

DATA MINING UNTUK MEMREDIKSI HASIL PRODUKSI BUAH SAWIT PADA PT BUMI SAWIT SUKSES (BSS) MENGGUNAKAN METODE *K-NEAREST NEIGHBOR*

Ahmad Yunus¹, Muhammad Akbar², Andri³

Dosen Universitas Bina Darma^{2,3}, Mahasiswa Universitas Bina Darma¹

Jalan Jenderal Ahmad Yani No.3 Palembang

Email : yunusgudang@gmail.com¹, muhamad.akbar@binadarma.ac.id²,
andrepro2007@gmail.com

ABSTRACT

PT. Bumi Sawit Sukses (BSS) is a company engaged in oil palm plantations. This study discusses Data Mining to Predict Palm Fruit Production Results at PT Bumi Sawit Sukses (BSS) Using Method (K-NN) K-Nearest Neighbor, where the amount of palm oil production is erratic every month. This study aims to use the method (K - NN) K-Nearest Neighbor in Predicting Palm Fruit Production Results for the next period at PT. Bumi Sawit Sukses (BSS). The object of this research is the amount of palm oil production while the data used is historical data from the amount of oil palm production in 2015-2017. By using the K-Nearest Neighbor (K - NN) method, the output from Rapidminer obtained an accuracy of 85.15%, which can be used as a reference for the following year so that the company can make the right decisions if the production of oil palm plants decreases and the company can continue to improve palm oil production.

Keywords: Data Mining, Algorithm, (K - NN) K-Nearest Neighbor, Rapidminer, Production Results

ABSTRAK

PT. Bumi Sawit Sukses (BSS) merupakan perusahaan yang bergerak dalam bidang perkebunan kelapa sawit. Penelitian ini membahas tentang Data Mining Untuk Memprediksi Hasil Produksi Buah Sawit Pada PT Bumi Sawit Sukses (BSS) Menggunakan Metode (K-NN) *K-Nearest Neighbor*, dimana jumlah produksi kelapa sawit tidak menentu setiap bulannya. Penelitian ini bertujuan untuk menggunakan metode (K - NN) *K-Nearest Neighbor* dalam Memprediksi Hasil Produksi Buah Sawit untuk periode berikutnya pada PT. Bumi Sawit Sukses (BSS). Objek penelitian ini adalah jumlah produksi kelapa sawit sedangkan data yang digunakan merupakan data historis dari jumlah produksi tanaman kelapa sawit tahun 2015-2017. Dengan menggunakan metode (K - NN) *K-Nearest Neighbor*, Output dari *Rapidminer* didapatkan hasil akurasi 85,15 %, dapat dijadikan sebagai acuan untuk tahun selanjutnya agar perusahaan dapat mengambil keputusan yang tepat apabila produksi tanaman kelapa sawit mengalami penurunan dan perusahaan dapat terus meningkatkan hasil produksi kelapa sawit.

Kata Kunci : Data Mining, Algoritma, (K - NN) *K-Nearest Neighbor*, *Rapidminer*, Hasil Produksi

I. PENDAHULUAN

Tanaman Kelapa Sawit saat ini merupakan salah satu jenis tanaman perkebunan yang menduduki posisi penting disektor pertanian umumnya, dan sektor perkebunan khususnya, hal ini disebabkan karena dari sekian banyak tanaman yang menghasilkan minyak atau lemak, kelapa sawit yang menghasilkan nilai ekonomi terbesar per hektarnya di dunia. Peramalan atau *forecasting* merupakan seni dan ilmu yang dilakukan untuk mengetahui atau memperkirakan apa yang terjadi pada masa yang akan datang. Peramalan menjadi sangat penting karena penyusunan suatu rencana diantaranya didasarkan pada suatu proyeksi atau ramalan.

Dalam melakukan peramalan yang menjadi faktor utama adalah pemilihan metode peramalan karena pemilihan metode berpengaruh terhadap hasil peramalan. Pada penelitian sebelumnya proses peramalan dilakukan dengan metode K-Nearest Neighbor yang diimplementasikan untuk peramalan harga saham. Penelitian tersebut melakukan prediksi harga saham dengan teknologi data mining untuk menganalisis volume data bisnis dan keuangan. Algoritma K-Nearest Neighbor digunakan karena memiliki akurasi yang tinggi dengan rasio kesalahan kecil. Hasil dari prediksi atau peramalan bermanfaat untuk membantu investor dan manajemen dalam pengambilan keputusan investasi. Hasil dari penelitian tersebut menunjukkan bahwa hasil prediksi dengan metode K- nearest neighbor mempunyai tingkat akurasi yang cukup tinggi dengan data harga saham sebenarnya.

Penelitian yang berkaitan dengan algoritma k -NN ada beberapa orang terdahulu yang telah melakukannya, salah satunya yaitu, Nobertus Krisandi. Dkk dengan judul “Algoritma *K-Nearest Neighbor* dalam Klasifikasi Data Hasil Produksi Kelapa Sawit pada PT. Minamas Kecamatan Parindu”. Data yang digunakan adalah data hasil produksi kelapa sawit (Tonase) dari 50 kelompok tani pada periode Juli-Desember 2011 pada PT. Minasa Kabupaten Sanggau. Nilai k yang digunakan adalah $k=1$, $k=3$, $k=5$ dan $k=7$. Hasil penelitian menunjukkan bahwa, hasil produksi yang dominan adalah dengan nilai $k=7$ sebesar yang memiliki nilai accuracy sebesar 34%. Hal tersebut juga mengindikasikan bahwa *K-Nearest Neighbor* (KNN) dipengaruhi oleh jumlah klustering data.

Dengan adanya prediksi maka perusahaan dapat mencapai tujuan serta pengambilan keputusan. Prediksi dilakukan dengan mengambil data jumlah produksi sebelumnya. Berdasarkan urai-uraian diatas, penulis bermaksud mengangkat skripsi dengan judul “Data Mining Untuk Memprediksi Hasil Produksi Buah Sawit Pada PT Bumi Sawit Sukses BSS Menggunakan Metode (K - NN) *K-Nearest Neighbor*”

II. METODE PENELITIAN

II.1. Tempat dan waktu Penelitian

Tempat dalam melakukan penelitian tugas akhir ini yaitu berlokasi di PT. Bumi Sawit Sukses (BSS) yang beralamat di Simpang Rimba Kecamatan Bangka Selatan. Penelitian dilakukan pada bulan Januari 2018 sampai dengan Juni 2018.

II.2. Prosedur Penelitian

Prosedur penelitian adalah langkah – langkah yang digunakan sebagai alat untuk mengumpulkan data dan menjawab pertanyaan – pertanyaan dalam penelitian. Didalam prosedur penelitian ini, penulis membahas tentang metode dan teknik pengumpulan data. Penelitian eksperimen menggunakan data dalam penelitiannya dan mengasilkan kesimpulan yang mampu dibuktikan dengan pengamatan atau percobaan. Penelitian ini termasuk penelitian eksperimen dengan menggunakan data dari PT. Bumi Sawit Sukses (BSS).

II.2.1 Metode Penelitian

Dalam melakukan penelitian ini penulis menggunakan metode deskriptif (*Deskriptive Research*). Metode deskriptif digunakan untuk mendeskripsikan atau menggambarkan data yang telah terkumpul sebagaimana adanya tanpa bermaksud membuat kesimpulan yang berlaku untuk umum atau generalisasi. Metode deskriptif merupakan metode yang meneliti secara sistematis, faktual dan akurat mengenai fakta-fakta dan sifat-sifat suatu objek. Maka dari itu penulis menggunakan metode deskriptif dikarenakan permasalahan yang sedang dilakukan penulis berdasarkan data yang sebenarnya atau bersifat fakta yaitu mengenai data jumlah produksi tanaman kelapa sawit pada PT. Bumi Sawit Sukses (BSS). Berikut ini adalah merupakan langkah-langkah pokok dalam metode deskriptif (*Deskriptive Research*), yaitu (Nazir 1988: 73-74) :

1. *Memilih dan merumuskan masalah* yang menghendaki konsepsi ada kegunaan masalah tersebut serta dapat diselidiki dengan sumber yang ada.
2. *Menentukan tujuan dari penelitian* yang akan dikerjakan. Tujuan dari penelitian harus konsisten dengan rumusan dan definisi dari masalah.
3. *Memberikan limitasi* dari area atau scope atau sejauh mana penelitian deskriptif tersebut akan dilaksanakan. Termasuk di dalamnya daerah geografis di mana penelitian akan dilakukan, batasan-batasan kronologis, ukuran tentang dalam dangkal serta seberapa utuh daerah penelitian tersebut akan dijangkau.
4. Pada bidang ilmu yang telah mempunyai teori-teori yang kuat, maka *perlu dirumuskan kerangka teori* atau kerangka konseptual yang kemudian diturunkan dalam bentuk hipotesa-hipotesa untuk diverifikasikan. Bagi ilmu sosial yang telah berkembang baik, maka kerangka analisa dapat dijabarkan dalam bentuk-bentuk model matematika.
5. *Menelusuri sumber-sumber kepustakaan* yang ada hubungannya dengan masalah yang ingin dipecahkan.
6. *Merumuskan hipotesa-hipotesa* yang ingin diuji, baik secara eksplisit maupun secara implisit.
7. *Melakukan kerja lapangan* untuk mengumpulkan data, gunakan teknik pengumpulan data yang cocok untuk penelitian.
8. *Membuat tabulasi serta analisa* statistik dilakukan terhadap data yang telah dikumpulkan. Kurangi penggunaan statistik sampai kepada batas-batas yang dapat dikerjakan dengan unit-unit pengukuran sepadan.
9. *Memberikan interpretasi* dari hasil dalam hubungannya dengan kondisi sosial yang ingin diselidiki serta dari data yang diperoleh serta referensi khas terhadap masalah yang ingin dipecahkan.
10. *Mengadakan generalisasi* serta deduksi dari penemuan serta hipotesa-hipotesa yang ingin diuji. Berikan rekomendasi-rekomendasi untuk kebijakan-kebijakan yang dapat ditarik dari penelitian.
11. *Membuat laporan* penelitian dengan cara ilmiah.

Dalam semua situasi dan kondisi peramalan itu mengandung derajat ketidakpastian. Kita mengenal fakta nyata ini dengan memasukkan unsur error atau kesalahan dalam perumusan sebuah peramalan. Sumber penyimpangan dalam peramalan bukan hanya disebabkan oleh unsur error tetapi ketidak mampuan suatu model peramalan mengenali unsur yang lain dalam deret data yang mempengaruhi besarnya penyimpangan dalam ramalan.

Besarnya penyimpangan hasil ramalan dapat disebabkan oleh besarnya faktor yang tidak diduga di mana tidak ada metode peramalan yang mampu menghasilkan peramalan yang akurat atau dapat juga disebabkan metode peramalan yang digunakan tidak dapat memprediksi dengan tepat komponen trend, komponen musiman atau komponen siklus yang mungkin terdapat dalam deret data dan itu berarti metode yang digunakan tidak tepat.

II.2.2. Teknik Pengumpulan Data

Pengumpulan data merupakan langkah awal pada suatu penelitian, dimana sumber data pada penelitian ini diambil dari data PT. Bumi Sawit Sukses (BSS). Data yang diperoleh berupa data hasil produksi tanaman kelapa sawit periode tahun 2015-2017. Menurut Sugiyono (2013:224) teknik pengumpulan data merupakan langkah yang paling strategis dalam penelitian, karena tujuan utama dari penelitian adalah mendapatkan data. Adapun teknik atau metode pengumpulan data dalam penelitian antara lain:

1. *Penelitian Kepustakaan (Library Research)*
Merupakan penelitian yang dilakukan penulis dengan mempelajari buku-buku, jurnal ilmiah, laporan penelitian yang memuat teori-teori Metode (K - Nn) *K-Nearest Neighbor* yang

diharapkan dapat memberikan referensi yang sesuai beserta sumber-sumber dari berbagai situs yang mendukung Tugas Akhir.

2. Penelitian Lapangan (*Field Research*)

Pengumpulan data melalui pengambilan dari pada objek penelitian dan pencatatan secara menyeluruh mengenai data-data yang ada pada perusahaan tersebut.

II.2.3. Identifikasi Masalah

Pada dasarnya semua penelitian akan selalu didahului dengan identifikasi masalah, hal ini berguna untuk peneliti agar fokus pada titik permasalahan. Dalam kasus ini peneliti menemukan perhitungan atau prediksi PT masih menggunakan Microsoft Excel dengan melihat hasil buah sawit setiap bulanya. Peneliti ingin menganalisa dengan melakukan observasi klasifikasi lama kerja di PT. Bumi Sawit Sukses (BSS) dengan menggunakan algoritma KNN sekaligus ingin mengetahui bagaimana pola dan cara kerja klasifikasi dalam *data mining* tersebut.

II.3. Analisis Data Mining

Analisis data mining merupakan proses menganalisis sumber data (*data Source*) sehingga menjadi data yang siap digunakan dalam proses *data mining* nantinya. Analisa *data mining* mengikuti tahapan dalam *Knowledge Discovery in Database* (KDD), KDD (*Knowledge Discovery in Database*) adalah keseluruhan proses non-trivial untuk mencari dan mengidentifikasi pola (*pattern*) dalam data, dimana pola yang ditemukan bersifat sah, baru, dapat bermanfaat dan dapat dimengerti. Serangkaian proses tersebut yang memiliki tahap. *data selection*, *data cleaning*, *data integration* dan *transformation*. Jika *data source* yang digunakan telah melalui proses data tersebut siap diolah dengan proses *data mining*.

II.3.1. Data source

Data Source merupakan data mentah yang bersumber dari objek penelitian. *Data source* bersifat murni, artinya data-data tersebut asli dan tidak dapat diubah/diedit ataupun dimanipulasi. Namun, *data source* masih terdapat *noise data* yang harus dibersihkan terlebih dahulu, sehingga data siap diolah dengan menggunakan teknik *data mining*.

II.3.2. Data Selection (Pemilihan Data)

Data Selection merupakan proses meminimalkan jumlah data yang digunakan untuk proses mining dengan tetap merepresentasikan data aslinya. Kumpulan data operasional Seleksi (*Selection*) Data darurat perlu dilakukan sebelum langkah penggalian informasi di KDD dimulai. Hasil seleksi data yang akan digunakan untuk proses data mining, disimpan dalam file terpisah dari database operasional.

- a) Menciptakan himpunan data target, pemilihan himpunan data, atau memfokuskan pada subset variabel atau sampel data, dimana penemuan (discovery) akan dilakukan.
- b) Pemilihan (seleksi) data dari sekumpulan data operasional perlu dilakukan sebelum tahap penggalian informasi dalam KDD dimulai. Data hasil seleksi yang akan digunakan untuk proses data mining, disimpan dalam suatu berkas, terpisah dari basis data operasional

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 di peroleh oleh penulis dari PT. Bumi Sawit Sukses (BSS) sudah memiliki kesesuaian format yang terdiri dari atribut Tahun Produksi, Bulan dan Jumlah Produksi.

II.3.3. Data Pre Processing atau Data Cleaning

Pada tahap ini akan dilakukan perbesihan data, yakni membuang data yang tidak konsisten dan noise/redundancy data. Pada data yang didapat, terdapat beberapa data yang tidak konsisten dan noise. 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 suatu perusahaan maupun hasil eksperimen, memiliki isi yang tidak sempurna seperti data yang hilang, data yang tidak valid atau juga hanya sekedar salah ketik. Selain itu, ada juga atribut-atribut data yang tidak relevan dengan hipotesa data mining yang dimiliki. Data cleaning juga akan mempengaruhi hasil informasi dari teknik data mining karena data yang ditangani akan berkurang jumlah dan kompleksitasnya.

II.3.4. Data Integration

Tahap *integrasi* data adalah tahap penggabungan data dari berbagai sumber. *Dataset* Hasil Produksi PT. Bumi Sawit Sukses (BSS) berasal dari satu sumber yaitu *manager* PT. Bumi Sawit Sukses (BSS), dan dari data yang ada tersebut dapat digunakan untuk proses pengolahan *data mining* dengan menggunakan metode *K-Nearest Neighbor* (KNN). Data ini terdiri dari data produksi untuk tahun 2015, 2016 dan 2017 yang kemudian digabungkan kedalam satu *dataset* untuk di ubah ke format CSV dengan menggunakan *Ms. Excel* yang nantinya akan di import ke dalam database *Mysql* untuk proses *Data Manipulation*.

II.3.5. Data Tranformasi

Ada pun pada tahap ini data akan diubah menjadi bentuk yang sesuai untuk proses data mining. Karena penelitian ini akan dilakukan uji coba secara teoritis dan menggunakan *software data mining* yaitu *RapidMiner*, maka data yang telah melalui proses sebelumnya akan di *tranformasi* agar dapat sesuai dengan algoritma yang dipakai yaitu algoritma *K-Nearest Neighbor* (KNN). Perubahan bahwa data yang sebelumnya berformat *Microsoft Excel* (.xls) telah di tranformasikan kedalam format (.CVS). Maka *dataset* tersebut dapat dibaca dan siap untuk diolah menggunakan *Software RapidMiner*.

III. Hasil Dan Pembahasan Persiapan Data

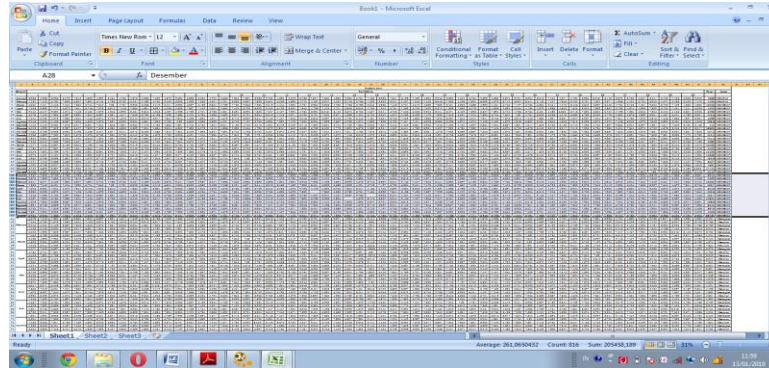
Tahap awal yang dilakukan dalam penelitian ini adalah persiapan data awal yang diperoleh dari pengumpulan data PT. Bumi Sawit Sukses (BSS), data yang diperoleh merupakan hasil produksi tahun 2015, 2016, dan 2017.

| PRODUKSI BUAH SAWIT PT BUMI SAWIT SUKSES PRATAMA | | | | | | | | | | | | |
|--|---------|----------|-------|-------|-------|-------|-------|---------|-----------|---------|----------|----------|
| TAHUN 2015 | | | | | | | | | | | | |
| | Januari | Februari | Maret | April | Mei | Juni | Juli | Agustus | September | Oktober | November | Desember |
| | 8,928 | 7,803 | 7,998 | 4,965 | 5,353 | 9,897 | 7,803 | 8,643 | 5,892 | 9,543 | 6,531 | 7,453 |
| | 5,928 | 7,876 | 6,784 | 6,789 | 7,232 | 8,887 | 4,672 | 6,826 | 7,839 | 6,457 | 6,871 | 8,792 |
| | 8,906 | 6,903 | 8,987 | 8,909 | 9,756 | 5,789 | 3,783 | 7,432 | 6,725 | 9,895 | 8,541 | 8,932 |
| | 6,578 | 6,897 | 8,886 | 5,894 | 5,782 | 8,456 | 6,723 | 6,731 | 7,431 | 5,675 | 8,912 | 9,542 |
| | 7,812 | 6,189 | 7,894 | 4,671 | 4,786 | 6,783 | 6,543 | 6,319 | 7,451 | 7,458 | 8,562 | 6,732 |
| | 4,896 | 9,854 | 7,894 | 4,679 | 6,789 | 9,854 | 6,782 | 5,829 | 6,058 | 5,423 | 7,482 | 8,035 |
| | 5,893 | 5,789 | 9,657 | 8,712 | 8,643 | 7,895 | 5,821 | 7,892 | 8,903 | 4,879 | 5,327 | 8,934 |
| | 9,564 | 7,765 | 9,886 | 8,907 | 6,826 | 6,432 | 8,691 | 8,431 | 8,392 | 5,842 | 7,923 | 8,932 |
| | 4,783 | 9,836 | 7,564 | 8,341 | 7,432 | 5,553 | 8,094 | 6,832 | 6,873 | 6,895 | 6,784 | 7,806 |
| | 5,872 | 7,984 | 9,459 | 9,142 | 6,731 | 5,432 | 7,832 | 3,689 | 5,675 | 8,561 | 7,985 | 7,853 |
| | 3,986 | 9,319 | 9,675 | 7,457 | 6,319 | 6,322 | 7,547 | 5,892 | 7,458 | 9,843 | 5,654 | 3,678 |
| | 7,329 | 5,445 | 8,954 | 7,846 | 5,829 | 4,442 | 5,437 | 4,521 | 5,423 | 4,048 | 7,456 | 5,678 |
| | 4,879 | 4,568 | 8,954 | 5,857 | 7,892 | 4,568 | 7,904 | 6,742 | 4,879 | 4,587 | 6,874 | 8,762 |
| | 5,673 | 7,893 | 9,785 | 7,349 | 7,923 | 5,534 | 8,954 | 7,197 | 5,842 | 4,982 | 6,732 | 7,894 |

Gambar 4.1. Data Produksi Buah Sawit Tahun (2015-2017)

Pada data Perusahaan yang ditampilkan pada gambar 4.1 terdapat atribut Bulan, Tanggal, Jumlah Produksi, dan keterangan. Namun tidak semua atribut digunakan karena harus melalui beberapa tahap pengolahan awal data (*Preparation data*) untuk mendapatkan data yang berkualitas, beberapa teknik yang dilakukan adalah sebagai berikut (Vercellis, 2009) :

1. *Data validation*, untuk mengidentifikasi dan menghapus data yang ganjil (*outlier/noise*), data yang tidak konsisten, dan data yang tidak lengkap (*missing value*). Data *Missing Value* bisa dilihat pada Gambar 4.2



Gambar 4.2 data missing value (Data yang Tidak Lengkap)

2. *Data integration and Transformation*, untuk meningkatkan akurasi dan efisiensi algoritma. Data yang digunakan dalam penulisan ini bernilai kategorikal. Data ditransformasikan ke dalam *software RapidMiner*.
3. *Data size reduction and dicrtization*, untuk memperoleh data set dengan jumlah atribut dan record yang lebih sedikit tetapi bersifat informatif.

Data Mining

Data mining adalah proses mencari pola atau informasi menarik dalam sebuah data dengan menggunakan teknik atau metode tertentu. Teknik, metode, atau algoritma data mining sangatlah bervariasi. Pemilihan metode atau algoritma yang tepat sangat bergantung pada tujuan dan proses *data mining*. Dalam penelitian ini akan dilakukan eksperimen dengan menggunakan metode klasifikasi data mining *K-Nearest Neighbor* terhadap data hasil produksi yang terkait dengan peningkatan dan penurunan hasil produksi. Data akan diolah dengan menggunakan algoritma *K-Nearest Neighbor* dan menghasilkan model, maka terhadap model yang dihasilkan tersebut dilakukan pengujian menggunakan *k-fold cross validation*, kemudian dilakukan evaluasi dan validasi hasil dengan *confusion matrix* dan kurva *ROC*.

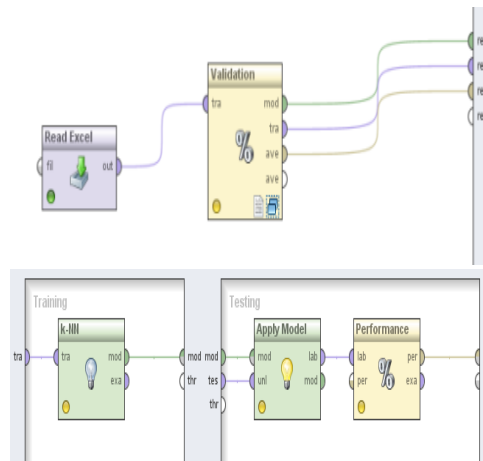
Proses Data Mining Menggunakan Rapid Miner

Setelah melakukan pengolahan data pada tahap *Data validation*, untuk mengidentifikasi dan menghapus data yang ganjil (*outlier/noise*), data yang tidak konsisten, dan data yang tidak lengkap (*missing value*), tahap *integration and Transformation*, untuk meningkatkan akurasi dan efisiensi algoritma, data ditransformasikan ke dalam *software RapidMiner* dan tahap terakhir yaitu *size reduction and dicrtization*, untuk memperoleh data set dengan jumlah atribut dan record yang lebih sedikit tetapi bersifat informatif di dapatkan 2 hasil kelas yaitu “Meningkat” dan “menurun”. Dan atribut yang dipakai dalam eksperimen yaitu : Bulan, Tanggal, Total Perbulan, dan Keterangan.

Pengujian Model

Pengujian model dalam penelitian ini menggunakan *Cross Validation* adalah teknik validasi dengan membagi data secara acak kedalam *k* bagian dan masing-masing bagian akan dilakukan proses

klasifikasi (Han & Kamber, 2006). Dengan menggunakan *cross validation* akan dilakukan percobaan sebanyak k . Data yang digunakan dalam percobaan ini adalah data *training* untuk mencari nilai *error rate* secara keseluruhan. Secara umum pengujian nilai k dilakukan sebanyak 10 kali untuk memperkirakan akurasi estimasi. Dalam penelitian ini nilai k yang digunakan berjumlah 10 atau *10-fold Cross Validation*.



Gambar 4.2 Ilustrasi 10 Fold Cross Validation Algoritma K-Nearest Neighbor

Pada data training (1) berfungsi untuk *example* data pada hasil produksi, pada data testing (2) berfungsi untuk *clustering* model pada sebuah data hasil produksi dimana data mulai menentukan jumlah *cluster* yang diperlukan pada tahap ini juga data mulai di *cluster*, *Validation* (3) berfungsi untuk memvalidasi data untuk proses *K-Nearest Neighbord* (KNN), pada *retrieve* berfungsi untuk *memformance* sebuah data dimana pada tahap ini menunjukkan nilai suatu data. Dengan menggunakan pemodelan *K-Nearest Neighbord* seperti Gambar 4.3 diatas.

| accuracy: 80,95% | | | |
|------------------|----------------|--------------|-----------------|
| | true Meningkat | true Menurun | class precision |
| pred. Meningkat | 12 | 2 | 85,71% |
| pred. Menurun | 2 | 5 | 71,43% |
| class recall | 85,71% | 71,43% | |

Gambar 4.3 Tingkat akurasi Algoritma K-Nearest Neighbor

Jumlah *True Positive* (TP) adalah 14 *record* diklasifikasikan sebagai Meningkat terpilih dan *False Negative* (FN) sebanyak 2 *record* diklasifikasikan sebagai Meningkat terpilih tetapi Menurun terpilih. Berikutnya 5 *record* untuk *True Negative* (TN) diklasifikasikan sebagai MENINGKAT terpilih, dan 2 *record* *False Positive* (FP) diklasifikasin sebagai MENURUN terpilih ternyata MENINGKAT. Berdasarkan Gambar 4.1 tersebut menunjukkan bahwa, tingkat akurasi dengan menggunakan algoritma tersebut adalah sebesar 85,15% dengan kondisi $k=5$. Dibawah ini adalah perbandingan tingkat akurasi dengan kondisi k yang berbeda-beda:

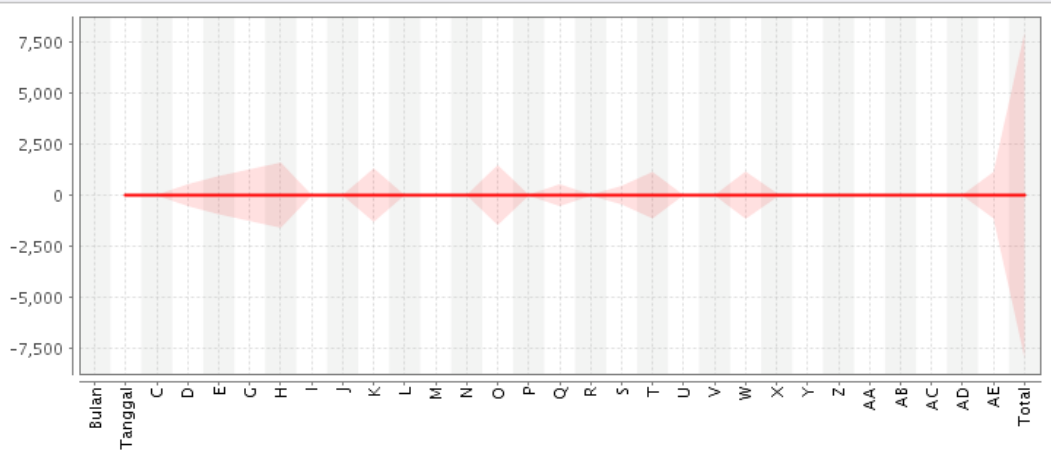
Tabel 4.1 Perbandingan Jumlah Cluster dan akurasi

| Cluster | Akurasi |
|---------|---------|
| K1 | 80,95% |
| K2 | 79,45% |
| K3 | 83,95% |
| K4 | 82,62% |
| K5 | 85,15% |

Dari perbandingan diatas, tingkat akurasi dapat dipengaruhi oleh jumlah kalstering data.

Evaluasi ROC curve

Untuk dapat melihat akurasi secara manual dilakukan perbandingan klasifikasi menggunakan *curva ROC* hasil ekspresi dari *confusion matrix*. Kurva ROC (*Receiver Operating Characteristic*) adalah cara lain untuk mengevaluasi akurasi dari klasifikasi secara visual (Vercellis, 2009). Dalam penelitian ini, setelah di evaluasi dengan *ROC curve*, menghasilkan nilai AUC (*Area Under Curve*) sebesar 0,888 dengan klatering data $k=5$.



Gambar 4.4 Grafik ROC Curve Algoritma K-Nearest Neighbor
Tabel 4.2 Perbandingan jumlah cluster dan akurasi dengan ROC Curve

| Cluster | Akurasi |
|---------|---------|
| K1 | 0,500 |
| K2 | 0,826 |
| K3 | 0,853 |
| K4 | 0,874 |
| K5 | 0,888 |

Tingkat akurasi dapat di diagnosa sebagai berikut (Gorunescu, 2011):

- Akurasi 0.90 – 1.00 = *Excellent classification*
- Akurasi 0.80 – 0.90 = *Good classification*
- Akurasi 0.70 – 0.80 = *Fair classification*
- Akurasi 0.60 – 0.70 = *Poor classification*
- Akurasi 0.50 – 0.60 = *Failure*

Sesuai dengan grafik dan tabel diatas, tingkat akurasi dengan ROC Curve menunjukkan akurasi yang tergolong *Good Classification* yaitu nilai AUC (*Area Under Curve*) sebesar 0,888 dengan klastering data $K=5$.

Kesimpulan

Berdasarkan hasil dan pembahasan penelitian mengenai Prediksi Hasil Produksi Buah Sawit Pada PT Bumi Sawit Sukses (BSS) dengan menggunakan proses data mining dengan metode *K-Nearest Neighbor* (K-NN) menghasilkan kesimpulan sebagai berikut :

- 1) Metode *K-Nearest Neighbor* (K-NN) dapat digunakan dalam klasifikasi data mining hasil produksi PT. Bumi Sawit Sukses (BSS) dengan menggunakan hasil produksi tahun 2015-2017.
- 2) Tahap yang dialkuakan dalam penelitian ini adalah *Data Validation*, dan *Data Integration and Tranformation*.

3) Model diuji dengan menggunakan *Software Rapidminer* dengan membagi data *Testing* dan data *Training*. Data yang digunakan dalam penelitian ini adalah data *Testing*.

4) Menetapkan *data training*

Data training yang digunakan merupakan data hasil produksi buah sawit PT Bumi Sawit Sukses (BSS) Tahun 2015-2017, *Data training* merupakan data yang sudah memiliki Klasifikasi. Klasifikasi ditetapkan berdasarkan hasil produksi setiap hari, dimana jika hasil produksi melebihi 5.000kg maka "MENINGKAT" dan jika kurang dari 5.000kg maka "MENURUN".

5) Menetapkan *data testing*

Data Testing merupakan data hasil produksi yang akan dilakukan klasifikasi, apakah melebihi hasil produksi yang ingin diprediksi.

DAFTAR PUSTAKA

- Asmadawati. 2014. "Menciptakan Pembelajaran Kreatif dan Kondusif". Jurnal Pendidikan IAIN Padangsidempuan. Diunduh Februari 2018 (<http://journal.iain.ac.id/index.php/judika/article/view/117/121>).
- Asri, Marwan dan Adi Saputro, Gunawan. 2000. *Anggaran Perusahaan*. Edisi 3. BPFE. Yogyakarta.
- Ahmad, Ridok dan Furcon, M.Tanzil. *Pengelompokan Bahasa Indonesia Menggunakan Metode k-NN* [Internet]. 2018 [Updated Februari 2018]. Available from: <http://mc4teknik.blogspot.com/2011/04/klasifikasi-citra-tekstur-menggunakan-k.html>.
- Dian Nuswantoro DKK Analisa Metode K-Nearest Neighbor Dalam Peramalan Penjualan Mobil Pada Pt Karya Zirang Utama Analysis Method Of K-Nearest Neighbor In Car Sales Forecasting At Pt Zirang [Internet]. Februari 2018 (<http://mahasiswa.dinus.ac.id/docs/skripsi/jurnal/19658.pdf>)
- Emerensye S. Y. Pandie, "Implementasi Algoritma Data Mining K-Nearest Neighbour (K-NN) Dalam Pengambilan Keputusan Pengajuan Kredit," *Seminar Nasional Sains dan Teknik 2012 (SAINSTEK 2012)*,.
- Ginting, Rosnani. 2007. *Sistem Produksi Edisi Pertama*, Yogyakarta : Graha Ilmu
- Gorunescu, F. 2011. *Data Mining Concept Model and Techniques*. Berlin: Springer. ISBN 978-3-642-19720-8
- Han, Jiawei dan Kamber, Micheline. (2006), *Data Mining : Concept and Techniques Second Edition*, Morgan Kaufmann Publishers.
- Heizer Jay, Render Barry. 2005. *Operations Management*. Jakarta: Salemba Empat.
- Herjanto, Eddy,. 2008:78. *Manajemen operasi*. Edisi ketiga..jakarta.grasindo
- Ishak, Aulia. 2010. *Manajemen Operasi, Edisi Pertama*. Yogyakarta: Graha Ilmu.
- Kotler,Philip.2007.*Manajemen Pemasaran*.Jakarta: PT.Indeks
- Kusrini, luthfi taufiq Emha, (2009), *Algoritma Data Mining*, Penerbit Andi, Yogyakarta.
- Krisandi, Nobertus., Helmi., Prihandono, Bayu (2013)"Algoritma K-Nearest Neighbor dalam Klasifikasi Data Hasil Produksi Kelapa Sawit pada PT. Minamas Kecamatan Parindu. Jurnal Buletin Ilmiah Math. Stat. dan Terapannya (Bimaster) Volume 02, No.1(2013), hal. 33-38. Internet Februari 2018 <http://jurnal.untan.ac.id/index.php/jbmstr/article/viewFile/1540/1484>
- Larose D, T., 2005, *Discovering knowledge in data : an introduction to data mining*, Jhon Wiley & Sons Inc.
- Larose D, T., 2006, *Data Mining Methods and Models*, Jhon Wiley & Sons, Inc.Hoboken New Jersey
- Luthfi, Taufiq Emha. 2009. Penerapan data Mining algoritma Asosiasi untuk meningkatkan penjualan. Jurnal DASI, Vol.10 No.1.
- M. Nazir, *Metode Penelitian*, Jakarta: Ghalia Indonesia, 1988.

- Pandie , Emerensye.S.Y. Sistem Informasi Pengambilan Keputusan Pengajuan Kredit Dengan Algoritma K-Nearest Neighbour (Studi Kasus: Koperasi Simpan Pinjam). Program Pascasarjana Universitas Diponegoro. Semarang. 2012.
- Pramudiono, I. 2006. Apa itu Data Mining? Dalam <http://datamining.japati.net/cgi-bin/indodm.cgi>.
Diakses tanggal Februari 2018
- Santoso H. Memahami Krisis Lanjut Usia. Jakarta: PT. Gunung Mulia; 2009
- Sugiarto dan Harijono, 2000. Peramalan Bisnis, PT Gramedia Pustaka Utama, Jakarta.
- Sugiyono. 2013. Metode Penelitian Pendidikan Pendekatan Kuantitatif, Kualitatif, dan R&D. Bandung: Alfabeta.
- Susanto S., Dedy Suryadi, 2010, Pengantar Data Mining – Menggali Pengetahuan dari Bongkahan Data, Andi, Yogyakarta
- Turban, E., dkk. 2005. Decision Support Systems and Intelligent Systems. Yogyakarta: Andi Offset.
- Widiarsana, I.G.A., 2011, *Metode Klasifikasi K-Nearest Neighbor (KNN)*, Fakultas Teknik, Universitas Udayana, Denpasar.