

Analisis Daftar Pemilih Tetap pada Hasil Rekapitulasi KPU Berdasarkan Usia Menggunakan Algoritma *K-Means* (Studi Kasus : Kota Bandar Lampung)

Gilang Fajriansyah¹, Gigih Forda Nama², Yessi Mulyani³

Program Studi Teknik Informatika, Jurusan Teknik Elektro, Universitas Lampung, Bandar Lampung

Jl. Prof. Sumantri Brojonegoro No.1 Bandar Lampung 35145

¹gilangfajriansyah@gmail.com

²gigih@eng.unila.ac.id

³yessi.mulyani@eng.unila.ac.id

Intisari — Pada pemilihan umum di Kota Bandar Lampung, masih banyak yang masuk ke dalam golongan putih (golput). Minimnya pengetahuan masyarakat mengenai tata cara pemilu dan pentingnya menggunakan hak pilih mereka terutama pemilih pemula, dewasa dan usia lanjut menjadi faktor rendahnya penggunaan hak pilih. Kurang efektifnya pendekatan sosialisasi dari panitia penyelenggara kepada masyarakat, tetapi jika harus melakukan sosialisasi secara acak dan menyeluruh ke semua daerah akan menghabiskan dana yang besar. Dari latar belakang tersebut bertujuan untuk menerapkan Teknik data mining dengan metode *clustering* dengan menggunakan algoritma *K-Means* dan memanfaatkan tools data mining RapidMiner 9.2 terhadap data yang ada untuk memperoleh informasi mengenai daerah mana yang banyak terdapat pemilih muda, dewasa dan lansia. Penelitian ini mengelompokkan data DPT dari Kecamatan Langkapura, Rajabasa dan Kemiling. Data di kelompokkan berdasarkan usia dan daerah. Algoritma yang digunakan untuk meng-*cluster* adalah *K-Means*, dengan menggunakan metode CRISP-DM (*Cross Industry Standart for Data Mining*). Hasil *cluster* dianalisa berdasarkan Kelurahan dan Kecamatan. Hasil analisa *cluster* Kecamatan Langkapura usia muda berjumlah 10167, usia dewasa berjumlah 9527, lansia berjumlah 4821 orang. Kecamatan Rajabasa usia muda berjumlah 12557, dewasa berjumlah 10930 dan lansia berjumlah 5097. Kecamatan Kemiling pada usia muda, dewasa dan lansia berjumlah 19442, 19086 dan 9394.

Kata kunci — pemilu, *K-Means*, *clustering*, data, usia, daerah

Abstract — General election in Bandar Lampung City, there were still many who entered the white group (abstentions). The lack of public knowledge about the procedures for elections and the importance of using their voting rights, especially first-time voters, adults and the elderly are a factor in the low use of voting rights. The organizing committee's lack of effectiveness approaches to the community, but if it has to carry out random and comprehensive socialization to all regions it will cost a lot of money. From this background the aim is to apply data mining techniques with clustering methods using the K-Means algorithm and to use RapidMiner 9.2 data mining tools on existing data to obtain information about which areas there are many young, adult and elderly voters. This study grouped the DPT data from Langkapura, Rajabasa and Kemiling Districts. Data is grouped by age and region. The algorithm used to cluster is K-Means, using the CRISP-DM (Cross Industry Standards for Data Mining) method. Cluster results are analyzed based on Kelurahan and Kecamatan. The results of the cluster analysis of Langkapura Subdistrict were 10,167 young people, 9527 adults, 48,21 elderly people. Rajabasa sub-district young people number 12557, adults number 10930 and elderly number 5097. Kemiling sub-district at young age, adults and elderly number 19442, 19086 and 9394.

Keywords - election, *K-Means*, *clustering*, data, age, region.

I. PENDAHULUAN

Di dalam negara yang menganut sistem demokrasi, pemilihan umum (pemilu) merupakan satu-satunya mekanisme pergantian kekuasaan yang sah. Pemilu menjamin berlangsungnya rotasi kekuasaan penyelenggara negara. Rotasi kekuasaan inilah yang merupakan hasil dari proses penyelenggaraan pemilu. Untuk Kota Bandar Lampung sendiri, yang dalam penelitian ini akan menjadi objek penelitian, tercatat pada tahun 2014 memiliki tingkat partisipasi memilih 75,73% dengan angka golput 24,27%. Lalu pada Pilkada tahun 2018 sebanyak 1.707.747 masyarakat Provinsi Lampung masuk ke dalam golongan putih (golput).[1]

Kurang efektifnya pendekatan sosialisasi dari panitia penyelenggara kepada masyarakat, tetapi jika harus melakukan sosialisasi secara acak dan menyeluruh ke semua daerah akan menghabiskan dana yang besar. Perkembangan teknologi saat ini menciptakan kondisi akan kaya data namun minim informasi.[2]

KPU Bandar Lampung memiliki data yang sangat besar, salah satunya yaitu data Daftar Pemilih Tetap. Sehingga diperlukan *data scientist* dalam menganalisa data dalam jumlah besar dalam sistem yang ada di KPU Bandar Lampung sehingga dapat ditemukan pola dan algoritma guna mendapatkan informasi baru dengan memanfaatkan teknik *data mining* dan menggunakan *tools* dari sistem *data mining* untuk membantu mengolah dan menganalisis data dalam jumlah yang besar. Salah satu algoritma dalam *data mining* yang dapat membagi dan meng-*cluster* data yang besar menjadi *record* yang lebih kecil yaitu Algoritma *K-Means*. [3]

II. TINJAUAN PUSTAKA

A. Pemilu

Pemilihan umum (pemilu) merupakan instrumen penting dalam negara demokrasi yang menganut sistem perwakilan. Pemilu berfungsi sebagai alat penyaring bagi “politikus” yang akan mewakili dan membawa suara rakyat didalam lembaga

perwakilan. Pemilu adalah sarana pelaksanaan kedaulatan rakyat yang dilaksanakan secara langsung, umum, bebas, rahasia, jujur, dan adil dalam Negara Kesatuan Republik Indonesia berdasarkan Pancasila dan Undang-Undang Dasar Negara Republik Indonesia Tahun 1945 (UU No. 10 Tahun 2008).[4]

B. Data Mining

Data mining adalah bidang dari beberapa bidang ilmu yang menyatukan teknik dari machine learning, pengenalan pola, statistik, basis data, dan visualisasi untuk penanganan permasalahan pengambilan informasi dari data dalam jumlah besar. Data mining merupakan proses mengekstraksi dan mengidentifikasi suatu informasi, menemukan hubungan, menghasilkan pattern atau pola dengan data dari basis data yang besar.[5] Berikut 5 teknik data mining, yaitu :

1. Estimasi

Model dibangun dari data dengan *record* yang lengkap, yang menyediakan nilai dari variabel sebagai prediktor, kemudian estimasi nilai dari *variable* target ditentukan berdasarkan nilai dari variabel prediktor.

2. Prediksi

Algoritma prediksi sama dengan algoritma estimasi di mana label/ target/ *class* bertipe numerik, bedanya adalah data yang digunakan merupakan data rentetan waktu (*data time series*). Sifat prediksi bisa menghasilkan *class* berdasarkan berbagai atribut yang kita sediakan.

3. Klasifikasi

Algoritma yang menggunakan data dengan target/class/label berupa nilai kategorikal (nominal). Pengelompokan data-data yang ada menjadi dalam kelompok yang sudah ditentukan nama kelompoknya.

4. Klastering

Klastering merupakan mengelompokkan data, hasil observasi dan kasus ke dalam *class* yang mirip, koleksi data yang mirip antara satu dengan yang lain, dan memiliki perbedaan bila dibandingkan dengan data dari klaster lain.

5. Asosiasi

Algoritma *association rule* (aturan asosiasi) adalah algoritma yang menemukan atribut yang “muncul bersamaan”. Pengelompokan data-data yang saling terkait dan berhubungan. Algoritma asosiasi akan mencari aturan yang menghitung hubungan diantara dua atau lebih atribut.[6]

Tahap - tahap dalam data mining adalah sebagai berikut :

- Pembersihan Data (*Data cleaning*)
Proses menghilangkan *noise* dan data yang tidak relevan atau konsisten.
- Integrasi Data (*Data Integration*)
Proses Penggabungan data dari berbagai basis data menjadi basis data baru.
- Seleksi Data (*Data Selection*)
Proses memilih data yang sesuai pada basis data untuk dianalisis.
- Transformasi Data (*Data Transformation*)
Proses mengubah data atau menggabung data ke dalam format yang sesuai untuk diproses menggunakan teknik *data mining*.
- Reduksi Data (*Data Reduction*)
Proses menguraikan data dalam bentuk yang lebih kecil namun tetap menghasilkan analisis yang sama.
- Proses *Mining*
Proses utama saat menerapkan metode untuk mendapatkan suatu informasi dari data.
- Evaluasi pola (*Pattern Evaluation*)
Proses mengidentifikasi pola ke dalam *knowledge based* yang didapatkan.
- Presentasi Pengetahuan (*Knowledge Presentation*)
Proses penyajian pengetahuan mengenai metode yang digunakan untuk memperoleh suatu pengetahuan yang didapatkan pengguna.[7]

C. K-Means

K-Means *clustering* adalah metode untuk mengelompokkan item ke dalam k kelompok (dimana k adalah jumlah kelompok yang diinginkan). Metode *K-Means* berusaha mengelompokkan data

yang ada ke dalam beberapa kelompok, dimana data dalam satu kelompok mempunyai karakteristik yang sama satu sama lainnya dan mempunyai karakteristik yang berbeda dengan data yang ada di dalam kelompok yang lain. Dengan kata lain, metode ini berusaha untuk meminimalkan variasi antar data yang ada di dalam suatu *cluster* dan memaksimalkan variasi dengan data yang ada di *cluster* lainnya. Algoritma ini bekerja pada atribut numerik. Algoritma *K-Means* pada awalnya mengambil sebagian dari banyaknya komponen dari populasi untuk dijadikan pusat *cluster* awal. Pada step ini pusat cluster dipilih secara acak dari sekumpulan populasi data. Berikutnya *K-Means* menguji masing-masing komponen didalam populasi data dan menandai komponen tersebut ke salah satu pusat *cluster* yang telah didefinisikan tergantung dari jarak minimum antar komponen dengan tiap-tiap pusat *cluster*. Posisi pusat cluster akan dihitung kembali sampai semua komponen data digolongkan kedalam tiap-tiap cluster dan terakhir akan terbentuk posisi *cluster* baru.[8]

Penarapan *K-Means clustering* dapat dilakukan dengan prosedur berikut :

- Inisialisasi, tentukan nilai K sebagai jumlah *cluster*.
- Tentukan K titik pusat *cluster* (*centroid*) awal yang dilakukan secara random.
- Menghitung jarak dari setiap objek ke masing – masing *centroid* dari masing – masing *cluster* menggunakan *Eucledian Distance*.
- Alokasikan dari masing – masing objek kedalam *centroid* paling dekat atau jarak paling minimum
- Ulangi langkah tiga jika posisi *centroid* baru tidak sama.[9]

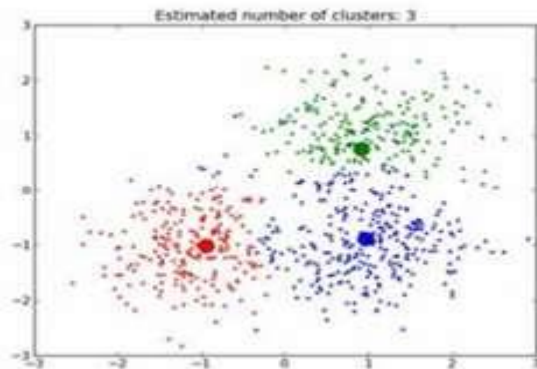
Eucledian distance berfungsi untuk mendapatkan nilai *centroid* baru hingga nilai *centroid* tetap (konvergen).

Euclidian Distance:

$$d_{ij} = \sqrt{[(x_i - x_j)^2 + (y_i - y_j)^2]}$$

x_i = koordinat x untuk fasilitas i
 y_i = koordinat y untuk fasilitas i
 d_{ij} = jarak antar fasilitas I dan j

Gbr. 1 Rumus Ecludian Distance.



Gbr. 2 Contoh Cluster K-Means.

Suatu data akan menjadi anggota dari *cluster ke-k* apabila jarak data tersebut ke pusat *cluster ke-k* bernilai paling kecil jika dibandingkan dengan jarak ke pusat *cluster* lain.[10]

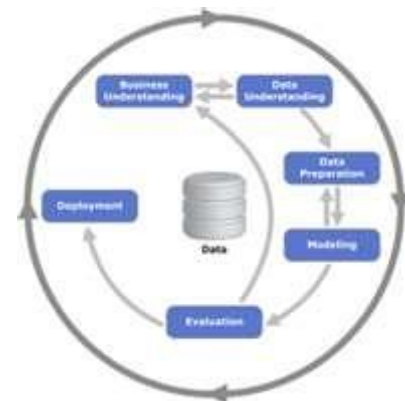
D. RapidMiner

RapidMiner adalah salah satu software untuk pengolahan data mining. Pekerjaan yang dilakukan oleh RapidMiner text mining adalah berkisar dengan analisis teks, mengekstrak pola-pola dari data set yang besar dan mengkombinasikannya dengan metode statistika, kecerdasan buatan, dan database. Tujuan dari analisis teks ini adalah untuk mendapatkan informasi bermutu tertinggi dari teks yang diolah.[11]

E. CRISP-DM

Cross Industry Standard Process for Data mining (CRISP-DM) merupakan standar proses pada data mining yang telah dikembangkan pada tahun 1996 yang ditujukan untuk melakukan proses analisis strategi pemecahan masalah secara umum dari bisnis atau untuk penelitian. Pada tahun 1997, CRISP-DM menjadi proyek Uni Eropa di bawah inisiatif pendanaan ESPRIT yang

dipimpin oleh lima perusahaan: *Integral Solutions Ltd* (ISL), *Teradata*, *Daimler AG*, *NCR Corporation* dan *OHRA*, sebuah perusahaan asuransi. Pada metode CRISP-DM terdapat enam fase yaitu *Business Understanding Phase* (Fase Pemahaman Bisnis), *Data Understanding Phase* (Fase Pemahaman Data), *Data Preparation* (Persiapan Data), *Modeling Phase* (Fase Pemodelan), *Evaluation Phase* (Fase Evaluasi), dan *Deployment Phase* (Fase Penyebaran).[12]



Gbr. 3 Tahapan Metode CRIPD-DM.

F. Penelitian Terkait

Berikut ini beberapa penelitian terdahulu yang terkait mengenai teknik data mining, Algoritma *K-Means* dan metode CRISP-DM yang akan digunakan dalam penelitian antara lain :

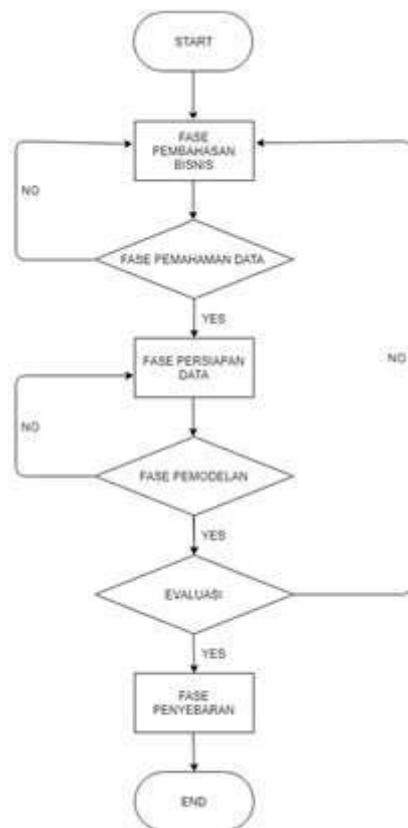
- "Implementasi Metode *K-Means* dan Naïve Bayes Classifier untuk Analisis Sentimen Pemilihan Presiden (Pilpres) 2019" oleh Imam Kurniawan dan Ajib Susanto pada tahun 2019. Penelitian tersebut bertujuan untuk memperoleh analisis dokumen text untuk mendapatkan sentimen positif atau negatif. Metode yang digunakan *K-Means* untuk melakukan klastering pada data latih dan Naive Bayes classifier untuk mengklasifikasi pada data testing. Hasil dari pembobotan ini berupa sentimen positif dan negatif. Data diambil dari *Twitter* mengenai pemilu presiden 2019 sebanyak 500 data tweet. Dari hasil pengujian 100 dan 150 data uji diperoleh akurasi rata-rata 93.35% dan *error rate* sebesar 6.66%.[13]

- “Implementasi Algoritma *K-Means* untuk Klasterisasi Kinerja Akademik Mahasiswa” yang dilakukan oleh Fajar Nur Rohmat Fauzan, Budi Darma Setiawan dan Issa Arwani pada tahun 2018. Pada penelitian ini data yang digunakan adalah data mahasiswa tahun masuk ajaran 2012 sebanyak 90 data, data pemilu tersebut akan diolah menggunakan model algoritma *K-Means* untuk mendapatkan nilai akurasi yang baik dan dapat digunakan sebagai klasterisasi kinerja akademik mahasiswa. Terdapat 3 *cluster* pada penelitian ini yaitu, cluster I, cluster II dan cluster III. Hasil penelitian tersebut menunjukkan bahwa pendapatan orang tua tidak mempengaruhi tingkat kinerja akademik mahasiswa dan mahasiswa yang masuk perguruan tinggi melalui jalur reguler dan jalur prestasi akademik mempunyai nilai IPK rata – rata tertinggi.[14]
- “Clustering Penentuan Potensi Kejahatan Daerah Di Kota Banjarbaru Dengan Metode K-Means” dilakukan pada tahun 2014 oleh Sri Rahayu, Dodon T Nugrahadhi dan Fatma Indriani. Pada penelitian ini, data didapatkan dari Polres di Kota Banjarbaru. Data yang tersedia meliputi data hukuman, bulan dan laporan. Metode yang digunakan dalam menggunakan teknik data mining yaitu metode clustering dengan Algoritma K-Means. Hasil akurasi yang diperoleh dengan menggunakan Delphi tingkat kejahatan pada tahun 2014 sebesar 63,5% dengan jumlah 1266 laporan.[15]
- “A *Dynamic Flooding Attack Detection System Based on Different Classification Techniques and Using SNMP MIB Data*” dilakukan pada tahun 2014 oleh Sahar Namvarasl dan Marzieh Ahmadzadeh . Penelitian tersebut memakai beberapa algoritma klasifikasi yaitu Neural Network, Bayesian Network dan K-Means. Untuk menganalisis sistem jaringan LAN penulis memakai 4 PC dan 1 switch menggunakan OS Linux Ubuntu 14. 04, Hping3 untuk mensimulasikan serangan dan Weka untuk aplikasi *Data Mining*. Data yang terekam sebesar 4600 *record*. Sistem berjalan dengan 3 step yaitu memilih variabel yang efektif, meng-generate model yang paling akurat dan mendeteksi serangan *real time* lalu meng-*update* dataset. Hasil akurasi yang diperoleh dengan menggunakan Neural Network yaitu 99,03%, sedangkan untuk Bayesian Network sebesar 98,83% dan *K-Means* yaitu 98,72%. [16].
- “An Enhanced *K-Means* Clustering Algorithm for Pattern Discovery in Big Data Analysis of 3-Phase Electrical Quantities” penelitian ini dilakukan pada tahun 2018 oleh Gigih Forda Nama dan Dikpride Despa. Penelitian tersebut mengembangkan pemantauan online sistem distribusi daya berbasis teknologi Internet of Things (IoT) di Jurusan Teknik Elektro Universitas Lampung (Unila). Sistem ini memantau listrik jumlah panel distribusi utama 3-fase di gedung H. Sistem pengukuran melibatkan beberapa sensor seperti sensor arus dan sensor tegangan, data pengukuran yang disimpan di server database dan menunjukkan informasi secara real-time melalui aplikasi berbasis web. Tujuan utama dari penelitian ini adalah untuk menangkap, menganalisis, dan mengidentifikasi pola pengetahuan pengukuran atau jumlah listrik, menggunakan Proses Standar Lintas-Industri untuk data mining (CRISP-DM). Berdasarkan analisis data, menunjukkan bahwa perbandingan faktor tegangan, arus, dan daya pada semua Fase tidak seimbang, sudah dilaporkan kepada para stakeholders dan merekomendasikan mereka untuk melakukan evaluasi total penggunaan perangkat listrik sehingga muatan listrik menyala. setiap fase dapat diseimbangkan, terutama transfer beban listrik pada Fase 0 yang terlalu besar dibandingkan dengan Fase 1 dan 2.[17]
- “Synthesizing CRISP-DM and Quality Management: A Data Mining Approach for Production Processes” dilakukan oleh Franziska Schäfer, Christian Zeiselmaier, Jonas Becker, Heiner Otten pada tahun 2018. Penelitian ini menyajikan QM-CRISP-DM untuk aplikasi data mining dalam konteks peningkatan proses produksi. Mengusulkan tools manajemen kualitas yang konkrit untuk setiap fase

CRISP-DM. Khusus untuk pemula, fase yang dijelaskan dari siklus CRISP-DM sering terlalu umum. Oleh karena itu, peneliti mensintesis alat manajemen kualitas dengan metodologi CRISP-DM untuk memberikan panduan pengguna pada pemula. QMCRISP-DM yang disajikan divalidasi dalam proyek data mining untuk pengembangan sistem error forecasting di bidang produksi elektronik. [18]

III. METODE PENELITIAN

Adapun metode yang digunakan dalam penelitian ini adalah CRISP-DM (*Cross Industry Standart Process for Data Mining*).



Gbr. 4 Tahapan Penggunaan Metode CRISP-DM.

A. Fase Pemahaman Bisnis

Pada tahapan pemahaman bisnis ada beberapa hal yang dilakukan antara lain, menentukan tujuan bisnis, menilai situasi, dan menentukan tujuan data mining.

1. Menentukan Tujuan Bisnis

Pada fase ini bertujuan untuk merancang model data mining yang dapat digunakan untuk meng-*cluster* daerah yang banyak

terdapat pemilih muda, pemilih dewasa dan pemilih lanjut usia. Hasil *clustering* ini diharapkan dapat dijadikan acuan atau referensi kepada KPU Kota Bandar Lampung dalam pendekatan sosialisasi pemilu yang lebih efektif.

2. Menilai Situasi

Rendahnya partisipasi pemilih muda, pemilih dewasa dan lansia dikarenakan kurangnya fasilitas pendukung yang seharusnya disediakan pemerintah untuk memudahkan mereka dalam memilih serta kurang efektifnya pendekatan sosialisasi dari panitia penyelenggara kepada masyarakat. Tetapi jika harus melakukan sosialisasi secara acak dan menyeluruh ke semua daerah akan menghabiskan dana yang besar. men-*cluster* daerah-daerah yang dimana banyak terdapat pemilih muda, pemilih dewasa dan pemilih lanjut usia. Untuk menghemat biaya sosialisasi dan juga akan lebih efisien untuk mendongkrak tingkat partisipasi pemilu.

3. Menentukan Tujuan Data Mining

Model data mining yang dibuat bertujuan untuk meng-*cluster* daerah mana saja yang banyak terdapat pemilih muda, pemilih dewasa dan pemilih lanjut usia. Pengetahuan ini dapat digunakan oleh KPU Kota Bandar Lampung untuk sosialisasi pemilu yang lebih efektif serta dapat menghemat cost dan waktu.

B. Fase Pemahaman Data

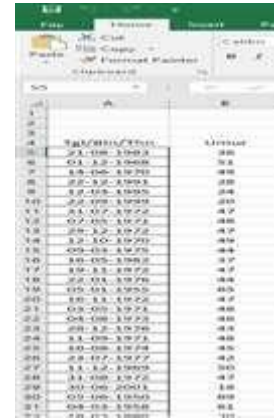
Tahap pemahaman data dimulai dengan pengumpulan data awal dan dilanjutkan dengan kegiatan untuk membiasakan diri dengan data, untuk mengidentifikasi masalah kualitas data, untuk menemukan wawasan pertama ke dalam data, atau untuk mendeteksi subset yang menarik untuk membentuk hipotesis untuk informasi tersembunyi.

C. Fase Persiapan Data

Pada tahap ini, ada beberapa hal yang dilakukan antara lain, deskripsi data set, memilih data, mengintegrasikan data dan membersihkan data dari 100.661 record data DPT.

D. Fase Pemodelan

Pemodelan adalah fase yang secara langsung melibatkan teknik data mining yaitu dengan melakukan pemilihan teknik data mining dan menentukan algoritma yang akan digunakan. Fase ini menggunakan algoritma *K-Means* untuk mengelompokkan data usia. Pada fase pemodelan juga akan dijelaskan perhitungan manual dari algoritma *K-Means*. Penelitian ini menggunakan software RapidMiner 9.2 untuk mengolah data dan mengelompokkan data tersebut.



Gbr. 5 Dataset Yang Telah Diolah.

E. Fase Evaluasi

Evaluasi adalah fase lanjutan terhadap tujuan data mining. Evaluasi dilakukan secara mendalam dengan tujuan agar hasil pada tahap pemodelan sesuai dengan sasaran yang ingin dicapai dalam tahap business understanding.

F. Fase Penyebaran

Pada tahap ini, pengetahuan atau informasi yang telah diperoleh akan dipresentasikan dalam bentuk khusus sehingga dapat diketahui atau dibaca oleh orang awam. Hasil dari penelitian ini berupa *clustering* daerah mana saja yang banyak terdapat pemilih muda, pemilih dewasa dan pemilih lanjut usia di Kota Bandar Lampung khususnya Kecamatan Langkapura, Kemiling dan Rajabasa dalam bentuk grafik dan infografik dapat digunakan oleh KPU Kota Bandar Lampung sebagai bahan pertimbangan dalam pendekatan sosialisasi serta penyediaan fasilitas agar meningkatnya partisipasi masyarakat saat pemilu.

B. Pemodelan

Pemodelan adalah fase yang secara langsung melibatkan teknik data mining yaitu dengan melakukan pemilihan teknik data mining dan menentukan algoritma yang akan digunakan. Pada penjelasan fase ini digunakan data Bilabong Jaya untuk penentuan jumlah *cluster* dan perhitungan manual dan menggunakan PHP.

1. Penentuan Jumlah *Cluster*

Berikut tabel dari masing – masing *cluster* dari $K = 4$ sampai dengan $K = 10$ dari data Kelurahan Bilabong Jaya.

K	Cluster0	Cluster1	Cluster2	Cluster3	Cluster4	Cluster5	Cluster6	Cluster7	Cluster8	Cluster9
2	36,74	47,67								
3	34,20	42,06	58,23							
4	22,66	38,52	49,52	65,15						
5	22,19	36,42	45,97	55,54	71,86					
6	21,45	33,81	42,93	51,50	62,52	77,65				
7	21,17	31,57	39,81	47,14	55,21	65,45	80,69			
8	20,36	28,06	35,43	41,97	48,51	55,75	65,45	80,69		
9	19,55	25,07	31,68	37,81	43,40	49,49	56,64	65,98	80,69	
10	20,06	26,32	30,79	32,67	44,40	49,78	55,38	62,82	70,40	81,72

Gbr. 6 Centroid $K = 2$ Sampai $K = 10$.

Pada Gbr. 6 merupakan titik centroid dari setiap *cluster* dari $K = 2$ sampai dengan $K=10$ pada data Kelurahan Bilabong Jaya. Pada penelitian ini parameter penentuan *cluster* adalah berdasarkan usia, yaitu usia muda, dewasa dan lansia. Ketika K melebihi tiga maka menjadi tidak ideal. Oleh karena itu penelitian ini menggunakan $K = 3$.

2. Modelling Data

Peneliti mengambil satu sample data yaitu Kelurahan Bilabong Jaya sebagai contoh perhitungan manual. Rumus yang dipakai untuk menemukan titik centroid adalah

IV. HASIL DAN PEMBAHASAN

A. Pemahaman dan Persiapan Data

Data DPT didapatkan dari KPU Kota Bandar Lampung, data awal dari KPU di seleksi dan di proses dan hanya di ambil data tanggal lahir dan umur.

dengan Ecludian Distance. Penetapan jumlah *cluster*, yaitu $k = 3$.

Centroid	Usia
C1	17.0
C2	47.0
C3	70.0

Gbr. 7 Penentuan Titik Centroid.

Hitung jarak setiap data ke pusat *cluster* antara objek ke centroid dengan perhitungan jarak Euclidean. Sebagai contoh, akan dihitung jarak dari data Usia pertama ke pusat *cluster* pertama dengan persamaan:

$$d(1,1) = \sqrt{(36 - 17)^2}$$

$$= 19$$

Dari hasil perhitungan di atas di dapatkan hasil bahwa jarak data usia pertama dengan *cluster* pertama adalah 19. Jarak dari data Usia pertama ke pusat *cluster* kedua dengan persamaan:

$$d(1,2) = \sqrt{(36 - 47)^2}$$

$$= 11$$

Dari hasil perhitungan di atas di dapatkan hasil bahwa jarak data usia pertama dengan *cluster* pertama adalah 11. Jarak dari data Usia pertama ke pusat *cluster* ketiga dengan perhitungan:

$$d(1,3) = \sqrt{(36 - 70)^2}$$

$$= 34$$

Dari hasil perhitungan di atas di dapatkan hasil bahwa jarak data usia pertama dengan *cluster* pertama adalah 34.

Berdasarkan hasil ketiga perhitungan di atas dapat disimpulkan bahwa jarak data usia pertama yang paling dekat adalah *cluster* 1, sehingga usia pertama dimasukkan ke dalam *cluster* 1. Hasil perhitungan selengkapnya dapat dilihat pada gambar berikut.

Gbr. 8 Hasil Perhitungan Perulangan 1

Setelah dihitung jarak terdekat menggunakan Eucludian Distance, kelompokan data ke dalam *cluster* terdekat.

Total	12139	46286	6857
Jumlah Data	511	1058	104
Centroid Baru	23.755382	43.748582	65.932692

Gbr. 9 Centroid Perulangan 1

Setelah mendapatkan label *cluster* untuk masing-masing data, yaitu pada perulangan 1 *cluster* 0 berjumlah = 511 data, *cluster* 1 = 1058 dan *cluster* 2 = 104. Maka dicari nilai rata-ratanya dengan menjumlahkan seluruh anggota masing-masing *cluster* dan dibagi jumlah anggotanya, lalu centroid baru pada perulangan pertama ini adalah 23.75, 42.74 dan 65.93. Karena $k = 3$, maka *cluster* dibagi menjadi *cluster* 0, *cluster* 1 dan *cluster* 2.

3. Perhitungan Menggunakan PHP

Pada perhitungan menggunakan php peneliti menggunakan data Kelurahan Bilabong Jaya. Terdapat tiga *source code* yang digunakan yaitu *excel_reader2.php* untuk membaca excel yang di import, *index.php* untuk meng-import excel ke *source code* selanjutnya dan *upload_aksi.php* untuk menghitung algoritma K-Means.



Gbr. 10 Tampilan Import Excel


```

<!--fungsi form untuk mengirim data ke file upload aksi.php-->
<form method="post" enctype="multipart/form-data" action="upload_aksi.php">
  Pilih File:
  <input name="datakelurahan" type="file" required="required">
  <input name="upload" type="submit" value="Import">
</form>

```

Gbr. 11 Source code Impor Excel

Pada perhitungan *K-Means* menggunakan PHP, digunakan fungsi form untuk mengirim data ke file upload_aksi.php dengan menggunakan metode post. Jadi pada tampilan, setelah memilih data lalu klik import untuk mengirim ke file upload_aksi.php.

```

<?php
//memanggil file excel_reader2.php
include "excel_reader2.php";
?>

<?php
// upload file xls
$target = basename($_FILES['datakelurahan']['name']);
move_uploaded_file($_FILES['datakelurahan']['tmp_name'], $target);

// beri permissi agar file xls dapat di baca
chmod($_FILES['datakelurahan']['name'], 0777);

// mengambil isi file xls
$data = new Spreadsheet_Excel_Reader($_FILES['datakelurahan']['name'], false);
// menghitung jumlah baris data yang ada
$jumlah_baris = $data->rowcount($sheet_index=0);

```

Gbr. 12 Source code upload_aksi.php

Pada Gbr. 12 terdapat include untuk memanggil file php, dimana nama file excel_reader2.php. Yang fungsi nya untuk membaca file excel. Lalu terdapat source code untuk meng-upload file excel.

```

for ($i=0; $i<$jumlah_baris; $i++) {

  // menangkap data dan memasukan ke variabel sesuai dengan kolomnya masing-masing
  //convert $data->val($i, 1) <<<< tanggal lahir menjadi date
  $tanggal = new DateTime('Y-m-d', strtotime($data->val($i, 1)));
  // mengambil tanggal sekarang
  $date2 = new DateTime(date:'Y-m-d');
  //mengambil umur dengan interval tanggal lahir dan tanggal sekarang
  $interval = $tanggal->diff($date2);
  // mengambil tahun dari hasil interval (umur)
  $umur[] = $interval->y;
  //menentukan jumlah array umur
  $k++; // $k=1;
}

```

Gbr. 13 Source code upload_aksi.php

Pada Gbr. 13 variabel k berfungsi untuk menghitung jumlah data tgl lahir. Perulangan dimulai dari baris ke 5, sampai dengan jumlah baris. Lalu menentukan data mana yang akan dipakai, dengan men-convert data awal yang sebelumnya string menjadi format

date. Lalu variabel date2 adalah mengambil data tanggal sekarang, setelah itu diselisihkan antara tanggal lahir dan tanggal sekarang. Pada variabel usia yang diambil hanya data tahun saja.

```

//menentukan tiga centroid
for ($j=0; $j<3; $j++) {
  //mendefinisikan total cluster dan jumlah data = 0
  $totcl1 = 0;
  $totcl2 = 0;
  $totcl3 = 0;
  $jum1 = 0;
  $jum2 = 0;
  $jum3 = 0;
  $c1 = 0;
  $c2 = 0;
  $c3 = 0;
  for ($i=0; $i<$jumlah_baris; $i++) {
    //menentukan jarak minimal
    $dist1 = abs($umur[$i] - $c1);
    $dist2 = abs($umur[$i] - $c2);
    $dist3 = abs($umur[$i] - $c3);
    //menentukan cluster mana yang terdekat
    if ($dist1 <= $dist2 && $dist1 <= $dist3) {
      $cluster = 1;
    } else if ($dist2 <= $dist1 && $dist2 <= $dist3) {
      $cluster = 2;
    } else {
      $cluster = 3;
    }
    //menentukan umur di variabel c1, c2, c3
    $c1 += $umur[$i];
    $c2 += $umur[$i];
    $c3 += $umur[$i];
    //menentukan total cluster
    $totcl1 += $dist1;
    $totcl2 += $dist2;
    $totcl3 += $dist3;
    //menentukan jumlah data
    $jum1++;
  }
}

```

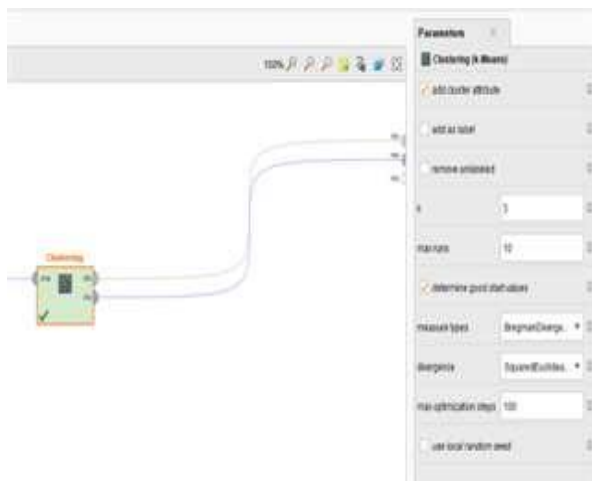
Gbr. 14 Source code upload_aksi.php

Pada Gbr. 14 menentukan tiga centroid, lalu perulangan for untuk menentukan batas iterasi sebanyak 10 kali. Selanjutnya mendefinisikan total cluster dan jumlah data = 0. Lalu masuk ke perhitungan Ecludian Distance, dimana \$c1[\$m][i]=abs(\$umur[\$i]- \$cent1) c1 = umur-centroid1.

Mencari jarak minimal yaitu dengan if(\$c1[\$m][i]<=\$c2[\$m][i]&&\$c1[\$m][i]<=\$c3[\$m][i]) yaitu jika cluster 1 kurang dari sama dengan cluster 2 dan cluster 1 kurang dari sama dengan cluster 3 maka jarak minimal = cluster 1 dan tuliskan umur di variabel c11, sedangkan c12 dan c13 tuliskan 0.

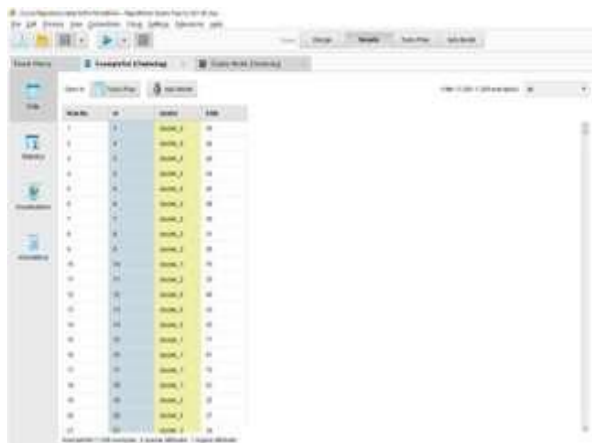
\$totcl1=\$totcl1+\$c1[\$m][i] berfungsi untuk menentukan total cluster dan variabel \$jum1++ untuk menghitung jumlah data. Jika fungsi if tidak terpenuhi maka akan masuk ke else if, dan jika tidak terpenuhi juga maka akan masuk ke else.

Gbr. 16 menunjukkan konfigurasi setelah memasukan node read excel yang digunakan untuk menentukan kolom yang akan di proses, yaitu kolom Usia.



Gbr. 18 Clustering dengan k=3

Pada Gbr. 17 menunjukan node *clustering* berfungsi untuk menemukan titik centroid dari dataset yang diberikan dimana *clustering* yang dibagi sebanyak tiga class atau $k = 3$ dimana class tersebut dibagi menjadi *cluster_0* adalah usia muda, *cluster_1* adalah usia dewasa, dan *cluster_2* adalah lanjut usia.



Gbr. 19 Hasil Konfigurasi RapidMiner

Hasil konfigurasi pada gambar 4.27 lalu diolah menggunakan MS. Excel dengan menggabungkan data dan memilih data yang akan digunakan saja.

C. Penyebaran (Deployment)

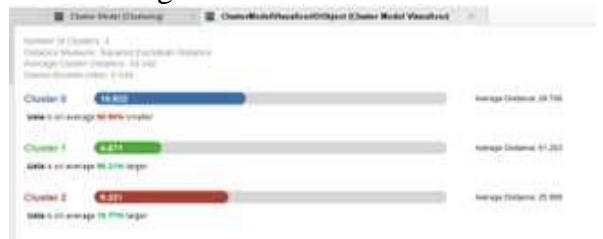
Pada tahap ini, pengetahuan atau informasi yang telah diperoleh akan dipresentasikan dalam bentuk khusus sehingga dapat diketahui atau dibaca oleh orang awam.

Tabel 1. Hasil *Clustering* dari setiap Kelurahan

Kecamatan	Kelurahan	Cluster 0	Titik Centroid	Cluster 1	Titik Centroid	Cluster 2	Titik Centroid	Jumlah Data
Langkapura	Bilabong Jaya	546	24,17	856	42,06	271	58,26	1673
	Gunung Agung	2068	26,12	1933	44,17	990	63,93	4991
	Gunung Terang	3379	26,38	2802	43,83	1406	62,25	7587
	Langkapura Baru	1994	26,25	1851	42,71	1055	61,69	4900
	Langkapura	2180	25,90	2085	42,47	1099	61,93	5364
Rajabasa	Gedong Meneng Baru	586	27,13	497	47,49	246	67,11	1329
	Gedong Meneng	1349	25,84	1314	44,92	512	67,08	3175
	Rajabasa Jaya	2708	26,62	2568	43,73	975	65	6251
	Rajabasa Nunyal	2390	26,03	1845	42,98	1155	61,05	5390
	Rajabasa	1340	25,68	1190	44,50	526	64,22	3056
	Rajabasa Penuka	1792	25,73	1551	42,90	822	59,28	4165
	Rajabasa Raya	2392	25,74	1965	44,92	861	63,69	5218
Kemiling	Beringin Jaya	2189	26,04	2144	44,18	934	61,96	5267
	Beringin Raya	3280	25,71	3164	43,21	1777	62,90	8221
	Kedaung	752	27,42	914	40,91	320	61,42	1986
	Kemiling Permal	3125	23,98	2709	41,05	1545	57,03	7379
	Kemiling Raya	2039	25,34	1789	42,84	866	61,47	4694
	Pinang Jaya	2075	26,04	2043	42,09	864	62,79	4982
	Sumber Agung	1089	25,19	1103	41,53	502	62,60	2694
	Sumber Rejo	2623	25,92	2183	44,08	1282	62,39	6088
	Sumber Rejo Sejahtera	2270	24,98	3037	41,83	1304	64,22	6611

1. Kecamatan Langkapura

Hasil *cluster* dari Kecamatan Langkapura adalah sebagai berikut:

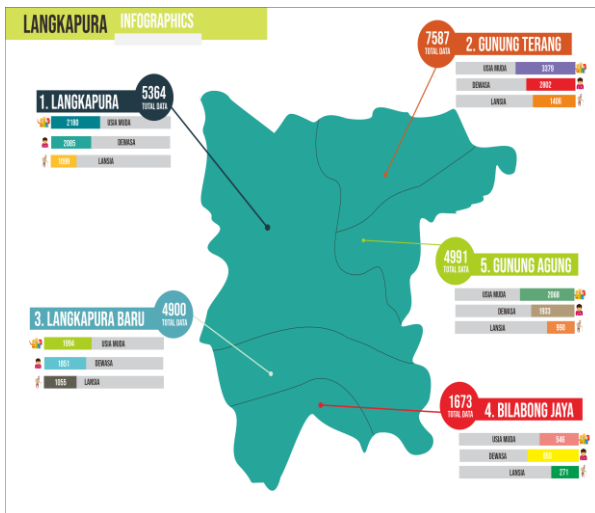


Gbr. 20 Hasil *Clustering* Kecamatan Langkapura.

Cluster	Titik Centroid
Cluster 0	25,394
Cluster 1	43,688
Cluster 2	62,531

Gbr. 21 Titik Centroid Kecamatan Langkapura

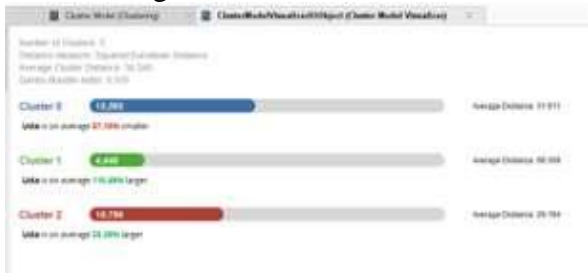
Pada Gbr. 19 dan Gbr. 20 adalah hasil dari *clustering* Kecamatan Langkapura dengan RapidMiner, pada Kecamatan Langkapura memiliki jumlah keseluruhan DPT yaitu 24515. Setelah di *cluster* dengan RapidMiner dan $k = 3$ maka, *cluster 0* berjumlah 10522 DPT dengan titik centroid 25.394, *cluster 1* berjumlah 4671 DPT dengan titik centroid = 62.531 dan *cluster 2* berjumlah 9321 DPT dengan titik centroid 43.688.



Gbr. 22 Infographic Kecamatan Langkapura

2. Kecamatan Rajabasa

Hasil *cluster* dari Kecamatan Rajabasa adalah sebagai berikut:

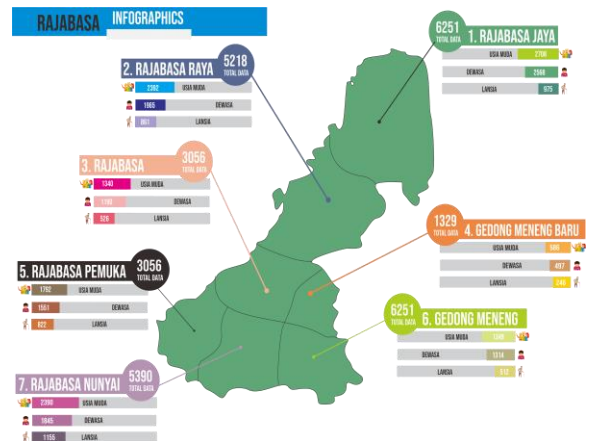


Gbr. 23 Hasil Clustering Kecamatan Rajabasa



Gbr. 24 Titik Centroid Kecamatan Rajabasa

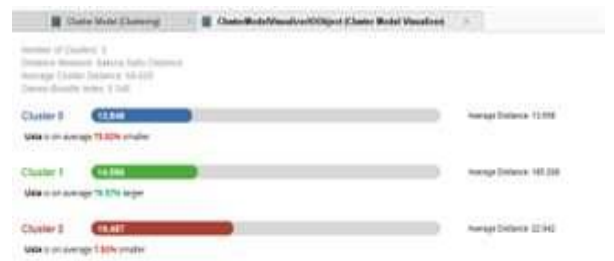
Pada Gbr. 22 dan Gbr. 23 adalah hasil dari *clustering* Kecamatan Rajabasa dengan RapidMiner, pada Kecamatan Rajabasa memiliki jumlah keseluruhan DPT yaitu 28583. Setelah di *cluster* dengan RapidMiner dan k = 3 maka, *cluster* 0 berjumlah 13353 DPT dengan titik centroid 23.079, *cluster* 1 berjumlah 4440 DPT dengan titik centroid = 26.696 dan *cluster* 2 berjumlah 10790 DPT dengan titik centroid 45.369.



Gbr. 25 Infographic Kecamatan Rajabasa

3. Kecamatan Kemiling

Hasil *cluster* dari Kecamatan Kemiling adalah sebagai berikut:

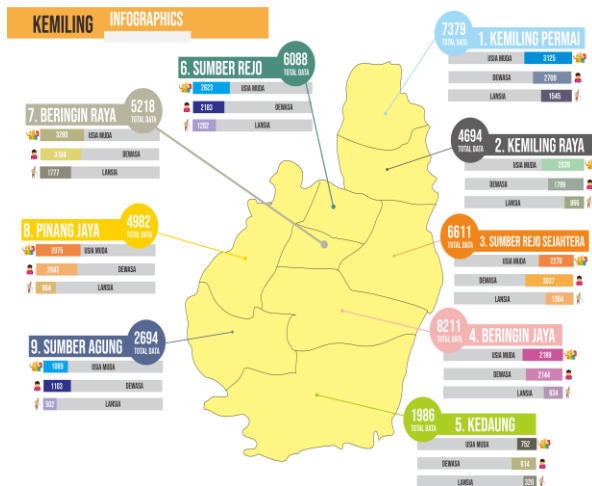


Gbr. 26 Hasil Clustering Kecamatan Kemiling



Gbr. 27 Titik Centroid Kecamatan Kemiling.

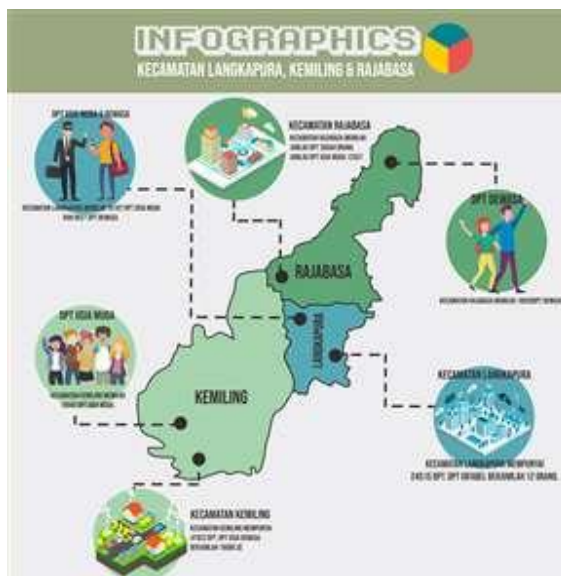
Pada Gbr. 25 dan Gbr. 26 adalah hasil dari *clustering* Kecamatan Kemiling dengan RapidMiner, pada Kecamatan Kemiling memiliki jumlah keseluruhan DPT yaitu 47921. Setelah di *cluster* dengan RapidMiner dan k = 3 maka, *cluster* 0 berjumlah 13848 DPT dengan titik centroid 23.079, *cluster* 1 berjumlah 14586 DPT dengan titik centroid = 26.696 dan *cluster* 2 berjumlah 19487 DPT dengan titik centroid 64.658.



Gbr. 28 Infographic Kecamatan Kemiling

4. Hasil Clustering dari 3 Kecamatan

Adapun hasil dari 3 Kecamatan adalah sebagai berikut:



Gbr. 29 Infographic Kecamatan Langkapura, Rajabasa dan Kemiling



Gbr. 30 DPT dari Kecamatan Langkapura, Rajabasa dan Kemiling

Pada Gbr. 28 dan Gbr. 29 adalah *infographic* dan grafik dari Kecamatan Langkapura, Rajabasa dan Kemiling. Pada Infographics tersebut menjelaskan data DPT dari setiap Kecamatan, di Kecamatan Langkapura tercatat memiliki jumlah 24515 DPT. Pada DPT usia muda berjumlah 10167, sedangkan DPT usia dewasa berjumlah 9527 lalu DPT lansia berjumlah 4821 orang. Selanjutnya pada Kecamatan Rajabasa memiliki jumlah keseluruhan DPT 28584, DPT usia muda berjumlah 12557. Lalu DPT usia dewasa berjumlah 10930, sedangkan DPT lansia berjumlah 5097. Pada Kecamatan Kemiling memiliki keseluruhan jumlah DPT 47922, paling banyak diantara 3 Kecamatan yang dibahas. DPT usia muda pada berjumlah 19442, DPT usia dewasa berjumlah 19086 dan DPT lansia berjumlah 9394.

V. KESIMPULAN

A. Kesimpulan

Berdasarkan hasil penelitian yang telah dilakukan, maka didapat beberapa kesimpulan :

1. Berdasarkan hasil data yang telah didapatkan, usia yang paling banyak dari setiap Kecamatan adalah *cluster* usia muda. Dimana pada Kecamatan Langkapura
2. Data *cluster* usia muda terbanyak terdapat pada Kecamatan Langkapura Kelurahan Gunung Terang dengan jumlah 3379 usia muda.
3. DPT Usia Dewasa terbanyak pada Kelurahan Beringin Raya dengan jumlah 3164 orang.
4. Daerah yang paling banyak lansia yaitu Kelurahan Beringin Raya dengan jumlah lansia 1777 orang.

B. Saran

Adapun saran dari penelitian ini adalah:

1. Diharapkan pada penelitian selanjutnya menggunakan *software* selain RapidMiner seperti Weka dan R, agar dapat membandingkan dan memberikan hasil yang lebih baik lagi.
2. Pada penelitian selanjutnya diharapkan dapat memasukan semua Kecamatan yang ada di Bandar Lampung.

DAFTAR PUSTAKA

- [1] Triyadi Isworo. (2018) 1.707.747 Warga Lampung Golput, *Lampost*. [Online]. Available: <https://www.lampost.co/berita-1-707-747-warga-lampung-golput.html>.
- [2] Tri Kurniawan. Adelia, Hendra Saputra, Sakdon. "Peran Badan Pengawas PEMILU Dalam Meningkatkan Partisipasi Lansia Pada PEMILU 2019 Di Kota PangkalPinang". *Journal of Chemical Information and Modeling*, PangkalPinang, 2019.
- [3] Informatikalogi. (2018) Algoritma *K-Means* Clustering, *www.informatikalogi.com*, [Online]. Available: <https://informatikalogi.com/algoritma-k-means-clustering/>.
- [4] F. Ramadhanil, V. Junaidi, S. Pramono, and R. Widyastuti. "Desain Partisipasi Masyarakat Dalam Pemantauan Pemilu". *Kemitraan bagi Pembaruan Tata Pemerintahan di Indonesia*, Jakarta Selatan, 2015.
- [5] Thi Bi Dan, S. Widya Sihwi, and R. Anggrainingsih. "Implementasi Iterative Dichotomiser 3 Pada Data Kelulusan Mahasiswa S1 Di Universitas Sebelas Maret". *Jurnal UNS, Jawa Tengah*, 2015.
- [6] Dedi Suryadi. Sani Susanto, "Pengantar Data Mining". ANDI OFFSET, Jogja, 2010.
- [7] Yuli Asriningtias, Rodyah Mardhiyah. "Aplikasi Data Mining Untuk Menampilkan Informasi Tingkat Kelulusan Mahasiswa". *Journal UAD*, Yogyakarta, 2014.
- [8] Yudi Agusta, "*K-Means* - Penerapan, Permasalahan dan Metode Terkait". *Jurnal Sistem dan Informatika*, Bali, 2007.
- [9] G. Gustientiedina, M. Hasmil Adiya, and Y. Desnelita. "Penerapan Algoritma *K-Means* Untuk Clustering Data Obat-Obatan". *Jurnal Nasional Teknologi & Sistem Informasi*, Pekanbaru, 2018.
- [10] Ade Bastian. "Penerapan Algoritma *K-Means* Clustering Analysis Pada Penyakit Menular Manusia (Studi Kasus Kabupaten Majalengka)". *Jurnal Sistem Informasi*, Majalengka, 2018.
- [11] Mardalius. "Pemanfaatan Rapid Miner Studio 8.2 Untuk Pengelompokan Data Penjualan Aksesoris Menggunakan Algoritma *K-Means*". *Jurnal Teknologi dan Sistem Informasi*, Sumatera Utara, 2018.
- [12] Rüdiger Wirth, Jochen Hipp. "CRISP-DM: Towards a Standard Process Model for Data Mining". University of Tübingen, Germany, 2000.
- [13] Imam Kurniawan, Ajib Susanto. "Implementasi Metode *K-Means* dan Naïve Bayes Classifier untuk Analisis Sentimen Pemilihan Presiden (Pilpres) 2019". *Jurnal Eksplora Informatika*, Semarang, 2019.
- [14] Fajar Nur, Budi Setiawan, and Issa Arwani. "Implementasi Algoritma *K-Means* untuk Klasterisasi Kinerja Akademik Mahasiswa". *Jurnal Universitas Brawijaya*, Malang, 2018.
- [15] S. Rahayu, D. T. Nugrahadi, and F. Indriani. "Clustering Penentuan Potensi Kejahatan Daerah Di Kota Banjarbaru Dengan Metode *K - Means*". *Jurnal KLIK.*, Banjarbaru, 2014.
- [16] Sahar Namvarasl and Marzieh Ahmadzadeh. "A Dynamic Flooding Attack Detection System Based on Different Classification Techniques and Using SNMP MIB Data," *International Journal of Computer Networks and Communications*, Iran, 2014.
- [17] Gigih Forda Nama, Dikpride Despa, "An Enhanced *K-Means* Clustering Algorithm for Pattern Discovery in Big Data Analysis of 3-Phase Electrical Quantities". *International Journal of Engineering & Technology (IJET)*, Indonesia, 2018.
- [18] F. Schafer, C. Zeiselmaier, J. Becker, and H. Otten. "Synthesizing CRISP-DM and Quality Management: A Data Mining Approach for Production Processes". *IEEE International*

Conference on Technology
Management, Operations and Decision,
Germany, 2018.