

PENERAPAN DATA MINING DENGAN METODE CLUSTERING UNTUK PENGELOMPOKAN DATA PENGIRIMAN BURUNG

Sri Mulyati¹⁾

¹Pascasarjana, Universitas Putra Indonesia “YPTK” Padang
e-mail: mulyati.sri52@gmail.com

Abstrak

Banyaknya data pengiriman burung pada Kantor Balai Karantina Pertanian Kelas I Jambi terjadi setiap tahunnya. Sehingga terjadi penumpukan data dan dapat diibaratkan seperti gunung data. Berlimpahnya data tersebut dapat kita manfaatkan untuk mencari informasi yang tersembunyi dari data-data tersebut. Untuk dapat mengetahui informasi yang tersembunyi dari data-data tersebut, maka perlu dilakukan pengolahan terhadap data-data tersebut. Proses pengolahan data tersebut disebut juga dengan *Data Mining*. Penelitian ini bertujuan untuk melakukan pengelompokan terhadap data pengiriman burung menggunakan teknik *Clustering*. Metode yang digunakan adalah *K-Means Clustering*. Dengan menggunakan metode ini data-data yang telah didapatkan dapat dikelompokkan ke dalam beberapa *cluster*, berdasarkan kemiripan dari data-data tersebut, sehingga data-data yang memiliki karakteristik yang sama dikelompokkan dalam satu *cluster* dan yang memiliki karakteristik yang berbeda dikelompokkan dalam *cluster* yang lain. Sehingga dengan adanya pengelompokan ini akan diketahui jenis burung apa saja yang sering dikirim oleh masyarakat sehingga dapat diketahui daerah persebaran burung tersebut. Hasil penelitian didapatkan bahwa penyebaran burung banyak berasal dari daerah Jambi dengan tujuan daerah di pulau Jawa. Selain Jambi sebagai daerah asal penyebaran burung Jambi juga sebagai transit penyebaran burung yang akan dikirim ke pulau Jawa. Sehingga apabila terdapat kasus berjangkitnya wabah penyakit Flu Burung (*Avian Influenza*) dapat segera dilakukan tindakan antisipasi salah satunya dengan melakukan metode penyuluhan pencegahan dan penanggulangan flu burung pada daerah asal pengiriman burung dan daerah tujuan pengiriman burung tersebut.

Kata Kunci : *Data Mining, Clustering, K Means*, Pengelompokan Burung, Balai Karantina Pertanian Kelas I Jambi

1. PENDAHULUAN

Saat ini banyak masyarakat menggemari *hobby* memelihara berbagai macam jenis burung berkicau seperti burung Murai Batu, Kacer, Cucak Hijau dan lain-lain. Selain suaranya yang merdu dan enak didengar, keindahan tubuhnya juga tak kalah menariknya. Dari segi nilai ekonomis, harga jual burung ini pun cukup tinggi di pasaran. Hal ini terlihat dari banyaknya data pengiriman burung pada Kantor Balai Karantina Pertanian Kelas I Jambi setiap tahunnya. Data tersebut meliputi data masuk dan keluarnya berbagai macam jenis burung.

Penelitian ini bertujuan untuk melakukan pengelompokan terhadap data pengiriman burung menggunakan teknik *Clustering*. Metode yang digunakan adalah *K-Means Clustering*. Dengan

menggunakan metode ini data-data yang telah didapatkan dapat dikelompokkan ke dalam beberapa *cluster*, berdasarkan kemiripan dari data-data tersebut, sehingga data-data yang memiliki karakteristik yang sama dikelompokkan dalam satu *cluster* dan yang memiliki karakteristik yang berbeda dikelompokkan dalam *cluster* yang lain.

Sehingga dengan adanya pengelompokan ini akan diketahui jenis burung apa saja yang sering dikirim oleh masyarakat sehingga dapat diketahui daerah persebaran burung tersebut. Sehingga apabila terdapat kasus berjangkitnya wabah penyakit Flu Burung (*Avian Influenza*) dapat segera dilakukan tindakan antisipasi pada daerah asal pengiriman burung dan daerah tujuan pengiriman burung tersebut.

Analisa *cluster* merupakan teknik multivariat yang mempunyai tujuan utama untuk mengelompokkan objek-objek berdasarkan karakteristik yang dimilikinya. Analisis *cluster* mengklasifikasikan objek sehingga setiap objek yang paling dekat kesamaannya dengan objek lain berada dalam *cluster* yang sama (Ediyanto *et al*, 2013).

2. KAJIAN LITERATUR

Data Mining merupakan ekstraksi informasi yang tersembunyi dari database yang besar. *Data Mining* adalah teknologi baru yang kuat dengan potensi besar untuk membantu perusahaan agar fokus pada informasi yang paling penting dalam gudang data mereka (Ramamohan *et al*, 2012).

Menurut Eko Prasetyo (2012), pekerjaan yang berkenaan dengan *Data Mining* dapat dibagi menjadi empat kelompok, yaitu model prediksi (*prediction modelling*), analisis kelompok (*Cluster analysis*), analisis asosiasi (*association analysis*) dan deteksi anomali (*anomaly detection*).

Analisa *cluster* merupakan teknik multivariat yang mempunyai tujuan utama untuk mengelompokkan objek-objek berdasarkan karakteristik yang dimilikinya. Analisis *cluster* mengklasifikasikan objek sehingga setiap objek yang paling dekat kesamaannya dengan objek lain berada dalam *cluster* yang sama (Ediyanto *et al*, 2013).

Langkah-langkah melakukan *Clustering* dengan metode *K Means* adalah sebagai berikut : (Ong, 2013)

- a. Pilih jumlah *cluster* K.
- b. Inisialisasi K pusat *cluster* ini bisa dilakukan dengan berbagai cara. Namun yang paling sering dilakukan adalah dengan cara random. Pusat-pusat *cluster* diberi

nilai awal dengan angka-angka random.

- c. Alokasi semua data/objek ke *cluster* terdekat. Kedekatan dua objek ditentukan berdasarkan jarak kedua objek tersebut. Demikian juga kedekatan suatu data ke *cluster* tertentu ditentukan jarak antara data dengan pusat *cluster*. Dalam tahap ini perlu dihitung jarak tiap data ke tiap pusat *cluster*. Jarak antara satu data dengan satu *cluster* tertentu akan menentukan suatu data masuk dalam *cluster* mana. Untuk menghitung jarak semua data ke setiap titik pusat *cluster* dapat menggunakan teori jarak *Euclidean* yang dirumuskan sebagai berikut :

$$D(ij) = \sqrt{(X_{1i} - X_{1j})^2 + (X_{2i} - X_{2j})^2 + \dots + (X_{ki} - X_{kj})^2} \dots\dots(1)$$

Di mana :

$D(ij)$ = jarak data ke (i) ke pusat *cluster* (j)

X_{ki} = Data ke (i) pada atribut data ke (k)

X_{kj} = Titik pusat (j) pada atribut (k)

- d. Hitung kembali pusat *cluster* dengan keanggotaan *cluster* yang sekarang. Pusat *cluster* adalah rata-rata dari semua data/ objek dalam *cluster* tertentu. Jika dikehendaki bisa juga menggunakan *median* dari *cluster* tersebut. Jadi rata-rata (*mean*) bukan satu-satunya ukuran yang bisa dipakai.

$$R_k = \frac{1}{N_k}(X_{1k} + X_{2k} + \dots + X_{nk}) \dots\dots(2)$$

di mana :

R_k = Rata-rata baru.

N_k = Jumlah *training*

pattern pada *cluster* (k).

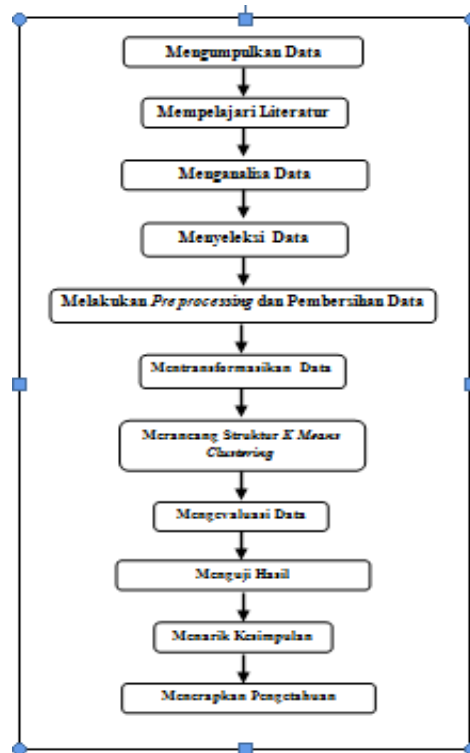
X_{nk} = Pola ke (n) yang menjadi bagian dari *cluster* (k).

- e. Tugaskan lagi tiap objek memakai pusat *cluster* yang baru. Jika pusat *cluster* tidak berubah lagi maka proses *clustering* selesai. Atau, kembali ke langkah nomor 3

sampai pusat *cluster* tidak berubah lagi.

3. METODE PENELITIAN

Kerangka kerja dalam penelitian ini dapat dilihat pada gambar 3.1 berikut ini :



Gambar 3.1 Kerangka Kerja Penelitian

Langkah-langkah penelitian dapat dijelaskan sebagai berikut :

1. Mengumpulkan Data
Pada tahap pengumpulan data ini, penulis menggunakan beberapa metode yaitu:
 - a. Metode wawancara, penulis melakukan tanya jawab langsung dengan staf kantor Balai Karantina Pertanian Kelas I Jambi mengenai data pengiriman burung yang meliputi data jenis burung, data pengirim, dan data penerima.
 - b. Metode observasi, penulis melakukan pengamatan langsung terhadap objek penelitian yaitu pada kantor Balai Karantina Pertanian Kelas I Jambi.
2. Mempelajari Literatur.

Penulis melakukan pengumpulan pengetahuan dari berbagai macam sumber literatur berupa buku-buku, jurnal dan karya ilmiah yang berkaitan dengan topik yang penulis angkat.

3. Menganalisa Data.
Pada tahap ini penulis melakukan analisa data menggunakan tahapan KDD (*Knowledge Discovery in Database*). Di mana tahapan KDD ini meliputi : Seleksi Data, *Preprocessing* dan Pembersihan Data, Transformasi Data, *Data Mining*, Evaluasi/Interpretation.
4. Menyeleksi Data (*Data Selection*)
Pada proses ini dilakukan pemilihan himpunan data, menciptakan himpunan data target, atau memfokuskan pada subset variabel (*sampel data*) di mana penemuan (*Discovery*) akan dilakukan. Hasil seleksi disimpan dalam suatu berkas yang terpisah dari basis data operasional *Preprocessing*.
5. Melakukan *Preprocessing* dan Pembersihan Data (*Cleaning Data*)
Pre processing dan *Cleaning Data* dilakukan dengan membuang data yang tidak konsisten dan *noise*, duplikasi data, memperbaiki kesalahan data dan bisa diperkaya dengan data eksternal yang relevan.
6. Mentransformasi Data (*Data Transformation*)
Proses ini mentransformasikan atau menggabungkan data ke dalam yang lebih tepat untuk melakukan proses *Mining* dengan cara melakukan peringkasan (*Agregasi*).
7. Merancang Struktur *K Means Clustering*
Proses *Data Mining* yaitu proses mencari pola atau informasi menarik dalam data terpilih dengan menggunakan teknik, metode atau algoritma tertentu sesuai dengan tujuan dari proses KDD secara keseluruhan. Algoritma yang digunakan dalam penelitian ini menggunakan *K Means Clustering*.
8. Mengevaluasi Data (*Evaluation Data*)
Tahap selanjutnya yaitu mengevaluasi data. Tahap ini

digunakan untuk menguji kualitas dari data apakah pola atau informasi yang ditemukan bersesuaian atau bertentangan dengan fakta sebelumnya.

9. Menguji Hasil

Tahap pengujian hasil adalah suatu teknik yang digunakan untuk menentukan bahwa penerapan metode yang digunakan telah mampu memecahkan masalah. Pada penelitian ini dilakukan pengujian dengan menggunakan aplikasi *Data Mining RapidMiner* versi 5.3. Berdasarkan pengetahuan yang diperoleh dari pola-pola yang terbentuk, maka pola yang terbentuk tersebut dapat dipresentasikan ke dalam bentuk visualisasi.

10. Menarik Kesimpulan

Setelah melakukan analisa data menggunakan tahapan KDD maka dapat ditarik suatu kesimpulan apakah informasi yang ditemukan berkesesuaian dengan fakta sebelumnya atau tidak, sehingga akan diperoleh pengetahuan baru.

11. Menerapkan Pengetahuan.

Langkah selanjutnya yaitu menerapkan pengetahuan yang diperoleh, sehingga diharapkan dapat membantu pimpinan organisasi/perusahaan dalam proses pembuatan keputusan yang akan datang.

4. HASIL DAN PEMBAHASAN

Hasil dari penelitian ini dapat dilihat pada tabel 4.1 berikut ini :

Tabel 4.1 Hasil Cluster Satu

Hasil Cluster Satu		
Cluster satu terdiri dari 345 item yang terdiri dari :		
Jenis Burung	Daerah Asal	Daerah Tujuan
a. Murai Batu = 111	a. Jambi = 344	a. Jawa Barat = 114
b. Kacer = 81	b. Sumatera Selatan = 1	b. Jawa Timur = 107
c. Cucak Hijau = 79		c. DI Yogyakarta = 45
d. Cililin = 31		d. Banten = 32
e. Love Bird = 29		e. Kepulauan Riau = 16
f. Tengkek Buto = 9		f. Sulawesi Selatan = 12
g. Kenari = 4		g. Sumatera Utara = 9
h. Cucak Lilin = 1		h. Bangka Belitung = 3
		i. Lampung = 2
		j. Kalimantan Timur = 3
		k. Aceh = 1
		l. Kalimantan Barat = 1

Dari hasil pengelompokan *cluster* satu di atas, dapat dianalisa bahwa penyebaran burung paling banyak di pulau Jawa, khususnya pada provinsi Jawa Barat dan Jawa Timur, sehingga apabila terjadi wabah flu burung dapat dilakukan antisipasi di kedua daerah tersebut.

Tabel 4.2 Hasil Cluster Dua

Hasil Cluster Dua		
Cluster dua terdiri dari 703 item yang terdiri dari :		
Jenis Burung	Daerah Asal	Daerah Tujuan
a. Murai Batu=316	a. Jambi = 581	a. DKI Jakarta = 315
b. Kacer = 142	b. Jawa Tengah = 43	b. Jawa Tengah = 261
c. Cucak Hijau = 108	c. Kalimantan Barat = 32	c. Jambi = 127
d. Cililin = 64	d. Jawa Timur = 12	
e. Love Bird = 73	e. DI Yogyakarta = 11	
	f. Jawa Barat = 10	
	g. Banten = 3	
	h. DKI Jakarta = 4	
	i. Kepulauan Riau = 2	
	j. Sumatera Utara = 2	
	k. Sumatera Selatan = 1	
	m. Bangka Belitung = 1	
	n. Sumatera Barat = 1	

Berdasarkan data hasil pengelompokan *cluster* dua, dapat dianalisa bahwa daerah asal penyebaran burung selain dari Jambi juga berasal dari provinsi di luar Jambi dengan tujuan paling banyak di provinsi DKI Jakarta, Jawa Tengah dan provinsi Jambi sendiri. Di sini provinsi Jambi berfungsi sebagai daerah transit pengiriman burung. Untuk itu, apabila terjadi wabah flu burung dapat dilakukan antisipasi di daerah Jambi dan daerah tujuan terutama daerah DKI Jakarta dan Jawa Tengah.

Tabel 4.3 Hasil Cluster Tiga

Hasil Cluster Tiga		
Cluster tiga terdiri dari 257 item yang terdiri dari :		
Jenis Burung	Daerah Asal	Daerah Tujuan
a. Tengkek Buto = 56	a. Jambi = 211	a. DKI Jakarta = 100
b. Kenari = 47	b. Jawa Tengah = 19	b. Jawa Tengah = 47
c. Kapas Tembak = 44	c. Kalimantan Barat = 5	c. Jambi = 44
d. Cucak Lilin = 37	d. Jawa Timur = 3	d. Jawa Barat = 29
e. Merpati = 17	e. DI Yogyakarta = 4	e. Jawa Timur = 21
f. Cucak Ranting = 13	f. Jawa Barat = 3	f. DI Yogyakarta = 10
g. Poksai = 13	g. Banten = 3	g. Banten = 2
h. Serindit = 9	h. DKI Jakarta = 1	h. Kepulauan Riau = 2
i. Beo = 10	i. Kepulauan Riau = 2	i. Sulawesi Selatan = 1
j. Ciblek = 11	j. Sumatera Utara = 3	
	k. Sumatera Selatan = 2	
	l. Bangka Belitung = 1	

Berdasarkan data hasil pengelompokan *cluster* tiga, dapat dianalisa bahwa daerah asal penyebaran burung selain dari Jambi juga berasal dari provinsi di luar kota Jambi. Disini provinsi Jambi juga berfungsi sebagai daerah transit pengiriman burung yang nantinya akan dikirimkan ke daerah tujuan di provinsi-provinsi lain yang kebanyakan berada di pulau Jawa. Untuk itu apabila terjadi wabah flu burung dapat diantisipasi di daerah Jambi dan daerah lain di pulau Jawa seperti daerah DKI Jakarta, Jawa Tengah dan Jawa Timur.

Berdasarkan data hasil dari *clustering* di atas, maka dapat dilakukan metode penyuluhan pencegahan flu burung yang dapat dilihat pada tabel 4.4 berikut ini :

Tabel 4.4 Metode Penyuluhan Pencegahan Flu Burung

No.	Metode Penyuluhan Pencegahan Flu Burung	Cluster Satu	Cluster Dua	Cluster Tiga
1.	Pendekatan Perorangan a. Cetak, CD dan brosur. b. Kunjungan ke rumah/ tempat usaha. c. Komunikasi melalui HP/ Telepon.	✓ ✓ ✓	✓ ✓	✓ ✓
2.	Pendekatan Kelompok a. Studi banding ke tempat komunitas penggemar burung. b. Loka karya/ kursus.	✓ ✓		
3.	Pendekatan Massal a. Kampanye/ceramah. b. Berita koran/ Majalah, poster dan spanduk. c. Siaran radio / Televisi, pemutaran film, pameran dan pertunjukan kesenian.	✓ ✓ ✓	✓ ✓	✓ ✓

5. KESIMPULAN

Dari hasil analisa pengelompokan data penyebaran burung menggunakan *K Means Clustering* maka dapat diambil kesimpulan sebagai berikut :

1. Terbentuk tiga *cluster* daerah penyebaran burung yaitu *cluster* satu dengan jumlah 345 *item* di mana penyebaran burung berasal dari daerah Jambi dengan tujuan daerah Jawa Barat dan Jawa Timur, *cluster* dua dengan jumlah 703 *item* di mana selain daerah asal penyebaran burung Jambi juga merupakan daerah transit pengiriman burung ke daerah lain seperti daerah DKI Jakarta dan Jawa Tengah dan *cluster* tiga dengan jumlah

257 *item* di mana penyebaran burung berasal dari daerah Jambi dan Jambi juga merupakan daerah transit pengiriman burung dengan daerah tujuan pengiriman DKI Jakarta, Jawa Tengah dan Jawa Timur.

2. Metode penyuluhan pencegahan flu burung dapat dilaksanakan dengan baik berdasarkan *cluster* yang terbentuk, sehingga masyarakat dapat terhindar dari penyakit flu burung dan dapat mengantisipasinya dengan baik.

6. REFERENSI

Amita Verma and Ashwani Kumar, (2014),” *Performance Enhancement of K Means Clustering Algorithms for High Dimensional Data Sets*”, International Journal of Advanced Research in Computer Science and Software Engineering, Vol.4, Issue 1, January 2014

Ediyanto *et al*, (2013). “*Pengklasifikasian Karakteristik Dengan Metode K Means Cluster Analysis*”, Buletin Ilmiah Mat. Stat. Terapannya (Bimaster), Volume 02, No.2, (2013)

Eko Prasetyo, (2012), “*Data Mining Konsep dan Aplikasi Menggunakan Matlab*”, Penerbit Andi, Yogyakarta

Er. Nikhil Chaturvedi & Er. Anand Rajavat, (2013).”*An Improvement in K Means Clustering Algorithm Using Better Time and Accuracy*”, International Journal Of Programming Languages and Applications (IJPLA), Vol.3, No.4, October 2013

Fajar Astuti Hermawati, (2009),” *Data Mining*”, Penerbit Andi, Yogyakarta

Feri Sulianta, (2014), “ *Customer Profiling Pada Supermarket Menggunakan Algoritma K Means Dalam Memilih Produk Berdasarkan Selera Konsumen Dengan Daya Beli Maksimum*”, Jurnal Ilmiah Teknologi Informasi Terapan, ISSN 2407-3911, Vol.1, No.1, Desember 2014

Johan Oscar Ong, (2013), “*Implementasi Algoritma K Means Clustering Untuk Menentukan Strategi Marketing President University*”, Jurnal Ilmiah Teknik Industri, Vol.12, No.1, Juni 2013, ISSN 1412-6869

Madlu Yedla *et al*, (2010),” *Enhancing K Means Clustering Algorithm with Improved Initial Center*”, International Journal of Computer Science and Information Technologiest (IJCSIT), Vol.1, No.2, 2010, 121-125

Oyelade O.J, *et al*, (2010), “*Application of K Means Clustering Algorithm For Prediction of Students Academic Performance*”, International Journal of Computer Science and Information Security (IJCSIS), Vol. 7, No. 1, 2010

Pratiyush Guleria and Manu Sood, (2014), “*Data Mining in Education : A Review On The Knowledge Discovery Perspective*”, International Journal Of Data Mining & Knowledge Management Process (IJDKP), Vol.4, No.5, September 2014

Rendy Handoyo, *et al*, (2014), “*Perbandingan Metode Clustering Menggunakan Metode Single Linkage dan K Means Pada Pengelompokan Dokumen*”, JSM STMIK Mikroskill, ISSN, 1412-0100, Vol.15, No.2, Oktober 2014

Simmi Bagga & G. N Singh, (2011). “*Three Phase Iterative Model of KDD*”, International Journal Of Information Technology and Knowledge Management, July-December 2011, Vol. 4, No 2, pp 695-697

Vipin Kumar, *et al*, (2013), “*K Means Clustering Approach to Analyze NSL-KDD Intrusion Detection Dataset*”, International Journal of Soft Computing and Engineering (IJSCE), ISSN : 2231-2307, Vol.3, Issue-4, September 2013

Y. Ramamohan *et al*, (2012), “*A Study of Data Mining Tools In Knowledge Discovery Process*”, International Journal of Soft Computing and Engineering (IJSCE), ISSN: 2231-2307, Vol.2, Issue-3, July 2012

[Online] Tersedia :[http:// ribhy.iniaja.com/just/bahasa-indonesia-just/metode-penelitian/](http://ribhy.iniaja.com/just/bahasa-indonesia-just/metode-penelitian/) diakses tanggal 1 April 2015

[Online] Tersedia :[https:// id.wikipedia.org/wiki/fluburung/](https://id.wikipedia.org/wiki/fluburung/) diakses tanggal 9 Juli 2015

[Online] Tersedia :[https:// id.wikipedia.org/wiki/burung/](https://id.wikipedia.org/wiki/burung/) diakses tanggal 30 Juli 2015

[Online] Tersedia : Penulis Inang Sariati, [https:// cybex.org/metode-penyuluhan-pencegahan-flu-burung/](https://cybex.org/metode-penyuluhan-pencegahan-flu-burung/) diakses tanggal 28 April 2015