PENERAPAN ALGORITMA *K-NEAREST NEIGHBOR* DALAM KLASIFIKASI DATA HASIL PRODUKSI KELAPA SAWIT PADA KUD TIRTA KENCANA

Mardeni[1], Susanti[2]

[1,2]STMIK Hang Tuah Pekanbaru
Email: mdn@htp.ac.id[1], santyfelosa@gmail.com[2]

Abstrak: KUD Tirta Kencan terletak di Desa Air Emas , Kecamatan Singingi, Kabupaten Kuantan Singingi, Riau, membantu pengolahan hasil perkebunan kelapa sawit milik warganya demi berkembangnya roda perekonomian di Desa Air Emas. KUD Tirta Kencana Beranggotakan 17 Kelompok Tani. Sistem pengolahan data di KUD Tirta Kencana masih menggunakan pencatatan secara konvensional dimana data hasil produksi kelapa sawit setiap kelompok tani di rekapitulasi kedalam Ms. Excel, sehingga data yang dimasukkan kedalam Ms. Excel semakin lama semakin menumpuk dan tidak teratur, yang menyebabkan data sulit dipahami dan kurangnya informasi yang bisa digunakan untuk perkembangan KUD Tirta Kencana pada bidang hasil produksi kelapa sawit dimasa mendatang. Dengan ini perlu dilakukan analisa dan pengolahan data hasil produksi kelapa sawit di KUD Tirta Kencana menggunakan teknik perhitungan data mining, Salah satu algoritma yang terdapat pada teknik data mining adalah algoritma *k-Nearest Neighbor (k-NN)*. Algoritma ini merupakan suatu metode yang menggunakan algoritma *supervised learning*, dimana hasil dari sampel uji yang baru diklasifikasikan berdasarkan mayoritas dari kategori pada *k-NN*. Dari penelitian ini diketahui hubungan kemiripan hasil produksi antar kelompok tani, dengan demikian dapat diprediksikan hasil produksi kelapa sawit dimasa mendatang, berkisar pada hubungan kesamaan hasil produksi antar kelompok-kelompok tani berdasarkan *clusternya* masing-masing. Hasil uji menggunakan nilai K=2, K=3, dan K=4, memiliki nilai akurasi K=2 dan K=3 yaitu 85,71%, untuk K=4 menghasilkan nilai akurasi 71,43%.

Kata kunci: *Data Mining,* Klasifikasi, Kelapa Sawit*, k-Nearest Neighbor, RapidMiner,*

Abstract: *KUD Tirta Kencana is located in Air Emas Village, Singingi Subdistrict, Kuantan Singingi Regency, Riau, assisting the processing of oil palm plantations owned by its citizens for the development of the economy in the Air Emas Village. KUD Tirta Kencana consists of 17 Farmers Groups.The data processing system in KUD TirtaKencana still uses conventional recording where the palm oil production data of each farmer group is recapitulated into Ms. Excel, so the data entered into Ms. Excel is increasingly piling up and irregular, which makes data difficult to understand and the lack of information that can be used for the development of KUD TirtaKencana in the field of palm oil production in the future.With this, it is necessary to analyze and process the data of palm oil production in KUD Tirta Kencana using data mining calculation techniques, one of the algorithms contained in the data mining technique is the k-Nearest Neighbor (k-NN) algorithm. This algorithm is a method that uses a supervised learning algorithm, where the results of the new test sample are classified based on the majority of the categories in k-NN.From this research, it is known that the similarity of production yields among farmer groups, thus it can be predicted that the yield of palm oil production in the future, revolves around the relationship between the similarity of production between farmer groups based on their respective clusters. The test results using the values of K = 2, K = 3, and K = 4, have an accuracy value of K = 2 and K = 3 that is 85.71%, for K = 4 produces an accuracy value of 71.43%.*

Keywords: *Data Mining, Classification, k-Nearest Neighbor, Palm Oil, Rapid Miner,*

1. PENDAHULUAN

Dinas Penanaman Modal Pelayanan Terpadu Satu Pintu (DPM PTSP) Riau mencatat luas kebun kelapa sawit di Riau tahun 2018 seluas 2.424.545 ha. Berdasarkan data tersebut Riau menjadi provinsi dengan lahan sawit terluas, yakni 19% dari total perkebunan sawit Indonesia. KUD Tirta Kencana terletak di Desa Air Emas , Kecamatan Singingi, Kabupaten Kuantan Singingi, Riau, membantu pengolahan hasil perkebunan kelapa sawit milik warganya demi berkembangnya roda perekonomian di Desa tersebut. Data hasil produksi

kelapa sawit yang di rekapitulasi kedalam Ms. Excel, semakin lama semakin menumpuk dan tidak teratur, yang menyebabkan data sulit dipahami dan kurangnya informasi yang bisadigunakan untuk perkembangan KUD Tirta Kencana pada bidang hasil produksi kelapa sawit dimasa mendatang. Dengan ini perlu dilakukan analisa dan pengolahan data hasil produksi kelapasawit dengan teknik perhitungan data mining menggunakan algoritma *k-Nearest Neighbor* agar dapat menghasilkan informasi yang bisa digunakan dalam mengambil kebijakan bagi perencanaan peningkatan hasil produksi kelapa sawit.

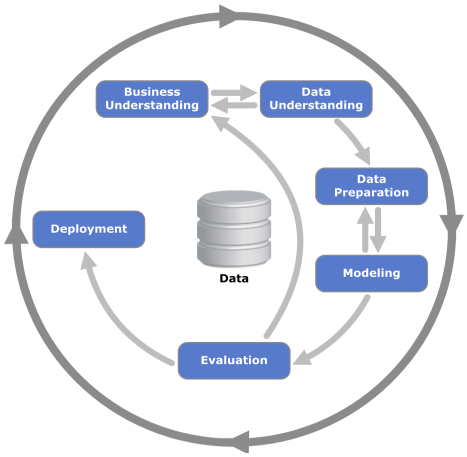
Permasalahan yang dihadapi adalah Kurangnya informasi yang bisa didapatkan dari tumpukan data hasil produksi kelapa sawit di KUD Tirta Kencana, Kurangnya analisa dan pengolahan data hasil produksi kelapa sawit di KUD Tirta Kencana., KUD Tirta kencana belum mampu memprediksi hasil produksi kelapa sawit dimasa mendatang. Untuk mneghindari pembahasan masalah yang lebih luas, maka penulis membatasi permasalahan Penerapan Algoritma *k-Nearest Neighbor* Dalam Klasifikasi DataHasil Produksi Kelapa Sawit Pada KUD Tirta Kencana**,** meliputi: Data yang diolah hanya seputar hasil produksi kelapa sawit periode Januari 2017 – Desember 2018, Pengolahan data menggunakan Aplikasi Rapid Miner., Penelitian ini dilakukan di KUD Tirta Kencana Air Emas, Kuantan Singingi.

Capaian dari penerapan algoritma ini yaitu Bagaimana tumpukan data hasil produksi kelapa sawit dapat menjadi informasi bagi KUD Tirta Kencana?. Bagaimana teknik data mining dapat menganalisa dan mengolah data hasil produksi kelapa sawit di KUD Tirta Kencana?. Bagaimana Algoritma *k-Nearest Neighbor* (*k-NN*) memprediksi hasil produksi kelapa sawit KUD Tirta Kencana dimasa mendatang?

2. METODE PENELITIAN

Metode*CRISP-DM*

CRISP-DM (*Cross Industry Standard Proses for Data Mining*) merupakan suatu metodologi data mining yang disusun oleh konsorsium perusahaan yang didirikan oleh Komisi Eropa pada tahun 1996 dan telah ditetapkan sebagai proses standar dalam data mining. Menurut Larose, data mining memiliki enam fase *CRISP-DM*, seperti yang tertera pada gambar 2.3 (Larose, 2006:6) :



Gambar 1 Tahapan Metode *CRISP-DM*

a. Fase Pemahaman Bisnis (*Bussiness Understanding Phase*)
Pada fase ini peneliti harus memahami tujuan proyek (penelitian) dan kebutuhan tujuan bisnis (penelitian). Tujuan proyek (penelitian) untuk mengetahui klasifikasi hasil produksi kelapa sawit berdasarkan klaster-klasternya. Kebutuhan tujuan bisnis (penelitian) ini untuk memprediksi hasil produksi kelapa sawit dimasa mendatang berdasarkan hasil klasifikasi yang telah dilakukan.

b. Fase Pemahaman Data (*Data Understanding Phase*)
Dalam penelitian ini data yang diperoleh dari KUD Tirta Kencana berupa data hasil produksi kelapa sawit periode 2017-2018 dengan atribut tanggal, supir, nopol, kelompok tani, tujuan,uang jalan supir, sisa uang jalan, sumber dana, jumlah tonase, harga kelapa sawit perkilo, dan lain-lain. Dalam fase ini peneliti harus memahami atribut-atribut apa saja yang bisa digunakan dalam proses pengolahan data, oleh karena itu hanya data yang sesuai untuk di analisis yang akan uji pada proses atau tahap berikutnya.

c. Fase Pengolahan Data (*Data Preparation Phase*)
Setelah peneliti memahami data yang diperoleh, maka selanjutnya dalam fase pengolahan data penelilti melakukan dua proses pengolahan data, yaitu:
1. Data *Selection* (Pemilihan Data)

Data *Selection* merupakan proses meminimalkan jumlah data yang digunakan untuk proses mining dengan tetap merepresentasikan data aslinya. Pada tahap ini data yang digunakan akan diseleksi dengan cara melihat kecenderungan data / kesesuaian data dengan topik / judul penelitian yang akan diteliti oleh penulis, dalam hal ini data yang diperoleh oleh penulis dari KUD Tirta Kencana sudah memiliki kesesuaian format yang terdiri dari atribut Kelompok Tani, Tahun Produksi, Bulan dan Jumlah Produksi.

2. Data *Pre Processing* atau Data *Cleaning*
   Pada tahap ini akan dilakukan pembersihan data, yakni membuang data yang tidak konsisten dan *noise / redundancy* data. Data *cleaning* merupakan proses membuang duplikasi data, memeriksa data yang tidak konsisten, dan memperbaiki kesalahan pada data, seperti kesalahan penulisan. Pada umumnya data yang diperoleh baik dari database maupun hasil eksperimen, memiliki isi yang tidak sempurna seperti data yang hilang, data yang tidak valid atau juga hanya sekedar salah ketik. Data *cleaning* juga akan mempengaruhi hasil informasi dari teknik data mining karena data yang ditangani akan berkurang jumlah kompleksitasnya.

d. Fase Pemodelan (*Modeling Phase*)

Merupakan suatu proses utama saat metode diterapkan untuk menentukan pengetahuan berharga dan tersembunyi dari data. Data hasil produksi dari 17 kelompok tani akan dibagi kedalam data *training* dan data *testing*. Berdasarkan pembagian tersebut maka diperoleh 10 kelompok tani pertama terletak pada data *training* dan 7 kelompok tani lainnya terletak pada data *testing*.

Setelah data dibagi menjadi dua, selanjutnya menghitung data training untuk menentukan banyaknya jumlah cluster yang akan digunakan dengan menggunakan rumus H.A.Sturges sebagai berikut :

1. Menentukan banyaknya jumlah cluster dan nilai rentang antar cluster.
   Diketahui :
   n (jumlah data) =10
   Data terbesar =223,680
   Data terkecil =16,524
   Range =Data terbesar – Data terkecil
   =223,680 – 16,524
   = 207,156
   Banyakkelas =1+3,3 log n
   =1+3,3 log 10
   =4,3 = 4
   Panjang interval =*Range* / banyakkelas
   =207,156 / 4
   =51,789
   Rentangjarakantar cluster =Data terkecil + Panjang interval
   =16,524 + 51,789
   = 68,313
   Makadiperoleh 4 buah cluster denganrentangjarak cluster sebagaiberikut :
   C1 = d < 68,313
   C2 =68,314≤ d < 136,626
   C3 = 136,627≤ d < 204,939
   C4 = 204,940≤ d > 273,252
2. Menghitung jarak *Euclidean* pada data traininig.

   $d(x_1, x_1) = \sqrt{\sum_{i=1}^{n} \left(a_{r(x,1)} - a_{r(x,1)}\right)^2}$

   $d(1, 2) = \sqrt{}$ ((127,550-86,195)^2) + ((109,790-86,825)^2) + ((91,484-74,985)^2) + ((91,612-88,470)^2) + ((158,620-119,110)^2) + ((95,560-64,900)^2) + ((101,310-74,920)^2) + ((131,120-102,815)^2) + (97,910-65,180)^2) + ((107,650-70,890)^2) + ((116,830-74,460)^2) + ((109,185-70,640)^2) + ((86,630-68,375)^2) + ((92,955-72,840)^2) + ((106,830-78,090)^2) + ((163,440-102,590)^2) + ((128,350-84,270)^2) + ((48,570-43,742)^2) + ((113,441-97,850)^2) + ((91,186-119,193)^2) + ((62,224-75,026)^2) + ((61,976-98,663)^2) + ((54,295-84,389)^2) + ((89,074-114,442)^2)
   d(1 ,2) =153,43 (C3)
   d(3 ,4) =361,27 (C4)
   d(5 ,6) =131,87 (C2)

d(7 ,8) =180,77 (C3)
d(9 ,10) =126,46 (C2).

Tabel 1Cluster Data Training

Data	Nilai	Cluster
d(1,2)	153.43	C3
d(3,4)	361.37	C4
d(5,6)	131.87	C2
d(7,8)	180.77	C3
d(9,10)	126.46	C2

Setelah penempatan *cluster-cluster* pada data training selesai, selanjutnya menentukan *cluster* data testing dengan menggunakan nilai k masing-masing k=2, k=3, k=4.

Tabel 2 Cluster Data Testing

Data	Cluster		
	K = 2	K = 3	K = 4
11	C3	C3	C3
12	C4	C2	C2
13	C2	C2	C3
14	C3	C3	C3
15	C3	C3	C3
16	C3	C3	C3
17	C4	C2	C2

Tabel 3 Rekapitulasi Hasil *Cluster*

Cluster	K = 2
	Anggota
C1	-
C2	5,6,9,10,13
C3	1,2,7,8,11,14,15,16
C4	3,4,12,17

Cluster	K = 3
	Anggota
C1	-
C2	5,6,9,10,12,13,17
C3	1,2,7,8,11,14,15,16
C4	3,4

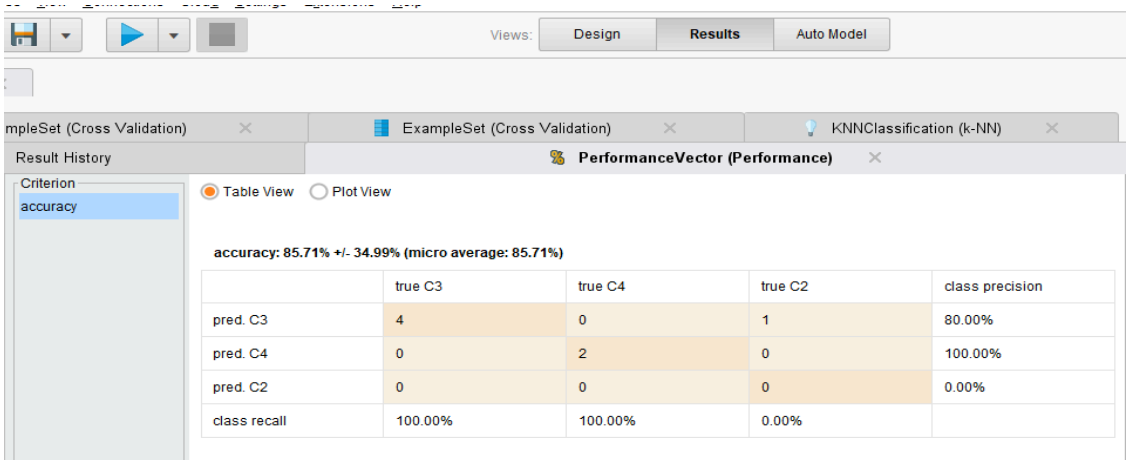
Cluster	K = 4
	Anggota
C1	-
C2	5,6,9,10,12,17

C3	1,2,7,8,11,13,14,15,16
C4	3,4

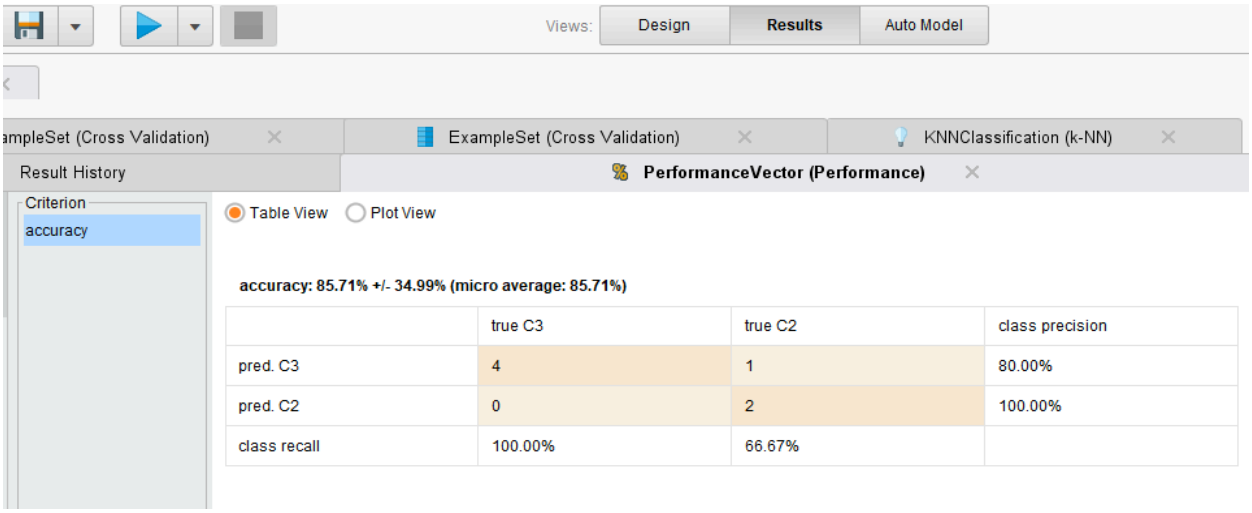
Setelah semua selesai dihitung dengan cara manual menggunakan Ms. excel, Selanjutnya membandingkan nilai akurasi dari setiap nilai k yang digunakan yaitu K=2, K=3 dan K=4 menggunakaan *Software RapidMiner*.

e. Fase Evaluasi (*Evaluation Phase*)

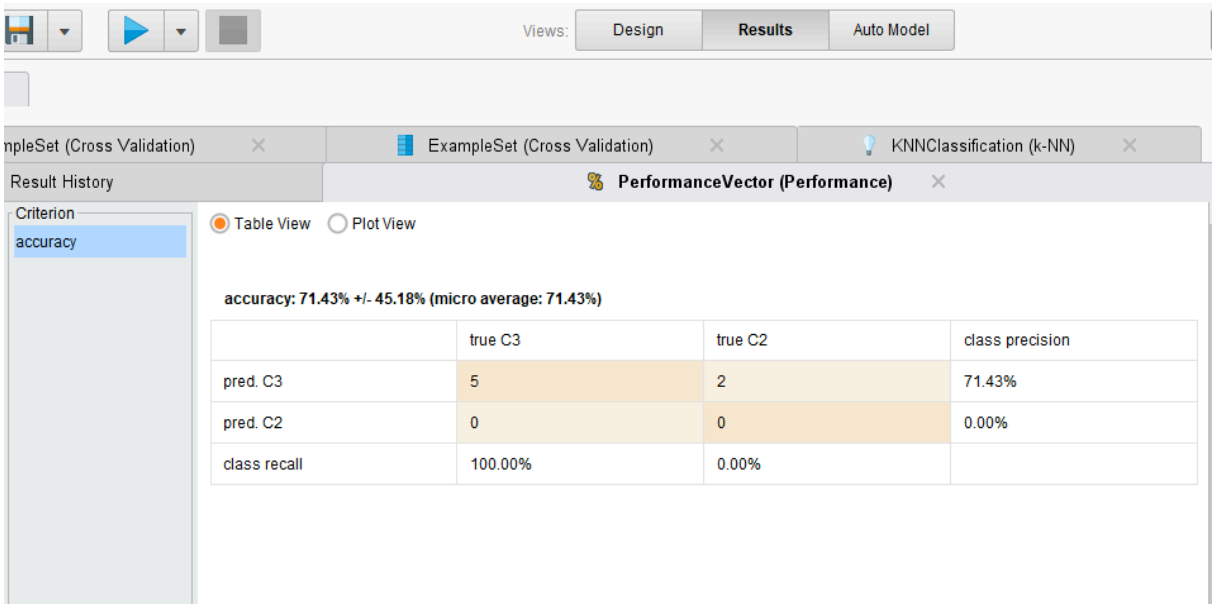
Fase Evaluasi ini akan mengevaluasi dan meneliti untuk menyakinkan kalau tahap pemodelan dari data hasil produksi kelapa sawit setiap kelompok tani yang digunakan memenuhi tujuan dari penelitian dengan menghitung nilai akurasi dari masing-masing nilai k.



Gambar 2 Nilai akurasi k=2 adalah 85,71%



Gambar 3 Nilai Akurasi k=3 adalah 85,71%



Gambar 4 Nilai Akurasi k=4 adalah 71,43%

3. HASIL DAN PEMBAHASAN

Tabel 5 Hasil *Klasifikasi* Kelompok Tani

Cluster	K = 2
	Anggota
C1	-
C2	-FajarPagi -Mekar Wangi -Sumber Jaya -Muda Karya -Mukti Tama
C3	-Harapan -Margo Mulyo -Tunas Harapan -Tunas Mekar -SekarMukti -SidoMulyoBakti -SumberRezeki -Tunas Tani
C4	-TaniMulya -SukaMakmur -SukaKarya -Bina Sejahtera
Cluster	K = 3
	Anggota
C1	-
C2	-FajarPagi -Mekar Wangi -Sumber Jaya -Muda Karya -SukaKarya -Mukti Tama

	-Bina Sejahtera
C3	-Harapan -Margo Mulyo -Tunas Harapan -Tunas Mekar -SekarMukti -SidoMulyoBakti -SumberRezeki -Tunas Tani
C4	-TaniMulya -SukaMakmur
Cluster	K = 4
	Anggota
C1	-
C2	-FajarPagi -Mekar Wangi -Sumber Jaya -Muda Karya -SukaKarya -Bina Sejahtera
C3	-Harapan -Margo Mulyo -Tunas Harapan -Tunas Mekar -SekarMukti -Mukti Tama -SidoMulyoBakti -SumberRezeki -Tunas Tani
C4	-TaniMulya -SukaMakmur

Dari penelitian ini diketahui hubungan kemiripan hasil produksi antar kelompoktani, dengan demikian dapat diprediksikan hasil produksi kelapa sawit dimasa mendatang, berkisar pada hubungan kesamaan hasil produksi antar kelompok-kelompok tani berdasarkan *cluster-clusternya* masing-masing

4. SIMPULAN DAN SARAN

Kesimpulan

1. Tumpukan data hasil produksi kelapa sawit yang diperoleh dari KUD Tirta Kencana dengan periode 2017 dan 2018 yang telah diolah menghasilkan informasi berupa klasifikasi data yang terklasifikasi kedalam 4 cluster yaitu : C1, C2, C3, C4.
2. Dengan Algoritma *k-Nearest Neighbor (k-NN)* setiap klaster diuji dengan nilai K=2, K=3 dan K=4, menghasilkan :
   a. Hasil uji dengan nilai K=2 pada cluster pertama (C1) tidak memiliki anggota cluster, maka persentase keanggotaanya 0%. Pada cluster kedua (C2) memiliki 5 anggota cluster dengan persentase 29,4%. Pada Cluster ketiga (C3) memiliki 8 anggota cluster dengan persentase 47%. Pada cluster keempat (C4) memiliki 4 anggota cluster dengan persentase 23,6%
   b. Hasil uji untuk K=3 pada cluster pertama (C1) tidak memiliki anggota cluster, maka persentase keanggotaanya 0%. Pada cluster kedua (C2) memiliki 7 anggota cluster dengan persentase 41,2%. Pada Cluster ketiga (C3) memiliki 8 anggota cluster dengan persentase 47%. Pada cluster keempat (C4) memiliki 2 anggota cluster dengan persentase 11,8% .
   c. Hasil uji untuk K=4 pada cluster pertama (C1) tidak memiliki anggota cluster, maka persentase keanggotaanya 0%. Pada cluster kedua (C2) memiliki 6 anggota cluster dengan persentase 35,2%.

Pada Cluster ketiga (C3) memiliki 9 anggota cluster dengan persentase 53%. Pada cluster keempat (C4) memiliki 2% anggota cluster dengan persentase 11,8%.

3. Hasil uji menggunakan nilai K=2, K=3, dan K=4, memiliki nilai akurasi K=2 dan K=3 yaitu 85,71%, untuk K=4 menghasilkan nilai akurasi 71,43%. Dari penelitian ini diketahui hubungan kemiripan hasil produksi antar kelompok tani, dengan demikian dapat diprediksikan hasil produksi kelapa sawit dimasa mendatang, berkisar pada hubungan kesamaan hasil produksi antar kelompok-kelompok tani berdasarkan *clusternya* masing-masing.

Saran

Berdasarkan dari pengkajian hasil penelitian diatas, maka penulis bermaksud memberikan saran yang mudah-mudahan dapat bermanfaat, sebagai berikut:

1. Bagi KUD Tirta Kencana

   Diharapkan dengan adanya penelitian penerapan Algoritma k-NN dalam klasifikasi data hasil produksi kelapa sawit, KUD Tirta kencana dapat terbantu dalam pengolahan data untuk dapat memprediksikan hasil produksi kelapa sawit dimasa mendatang, dan mengetahui akibat-akibat dari perbedaan yang mencolok dari hasil produksi (tonase) kelompok tani, dan jika perlu dapat dikembangkan lagi baik dengan mengimplementasikan ke dalam sebuah program agar lebih memudahkan proses pengolahannya.

2. Bagi peneliti selanjutnya

   Untuk peneliti selanjutnya diharapkan dapat menggunakan algoritma klasifikasi lainnya agar dapat dijadikan perbandingan dari penelitian sebelumnya.

DAFTAR PUSTAKA

[1] Aris, F. (2019). *Penerapan Data Mining untuk Identifikasi Penyakit Diabetes Melitus dengan Menggunakan Metode Klasifikasi*. *1*(1), 1–6.

[2] Badu, Z. S. (2016). *Penerapan Algoritma K-Nearest Neighbor Untuk Klasifikasi Dana Desa*. *November*.

[3] Dzikrulloh, N. N., & Setiawan, B. D. (2017). Penerapan Metode K – Nearest Neighbor (KNN) dan Metode Weighted Product (WP) Dalam Penerimaan Calon Guru Dan Karyawan Tata Usaha Baru Berwawasan Teknologi (Studi Kasus : Sekolah Menengah Kejuruan Muhammadiyah 2 Kediri). *Pengembangan Teknologi Informasi Dan Ilmu Komputer*, *1*(5), 378–385.

[4] Fatmawati, F. (2016). Perbandingan Algoritma Klasifikasi Data Mining Model C4.5 Dan Naive Bayes Untuk Prediksi Penyakit Diabetes. *None*, *13*(1), 50–59.

[5] Irawan, Y. (2019). Penerapan Data Mining Untuk Evaluasi Data Penjualan Menggunakan Metode Clustering Dan Algoritma Hirarki Divisive Di Perusahaan Media World Pekanbaru. Jurnal Teknologi Informasi Universitas Lambung Mangkurat (JTIULM), 4(1), 13-20.

[6] Kusrini dan Emha Taufiq. (2015). Proses Data Mining. *Data Mining*, 1–143.

[7] Pramono, F., Saputra, S. A., Burhanuddin, & Ade, K. (2018). Komparasi Klasifikasi Penentuan Keterlambatan Siswa SMA Datang Upacara Menggunakan Algoritma C4.5. *Seminar Nasional Teknologi Informasi Dan Komunikasi*, *2018*(Sentika), 80–86.

[8] Widiastuti, Y., Sihwi, S. W., & Sulistyo, M. E. (2016). Decision Support System for House Purchasing Using Knn (K-Nearest Neighbor) Method. *Jurnal Itsmart*, *5*(1), 43–49