

# PENERAPAN DATA MINING UNTUK MENGETAHUI TINGKAT PRODUKSI PERKEBUNAN KELAPA SAWIT DENGAN METODE *CLUSTERING* PADA PT BAILANGU CAPITAL INVESTMENT

Hervanus<sup>1</sup> A. Haidar Mirza, S.T., M.Kom<sup>2</sup>.NovriHadinata, M.Kom<sup>3</sup>.

<sup>1</sup>)Mahasiswa Teknik Informatika Universitas Bina Darma, <sup>2</sup>) Dosen Ilmu  
Komputer, <sup>3</sup>) Dosen Ilmu Komputer Jln, Jend A Yani No. 12 Plaju  
Palembang

Email :hervanus88@yahoo.com<sup>1</sup>, haidarmirza@mail.binadarma.ac.id<sup>2</sup>,  
novri@mail.binadarma.ac.id<sup>3</sup>

**Abstrak** :PT  
BailanguKapitalInvesmentmerupakansalahsaturperusahaanmilikswasta  
yang bergerakdalambidangperkebunankelapasawit. Perusahaan  
saatinimemilikilahanperkebunan yang  
cukupluasdimanadalampengelolaanoperasionalperkebunanmemerlukanp  
enanganan yang serius.Permasalahannyaadalahjumlahlahan yang  
luasdenganperbedaanumurtanamandan proses  
perawatannyasangatmerepotkanbagianadministrasiuntukpengolahandata  
nya. Metodedata query yang  
saatinidigunakanterasalambatdikarenakansemakinbertambahbanyaknya  
data yang tersimpan. Lamanya proses  
tanamhinggamenunggusaatpanenmemerlukanpengarsipan data yang  
cukupbesar.  
PermasalahaninimenyebabkanSulitnyamengetahuitingkatproduksisawitp  
adaperusahaan.Keragamanmasatanamdanluaslahanserta data yang  
tersimpancukup lama memerlukanmetodepengolahan yang tepat.Data  
dapattersimpanselamabeberapatahundanselamatanamanmasihberproduk  
si data tersebutmasihdigunakan.Besarnyajumlah data yang  
adadiperlukanteknikkhususuntukmemanfaatkannya.Teknik yang paling  
tepat yang digunakanuntukpemanfatan data yang  
besaryaitumenggunakan data mining.Metode yang  
digunakandalampenelitianiniyaitumenggunakanmetode*clustering*dengan  
algoritma*K-means*.

## 1. Pendahuluan

### 1.1 Latar Belakang

Perkembanganduniateknologiinformasitelahmenyebabkanbanyak orang  
dapatmemperoleh data  
denganmudahdncepat.Penerapandanpenguasaanteknologimenjadifaktor yang  
menentukanbagiperusahaanuntukmemperolehkeunggulandalambersaing,  
khususnyadalambidangteknologiinformasi.Perubahanteknologi yang  
demikianpesatsepertiotomasisidankomputerisasiadalahdalamsegalabidangusahamenuntuk

alangan ilmuwan bidang teknologi informasi dan para praktisi yang berkecimpung di dalamnya untuk lebih siap menghadapi kemajuan dan tantangan yang ada.

PT Bailang Kapital Investment yang merupakan salah satu perusahaan milik swasta yang bergerak dalam bidang perkebunan kelapa sawit saat ini memiliki lahan perkebunan yang cukup luas di mana dalam pengelolaan operasional perkebunan memerlukan penanganan yang serius. Dengan jumlah lahan yang luas dengan perbedaan umur tanam dan proses perawatannya cukup merepotkan bagian administrasi untuk pengolahannya. Lamanya proses tanam hingga panen memerlukan pengarsipan data yang cukup besar.

Sulitnya mengetahui tingkat produksi sawit pada perusahaan ini merupakan salah satu permasalahan yang harus dipecahkan. Keragaman masatan dan luas lahan serta data yang tersimpan cukup lama memerlukan metode pengolahan yang tepat. Metode *data query* yang saat ini digunakan teras lambat dikarenakan semakin bertambah banyaknya data yang tersimpan. Data dapat tersimpan selama beberapa tahun dan saat ini masih berproduksi data tersebut masih digunakan. Besarnya jumlah data yang ada diperlukan teknik khusus untuk memanfaatkannya. Teknik yang paling tepat yang digunakan untuk pemanfaatan data yang besar yaitu menggunakan data mining dengan menggunakan metode *clustering*.

Algoritma yang digunakan dalam data mining ini menggunakan *k-mean* yang merupakan salah satu algoritma dalam metode clustering di mana dalam algoritma ini implementasinya mudah dan waktu untuk menjalankan algoritma ini lebih cepat. Tools yang digunakan untuk implementasi menggunakan Weka, dengan memanfaatkan format file *CSV* yang dapat diterjemahkan oleh aplikasi Weka.

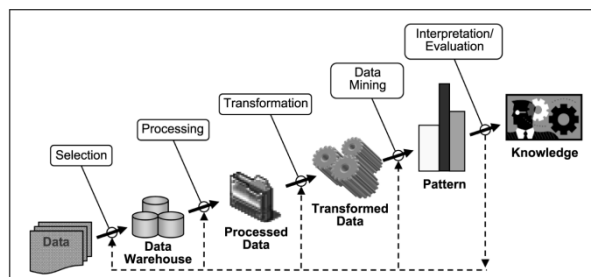
## 2. Metode dan Perancangan

### 2.1 Metode

Metode yang digunakan dalam penelitian ini adalah clustering dengan menggunakan algoritma *k-means*.

### 2.2 Proses Data Mining

#### A. Tahapan Data Mining



Gambar 1. Tahapan Data Mining sebagai Knowledge Discovery

#### 1. Data Selection

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.

## 2. *Pre-processing/ Cleaning*

Sebelum proses data mining dapat dilaksanakan, perlu dilakukan proses *cleaning* pada data yang menjadi fokus KDD. proses *cleaning* mencakup antara lain membuang duplikasi data, memeriksa data yang inkonsisten, dan memperbaiki kesalahan.

## 3. *Transformation*

*Coding* adalah proses transformasi pada data yang telah dipilih, sehingga data tersebut sesuai untuk proses data mining. Proses *coding* dalam KDD merupakan proses kreatif dan sangat tergantung pada jenis data dan informasi yang akan dicari dalam basis data.

## 4. *Data mining*

*Data mining* adalah proses mencari pola atau informasi menarik dalam data terpilih dengan menggunakan teknik atau metode tertentu. Teknik, metode, atau algoritma dalam data mining sangat bervariasi. Pemilihan metode atau algoritma yang tepat sangat bergantung pada tujuan dan proses KDD secara keseluruhan.

## 5. *Interpretation/ Evaluation*

Pola informasi yang dihasilkan dari proses Data mining perlu ditampilkan dalam bentuk yang mudah dimengerti oleh pihak yang berkepentingan. Tahap ini merupakan bagian dari proses KDD yang disebut *interpretation*. Tahap ini mencakup pemeriksaan apakah pola atau informasi yang ditemukan bertentangan dengan fakta atau hipotesis yang ada sebelumnya.

### B. *Metode Clustering*

*Clustering* adalah salah satu teknik *unsupervised learning* dimana kita tidak perlu melatih metode tersebut dengan data lain, tidak ada fase *learning*. Tujuan dari metode *clustering* adalah untuk mengelompokkan sejumlah data atau objek ke dalam klaster sehingga setiap *cluster* akan terdiri dari data yang semirip mungkin (Santosa, 2007).

### C. *Algoritma K-Means*

Menurut Hasan & Kamber algoritma *K-Means* bekerja dengan membagi data ke dalam k buah cluster yang telah ditentukan. Algoritma k-means merupakan model centroid. Model centroid adalah model yang menggunakan centroid untuk membuat cluster. Centroid adalah "titik tengah" suatu cluster.

## 3. Hasil dan Pembahasan

### 3.1 Hasil

Flowchart algoritma K-means :

- Penetapan jumlah cluster (K) pada penelitian ini yaitu berjumlah 3 (tiga) cluster.

Tabel 1. Titik pusat awal tiap cluster, luas, produksi 2013, produksi 2014 dan produksi 2015.

Tabel 1. Titik pusat awal tiap cluster.

Titik pusat awal	Afdeling	Luas	Prod2013	Prod2014	Prod2015
Cluster 1	5	154	29658	31292	27617
Cluster 2	13	205	29493	31508	27454

Cluster 3	25	233	29519	31442	26936
-----------	----	-----	-------	-------	-------

b. Hitung jarak setiap data ke pusat cluster antara objek ke centroid dengan perhitungan jarak Euclidean.

Persamaan yang digunakan adalah :

$$d(p, q) = \sqrt{(p_1 - q_1)^2 + (p_2 - q_2)^2 + (p_3 - q_3)^2 + (p_4 - q_4)^2 + (p_5 - q_5)^2}$$

Persamaan di atas digunakan karena atribut yang digunakan berjumlah 5.

Sebagai contoh, akan dihitung jarak dari data produksi pertama ke pusat cluster pertama dengan persamaan :

$$\begin{aligned} d(1,1) &= \sqrt{(1 - 5)^2 + (232 - 154)^2 + (29582 - 29658)^2 + (31119 - 31292)^2 + (27579 - 27619)^2} \\ d(1,1) &= \sqrt{(-4)^2 + (78)^2 + (-76)^2 + (-173)^2 + (-40)^2} \\ d(1,1) &= \sqrt{16 + 6084 + 5776 + 29929 + 1600} \\ d(1,1) &= \sqrt{43405} \\ d(1,1) &= 208,339 \end{aligned}$$

Dari perhitungan hasil di atas didapatkan hasil bahwa jarak data produksi pertama dengan cluster pertama adalah 208,339.

Jarak dari data produksi pertama dengan pusat cluster kedua dengan persamaan :

$$\begin{aligned} d(1,2) &= \sqrt{(1 - 13)^2 + (232 - 205)^2 + (29582 - 29493)^2 + (31119 - 31508)^2 + (27579 - 27454)^2} \\ d(1,2) &= \sqrt{(-12)^2 + (27)^2 + (89)^2 + (-389)^2 + (-125)^2} \\ d(1,2) &= \sqrt{144 + 279 + 7921 + 151321 + 15625} \\ d(1,2) &= \sqrt{175740} \\ d(1,2) &= 419,214 \end{aligned}$$

Dari perhitungan hasil di atas didapatkan hasil bahwa jarak data produksi pertama dengan cluster kedua adalah 419,214.

Jarak dari data produksi pertama dengan pusat cluster ketiga dengan persamaan :

$$\begin{aligned} d(1,2) &= \sqrt{(1 - 24)^2 + (232 - 233)^2 + (29582 - 29519)^2 + (31119 - 31442)^2 + (27579 - 26936)^2} \\ d(1,2) &= \sqrt{(-24)^2 + (-1)^2 + (63)^2 + (-323)^2 + (643)^2} \\ d(1,2) &= \sqrt{576 + 1 + 3969 + 104329 + 413449} \\ d(1,2) &= \sqrt{418428} \\ d(1,2) &= 646,860 \end{aligned}$$

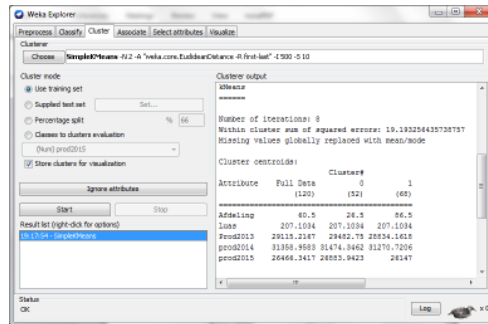
Dari perhitungan hasil di atas didapatkan hasil bahwa jarak data produksi pertama dengan cluster ketiga adalah 646,860.

Berdasarkan hasil ketiga perhitungan di atas maka dapat ditarik kesimpulan bahwa jarak produksi pertama yang paling dekat adalah cluster 1 sehingga data produksi pertama dimasukkan ke dalam cluster 1.

Langkah selanjutnya adalah menentukan metode serta algoritma, namun langkah ini harus disesuaikan dengan kebutuhan dalam pembahasan penelitian. Dalam hal ini teknik yang dipilih adalah teknik *clustering* serta algoritma yang dipilih adalah algoritma *K-means*. Algoritma *K-means* merupakan suatu penganalisaan data atau metode *data mining* yang melakukan proses permodelan *supersive*

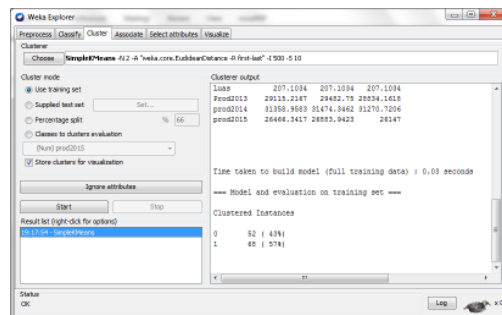
(*unspesived*) dan merupakan salah satu metode yang melakukan pengelompokan data dengan sistem partisi.

Selanjutnya kita tentukan parameter jumlah cluster yang ingin kita clusterisasi dalam hal ini diubah menjadi 3 cluster. Hasil *clusterer* output seperti ditunjukkan pada gambar 1. berikut ini.



Gambar 2. Clusterer Output

Nilai *cluster* yang telah diproses menggunakan tools WEKA (algoritma *K-Means*) Dari penjelasan tabel di atas bahwa tabel ini hasil akhir yang telah diproses dalam data mining. Di dalam tabel mempunyai beberapa atribut yaitu atribut *afdeling*, *luas*, *prod2013*, *prod2014* dan *prod2015*. Untuk *cluster* terdapat dua *cluster* yaitu *cluster 0* dan *cluster 1*.



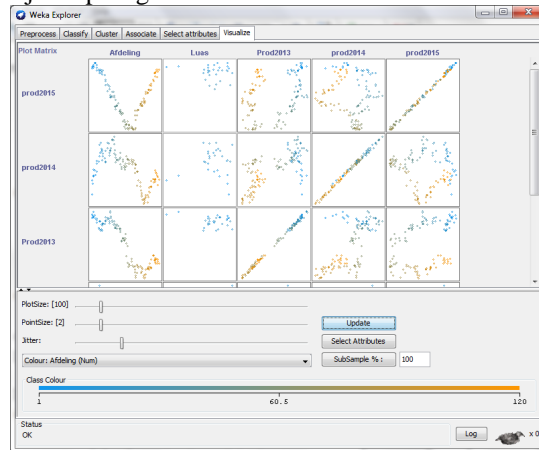
Gambar 3. Jumlah Pengelompokan cluster

Dari pengelompokan *clustering* pada tabel di atas untuk *cluster 0* yaitu 56%, dan *cluster 1* sebanyak 44%.

#### 4.3. Interpretation / Evaluation

Interpretation atau Evaluation merupakan polainformasi yang dapat dihasilkan dari proses *data mining* yang telah dilakukan sebelumnya, maka akan menghasilkan berupa grafik. Pada grafik ini merupakan hasil akhir tabel fitted mode yang dihasilkan . Di

grafik ini memberikan informasi jumlah produksi berdasarkan blok lahan dan tahun produksi. Seperti yang ditunjukkan pada gambar 4.10. di bawah ini.



**Gambar 4.** Grafik Jumlah Produksi Berdasarkan Tahun

## 4. Kesimpulan

### 4.1 Kesimpulan

Setelah melewati beberapa proses analisis data dan penerapan data mining dengan menggunakan aplikasi Weka maka dapat ditarik kesimpulan sebagai berikut:

1. Penerapan data mining dengan teknik *clustering* dan algoritma *K-Means* yang telah dilakukan menghasilkan sebuah informasi mengenai data produksi kelapa sawit pada PT Bailangu Capital Investment.
2. Informasi mengenai data produksi yang dihasilkan dari penerapan data mining yang dilakukan.
3. Semakin kecil batasan nilai data transaksi penjualan yang ditentukan maka semakin sedikit pula pola/aturan yang dapat dihasilkan dan waktu yang diperlukan lebih sedikit.
4. Waktu yang diperlukan dalam proses data mining relatif kecil untuk jumlah data yang ada.

### Referensi

1. Dimov, R, (2007), *Software Weka – Waikato Environment for Knowledge Analysis*, University of Saarland
2. Hermawati, F.A. (2013), *Data Mining*, Yogyakarta : Penerbit Andi
3. Kusriani, L, (2009), *Algoritma Data Mining*, Yogyakarta : Penerbit Andi
4. Lee, F S dan Santana, J. (2010). *Data Mining Meramalakan Bisnis Perusahaan*, Jakarta : PT Elex Media Komputindo
5. Nugroho, A, (2011), *Perancangan dan Implementasi Sistem Basis Data*, Yogyakarta : Penerbit Andi.
6. Novianti, D, (2012), *Penerapan Algoritma K-Means untuk Clustering Data Anggaran Pendapatan Belanja Daerah Kabupaten XYZ*, Gorontalo : Universitas Negeri Gorontalo.
7. Santosa, I, (2007), *“Komunikasi Data”*, Yogyakarta : Penerbit Andi.

8. Sirait, T.H., (2011), *Analisis Keberhasilan Mahasiswa dengan Metode Clustering K-Means*. Surabaya : Universitas Surabaya.