dan *database systems* yang disebut dengan data mining[7]. Data mining terdapat 5 peran utama salah satunya adalah klasifikasi. Klasifikasi memiliki beberapa model yang dapat digunakan dalam penyelesaian masalah yaitu *neural network*, *support vector machine*, algoritma genetik, *memory-based reasoning*, *rough set*, *K-nearest neighbor (K-NN)*, *decision tree C4.5*, dan metode berbasis aturan[8].

Beberapa penelitian yang menggunakan teknik data mining klasifikasi telah dilakukan terkait pemberian bantuan seperti pada penelitian Ari Rohmawati dan Damuri et al. yang telah melakukan penelitian terkait pemberian bantuan kepada UMKM di Kabupaten Pesawaran Provinsi Lampung menggunakan algoritma *naïve bayes* dengan menggunakan sampel data dari penerima UMKM. Hasil dari penelitian tersebut yaitu prediksi dari akurasi penerima bantuan yang sesuai kriteria dengan algoritma C4.5 mendapatkan nilai akurasi hingga 98,68% dan 86% [9][10]. Pada penelitian Sastri Yani et al. yang menganalisis penerimaan Bantuan Pangan Non Tunai (BPNT) di Kelurahan Karya Jaya, Palembang menggunakan algoritma *K-nearest neighbor (K-NN)* dikarenakan penyaluran bantuan yang dilakukan oleh pemerintah kurang maksimal. Hasil dari analisis tersebut adalah dari 511 data dapat diprediksi bahwa tingkat akurasi dari penerimaan bantuan yang sesuai kriteria pada penggunaan algoritma K-NN ini adalah 75,79% [11]. Pada penelitian lainnya, menggunakan perbandingan algoritma *decision tree C4.5* dan algoritma *decision tree C4.5* berbasis *Particle Swarm Optimization (PSO)* tanpa pruning. Hasil dari analisis tersebut bahwa algoritma C4.5 mendapatkan tingkat akurasi 98,56% dan AUC sebesar 0,979. Sedangkan setelah dilakukan optimasi dengan algoritma *decision tree C4.5* berbasis *Particle Swarm Optimization (PSO)* tanpa pruning menghasilkan tingkat akurasi 98,92% dan AUC sebesar 0,98. Sehingga dapat disimpulkan bahwa algoritma *decision tree C4.5* berbasis *Particle Swarm Optimization (PSO)* tanpa pruning dapat meningkatkan akurasi [12]. Pada penelitian yang dilakukan oleh Hasanah et al yang menggunakan algoritma *K-Nearest Neighbors (K-NN)* dengan pengujian *K-Fold Cross Validation* untuk penerimaan dana bantuan desa. Data yang digunakan sejumlah 159 record dan menggunakan 4 atribut. Hasil dari analisis tersebut dengan menggunakan perhitungan *Euclidean Distance* pada 5 nilai K yang berbeda, menghasilkan tingkat akurasi sebesar 81,25% [13]. Pada penelitian yang dilakukan oleh penulis, akan melakukan penelitian dengan menggunakan algoritma K-NN dan *decision tree C4.5* karena masing-masing algoritma tersebut memiliki tingkat akurasi yang berbeda pada *study case* yang berbeda, namun belum ada penelitian yang melakukan perbandingan antara algoritma tersebut pada *study case* yang sama. Penggunaan algoritma K-NN dikarenakan mempunyai kelebihan dapat menganalisis data training yang memiliki banyak noisy dan efektif apabila memiliki data training yang cukup banyak [14]. Selain itu penggunaan algoritma *K-nearest neighbor (K-NN)* juga telah dilakukan oleh penelitian sebelumnya terhadap bantuan UMKM namun belum ada perbandingan terhadap algoritma *decision tree C4.5*. Sedangkan algoritma *decision tree C4.5* memiliki keunggulan pada hasil *decision tree* yang mudah dipahami serta memiliki tingkat akurasi dan efisiensi yang baik dalam analisis data bertipe diskrit maupun numerik [15]. Hal tersebut membantu peneliti dalam memprediksi bantuan UMKM yang memiliki data numerik sehingga hasil analisis dapat menjadi lebih akurat. Hasil dari penelitian ini akan memberikan informasi perbandingan terkait prediksi bantuan UMKM di Kab. Pekalongan yang lebih efektif dengan menggunakan algoritma K-NN atau menggunakan algoritma *decision tree C4.5*.

2. METODOLOGI PENELITIAN

2.1 Tahapan Penelitian

Penelitian ini menggunakan metode kuantitatif yang menggunakan data UMKM yang didalamnya terdapat atribut bidang usaha, omset, kriteria UMKM dan lainnya. Data tersebut akan diolah melalui *preprocessing* data, kemudian akan di bagi menjadi data training dan data testing untuk diuji coba menggunakan algoritma K-NN dan C4.5 sehingga dapat dibandingkan dan hasilnya akan membentuk suatu rekomendasi untuk pemberian bantuan tersebut. Berikut adalah alur yang digunakan dalam penelitian ini yang dapat ditunjukkan pada Gambar 1.



Gambar 1. Alur Penelitian

Pada Gambar 1 merupakan tahapan pada penelitian yang dimulai dengan pengumpulan dataset, preprocessing, analisis algoritma hingga rekomendasi bantuan UMKM :

a. Pengumpulan Data

Tahap ini dilakukan pengambilan data UMKM di dinas koperasi dan UKM Kabupaten Pekalongan atau tingkat Provinsi Jawa Tengah. Data yang digunakan yaitu jumlah UMKM di Kab. Pekalongan, bidang usaha, omzet per tahun, kriteria usaha dan lainnya selama tahun 2020 hingga 2021.

a. Preprocessing data

Data yang telah dikumpulkan maka dilakukan *preprocessing* data. Pada proses ini, data yang mengandung duplikasi, tidak sesuai format, tidak konsisten maupun data kosong akan di proses menggunakan teknik *data cleaning* sehingga data tersebut sesuai dengan kebutuhan analisis.

b. Analisis Algoritma

Analisis menggunakan 2 algoritma yaitu *K-Nearest Neighbor* (K-NN) dan *Decision Tree C4.5*. Masing-masing algoritma akan melakukan perhitungan menggunakan data hasil *preprocessing* yang dipecah menjadi 2 bagian yaitu data training dan data testing. Hasil dari perhitungan algoritma merupakan nilai *accuracy*, *recall* dan *precision*.

c. Perbandingan algoritma

Pada tahap ini akan dilakukan perbandingan algoritma K-NN dan C4.5 dalam tingkat akurasi menggunakan fitur *performance* yang mencari nilai *accuracy*, *recall* dan *precision* di setiap algoritma klasifikasi yang digunakan. Perhitungan nilai tersebut menggunakan perbandingan antara nilai *actual values* dan *predicted values*. *Actual values* merupakan nilai asli sebelum dilakukan prediksi sedangkan *predicted values* merupakan nilai hasil analisis atau nilai prediksi . berikut adalah konsep dari perhitungan *accuracy*, *recall* dan *precision* yang ditunjukkan pada gambar 2.

		Actual Values	
		Positive (1)	Negative (0)
Predicted Values	Positive (1)	TP	FP
	Negative (0)	FN	TN

Gambar 2. Perbandingan nilai *accuracy*, *recall* dan *precision*

Perhitungan nilai *accuracy*, *recall* dan *prediction* adalah sebagai berikut :

$$accuracy = (TP + TN) / (TP + FP + FN + TN) \tag{1}$$

$$recall = (TP) / (TP + FN) \tag{2}$$

$$prediction = (TN) / (TN + FP) \tag{3}$$

Keterangan :

TP : True Positive

FP : False Positive

FN : False Negative

TN : True Negative

d. Rekomendasi bantuan UMKM

Pada tahap ini adalah hasil dari analisis algoritma yang terbaik sehingga dapat menjadi rekomendasi dari Dinas Koperasi dan UMKM dalam meningkatkan efektifitas bantuan pemerintah kepada UMKM.

2.2 Algoritma K-Nearest Neighbor (K-NN)

K-Nearest Neighbor (K-NN) merupakan salah satu peran utama dalam data mining yaitu klasifikasi yang cara kerjanya mencari jarak yang paling dekat diantara data yang akan di evaluasi melalui k neighbor yang terdekat di suatu data. Kedekatan didefinisikan pada jarak Euclidean atau jarak metrik. Untuk mengukur jarak Euclidean tersebut dapat dilakukan melalui persamaan berikut :

Dxy = sqrt(sum_{i=1}^n (x_i - y_i)^2) (4)

Keterangan :

: jarak kedekatan

: jumlah atribut individu

: training data

: testing data

: Atribut individu

Terdapat beberapa langkah untuk menghitung metode K Nearest Neighbor antara lain :

- 1. Memilih K neighbor yang terdekat
2. Menghitung jarak Euclidean dari jumlah K yang terdekat
3. Menghitung jumlah titik data dan membuat kelompok dari objek yang diurutkan sehingga mempunyai jarak Euclidean yang paling kecil.
4. Mengumpulkan kategori Y (Klasifikasi Nearest Neighbor)
5. Memprediksi nilai query instance melalui kategori dari nearest neighbor.

2.3 Algoritma Decision Tree C4.5

Metode klasifikasi yang menggunakan pohon struktur dari sekumpulan keputusan yang telah disusun[16]. Algoritma C4.5 menggunakan prinsip kerja dari algoritma ID3 yang dapat mengatasi continue data, pruning dan missing value pada data[17]. Penggunaan algoritma ini mempunyai beberapa langkah yang dapat ditunjukkan pada rumus-rumus berikut[18] :

- 1. Menghitung nilai entropy dengan melakukan persamaan:

Entropy(S) = sum_{i=1}^n p_i * log_2 p_i (5)

Keterangan:

S = jumlah kasus

n = total himpunan S

p_i = proporsi ke-i

- 2. Menghitung nilai Gain dengan melakukan persamaan:

Gain(S, A) = Entropy(S) - sum_{i=1}^n (|S_i|/|S|) * Entropy(S_i) (6)

Keterangan:

S : jumlah kasus

A : karakter

n : total A

|S_i| : jumlah kasus pada partisi ke i

|S| : Jumlah kasus dalam S

- 3. Menghitung nilai Split Info melalui persamaan:

SplitInfo(S, A) = - sum_{i=1}^n (|S_i|/|S|) * log_2 (|S_i|/|S|) (7)

Keterangan:

S = jumlah kasus

A = karakter

S_i = jumlah sampel karakter i

- 4. Untuk menentukan sebuah atribut dapat dijadikan akar maupun cabang suatu pohon keputusan didapat dari nilai Gain Ratio dengan persamaan:

$$GainRatio(S, A) = \frac{Gain(S,A)}{SplitInfo(S,A)} \quad (8)$$

Keterangan:

$Gain(S,A)$ = info gain pada karakter A

$SplitInfo(S,A)$ = split info pada karakter A

Beberapa kelebihan menggunakan algoritma C4.5 yaitu mampu mengatasi atribut dengan type data diskrit, hasil akurasi yang baik, dan dapat menginterpretasikan *decision tree* dengan mudah[19]

3. HASIL DAN PEMBAHASAN

Pada bagian ini berisi hasil dan pembahasan dari topik penelitian berdasarkan tahapan penelitian yang telah dibuat sebelumnya :

3.1 Pengumpulan Data

Pada penelitian ini menggunakan data UMKM Kab. Pekalongan dari tahun 2020 hingga tahun 2021 yang berjumlah 312 UMKM. Data yang didapatkan terdiri atribut nama usaha, no ktp, nama pemilik, nama desa, jenis usaha, klasifikasi, jumlah aset dan omset per bulan. Data tersebut dapat ditunjukkan pada Tabel 1.

Tabel 1. Data UMKM

No	Nama Usaha	No KTP	Nama Pemilik	...	Omset
1	Mekar Sari Abadi	xxx	Nurdin	...	57.000.000
2	Obet Farm	xxx	Ubaidillah	...	8.250.000
3	Kain Perca Batik	xxx	M.zamroni	...	8.000.000
4	Ud.Syafella Batik	xxx	M.aris musyafa'	...	14.000.000
5	Perdagangan Mori	xxx	Mustaqim	...	16.000.000
6	Konveksi	xxx	Muniroh	...	10.000.000
7	Nadia Snack	xxx	Wiwik sulistyani	...	6.250.000
8	Citras	xxx	Ika sagita	...	8.000.000
9	Ud.Muncul Jaya	Xxx	Dulbari	...	15.000.000
...
312	Toko Ropis	Xxx	Zaenal	...	5.000.000

Pada Tabel 1 menjelaskan bahwa terdapat berbagai jenis sektor usaha dan omset maupun aset yang bervariasi. Data tersebut kemudian akan dibagi menjadi 2 yaitu data training dan data testing. Data training yang digunakan yaitu 200 data UMKM yang berfungsi untuk melatih model sehingga mendapatkan pola dari hubungan algoritma dengan data yang diinputkan[20]. Sedangkan data testing yang digunakan adalah 79 data UMKM. Data ini digunakan setelah pola yang didapatkan dari model yang dilatih menggunakan data training, kemudian dilakukan evaluasi menggunakan data testing untuk mengukur seberapa baik model atau pola yang ada[21].

3.2 Data Preprocessing

Pada proses *preprocessing* dilakukan fase *feature extraction* yang menghapus data ataupun atribut yang tidak diperlukan pada analisis selain itu juga dilakukan fase data *cleaning* terhadap data-data yang tidak sesuai dengan format maupun data yang kosong. Seperti pada dataset yang memiliki atribut nama usaha, no ktp, nama pemilik, nama desa, jenis usaha, klasifikasi, jumlah aset dan omset per bulan, namun hanya atribut nama usaha, sektor usaha, aset dan omset yang diperlukan untuk proses analisis sehingga atribut no ktp, nama desa dan klasifikasi perlu dilakukan proses *feature extraction* yang bertujuan menghapus atribut yang tidak digunakan tersebut. Selain itu terdapat data yang tidak lengkap dan *noisy* pada beberapa UMKM yang tidak memiliki data omset atau aset padahal data tersebut merupakan data primer yang diperlukan untuk analisis tersebut maka dari itu data yang tidak lengkap tersebut perlu dilakukan *data cleaning* untuk memperbaiki atau menghapus data yang tidak sesuai sehingga tersisa data-data yang sudah lengkap dan sesuai untuk di proses pada tahap analisis. Berikut adalah data hasil preprocessing yang dapat ditunjukkan pada Tabel 2.

Tabel 2. Data hasil data *preprocessing*

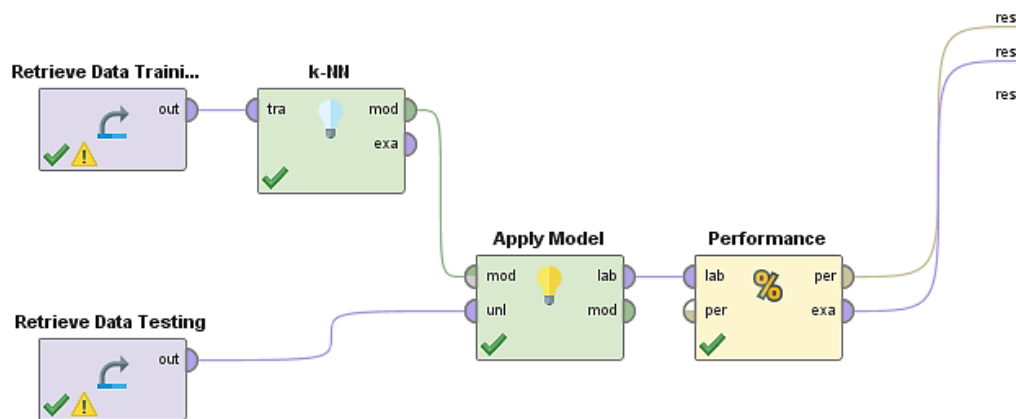
No	Nama Usaha	Sektor Usaha	Aset	Omset
1	Mekar Sari Abadi	Perdagangan & Reparasi	200.000.000	57.000.000
2	Obet Farm	Perikanan	21.000.000	8.250.000
3	Kain Perca Batik	Lainnya	25.000.000	8.000.000
4	Ud.Syafella Batik	Lainnya	50.000.000	14.000.000
5	Perdagangan Mori	Lainnya	35.000.000	16.000.000
6	Konveksi	Perdagangan dan Reparasi	23.000.000	10.000.000

No	Nama Usaha	Sektor Usaha	Aset	Omset
7	Nadia Snack	Lainnya	10.000.000	4.250.000
8	Citras	Lainnya	25.000.000	8.000.000
9	Ud.Muncul Jaya	Perdagangan dan Reparasi	50.000.000	15.000.000
...
279	Jajalen	Makanan/ Minuman	60.000.000	10.000.000

Pada tabel 2 menunjukan hasil yang berjumlah 279 dari 312 data UMKM Kab. Pekalongan yang diambil pada tahun 2020 hingga 2021. Untuk atribut hasil dari *feature extraction* berjumlah 4 yaitu nama usaha, sektor usaha, aset dan omset. Data tersebut kemudian dibagi menjadi 2 yaitu data training dan data testing yang akan digunakan pada proses analisis algoritma K-NN maupun C4.5.

3.3 Analisis algoritma K-NN

Pada algoritma K-NN, analisis data menggunakan *tools rapid miner* yang dapat membantu melakukan perhitungan algoritma K-NN secara cepat sesuai dengan rumus yang semestinya. Analisis algoritma K-NN dapat ditunjukkan pada Gambar 3.



Gambar 3. Analisis algoritma K-NN

Pada gambar 3 analisis menggunakan 2 data yaitu data training dan data testing. Data training yang berjumlah 200 data UMKM membentuk pola atau model dengan algoritma K-NN. Pola tersebut kemudian diuji coba dengan data testing yang berjumlah 79 melalui fitur apply model. Hasil dari data testing merupakan prediksi kelayakan dan tidak untuk bantuan UMKM. Kemudian hasil prediksi tersebut dilakukan analisis untuk mengukur tingkat keakuratan model prediksi melalui fitur performance yang dapat memberikan nilai akurasi, presisi dan recall pada model tersebut. Hasil dari analisis K-NN dapat ditunjukkan pada Tabel 3.

Tabel 3. Hasil analisis K-NN

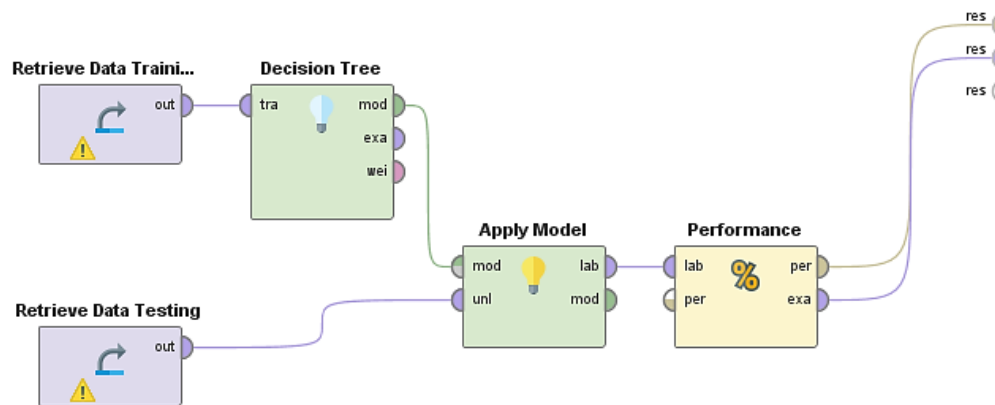
No	Nama Usaha	Sektor Usaha	Aset	Omset	Aktual	Prediksi
1	Sale Mba' Dah	Industri Pengolahan	35.000.000	49.000.000	Not approved	Not approved
2	Syakira Brownis & Cake	Perdagangan, Hotel, Resto	50.000.000	66.000.000	Not approved	Not approved
3	Batik Mas Agunx-Q	Perdagangan, Hotel, Resto	75.000.000	67.000.000	Not approved	Not approved
4	Aura Snack	Industri Pengolahan	35.000.000	36.000.000	Not approved	Not approved
5	Abid Snack	Industri Pengolahan	25.000.000	34.000.000	Not approved	Not approved
6	Mr Bronzee	Industri Pengolahan	50.000.000	75.000.000	Not approved	Not approved
7	Bawang Goreng	Industri Pengolahan	35.000.000	20.000.000	Approved	Approved
8	Fashion	Perdagangan, Hotel, Resto	70.000.000	73.000.000	Not approved	Not approved
9	Casmini Batik	Industri Pengolahan	25.000.000	35.000.000	Not approved	Not approved
10	Concept Rumah	Listrik, Air, Gas	15.000.000	12.750.000	Approved	Approved

No	Nama Usaha	Sektor Usaha	Aset	Omset	Aktual	Prediksi
Kreatif						
...
79	Isnaini	Makanan/ Minuman	10.000.000	8.500.000	Approved	Approved

Pada tabel 3 merupakan hasil dari analisis algoritma K-NN yang menunjukkan perbandingan antara nilai aktual dan prediksi pada atribut ke-5 dan 6. Pada atribut aktual merupakan hasil dari penyaluran bantuan yang telah diberikan pemerintah pada tahun 2020 hingga 2021 di UMKM Kab. Pekalongan. Sedangkan attribute prediksi merupakan hasil dari analisis menggunakan algoritma K-NN yang menghasilkan tingkat *accuracy* 94%, *precision* 94,74% dan *recall* 94,7%. Sehingga dari hasil tersebut dapat disimpulkan bahwa klasifikasi bantuan UMKM dengan menggunakan teknik data mining algoritma K-NN menghasilkan *accuracy* sebesar 94% dan tingkat error sebesar 5,06% atau dari 79 data yang di analisis terdapat 4 data yang tidak sesuai dengan kondisi aslinya dan 75 data telah dilakukan prediksi dengan tepat.

3.4 Analisis algoritma C4.5

Pada analisis algoritma C4.5 akan dilakukan analisis dengan data yang sama pada analisis pada algoritma K-NN yang menggunakan data training sejumlah 200 data dan data testing yang berjumlah 79 data. Analisis ini menggunakan bantuan *tools rapid miner* yang dapat ditunjukkan pada Gambar 4.



Gambar 4. Analisis Algoritma C4.5

Pada gambar 4 menunjukkan data training sebagai input untuk analisis algoritma C4.5. Pada prosesnya *tools rapid miner* akan melakukan perhitungan sesuai dengan rumus algoritma sehingga membentuk pohon keputusan atau pola yang kemudian diimplementasikan menggunakan fitur *apply model* dengan data testing yang berjumlah 79. Hasil Implementasi tersebut menghasilkan data prediksi yang dapat ditunjukkan pada Tabel 4. Sedangkan pada fitur *performance* akan memberikan informasi tingkat akurasi, precision dan recall dari analisis ini.

Tabel 4. Hasil analisis C4.5

No	Nama Usaha	Sektor Usaha	Aset	Omset	Aktual	Prediksi
1	Sale Mba' Dah	Industri Pengolahan	35.000.000	49.000.000	Not approved	Not approved
2	Syakira Brownis & Cake	Perdagangan, Hotel, Resto	50.000.000	66.000.000	Not approved	Not approved
3	Batik Mas Agunx-Q	Perdagangan, Hotel, Resto	75.000.000	67.000.000	Approved	Not approved
4	Aura Snack	Industri Pengolahan	35.000.000	36.000.000	Not approved	Not approved
5	Abid Snack	Industri Pengolahan	25.000.000	34.000.000	Not approved	Not approved
6	Mr Bronzee	Industri Pengolahan	50.000.000	75.000.000	Not approved	Not approved
7	Bawang Goreng	Industri Pengolahan	35.000.000	20.000.000	Approved	Approved
8	Fashion	Perdagangan, Hotel, Resto	70.000.000	73.000.000	Not approved	Not approved
9	Casmini Batik	Industri	25.000.000	35.000.000	Not approved	Not approved

No	Nama Usaha	Sektor Usaha	Aset	Omset	Aktual	Prediksi
		Pengolahan				
10	Concept Rumah Kreatif	Listrik, Air, Gas	15.000.000	12.750.000	Approved	Approved
...
79	Isnaini	Makanan/ Minuman	10.000.000	8.500.000	Approved	Approved

Pada tabel 4 menunjukkan hasil dari analisis menggunakan algoritma *decision tree* C4.5 yang menggunakan 79 data testing. Hasil pengujian tersebut membandingkan data aktual yang merupakan data dari bantuan yang pemerintah pada tahun 2020 hingga 2021 dan data prediksi merupakan hasil dari analisis algoritma C4.5. Hasil dari prediksi tersebut dari 79 data terdapat 11 data yang tidak sesuai dengan data aktual atau memiliki tingkat error sebesar 13,92%. Namun terdapat 68 data yang sesuai dengan data aktual yang berarti memiliki tingkat akurasi sebesar 86%. Sedangkan untuk nilai *precision* berjumlah 87% dan *recall* sebesar 88%.

3.5 Perbandingan algoritma

Berdasarkan hasil dari analisis data mining menggunakan algoritma *K-Nearest Neighbor* (K-NN) dan algoritma *Decision Tree C4.5* menunjukkan bahwa K-NN menghasilkan tingkat menghasilkan *accuracy* 94%, *precision* 94,74% dan *recall* 94,7%. Hasil tersebut menggambarkan bahwa dari 79 data yang di analisis terdapat 4 data yang not approved dengan kondisi aslinya. Sedangkan untuk algoritma *Decision Tree C4.5* memiliki *accuracy* 86%, *precision* 87%, *recall* 88% yang menunjukkan bahwa dari 79 data testing maka terdapat 11 data yang not approved dengan kondisi aslinya. Oleh sebab itu hasil dari perbandingan antara algoritma *K-Nearest Neighbor* (K-NN) dan algoritma *Decision Tree C4.5* menyatakan bahwa algoritma K-NN memiliki tingkat akurasi yang lebih baik yaitu sebesar 94% dibandingkan algoritma C4.5 yang hanya memiliki tingkat akurasi sebesar 86%

4. KESIMPULAN

Berdasarkan hasil analisis menggunakan algoritma algoritma *K-Nearest Neighbor* (K-NN) dan algoritma *Decision Tree C4* bahwa untuk prediksi pemberian bantuan dari pemerintah ke UMKM Kab. Pekalongan yang terbaik adalah menggunakan algoritma *K-Nearest Neighbor* (K-NN) karena mempunyai nilai akurasi yang paling tinggi yaitu 94%. Penelitian ini dapat ditingkatkan dengan berbagai variabel lainnya seperti legalitas maupun jumlah pegawai. Selain itu hasil perbandingan algoritma dapat ditingkatkan dengan menggunakan konsep *deep learning* untuk meningkatkan kualitas penelitian dengan mengarah ke *artificial intelligent* (AI) dan dapat diimplementasikan dalam bentuk sistem informasi sehingga dapat digunakan dengan mudah dari pengguna dengan berbagai data yang berbeda.

REFERENCES

- [1] Anggara Cahya Putra and K. Dwi Hartomo, "Optimalisasi Penyaluran Bantuan Pemerintah Untuk UMKM Menggunakan Metode Fuzzy C-Means," *Jurnal RESTI (Rekayasa Sistem dan Teknologi Informasi)*, vol. 5, no. 3, pp. 474–482, Jun. 2021, doi: 10.29207/resti.v5i3.2980.
- [2] Windia Ambar Sari, Erlin Khotijah, and Siti Sunarmintyastuti, "Pemodelan Reward Rule Game Streamer Indonesia Tingkat Amatir dengan Orange Data Mining," Aug. 2019.
- [3] U. Kurniasih and A. T. Suseno, "Analisis Sentimen Terhadap Bantuan Subsidi Upah (BSU) pada Kenaikan Harga Bahan Bakar Minyak (BBM)," *Jurnal Media Informatika Budidarma*, vol. 6, no. 4, Oct. 2022.
- [4] L. B. Putong and K. E. Wahyudi, "Implementasi Program Bantuan Sosial Tunai Pada Umkm," *Jurnal Kebijakan Publik*, vol. 13, no. 4, pp. 338–345, 2022.
- [5] A. Majid, D. D. Kurniawan, and K. N. Sigit, "Pengaruh Bantuan Presiden BLT UMKM Terhadap Produktivitas Usaha Mikro Kecil Dan Menengah Di Kabupaten Batang," *Jurnal Ilmu Manajemen Dan Akuntansi Terapan (JIMAT)*, vol. 12, no. 3, pp. 333–341, 2021.
- [6] H. C. Maharani and J. Jaeni, "Determinan kebijakan pemerintah sebuah solusi keberlangsungan usaha UMKM di tengah pandemi covid-19," *AKSES: Jurnal Ekonomi Dan Bisnis*, vol. 16, no. 1, 2021.
- [7] D. T. S. Kumar, "Data mining based marketing decision support system using hybrid machine learning algorithm," *Journal of Artificial Intelligence and Capsule Networks*, vol. 2, no. 3, pp. 185–193, 2020.
- [8] R. S. Asa, "Identifikasi Penyaluran Zakat Menggunakan Algoritma C4.5 (Studi Kasus Di Baznas Kabupaten Agam)," *Jurnal Sains Dan Informatika*, vol. 5, no. 1, 2019.
- [9] Ari Rohmawati, "PEMETAAN DAN REKOMENDASI PENINGKATAN KINERJA UMKM DI KABUPATEN PESAWARAN LAMPUNG MENGGUNAKAN ALGORITMA NAIVE BAYES DAN NORMALIZED RATING FREQUENCY(NRF)," *Masters thesis, IIB Darmajaya*, 2022.

- [10] A. Damuri, U. Riyanto, H. Rusdianto, and M. Aminudin, "Implementasi Data Mining dengan Algoritma Naïve Bayes Untuk Klasifikasi Kelayakan Penerima Bantuan Sembako," *JURIKOM (Jurnal Riset Komputer)*, vol. 8, no. 6, p. 219, Dec. 2021, doi: 10.30865/jurikom.v8i6.3655.
- [11] S. Yani, F. S. Jumeilah, and M. Kadafi, "Algoritma K-Nearest Neighbor Untuk Menentukan Kelayakan Keluarga Penerima Bantuan Pangan Non Tunai (Studi Kasus: Kelurahan Karya Jaya)," *Journal of Information Technology Ampera*, vol. 1, no. 2, pp. 75–87, 2020.
- [12] E. Ermawati, "Algoritma Klasifikasi C4.5 Berbasis Particle Swarm Optimization Untuk Prediksi Penerima Bantuan Pangan Non Tunai," *Sistemasi (jurnal sistem informasi)*, vol. 8, no. 3, 2019.
- [13] R. L. Hasanah *et al.*, "KLASIFIKASI PENERIMA DANA BANTUAN DESA MENGGUNAKAN METODE KNN (K-NEAREST NEIGHBOR)," *Jurnal TECHNO Nusa Mandiri*, vol. 16, no. 1, p. 1, 2019, [Online]. Available: <http://nusamandiri.ac.id/>
- [14] S. Ulya, M. A. Soeleman, and F. Budiman, "Optimasi Parameter K Pada Algoritma K-NN Untuk Klasifikasi Prioritas Bantuan Pembangunan Desa," *Techno. Com*, vol. 20, no. 1, pp. 83–96, 2021.
- [15] R. Rusbandi, F. A. Ma'ruf, A. Pratama, I. Sholihin, and A. Rinaldi, "Penerapan Model Prediksi Menggunakan Algoritma C. 45 Untuk Prediksi Kelulusan Siswa SMK Wahidin," *JURNAL DATA SCIENCE & INFORMATIKA*, vol. 1, no. 1, pp. 16–20, 2021.
- [16] A. P. Wibawa, M. Guntur, A. Purnama, M. F. Akbar, and F. A. Dwiyanto, "Metode-metode Klasifikasi," *Prosiding Seminar Ilmu Komputer dan Teknologi Informasi*, vol. 3, no. 1, 2018.
- [17] D. A. Nawangnugraeni, M. Z. Abdillah, and A. T. Suseno, "C4. 5 Algorithm Implementation For Public Sentymnet Analysis Covid-19 Vaccine," *Digital Zone: Jurnal Teknologi Informasi dan Komunikasi*, vol. 13, no. 2, pp. 151–160, 2022.
- [18] A. Saifur Rohman Cholil and T. Ardianita, "Prediksi Penyakit Demam Berdarah Di Puskesmas Ngemplak Simongan Menggunakan Algoritma C4.5," *Sistemasi*, pp. 529–542, Jul. 2020.
- [19] I. Junaedi, N. Nuswantari, and V. Yasin, "Perancangan dan Implementasi Algoritma C4.5 Untuk Data Mining Analisis Tingkat Risiko Kematian Neonatum Pada Bayi," *JISICOM (Journal of Information System, Informatics and Computing)*, vol. 3, no. 1, pp. 29–44, 2019.
- [20] J. Smith and J. Doe, "Image Classification using Convolutional Neural Networks: An Experimental Study," *Int J Comput Vis*, vol. 20, no. 5, 2020.
- [21] D. Lee and E. Chen, "Predicting Stock Market Trends with Machine Learning Algorithms," *IEEE Transactions on Financial Engineering*, vol. 15, no. 3, 2021.