

# ANALISA ALGORITMA K-MEANS CLUSTERING PEMETAAN JUMLAH TINDAK PIDANA

<sup>1</sup>Jajang Jaya Purnama, <sup>2</sup>Ridan Nurfalalah, <sup>3</sup>Sri Rahayu, <sup>4</sup>Hafifah Bella Novitasari

<sup>1,2,3,4</sup>Ilmu Komputer, STIMIK Nusa Mandiri Jakarta, Indonesia

Jl. Kramat Raya No. 18 Kelurahan Kwitang Kec. Senen

Telp. (021) 31908575 Fax. (021) 31908565

E-mail: <sup>1</sup>[Jajangja2412@bsi.ac.id](mailto:Jajangja2412@bsi.ac.id), <sup>2</sup>[ridannur1312@bsi.ac.id](mailto:ridannur1312@bsi.ac.id), <sup>3</sup>[srahayu110527](mailto:srahayu110527)

<sup>4</sup>[hafifah.bella@gmail.com](mailto:hafifah.bella@gmail.com)

## Abstract

*Marine fisheries is an effort to catch marine fish, fishermen catch fish in the sea using two kinds of ways, namely through traditional and modern methods. to support the daily lives of fishermen looking for sea fish. the abundance of marine fish in the Indonesian sea means that the processed food is also very diverse. The high level of crime in the Indonesian sea is a mirror of the quality of the Indonesian navy's military defense, considering that the Indonesian sea is very rich in marine resources which makes fish thieves from neighboring countries tempted to catch fish in the Indonesian sea. in general it can be called a crime if unlicensed fishermen, illegal fishing gear, without permission and illegal fishing gear, falsification of documents, incomplete documents, shocking (ACCU), carrying explosives / bombs, fishing ground, fishing ground and illegal logging equipment , fish transportation / transshipment, without information on criminal types of fisheries, transshipment and fishing gear, no transmitter, theft of coral reefs, unsuitable fishing gear (SIPI), incomplete documents and fishing ground, foreign crew members not suitable for SIPI, not fishing in accordance with SIKPI, documents are incomplete and there are no transmitters, SIB is not valid, SLO (SIB is not in accordance with SIPI), without permits and fake documents, sea sand without documents, do not have SLO, loading and unloading is not SIPI, uses chemical / biological / explosives, fishing in the Gray Area / illegal fishing equipment / returned to the country of origin related to the MoU. Based on the background described there are problems that occur, the formulation of the problem in this study are: Analyzing k-means clustering by using proximity euclidean distance distance, How to group data using K-means clustering for illegal fishing crime into the category of illegal crime fishing the highest, medium, and sufficient cases in 266 data with the euclidean distance calculation.*

**Keywords :** *criminal act, clustering, k-means*

## Abstrak

*Perikanan laut merupakan usaha menangkap ikan laut, para nelayan menangkap ikan di laut menggunakan dua macam cara yaitu melalui cara tradisional dan modern. untuk menunjang kehidupan sehari-hari nelayan mencari ikan kelaut. melimpahnya ikan laut di laut Indonesia berarti menjadikan olahan masakannya juga sangat beragam. Tingginya tingkat kejahatan di laut Indonesia merupakan cermin kualitas pertahanan militer angkatan laut Indonesia, mengingat laut Indonesia sangat kaya akan baharinya yang membuat para pencuri ikan dari negara-negara tetangga menjadi tergiur untuk menangkap ikan di laut Indonesia. secara umum bisa disebut tindak kejahatan bila mana nelayan tanpa ijin, alat tangkap terlarang, tanpa ijin dan alat tangkap terlarang, pemalsuan dokumen, dokumen tidak lengkap, penyetruman (ACCU), membawa bahan peledak/bom, fishing ground, fishing*

*ground dan alat tagkap terlarang, pengangkutan ikan/transshipment, tanpa keterangan jenis pidana perikanan, transshipment dan alat tangkap, tidak ada transmitter, pencurian terumbu karang, alat tangkap tidak sesuai ijin (SIPI), dokumen tidak lengkap dan fishing ground, ABK asing tidak sesuai SIPI, menampung ikan tidak sesuai SIKPI, Dokumen tidak lengkap dan tidak ada transmitter, SIB tidak berlaku, SLO (SIB tidak sesuai dengan SIPI), tanpa ijin dan dokumen palsu, pasir laut tanpa dokumen, tidak memiliki SLO, bongkar muat tidak sesuai SIPI, menggunakan bahan kimia/biologis/peledak, penangkapan ikan di daerah Grey Area/alat tangkap terlarang/dikembalikan ke negara asal terkait MoU. Berdasarkan latar belakang yang telah diuraikan terdapat permasalahan yang terjadi, rumusan permasalahan dalam penelitian ini adalah : Menganalisa k-means clustering dengan menggunakan kedekatan jarak euclidean distance, Bagaimana melakukan pengelompokan data menggunakan K-means clustering bagi tindak pidana ilegal fishing kedalam kategori tindak pidana ilegal fishing paling tinggi, menengah, dan cukup.studi kasus pada 266 data dengan perhitungan euclidean distance.*

**Kata Kunci :** *tindak pidana, clustering, k-means*

## **1. PENDAHULUAN**

### **1.1 Latar Belakang**

Perikanan laut merupakan usaha menangkap ikan laut, para nelayan menangkap ikan di laut menggunakan dua macam cara yaitu melalui cara tradisional dan modern. Untuk menunjang kehidupan sehari-hari nelayan mencari ikan kelaut. Melimpahnya ikan laut di laut Indonesia berarti menjadikan olahan masakannya juga sangat beragam, seperti Jepang yang dikenal dengan sushi-sashimi Salmon, beberapa daerah di Indonesia juga memiliki laut unggulan, seperti ikan tongkol, ikan cakalang, ikan tuna, ikan teri, ikan kakap, ikan kembung, ikan makarel, ikan baronang, ikan tenggiri, ikan belanak.

*Knowledge discovery in Database* (KDD) didefinisikan sebagai ekstraksi informasi potensial, implisit dan tidak dikenal dari sekumpulan data. Proses *knowledge discovery* melibatkan hasil dari proses data mining (proses mengekstrak kecenderungan pola suatu data). Kemudian mengubah hasilnya secara akurat menjadi informasi yang mudah dipahami. Ada beberapa macam pendekatan berbeda yang diklasifikasikan sebagai teknik pencarian informasi/pengetahuan dalam KDD. Ada pendekatan kuantitatif, seperti pendekatan probabilistik dan statistik. Beberapa pendekatan memanfaatkan teknik visualisasi, pendekatan klasifikasi seperti logika induktif, pencarian pola dan analisis. Pendekatan yang lain meliputi deviasi, analisis kecenderungan, algoritma genetik, jaringan syaraf tiruan dan pendekatan campuran dua atau lebih dari beberapa pendekatan yang ada.

Tingginya tingkat kejahatan di laut Indonesia merupakan cermin kualitas pertahanan militer angkatan laut Indonesia, mengingat laut Indonesia sangat kaya akan baharinya yang membuat para pencuri ikan dari negara-negara tetangga menjadi tergiur untuk menangkap ikan di laut Indonesia. TNI Angkatan Laut dan Kementerian kelautan dan perikanan harus bekerja keras menjaga kelestarian ikan dari orang-orang yang tidak bertanggung jawab. Sistem informasi dapat digunakan untuk mendapatkan, mengolah dan menyebarkan informasi untuk menunjang kegiatan pengambilan keputusan strategis.

Metode pengambilan keputusan konvensional yang ada, tidak dapat menangani data dalam jumlah yang sangat besar. Hal ini mendorong munculnya

ilmu baru untuk mengatasi masalah penggalian informasi yang penting dari kumpulan data, yang disebut dengan data mining.

Secara umum bisa disebut tindak kejahatan bila mana nelayan tanpa ijin, alat tangkap terlarang, tanpa ijin dan alat tangkap terlarang, pemalsuan dokumen, dokumen tidak lengkap, penyetruman (ACCU), membawa bahan peledak / bom, *fishing ground*, *fishing ground* dan alat tangkap terlarang, pengangkutan ikan / *transshipment*, tanpa keterangan jenis pidana perikanan, *transshipment* dan alat tangkap, tidak ada *transmitter*, pencurian terumbu karang, alat tangkap tidak sesuai ijin (SIPI), dokumen tidak lengkap dan *fishing ground*, ABK asing tidak sesuai SIPI, menampung ikan tidak sesuai SIKPI, Dokumen tidak lengkap dan tidak ada *transmitter*, SIB tidak berlaku, SLO (SIB tidak sesuai dengan SIPI), tanpa ijin dan dokumen palsu, pasir laut tanpa dokumen, tidak memiliki SLO, bongkar muat tidak sesuai SIPI, menggunakan bahan kimia/biologis/peledak, penangkapan ikan di daerah *Grey Area*/alat tangkap terlarang/dikembalikan ke negara asal terkait MoU.

Seiring dengan terus bertambahnya jumlah data tindak pidana setiap tahun, maka jumlah data yang tindak pidana terus meningkat sehingga penumpukan data yang belum diolah dengan optimal untuk menggali informasi dan pengetahuan baru melalui pola-pola yang terbentuk dari penumpukan data tersebut. Jumlah data yang terus meningkat ini memerlukan beberapa teknik ataupun metode untuk mengolahnya menjadi sebuah informasi dan pengetahuan yang dapat digunakan sebagai bahan pertimbangan kementerian kelautan dan perikanan dalam proses pengambilan kebijakan dan keputusan juga sebagai peringatan dini (*early warning*) bagi tindak pidana tertentu yang berdasarkan hasil pengelompokan tahun tindak pidana terendah.

## 1.2 Identifikasi Masalah

Berdasarkan latar belakang yang telah diuraikan terdapat permasalahan yang terjadi, pengelompokan tindak pidana merupakan hal yang sangat penting dalam sebuah pemerintahan karena hal tersebut bisa meminimalisir tindak pidana dalam *illegal fishing*, sehingga diperlukan metode untuk mengelompokkan tindak pidana sehingga bisa membantu TNI AL dan kementrerian Kelautan dan Perikanan untuk membasmi dan menangkap para tindak pidana *illegal fishing*. Rumusan permasalahan dalam penelitian ini adalah :

- a. Menganalisa K-Means Clustering dengan menggunakan kedekatan jarak *euclidean distance*.
- b. Bagaimana melakukan pengelompokan data menggunakan K-means clustering bagi tindak pidana ilegal *fishing* kedalam kategori tindak pidana *illegal fishing* paling tinggi, menengah, dan cukup.

## 1.3 Tujuan

Tujuan dari penelitian ini adalah memberikan penjelasan bagaimana mengelompokkan data dengan menggunakan metode K-means clustering

Manfaat Penelitian

- a. Penelitian ini diharapkan dapat digunakan oleh Kementerian Kelautan dan Perikanan dalam pengelompokan tindak pidana *illegal fishing*.
- b. Peringatan bagi TNI Angkatan Laut, daerah mana saja dan kejahatan mana yang sering dilakukan

- c. Data yang sudah dimining dan menghasilkan pengelompokan tindak pidana *illegal fishing*, maka informasi tersebut dapat dipergunakan sebagai patokan dalam meminimalisir tindak pidana *illegal fishing*.

#### 1.4 Metode Penelitian

Metode penelitian yang digunakan dalam penelitian ini adalah metode kuantitatif untuk mengelompokan tindak pidana *illegal fishing* kategori tindak kriminal paling tinggi, cukup dan menengah. Data yang digunakan sebanyak 267 data.

a. Jenis Penelitian

Jenis penelitian yang digunakan dalam penelitian ini adalah model penelitian eksperimen. Model eksperimen merupakan metode penelitian yang digunakan untuk mencari pengaruh perlakuan tertentu (Sugiyono, 2010). Jenis penelitian yang dilakukan pada jurnal ini yaitu menggunakan *k-means* untuk mengelompokan tindak pidana *illegal fishing*. Jenis ini mengambil algoritma *k-means* sebagai fungsi pengelompokan, yang merupakan ekstraksi pola yang menarik dari data dalam jumlah besar.

1) Studi Pendahuluan

Kegiatan yang dilakukan pada saat studi pendahuluan yaitu mengumpulkan materi-materi kepustakaan yang berhubungan dengan pengambilan judul. Dari penelitian tersebut akan di dapat beberapa kriteria yang digunakan untuk penelitian lebih lanjut.

2) Pengumpulan Data

Melakukan pengambilan data dari dataset global diambil dari data.go.id yang terdiri dari beberapa atribut dan berjumlah 267 data.

3) Pengolahan Data

Setelah mendapatkan data maka data diolah menggunakan metode data mining, yaitu pertama kali menggunakan Microsoft Excel.

4) Laporan

Setelah tahap pendahuluan, data, dan pengolahan data tersebut dilakukan maka disusunlah laporan penelitian ini kedalam bentuk jurnal ilmiah

b. Metode Pengumpulan Data

Data yang diolah dalam penelitian ini merupakan data sampel yang diambil dari dataset global tindak pidana, variabel didalamnya terdiri dari id, tindak pidana, tahun, jumlah.

c. Ruang Lingkup

Sejumlah permasalahan yang dibahas dalam usulan penelitian ini akan dibatasi ruang lingkup pembahasannya, antara lain:

1) Dataset global merupakan tindak pidana dari tahun 2003 sampai dengan tahun 2013 data dari Kementerian Kelautan dan Perikanan.

2) Variabel yang digunakan dalam penelitian ini merupakan atribut id, tindak pidana, tahun, jumlah

## 2. LANDASAN TEORI

### 2.1. Tinjauan Pustaka

#### 2.1.1 Data Mining

Data Mining adalah suatu istilah yang digunakan untuk menguraikan pengetahuan dalam database. Beberapa pengertian data mining menurut para ahli dalam buku (kusrini.2009).

- a. "*Data mining* adalah proses yang menggunakan teknik statistik, matematika, kecerdasan buatan, dan *machine learning* untuk mengekstraksi dan mengidentifikasi informasi yang bermanfaat dan pengetahuan yang terkait dari berbagai database besar" (Turban, dkk.2005).
- b. "Data mining adalah serangkaian proses untuk menggali nilai tambah dari suatu kumpulan data berupa pengetahuan yang selama ini tidak diketahui secara manual". (Pramudiono, 2006)
- c. "Data Mining adalah analisis otomatis dari data yang berjumlah besar atau kompleks dengan tujuan menemukan pola atau kecenderungan yang penting yang biasanya tidak disadari keberadaannya." (Pramudiono, 2006)
- d. "Data mining merupakan bidang dari beberapa bidang keilmuan yang menyatukan teknik dari pembelajaran mesin, pengenalan pola, statistik, database, dan visualisasi untuk penanganan permasalahan pengambilan informasi dari database yang besar." (Larose,2005)
- e. "*Data mining* merupakan analisis dari peninjauan kumpulan data untuk menemukan hubungan yang tidak di duga dan meringkas data dengan cara yang berbeda dengan sebelumnya, yang dapat dipahami dan bermanfaat bagi pemilik data" (Larose,2005)

#### 2.1.2 Clustering

Metode *hierarchical clustering* berbeda dengan metode non *hierarchical clustering*, metode ini justru dimulai dengan menentukan terlebih dahulu jumlah *cluster* yang diinginkan ( dua cluster, tiga cluster, atau lain sebagainya). Setelah jumlah cluster itu diketahui, baru proses cluster dilakukan tanpa mengikuti proses hierarki. Metode ini disebut dengan *K-Means Clustering* (Santoso, 2010).

#### 2.1.3 K-Means

Algoritma K-Means merupakan algoritma pengelompokan iteratif yang melakukan partisi set kedalam sejumlah K cluster yang sudah ditetapkan diawal. Algoritma K-Means sederhana untuk diimplmentasikan dan dijalankan, relatif cepat, mudah beradaptasi, umum penggunaannya dalam praktek secara historis, K-Means menjadi salah satu algoritma yang paling penting dalam (Wu dan Kumar 2009) dalam buku Eko Prasetyo.

### 2.2. Penelitian Terkait

Berikut ini adalah tinjauan studi terdahulu yang terkait dengan tema K-Means Clustering digunakan untuk membantu menyelesaikan tugas ini meliputi :

- a. Totok Suprawoto 2016 melakukan penelitian mengenai mengidentifikasi suatu kelompok dari objek yang memiliki karakteristik yang sama. Clustering data mahasiswa menggunakan metode K-Means, terdiri dari nilai rerata ujian nasional (UN) dan indeks prestasi kumulatif (IPK) mahasiswa. Penelitian ini

menggunakan data mahasiswa angkatan 2014/2015. Kemudian diperoleh kesimpulan bahwa kelompok mahasiswa dengan nilai rerata UN yang rendah pada jenjang diploma-3 (D-3) dan strata-1 (S-1). Jika mahasiswa memiliki nilai UN yang tinggi maka prestasi akademik mahasiswa juga tinggi pada semua jenjang. Dari hasil pengelompokan berdasarkan daerah asal sekolah IPK rata-rata yang tertinggi berasal dari provinsi Daerah Istimewa Yogyakarta (DIY) dan Jawa Tengah (Jateng).

- b. Ai Ilah Warnilah 2016 melakukan penelitian mengenai tingginya tingkat keberhasilan siswa dan rendahnya tingkat kegagalan siswa sebagai kualitas dunia pendidikan. Data mining mampu menganalisa jumlah data yang besar menjadi informasi yang mempunyai arti bagi pendukung keputusan. Salah satu proses data mining adalah *clustering*. Atribut yang digunakan dalam pengelompokan prestasi siswa adalah nama, ekstrakurikuler, nilai pengetahuan yang meliputi nilai keterampilan, nilai sikap, dan jumlah ketidakhadiran siswa. Studi kasus pada 173 siswa dengan perhitungan jarak menggunakan *manhattan distance*, *chbychep distance* *euclidian distance* menghasilkan akurasi sebesar 67%.
- c. Hongyuan Zha & Xiaofeng He melakukan penelitian mengenai metode turunan koordinat kemudian digunakan untuk mencari minimal lokal. dalam tulisannya menunjukkan bahwa minimalisasi dapat dirumuskan ulang sebagai masalah maksimasi jejak terkait dengan matriks Gram dari vektor data. Selanjutnya menunjukkan bahwa versi yang santai dari masalah maksimasi jejak memiliki solusi optimal global yang dapat diperoleh dengan komputasi eigendecomposisi parsial dari matriks Gram, dan tugas cluster untuk setiap vektor data dapat ditemukan dengan komputasi dekomposisi QR pivot dari matriks eigenvektor. Sebagai oleh produk juga menurunkan batas bawah untuk minimum fungsi biaya *sum-of-squares*.
- d. Kiri Wagstaff 2001 melakukan penelitian mendemonstrasikan bagaimana k-means yang populer algoritme pengelompokan dapat dilakukan secara terprogram ed untuk memanfaatkan untuk informasi ini, dalam percobaan dengan batasan arti pada enam set data, yang diamati peningkatan dalam pengelompokan ketepatan. penulis menerapkan metode ini ke masalah dunia nyata secara otomatis mendeteksi jalur jalan dari data GPS dan mengamati peningkatan dramatis dalam kinerja.
- e. Ming hua 2009 melakukan penelitian memantau k-means pengelompokan pembacaan sensor dalam jaringan sensor besar, diberikan satu set sensor yang pembacaannya berkembang dari waktu ke waktu, penulis ingin mempertahankan k-means dari pembacaan terus menerus, sasaran pengoptimalan adalah untuk mengurangi biaya pelaporan dalam jaringan, sensor sedikit-sedikit mungkin melaporkan pembacaan mereka saat ini ke pusat data dalam perjalanan pemeliharaan, untuk mengatasi masalah, penulis mengusulkan pohon pelaporan pembacaan, kumpulan data hirarkis, dan kerangka analisis, bahkan penulis mengembangkan beberapa pelaporan metode hemat biaya menggunakan pohon pelaporan pembacaan dalam pemantauan k-means secara berkelanjutan. Pertama, metode pengambilan sampel yang seragam menggunakan pohon pelaporan pembacaan dapat mencapai pendekatan kualitas yang baik dari k-means. Kedua, penulis mengusulkan

metode ambang pelaporan yang dapat menjamin kualitas perkiraan, kami mengeksplorasi pendekatan malas yang dapat mengurangi perhitungan antara secara substansial, kami melakukan evaluasi simulasi sistematis menggunakan set data sintesis untuk memeriksa karakteristik metode yang diusulkan.

### 2.3. Objek Penelitian

Sejak era reformasi bergulir di tengah percaturan politik Indonesia, sejak itu pula perubahan kehidupan mendasar berkembang di hampir seluruh kehidupan berbangsa dan bernegara. Presiden ke 4 Abdurrahman Wahid dengan keputusan Presiden No.355/M Tahun 1999 tanggal 26 Oktober 1999 dalam kabinet periode 1999-2004 mengangkat Ir. Sarwono Kusumaatmaja sebagai Menteri Eksplorasi Laut. Selanjutnya pengangkatan tersebut diikuti dengan pembentukan Departemen Eksplorasi Laut (DEL) beserta rincian tugas dan fungsinya melalui keputusan Presiden Nomor 136 Tahun 1999 tanggal 10 November 1999 tentang kedudukan, Tugas, Fungsi, Susunan Organisasi, dan Tata Kerja Departemen. Ternyata penggunaan nomenklatur DEL tidak berlangsung lama karena berdasarkan usulan DPR dan berbagai pihak, telah dilakukan perubahan penyebutan dari Menteri Eksplorasi Laut menjadi Menteri Eksplorasi Laut dan Perikanan berdasarkan Keputusan Presiden Nomor 145 Tahun 1999 tanggal 1 Desember 1999. Kemudian berubah menjadi kementerian Kelautan dan Perikanan sesuai dengan Peraturan Presiden No. 47 tahun 2009 tentang pembentukan dan organisasi Kementerian Negara, maka Nomenklatur Departemen Kelautan dan Perikanan menjadi Kementerian Kelautan dan Perikanan, sedangkan struktur organisasi pada kementerian kelautan dan perikanan tidak mengalami perubahan.

## 3. PEMBAHASAN

### 3.1 Metode Pemilihan Data

Data yang diolah dalam penelitian ini merupakan data sampel yang diambil dari data global dataset tindak pidana ilegal fishing dari tahun 2003-2013. Dataset tindak pidana ilegal fishing terdiri dari atribut id, tindak pidana, tahun, jumlah. Data yang akan diuji terdiri dari 266 data tindak pidana *illegal fishing*.

Tabel 1. Data awal tersebut tampak pada

<b>Id</b>	<b>Tindak Pidana</b>	<b>Tahun</b>	<b>Jumlah</b>
1	Tanpa Ijin	2003	91
26	Tanpa Ijin	2004	53
51	Tanpa Ijin	2005	26
76	Tanpa Ijin	2006	29
101	Tanpa Ijin	2007	65
126	Tanpa Ijin	2008	35
151	Tanpa ijin	2009	59
176	Tanpa ijin	2010	45
201	Tanpa ijin	2011	17
226	Tanpa ijin	2012	15
247	Tanpa ijin	2013	18
.	.	.	.
.	.	.	.
.	.	.	.

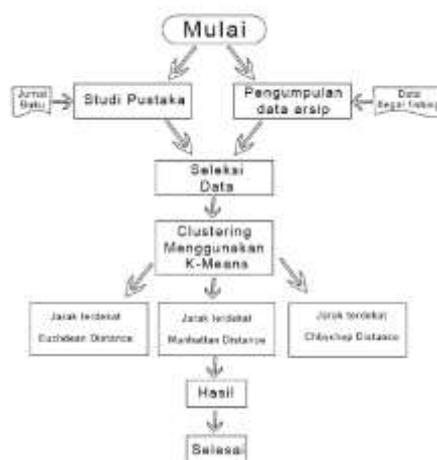
Id	Tindak Pidana	Tahun	Jumlah
265	Bongkar muat tidak sesuai SIPI	2013	0
244	Bongkar muat tidak sesuai SIPI	2012	0
225	Bongkar muat tidak sesuai SIPI	2011	1
239	Menggunakan Bahan Kimia/biologis/peledak	2012	0
260	Menggunakan Bahan Kimia/biologis/peledak	2013	2
246	Penangkapan ikan di daerah Grey Area/alat tangkap terlarang/dikembalikan ke negara asal terkait MoU	2012	4
267	Penangkapan ikan di daerah Grey Area/alat tangkap terlarang/dikembalikan ke negara asal terkait MoU	2013	2

### 3.2 Metode Analisis Dan Pengujian

Pengelompokan tindak pidana ilegal fishing dari data yang diperoleh dalam penelitian ini menggunakan algoritma *k-means clustering*. Data *clustering* merupakan salah satu metode data mining yang bersifat tanpa arahan (*unsupervised*). Ada dua jenis data clustering yang sering dipergunakan dalam proses pengelompokan data yaitu hierarchial (hirarki) data *clustering* dan *non-hierarchical* (non hirarki) data *clustering*. *K-means* merupakan salah satu metode data *clustering* non hirarki yang berusaha mempartisi data yang ada ke dalam bentuk satu atau lebih *cluster*/kelompok. Metode ini mempartisi data ke dalam *cluster*/kelompok sehingga data yang memiliki karakteristik yang sama dikelompokkan ke dalam satu cluster yang sama dan data yang mempunyai karakteristik yang berbeda dikelompokkan ke dalam kelompok yang lain.

Tujuan dari data *clustering* ini adalah untuk meminimalisirkan *objective function* yang diset dalam proses clustering. Yang pada umumnya berusaha meminimalisirkan variasi di dalam suatu *cluster* dan memaksimalkan variasi antar cluster.

### 3.3 Langkah-Langkah Penelitian



Gambar 1. Langkah-langkah penelitian

Langkah langkah penelitian ini sebagai berikut :

- Studi pustaka dan studi penelitian sebelumnya, studi terhadap penelitian-penelitian yang pernah dilakukan sebelumnya, terutama yang menggunakan metode *k-mens*.

- b. Pengumpulan data, mengumpulkan data peserta didik dalam beberapa kelas dengan variasi nilai dan kreatifitas.
- c. Eksperimen, Pada tahap ini data training dilatihkan pada algoritma data mining k-means secara manual dan bertahap sampai iterasi terakhir yang memungkinkan data dan tidak berpindah lagi.
- d. Analisa data hasil dari proses eksperimen dilakukan analisa, dengan menganalisa hasil awal *cluster* pusat yang ditentukan secara random dan pengelompokan data hasil proses algoritma dengan menghubungkan atribut-atribut yang sebelumnya telah disederhanakan.
- e. *Reporting*, menuliskan hasil penelitian dalam bentuk jurnal dari pendahuluan sampai dengan kesimpulan.

### 3.3.1 Algoritma K-Means

Langkah pertama algoritma *K-Means* adalah menentukan jumlah cluster, pada penelitian ini ditetapkan 4 Cluster, yang dipilih secara random . dengan variable tahun dan jumlah. Data yang dihasilkan tampak pada tabel dibawah ini

Tabel 2. Cluster Awal

	Tahun	Jumlah
C1	2004	53
C2	2007	65
C3	2010	45
C4	2013	18

$$d(x,y) = \sqrt{\sum_{i=1}^n (x_i - y_i)^2}$$

Setelah didapatkan data *cluster* kemudian langkah kedua menghitung jarak setiap data ke pusat *cluster* antara objek ke centroid. Untuk menghitung jarak pada penelitian ini digunakan  $L_2$  (Euclidean) *distance space*, jarak antara dua titik dihitung menggunakan rumus sebagai berikut :

$$= \sqrt{(x_1 - x_2)^2 + (y_1 - y_2)^2}$$

Dimana p = dimensi data, berdasarkan data p berjumlah 2. Maka dijabarkan rumus menjadi sebagai berikut

Sebagai contoh perhitungan jarak beberapa data adalah sebagai berikut :

Tabel 3. Jarak pertama dengan pusat Cluster pada iterasi 1

Data ke 1	Tahun	Jumlah	Cluster 1	Cluster 2	Cluster 3	Cluster 4	C1	C2	C3	C4
1	2003	91	38,01316	26,30589	46,52956	73,68175		*		
2	2004	53	0	12,36932	10	36,13862	*			
3	2005	26	27,01851	39,05125	19,64688	11,31371				*
4	2006	29	24,08319	36,01389	16,49242	13,0384				*
5	2007	65	12,36932	0	20,22375	47,38143		*		
6	2008	35	18,43909	30,01666	10,19804	17,72005			*	
7	2009	59	7,81025	6,324555	14,03567	41,19466		*		
8	2010	45	10	20,22375	0	27,16616			*	
9	2011	17	36,67424	48,16638	28,01785	2,236068				*
10	2012	15	38,83298	50,24938	30,06659	3,162278				*
11	2013	18	36,13862	47,38143	27,16616	0				*

Data ke 1	Tahun	Jumlah	Cluster 1	Cluster 2	Cluster 3	Cluster 4	C1	C2	C3	C4
12	2013	0	53,75872	65,27634	45,09989	18				*
13	2012	15	38,83298	50,24938	30,06659	3,162278				*
14	2011	0	53,46027	65,12296	45,01111	18,11077				*
15	2010	6	47,38143	59,07622	39	12,36932				*
16	2009	4	49,25444	61,03278	41,01219	14,56022				*
17	2008	4	49,16299	61,0082	41,04875	14,86607				*
18	2007	5	48,09366	60	40,11234	14,31782				*
19	2006	19	34,05877	46,01087	26,30589	7,071068				*
20	2005	36	17,02939	29,06888	10,29563	19,69772			*	
.	.	.	.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.	.	.	.
248	2007	0	53,08484	65	45,09989	18,97367				*
249	2006	0	53,03772	65,00769	45,17743	19,31321				*
250	2005	0	53,00943	65,03076	45,27693	19,69772				*
251	2004	0	53	65,06919	45,39824	20,12461				*
252	2003	0	53,00943	65,12296	45,54119	20,59126				*
253	2003	0	53,00943	65,12296	45,54119	20,59126				*
254	2004	0	53	65,06919	45,39824	20,12461				*
255	2005	0	53,00943	65,03076	45,27693	19,69772				*
256	2006	0	53,03772	65,00769	45,17743	19,31321				*
257	2007	0	53,08484	65	45,09989	18,97367				*
258	2008	0	53,15073	65,00769	45,04442	18,68154				*
259	2009	0	53,23533	65,03076	45,01111	18,43909				*
260	2010	0	53,33854	65,06919	45	18,24829				*
261	2013	0	53,75872	65,27634	45,09989	18				*
262	2012	0	53,60037	65,19202	45,04442	18,02776				*
263	2011	1	52,46904	64,12488	44,01136	17,11724				*
264	2012	0	53,60037	65,19202	45,04442	18,02776				*
265	2013	2	51,78803	63,28507	43,10452	16				*
266	2012	4	49,64877	61,20457	41,04875	14,03567				*
267	2013	2	51,78803	63,28507	43,10452	16				*

Setelah semua data ditempatkan ke dalam *cluster* yang terdekat, kemudian hitung kembali pusat *cluster* yang baru berdasarkan rata-rata anggota yang ada pada *cluster* tersebut. Contoh perhitungannya adalah sebagai berikut : Rata-rata tahun yang ada pada cluster 1 :

$$\begin{aligned}
 &= \frac{\sum C1}{n} \\
 &= \frac{2004+2012+\dots+2003}{3} \\
 &= 2006,333
 \end{aligned}$$

Rata-rata tahun yang ada pada cluster 2 :

$$\begin{aligned}
 &= \frac{\sum C2}{n} \\
 &= \frac{2003+2007+\dots+2009+2004}{5} \\
 &= 2006,6
 \end{aligned}$$

Rata-rata tahun yang ada pada cluster 3 :

$$= \frac{\sum C3}{n}$$

$$= \frac{2008+2010+\dots+2005+2005}{11}$$

$$= 2006,545$$

Rata-rata tahun yang ada pada cluster 4 :

$$= \frac{\sum C4}{n}$$

$$= \frac{2005+2006+\dots+2011+2012}{248}$$

$$= 2007,968$$

Rata-rata jumlah yang ada pada cluster 1

$$= \frac{\sum C1}{n}$$

$$= \frac{53+55+\dots+54}{3}$$

$$= 54$$

Rata-rata jumlah yang ada pada cluster 2

$$= \frac{\sum C2}{n}$$

$$= \frac{91+65+59+\dots+70+116}{5}$$

$$= 80,2$$

Rata-rata jumlah yang ada pada cluster 3

$$= \frac{\sum C3}{n}$$

$$= \frac{35+45+36+\dots+44+37+33}{11}$$

$$= 37,63636$$

Rata-rata jumlah yang ada pada cluster 4

$$= \frac{\sum C4}{n}$$

$$= \frac{26+29+17+\dots+15+18+15}{83}$$

$$= 7,759036$$

Jadi hasil centroid baru *cluster 1, cluster 2, Cluster 3, Cluster 4* adalah centroid baru Setelah didapatkan titik pusat baru dari tiap cluster, hitung kembali data dengan pusat cluster yang baru, ulangi sampai didapatkan pola terakhir yang sudah tidak berpindah, dalam penelitian ini, data dihitung sampai iterasi ke 3, dimana setiap *cluster* tidak berubah lagi dan tidak ada lagi data yang berpindah dari satu *cluster* ke *cluster* yang lain. Hasil dan pola terakhir dapat dilihat tabel dibawah.

Tabel 4. Centroid Baru

	Cluster	
Cluster 1	2006,5	59,33333
Cluster 2	2006,5	103,5
Cluster 3	2006,824	33,94118
Cluster 4	2007,983	6,246753

Tabel 5. Hasil dan Pola terahir jarak antara centroid dan pusat cluster

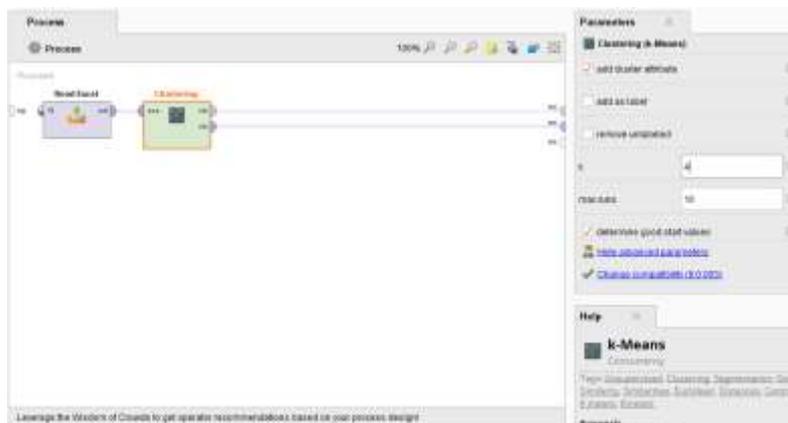
Data ke 1	Tahun	Jumlah	Cluster 1	Cluster 2	Cluster 3	Cluster 4	C1	C2	C3	C4
1	2003	91	31,8595	12,98075	57,18679	84,89963		*		
2	2004	53	6,808899	50,56184	19,26684	46,92264	*			
3	2005	26	33,36707	77,51451	8,147855	19,97728			*	
4	2006	29	30,33745	74,50168	5,009334	22,83954			*	
5	2007	65	5,688683	38,50325	31,05932	58,76148	*			
6	2008	35	24,37952	68,51642	1,582779	28,75325			*	
7	2009	59	2,522124	44,57017	25,15316	52,76304	*			
8	2010	45	14,75447	58,60461	11,50598	38,80568			*	
9	2011	17	42,57183	86,61697	17,44839	11,16834				*
10	2012	15	44,6732	88,67074	19,63578	9,630775				*
.	.	.	.	.	.	.				
.	.	.	.	.	.	.				
258	2008	0	59,35229	103,5109	33,96156	6,246775				*
259	2009	0	59,38598	103,5302	34,01089	6,328922				*
260	2010	0	59,43647	103,5592	34,08949	6,564169				*
261	2013	0	59,68831	103,7039	34,49858	8,011709				*
262	2012	0	59,5877	103,646	34,33365	7,426603				*
263	2011	1	58,50665	102,5987	33,20488	6,052096				*
264	2012	0	59,5877	103,646	34,33365	7,426603				*
265	2013	2	57,70062	101,7079	32,53287	6,572707				*
266	2012	4	55,606	99,65189	30,38536	4,602217				*
267	2013	2	57,70062	101,7079	32,53287	6,572707				*

Data yang dikelompokkan pada cluster 1 berjumlah 6 tindak pidana, di cluster 2 berjumlah 2 tindak pidana, di cluster 3 berjumlah 17 tindak pidana dan di cluster 4 berjumlah 242 tindak pidana. Dengan criteria data bahwa pada cluster 4 memiliki criteria tindak pidana *ilegal fishing* yang paling tinggi, cluster 3 menduduki peringkat ke 2, cluster ke 1 menduduki peringkat 3 dan cluster 2 memiliki criteria peringkat terakhir.

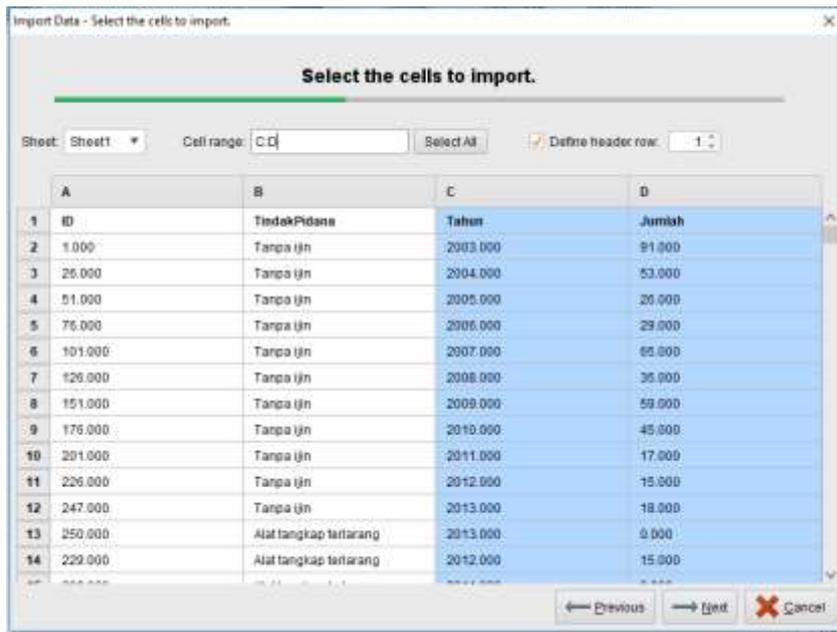
### 3.4 Evaluasi Dan Validasi Hasil

Pengujian jarak kedekatan dengan *euclidean distance*, *mahattan distance*, *chebifhap Distance*

a. Cluster model dengan kedekatan jarak *Euclidean Distance*

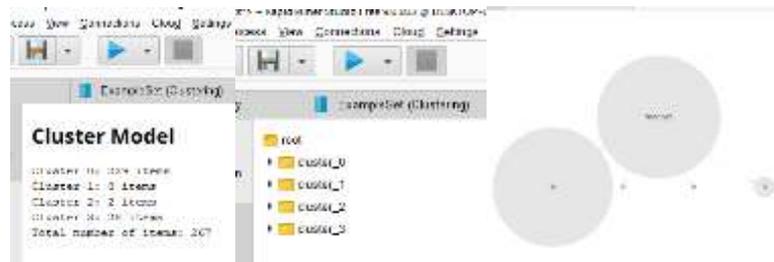


Gambar 1. Penentuan Jarak terdekat dengan *Euclidean distance*



Gambar 2. Seleksi atribut

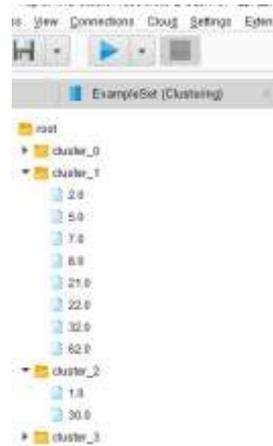
Cluster model yang diperoleh dari hasil pengujian terhadap data menggunakan k-means clustering terlihat pada gambar 3.



Gambar 3. Cluster Model

Dapat dilihat cluster model yang dihasilkan dari 0 229 items, cluster 1 8 items, cluster 2 2 items dan cluster 3 28 items. Dalam bentuk folder root tampak bahwa root memiliki 4 cluster yaitu folder cluster 0, cluster 1, cluster 2, cluster 3. Cluster model juga dapat digambarkan dalam tree. Induk adalah root yang memiliki child 0,1,2,3

Untuk memudahkan melihat tindak pidanan yang dimiliki oleh setiap folder cluster, gambar 4 menampilkan membership dari masing-masing cluster.



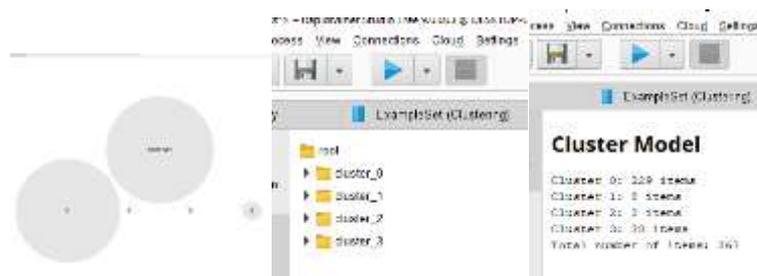
Gambar 4. Tampilan Member masing-masing cluster

Pada gambar 4 ditampilkan membership masing-masing cluster dimana setiap anggota diwakili oleh nomor barisnya. Cluster 0 memiliki anggota nomor baris 12.0, cluster 1 memiliki anggota nomor baris 2.0, cluster 2 memiliki anggota nomor baris 1.0 dan cluster 3 memiliki anggota nomor baris 3.0

Cluster	Value	Count	Total
cluster_0	12.0	12.0	12.0
cluster_1	2.0	2.0	2.0
cluster_2	1.0	1.0	1.0
cluster_3	3.0	3.0	3.0
			18.0

Gambar 5. Tampilan Member masing-masing cluster Hasil analisa *k-means clustering* clustering pada nilai

- b. Pengujian dengan jarak kedekatan *Mahattan Distance Cluster* model yang diperoleh dari hasil pengujian terhadap data menggunakan *k-means clustering* terlihat digambar 6.



Gambar 6. Cluster Model

#### 4. SIMPULAN

Penelitian ini dibuat untuk mengkaji penerapan model data mining dalam mengelompokkan *illegal fishing*, pada penelitian ini diimplementasikan algoritma data mining *k-means*. dari hasil analisis diketahui bahwa :

- Dengan analisis algoritma *k-means* dengan penentuan jarak terdekat menggunakan *eulclidean distance*
- Lebih optimal dibandingkan dengan menggunakan *mahattan distance* dan *chbchep distance* dalam mengelompokkan prestasi siswa

- c. Penentuan centroid (titik pusat) pada tahap awal algoritma k-means sangat berpengaruh pada hasil cluster seperti pada hasil pengujian yang dilakukan dengan menggunakan 267 record dengan centroid yang berbeda menghasilkan hasil cluster yang berbeda juga.
- d. Diperoleh suatu model *clustering* yang dapat digunakan *illegal fishing* dalam pengambilan keputusan tindak pidana *illegal fishing* tinggi, menengah, cukup.

#### DAFTAR PUSTAKA

- [1] Hongyuan Zha, X. H. (n.d.). Spectral Relaxation For K-Means Clustering.
- [2] Kiri Wagstaff, C. C. (2001). Constrained K-Means Clustering With Background Knowledge. *Proceedings Of The Eighteenth International Conference on Machine Learning*, 577-584.
- [3] Kusrini, & Luthfi, T. E. (2009). Algoritma Data Mining. Yogyakarta: Penerbit Andi
- [4] Ming Hua, M. K. (2009). Continuous K-Means Monitoring With Low Reporting Cost in Sensor Networks . *IEEE Transactions on Knowledge and Data Engineering*, 1679-1691.
- [5] Prasetyo, Eko. (2014). Data Mining Mengolah Data Menjadi Informasi Menggunakan Matlab. Yogyakarta: Andi.
- [6] Santoso, S. (2010). Statistik Multivariat. Jakarta: Elex Media Komputindo.
- [7] Suprawoto, T. (2016). KLASIFIKASI DATA MAHASISWA MENGGUNAKAN METODE K-MEANS UNTUK MENUNJANG PEMILIHAN STRATEGI PEMASARAN. *Jurnal Informatika dan Komputer (JIKO)*, 12-18.
- [8] Warnilah, A. I. (2016). Analisa Algoritma K-Means Clustering Untuk Pemetaan Prestasi Siswa Studi Kasus SMP Negeri 1 Sukahening. *Indonesian Journal on Computer and Information Technology*, 83-95.