

ALGORITMA PARTITIONAL CLUSTERING UNTUK PENGKLASIFIKASIAN TINGKAT PENGANGGURAN DI KABUPATEN LAHAT

Danang Nugraha¹, Megawaty², Dinny Komalasari³

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

Jl. Jendral Ahmad Yani No.12 Palembang

Pos-el : Danangnugraha19@gmail.com

ABSTRACT

Unemployment is a situation where a person belonging to the labor force wants to get a job but cannot get it. Unemployment causes various economic and social problems to those who experience it. In addition, it can interfere with the level of family health. Prolonged unemployment has a bad psychological effect on the unemployed and their families. Unemployment is a very bad problem for the economy and society, and therefore efforts must be made to overcome them. But based on its characteristics, unemployment can be classified into several groups. Therefore, using partitional clustering algorithm will classify unemployment based on its characteristics. So that the number of unemployed groups will be obtained based on their characteristics, so that their hopes make it easier to find a solution to overcome unemployment.

Keywords: partitional clustering algorithm, unemployment

ABSTRAK

Pengangguran adalah suatu ke-adaan di mana seseorang yang tergolong angkatan kerja ingin mendapatkan pekerjaan tetapi belum dapat memperolehnya. Pengangguran menimbulkan berbagai masalah ekonomi dan sosial kepada yang mengalaminya. Di samping itu ia dapat mengganggu taraf kesehatan keluarga. Pengangguran yang berkepanjangan menimbulkan efek psikologi yang buruk ke diri penganggur dan keluarganya. Pengangguran adalah masalah yang sangat buruk efeknya kepada perekonomian dan masyarakat, dan oleh sebab itu secara terus menerus usaha-usaha harus dilakukan untuk mengatasinya. Akan tetapi berdasarkan ciri-cirinya, pengangguran dapat digolongkan dalam beberapa golongan. Maka dari itu dengan menggunakan algoritma *partitional clustering* akan mengelompokkan pengangguran berdasarkan ciri-cirinya. Sehingga akan didapat jumlah kelompok-kelompok pengangguran berdasarkan ciri-cirinya, agar harapannya mempermudah untuk mendapatkan solusi tentang mengatasi pengangguran.

Kata Kunci: algoritma *partitional clustering*, pengangguran

1. PENDAHULUAN

1.1 Latar Belakang

Kehadiran teknologi informasi terutama basis data sudah menjadi kebutuhan pokok. Teknologi komputer yang diterapkan di BPS (Badan Pusat Statistik) selalu disesuaikan dengan perkembangan teknologi informasi dan juga mengacu kepada kebutuhan. Personal komputer yang secara umum lebih murah dan efisien. Sejak 1980-an, personal komputer telah digunakan di seluruh kantor BPS provinsi, diikuti dengan penggunaan komputer di seluruh BPS kabupaten dan kota sejak 1992. Data penduduk di Kabupaten Lahat yang tercatat di BPS pada tahun 2016 sebanyak 397.424 jiwa, yang tersebar di seluruh kecamatan yang ada di Kabupaten Lahat. Kabupaten Lahat sendiri memiliki 22 Kecamatan yang tersebar di seluruh Kabupaten Lahat, kecamatan-kecamatan tersebut antara lain: Kecamatan Tanjung Sakti Pumi, Kecamatan Tanjung Sakti Pumu, Kecamatan Kota Agung, Kecamatan Mulak Ulu, Kecamatan Tanjung Tebat, Kecamatan Pulau Pinang, Kecamatan Pagar Gunung, Kecamatan Gumay Ulu, Kecamatan Jarai, Kecamatan Pajar Bulan, Kecamatan Muara Payang, Kecamatan Suka Merindu, Kecamatan Kikim Barat, Kecamatan Kikim Timur, Kecamatan Kikim Selatan, Kecamatan Kikim Tengah, Kecamatan Lahat, Kecamatan Gumay Talang, Kecamatan Pseksu, Kecamatan Merapi Barat, Kecamatan Merapi Timur, kecamatan Merapi Selatan.

Dari seluruh kecamatan yang ada tersebut BPS mempunyai data per kecamatan tersebut mulai dari data jumlah penduduk per kecamatan, data jenis kelamin, keadaan geografi dan iklim perkecamatan, data desa per kecamatan meliputi luas desa serta persentase

penduduk yang menempati desa tersebut, dan data jumlah tempat-tempat pendidikan.

Di BPS sejauh ini menyediakan data jumlah pengangguran di Kabupaten Lahat hanya *field* jenis kelamin (Laki-laki dan Perempuan). Data jumlah pengangguran sangat membantu pemerintah dalam memantau perkembangan pengangguran yang ada di Kabupaten Lahat. Jika data pengangguran didapat mulai dari jenis kelamin, rentan umur, serta kecamatan apa saja yang mempunyai tingkat pengangguran tertinggi, lebih mempermudah perangkat pemerintahan dalam mengambil keputusan untuk mengatasi masalah pengangguran tersebut.

2. TINJAUAN PUSTAKA

2.1 Data

Menurut Wahyudi, Bambang. (2004:1) Makna kata data bagi manusia adalah segala sesuatu yang dapat diterima oleh indera manusia dari rangsangan-rangsangan yang ada di sekitarnya, baik secara tersurat maupun yang tersirat.

Sedangkan makna kata data di komputer adalah segala sesuatu yang bisa dikodekan atau disimbolkan atau dilambangkan dengan kode-kode, simbol-simbol, atau lambang-lambang yang telah disediakan di setiap komputer.

2.2 Data Mining

Vulandari, Retno Tri. (2017:1) Data *mining* adalah serangkaian proses untuk menggali nilai tambah berupa informasi yang selama ini tidak diketahui secara manual dari suatu basis data. Informasi yang dihasilkan diperoleh dengan cara mengekstraksi dan mengenali pola yang penting atau menarik dari data yang terdapat pada basis data. Data *mining* terutama digunakan untuk mencari

pengetahuan yang terdapat dalam basis data yang besar sehingga sering disebut *Knowledge Discovery Databases (KDD)*.

2.2.1 Tahapan Data Mining

Menurut Hermawati, Fajar Astuti (2013:5) Operasi data *mining* menurut sifatnya dibedakan menjadi dua, yaitu bersifat (1) prediksi (*Prediction driven*) untuk menjawab pertanyaan apa dan sesuatu yang bersifat remang-remang atau transparan. Operasi prediksi digunakan untuk validasi hipotesis, *querying* dan pelaporan (misal : *spreadsheet* dan *pivot* tabel), analisis statistik. (2) Penemuan (*discovery driven*) bersifat transparan dan untuk menjawab pertanyaan “mengapa?”. Operasi penemuan digunakan untuk analisis data eksplorasi, pemodelan prediktif, segmentasi *database*, analisis keterkaitan (*link analysis*) dan deteksi deviasi.

Menurut Vulandari, Retno Tri. (2017:2) *Data mining* merupakan salah satu dari rangkaian *Knowledge Discovery in Database (KDD)*. *KDD* berhubungan dengan teknik integrasi dan penemuan ilmiah, interpretasi dan visualisasi dari pola-pola sejumlah data.

Adapun untuk menganalisis data dalam penerapan *data mining* ini menggunakan tahapan *Knowledge Discovery in Database (KDD)* yang terdiri dari beberapa tahapan yaitu sebagai berikut :

1. Data Cleaning (Pembersihan Data)

Ini adalah sebuah fase dimana data-data tidak lengkap, mengandung error dan tidak konsisten dibuang dari koleksi data, sehingga data yang telah bersih dan relevan dapat digunakan untuk diproses ulang untuk penggalian pengetahuan (*discovery knowledge*).

2. Data Integration (Data Integrasi)

Pada tahap ini terjadi integrasi data, dimana sumber-sumber data yang berulang (*multiple data*), *file-file* yang berulang (*multiple file*), dapat dikombinasikan dan digabungkan ke dalam suatu sumber.

3. Data Selection (Seleksi Data)

Pada langkah ini, data yang relevan terhadap analisis dapat dipilih dan diterima dari koleksi data yang ada.

4. Data Transformation (Transformasi Data)

Juga dikenal sebagai data *consolidation*. Pada tahap ini, dimana data-data yang telah terpilih, ditransformasikan ke dalam bentuk-bentuk yang cocok untuk prosedur penggali (*mining porcedure*) dengan cara melakukan normalisasi dan agregasi data.

5. Data mining

Tahap ini adalah tahap yang paling penting, dengan menggunakan teknik-teknik yang diaplikasikan untuk mengekstrak pola-pola potensial yang berguna.

6. Evaluation

Pada tahap ini, pola-pola menarik dengan jelas merepresentasikan pengetahuan telah diidentifikasi berdasarkan measure yang telah diberikan.

7. Presentation

Ini merupakan tahap terakhir, dimana pengetahuan yang telah ditemukan secara visual ditampilkan kepada user. Tahap ini menggunakan teknik visualisasi untuk membantu user dalam mengerti dan menginterpretasikan hasil dari data mining.

2.3 Algoritma

Algoritma adalah sekumpulan langkah yang rinci yang ditujukan untuk menyelesaikan suatu masalah. Langkah-langkah yang dimaksud adalah agar bisa dituangkan ke dalam program sehingga bisa dieksekusi oleh sistem komputer. Namun, tanpa harus menuangkan ke dalam program, langkah-langkah yang terdapat di algoritma bisa diuji secara manual. (Kadir, 2013:10).

2.4 Clustering

Menurut Hermawati, Fajar Astuti (2013:16) *Clustering* adalah mempartisi data-set menjadi beberapa sub-set atau kelompok sedemikian rupa sehingga elemen-elemen dari suatu kelompok tertentu memiliki set properti yang dishare bersama, dengan tingkat similaritas yang tinggi dalam satu kelompok dan tingkat similaritas antar kelompok yang rendah. Disebut juga dengan 'unsupervised learning'.

Menurut Vlandari, Retno Tri. (2017:53) *Clustering* juga disebut sebagai *segmentation*. Metode ini digunakan untuk mengidentifikasi kelompok alami dari sebuah kasus yang didasarkan pada sebuah kelompok atribut, mengelompokkan data yang memiliki kemiripan atribut.

2.4.1 Partitional Clustering

Hermawati, Fajar Astuti (2013:125) *Partitional Clustering* adalah pembagian objek data ke dalam sub-himpunan (*cluster*) yang tidak overlap sedemikian hingga tiap objek data berada dalam tepat satu sub-himpunan.

2.4.2 K-Means

Menurut Vlandari, Retno Tri. (2017:60) *K-Means* merupakan algoritma clustering yang berulang-ulang. Algoritma K-Means menetapkan nilai-nilai cluster (K)

secara random, untuk sementara nilai tersebut menjadi pusat dari cluster atau bisa disebut dengan centroid, mean atau means. Kemudian menghitung jarak setiap data yang ada terhadap masing-masing centroid menggunakan rumus Euclidian hingga ditemukan jarak yang paling dekat dari setiap data dengan centroid. Klasifikasi setiap data berdasarkan kedekatannya dengan centroid. Lakukan langkah tersebut hingga nilai centroid tidak berubah (stabil).

Hermawati, Fajar Astuti (2013:127) *K-Means* merupakan pendekatan *partitional clustering*. Tiap *cluster* dihubungkan dengan sebuah *centroid* (titik pusat). Tiap titik ditempatkan ke dalam *cluster* dengan *centroid* terdekat. Jumlah *cluster*, K, harus ditentukan. Algoritma dasarnya sangat sederhana, yaitu:

1. Pilih K titik sebagai *centroid* awal
2. Ulangi
3. Bentuk K *cluster* dengan menempatkan semua titik yang terdekat.
4. Ulangi perhitungan *centroid* dari tiap *cluster*.
5. Sampai *centroid* tidak berubah.

$$d_{ij} = \sqrt{\{x - x\}^2}$$

Dimana:

- d_{ij} = pusat *cluster*
 X = data
 X = data

2.5 Pengangguran

Menurut Sukirno, Sadono. (2017:13) Pengangguran merupakan suatu keadaan di mana seseorang yang tergolong dalam angkatan kerja ingin mendapatkan pekerjaan tetapi mereka belum dapat memperoleh pekerjaan tersebut.

Sukirno, Sadono. (2017:330)
 Pengangguran dibagi menjadi empat kelompok:

1. Pengangguran terbuka

Pengangguran ini tercipta sebagai pertambahan lowongan pekerjaan yang lebih rendah dari pertambahan tenaga kerja. Pengangguran terbuka dapat pula wujud sebagai akibat dari kegiatan ekonomi yang menurun, dari kemajuan teknologi yang mengurangi penggunaan tenaga kerja, atau sebagai akibat dari kemunduran perkembangan sesuatu industri.

2. Pengangguran Bermusim

Pengangguran ini terutama terdapat di sektor pertanian dan perikanan. Pada musim hujan penyadap karet dan nelayan tidak dapat melakukan pekerjaan mereka dan terpaksa menganggur. Pengangguran seperti ini digolongkan sebagai pengangguran bermusim.

3. Setengah Menganggur

Di negara-negara berkembang migrasi dari desa ke kota adalah sangat pesat. Sebagai akibatnya tidak semua orang yang pindah ke kota dapat memperoleh pekerjaan dengan mudah. Sebagiannya terpaksa menjadi penganggur sepenuh waktu. Di samping itu ada pula yang tidak menganggur, tetapi tidak pula bekerja sepenuh waktu, dan jam kerja mereka adalah jauh lebih rendah dari yang normal.

3. Analisis Data Mining

3.1 Metode Analisis

3.1.1 Data Cleaning

Pembersihan data sebenarnya merupakan tahap awal dari proses *KDD*. merupakan proses menghilangkan *noise* dan data yang telah digabung akan dilakukan pembersihan, membuang data yang kosong dan memastikan data tersebut relevan atau terkait satu sama lain. Pembersihan data juga akan mempengaruhi performansi dari

teknik *data mining* karena data yang ditangani akan berkurang jumlah dan kompleksitasnya., pembersihan data terhadap *noise* yang ditemukan berupa *missing value*, *inkonsisten data*, dan *redundant data*.

Sebelum melanjutkan ke dalam data mining terlebih dahulu kita harus melakukan langkah-langkah yang berguna untuk meningkatkan kualitas hasil dari algoritma selanjutnya. Proses data mining yang pertama adalah *cleaning* data yaitu proses pembersihan data yang diawali dari prosedur pengambilan data. Berikut adalah data awal yang didapat dari BPS (Badan Pusat Statistik) data dalam bentuk excel.

Data Penduduk Tahun 2015							
Kode	Nama	Umur	Jenis Kelamin	Kecamatan	Agama	Pendidikan Terakhir	Jenjang Pemasangan
1	KHAFIUL NAZAR	44	Laki-Laki	Prabon	Islam	SDIA	48
2	NA KETUY SUWEDI	48	Pemrosesan	Sukoharjo	Islam	SD	27
3	CHORRAN	34	Pemrosesan	Ujung Timur	Islam	SDIA	24
4	MARTIA SAPUTRA	34	Laki-Laki	Sukoharjo	Islam	SDIP	32
5	ERNA NINGSIH	42	Pemrosesan	Ujung Timur	Islam	SD	40
6	ZULKARNEN	47	Laki-Laki	Gunung Kidul	Islam	SDIA	37
7	Prabon	48	Pemrosesan	Islam	Islam	Diploma III	48
8	SUCI LUCI	48	Laki-Laki	Luhur	Islam	SDIA	40
9	ANTON SAPUTRA	38	Laki-Laki	Gunung Ulu	Islam	SDIP	48
10	Elen Melya Sika	31	Pemrosesan	Ujung Barat	Islam	SD	7
11	SUPREADI	38	Laki-Laki	Luhur	Islam	SDIP	40
12	PEMILIA SULTANAWATI	33	Pemrosesan	Gunung Ulu	Islam	SI	37
13	MAJZANA S.Pp	46	Pemrosesan	Luhur	Islam	SI	40
14	FAZRI HARTADINATA	33	Laki-Laki	Ekonomi Baru	Islam	SI	40
15	RITTI RICHARD AGENTINI	42	Pemrosesan	Gunung Ulu	Islam	SDIA	37
16	RIJAM SUKELTI ADIPUTRA	32	Laki-Laki	Luhur	Islam	SDIA	48
17	Peg Wahidinda	20	Pemrosesan	Islam	Islam	SDIA	40
18	Novi Ranyana Tasa	49	Pemrosesan	Ekonomi Baru	Protestan	SI	48
19	SUDHARDIANSAH	34	Laki-Laki	Luhur	Islam	SDIA	40
20	ERFPA TULIANA	27	Pemrosesan	Tanjung Sakti Purno	Islam	SDIA	7
21	BEPI LAPRI	38	Laki-Laki	Tanjung Sakti Purno	Islam	SDIA	24
22	HENDRA GUNAWAN	39	Laki-Laki	Gunung Ulu	Islam	SDIA	40
23	Solihana	39	Pemrosesan	Luhur	Islam	SI	37
24	ARDIANTAH	25	Laki-Laki	Tanjung Sakti Purno	Islam	SDIA	48
25	HAZELON	43	Pemrosesan	Gunung Ulu	Islam	SDIA	0
26	PAFFINIS Sed	35	Pemrosesan	Luhur	Buddha	SI	24
27	ABDUL MUBTAKIN dan (DUL)	22	Laki-Laki	Prabon	Islam	SD	32

Gambar 3.1 Data penduduk tahun 2015

Pada gambar 3.2 adalah data penduduk tahun 2016 yang didapat dari Badan Pusat Statistik kabupaten Lahat.

Data Penduduk Tahun 2016							
Kode	Nama	Umur	Jenis Kelamin	Kecamatan	Agama	Pendidikan Terakhir	Jenjang Pemasangan
1	HERAN	33	Laki-Laki	Ekonomi Timur	Islam	SDIP	32
2	Marnab	42	Laki-Laki	Ekonomi Timur	Islam	SDIA	27
3	Ilmu Yusuf	36	Laki-Laki	Korona Agung	Islam	SDIA	40
4	IRACHETIAR	37	Laki-Laki	Tanjung Sakti Purno	Islam	SDIA	32
5	Nizar	27	Laki-Laki	Korona Agung	Islam	SDIA	0
6	ISABRI	28	Laki-Laki	Ujung Barat	Islam	SDIP	32
7	Shah Pujiatun Gan	40	Pemrosesan	Korona Agung	Islam	SDIA	24
8	Wanck Indah	36	Laki-Laki	Ekonomi Tengah	Islam	SDIP	18
9	Nidala Purni SE	33	Pemrosesan	Korona Agung	Islam	SI	48
10	Ilmuwan	23	Laki-Laki	Korona Agung	Islam	SDIA	32
11	JA RONI	30	Laki-Laki	Ekonomi Timur	Islam	SDIP	40
12	BEKA SANJAYA	43	Laki-Laki	Korona Agung	Islam	SDIA	0
13	NATYIA UTAMA	26	Laki-Laki	Ekonomi Tengah	Islam	SDIA	24
14	Indah	21	Laki-Laki	Korona Agung	Islam	SDIA	32
15	Nidala Anugrah	22	Pemrosesan	Korona Agung	Islam	SDIA	0
16	Ribang	38	Laki-Laki	Ekonomi Timur	Islam	SDIA	40
17	Ilmuwan	24	Laki-Laki	Tanjung Sakti Purno	Islam	SDIA	32
18	Ilmu	34	Laki-Laki	Ekonomi Tengah	Islam	SDIA	27
19	Ujung Zainun	32	Laki-Laki	Korona Agung	Islam	SDIP	32
20	ALZABRI	47	Laki-Laki	Ekonomi Tengah	Islam	SDIA	27
21	MAHSYUR	34	Laki-Laki	Tanjung Sakti Purno	Islam	SDIA	48
22	Ilmu Nidala Purno	26	Pemrosesan	Korona Agung	Islam	SDIP	27
23	Hendang Suparna	33	Laki-Laki	Gunung Ulu	Islam	SDIA	32
24	M. Achmady Mub	29	Laki-Laki	Ekonomi Timur	Islam	SDIA	32
25	ANAWATI PAULI ALAMI	31	Laki-Laki	Ekonomi Tengah	Islam	SDIA	32
26	Rita Ad	33	Pemrosesan	Korona Agung	Islam	SDIP	32
27	Abdi Mubak	34	Laki-Laki	Gunung Ulu	Islam	SDIP	0

Gambar 3.2 Data penduduk tahun 2016

Pada gambar 3.3 merupakan data penduduk kabupaten lahat tahun 2017.

Kode	Nama	Umur	Jenis Kelamin	Kecamatan	Agama	Pendidikan Terakhir	Jan Kerja Perunggu	Jan Kerja Perak
1	HANAN M	41	Laki-Laki	Pagar Gunung	Islam	SD	31	1482
2	DORA SAPITRI	31	perempuan	Kikim Tengah	Islam	SD	31	1482
3	Dahlan Tanjung	33	Laki-Laki	Pagar Gunung	Islam	SMP	27	1304
4	Kaga ISMAIL	30	Laki-Laki	Kikim Selatan	Islam	SD	16	812
5	ENDRI ANGGARA RITA	30	perempuan	Kikim Tengah	Islam	SD	40	1712
6	Almas Dga	23	Laki-Laki	Pagar Bulan	Islam	SD	24	1168
7	DEWI SELAR WANGI	22	perempuan	Kikim Selatan	Islam	SD	16	812
8	METLIN HERLINA	27	perempuan	Pagar Gunung	Islam	SD	24	1168
9	Rosa Budiman	48	Laki-Laki	Pagar Bulan	Islam	SD	32	1482
10	Ibrahim	32	Laki-Laki	Kikim Selatan	Islam	SMP	0	0
11	RISKA PERNAMA SARI	22	Perempuan	Kikim Tengah	Islam	SD	40	1712
12	UTULIS KINCY APRADA	22	perempuan	Pagar Bulan	Islam	SD	40	1712
13	ASHELILA MITZI	28	perempuan	Muara Prang	Islam	SD	0	0
14	RADUS SAPUTRA	38	Laki-Laki	Pagar Bulan	Islam	SD	16	812
15	ULFA SASMITA	37	perempuan	Kikim Tengah	Islam	SD	27	1304
16	WIDIA FITRIYANTI	32	Perempuan	Pulau Pinang	Islam	SD	32	1482
17	PUTRA JAYA	28	Laki-Laki	Merapi Barat	Islam	SD	0	0
18	RINI SANTIKA	25	Perempuan	Kikim Selatan	Islam	SMP	40	1712
19	WAMA JAKA	22	Laki-Laki	Kikim Tengah	Islam	SD	16	812
20	SEPTI NANTI	41	perempuan	Pagar Gunung	Islam	SD	24	1168
21	RISA IMAM	34	perempuan	Tanjung Sakti Pura	Islam	SD	16	812
22	WANDA SULASTRI	36	perempuan	Pagar Bulan	Islam	SD	27	1304
23	TEHU ALDIANSAR	39	Laki-Laki	Muara Prang	Islam	SMP	40	1712

Gambar 3.3 Data penduduk tahun 2017

Agar data penduduk dapat diolah dengan metode *k-means clustering*, maka data yang berjenis data char seperti jenis kelamin, kecamatan, agama, dan pendidikan terakhir harus di inialisasikan terlebih dahulu dalam bentuk angka.

Untuk melakukan inialisasi jenis kelamin, kecamatan, agama, dan pendidikan dilakukan dengan rumus =VLOOKUP di excel langkah-langkah sebagai berikut :

1. Buat tabel referensi di excel untuk inialisasi dengan menggunakan rumus =VLOOKUP seperti contoh dibawah.

Pada gambar 3.4 merupakan tabel diexcel untuk inialisasi jenis kelamin yang terdiri dari laki-laki dan perempuan.

Jenkel	Terbilang Angka
Laki-Laki	1
Perempuan	2

Gambar 3.4 Inialisasi jenis kelamin

Dibawah ini gambar 3.5 merupakan gambar tabel inialisasi kecamatan yang menginialisalkan nama kecamatan ke angka agar dapat diproses dengan menggunakan algoritma *k-means*.

Kecamatan	Terbilang Angka
Tanjung Sakti Pu	1
Tanjung Sakti Pu	2
Kota Agung	3
Mulak Ulu	4
Tanjung Tebat	5
Pulau Pinang	6
Pagar Gunung	7
Jarai	8
Pajar Bulan	9
Muara Payang	10
Suka Merindu	11
Kikim Barat	12
Kikim Timur	13
Kikim Selatan	14
Kikim Tengah	15
Lahat	16
Gumay Talang	17
Gumay Ulu	18
Pseksu	19
Merapi Barat	20
Merapi Timur	21
Merapi Selatan	22

Gambar 3.5 Inialisasi kecamatan

Gambar tabel inialisasi agama dapat dilihat pada gambar 3.6 dibawah ini.

Agama	Terbilang Angka
Islam	1
Protestan	2
Kristen	3
Hindu	4
Budha	5
Katolik	6

Gambar 3.6 Inialisasi agama

Gambar 3.7 merupakan gambar tabel inialisasi pendidikan mulai dari kategori tidak menempuh pendidikan sampai dengan yang menempuh pendidikan sampai S2.

Pendidikan	Terbilang Angka
Tidak Menempuh	1
Tidak Tamat SD	2
SD	3
SMP	4
SLTP	4
SMK	5
SMA	5
SLTA	5
Diploma II	6
Diploma III	7
S1	8
S2	9

Gambar 3.7 Inialisasi pendidikan

2. Setelah semua tabel inialisasi dibuat maka selanjutnya menggunakan rumus =VLOOKUP untuk mendapatkan hasil yang diinginkan. Keterangan rumus dijelaskan seperti gambar di bawah ini.

Setelah field jenis kelamin, kecamatan, agama, dan pendidikan telah diinisialisasikan dan digabung dalam satu sheet maka akan terlihat seperti gambar 3.9.

Gambar 3.8 Data yang sudah diinisialisasi

Total item data yang digunakan sebanyak 3526 item. Field data awal berjumlah 9 field yaitu field Kode, Nama, Umur, Jenis kelamin, kecamatan, agama, pendidikan terakhir, jam kerja perminggu, jam kerja pertahun. Pada gambar 3.9 ada field yang dikotak merah adalah field tambahan kode inialisasi field tambahan tersebut yaitu field kode jenis kelamin, kode kecamatan, kode agama, dan kode pendidikan sebagai field inialisasi. Jadi total field setelah field awal di tambah dengan field inialisasi berjumlah total 13 field, data tersebut bisa dilihat seperti gambar 3.8. Selanjutnya yaitu data di *cleaning*, ini adalah sebuah fase dimana data-data tidak lengkap, mengandung error, dan tidak konsisten dibuang dari koleksi data, sehingga data yang telah bersih dan relevan dapat digunakan untuk diproses

ulang untuk penggalian pengetahuan (*Discovery Knowledge*). Dan data yang mengandung error serta tidak konsisten akan digantikan.

Pada gambar 3.9 adalah data penduduk yang telah di masukkan kedalam *rapidminer* data berasal dari data penduduk yang berformat *Excel*. Setelah data di masukkan ke dalam *rapidminer* langkah selanjutnya adalah *cleaning* data atau membersihkan data yang kosong/*noise* dan data yang error akan dihilangkan dari koleksi data.

Gambar 3.9 Data telah di masukkan di *rapidminer*

Name	Type	Missing	Statistics
Kode	Integer	0	Min: 1, Max: 1208, Average: 580.847
Nama	Polynomial	0	Least: wart (1), Most: Mukaly (4), Values: Mukaly (4), SUITSRI
Umur	Integer	0	Min: 10, Max: 56, Average: 31.936
Jenis Kelamin	Polynomial	0	Least: Laki-laki (65), Most: Laki-Laki (2234), Values: Laki-Laki (2234), Per
Kode Jenis Kelamin	Integer	0	Min: 1, Max: 2, Average: 1.348
Kecamatan	Polynomial	0	Least: Kilm Selatan (119), Most: Tanjung Sakti Pumu (282), Values: Tanjung Sakti Pumu
Kode Kecamatan	Integer	0	Min: 1, Max: 22, Average: 11.402
Agama	Polynomial	0	Least: agama (1), Most: Islam (1171), Values: Islam (1171), Kristen

Gambar 3.10 Laporan tanpa data *missing*

Pada data ini di kolom *missing* tidak mengandung data error berarti data ini tidak mengandung data yang error atau tidak konsisten. Setelah data telah di *cleaning* maka dapat dilanjutkan ke langkah berikutnya.

3.1.1 Data Integration

Pada tahap data mining selanjutnya setelah dilakukan pembersihan data adalah integrasi data dimana data yang berulang, file yang berulang akan digabungkan dan dihapuskan salah satunya. Hasil dari penghilangan data yang telah dilakukan dapat dilihat pada gambar 3.11 berikut ini. Setelah dilakukan penghapusan data yang berulang maka didapatkan hasil data setelah dilakukan proses *integration* data.

Row No.	Kode	Nama	Umur	Jenis Kelamin	Kode Jenis ...	Kecamatan	Kode Keca...	Agama	Kode Agan
1	1	KHAFIKUL HA...	44	Laki-Laki	1	Panrau	19	Islam	1
2	2	NI KETUT S...	48	Perempuan	2	Suka Melindu	11	Islam	1
3	3	CHODUHI	44	Perempuan	2	Wenag Timur	21	Islam	1
4	4	IMARTA SAPU...	34	Laki-Laki	1	Suka Melindu	11	Islam	1
5	5	ERWAWINGS...	42	Perempuan	2	Wenag Timur	21	Islam	1
6	6	ZULKARNAE...	47	Laki-Laki	1	Gumaj Takang	17	Islam	1
7	7	Rohmah	48	Perempuan	2	Jara	8	Islam	1
8	8	MULYADI	49	Laki-Laki	1	Lahat	16	Islam	1
9	9	ANTON SAP...	30	Laki-Laki	1	Gumaj Ulu	18	Islam	1
10	10	Elen Waga St...	31	Perempuan	2	Wenag Barat	20	Islam	1
11	11	SUPRADI	38	Laki-Laki	1	Lahat	16	Islam	1
12	12	PENJAL SILU...	35	Perempuan	2	Gumaj Ulu	18	Islam	1
13	13	MARDIANA S...	48	Perempuan	2	Lahat	16	Islam	1
14	14	FARIDIMART...	33	Laki-Laki	1	Kilim Barat	12	Islam	1
15	15	SITI ROHAN...	45	Perempuan	2	Gumaj Ulu	18	Islam	1
16	16	IMAM SUBEK...	25	Laki-Laki	1	Lahat	16	Islam	1

Gambar 3.11 Hasil penghilangan data duplikat

Dari hasil tersebut jumlah item data masih berjumlah 3526 item data, berarti tidak ada data yang dihilangkan atau dihapus selama proses *integration* berlangsung.

3.1.2 Data Selection

Selection data berfungsi untuk memilih *field* yang sesuai kebutuhan untuk mendapatkan *knowledge*, *field* yang dipilih

adalah *field* yang benar-benar dibutuhkan pada penelitian saat ini *field* data yang sesuai untuk mendapatkan hasil yang diinginkan dan sesuai dengan prosesnya adalah umur, kode jenis kelamin,

kode kecamatan, kode pendidikan akhir, jam kerja perminggu, dan jam kerja pertahun. Pada gambar 3.12 adalah hasil *field* yang telah diseleksi.

Row No.	Jam Kerja P.	Umur	Kode Jenis ...	Kode Keca...	Kode Pendid...	Jam Kerja P.
1	2368	44	1	19	5	48
2	1904	48	2	11	3	27
3	1188	44	2	21	5	24
4	1844	34	1	11	4	32
5	1712	42	2	21	9	40
6	1835	47	1	17	5	37
7	1712	48	2	8	7	48
8	1712	49	1	16	5	40
9	2368	30	1	16	4	48
10	344	31	2	20	3	7
11	1712	38	1	16	4	40
12	1835	35	2	18	8	37
13	2060	46	2	16	8	40
14	1712	33	1	12	8	40
15	1835	45	2	18	5	37
16	1712	25	1	16	5	48
17	1712	20	2	8	5	40

Gambar 3.12 Hasil Seleksi

Jadi total *field* yang akan diproses selanjutnya agar mendapatkan *knowledge* yang dibutuhkan berjumlah 6 *field*, dari total keseluruhan *field* awal yang belum diseleksi berjumlah 13 *field*.

3.1.3 Data Transformation

Data Transformation juga dikenal sebagai data *consolidation*. Pada tahap ini, dimana data-data yang telah terpilih, ditransformasikan kedalam bentuk-bentuk yang cocok untuk prosedur penggalian (*mining procedure*) dengan cara melakukan normalisasi dan agregasi data.

Hasil dari normalisasi dan agregasi data dapat dilihat pada gambar 3.13

Row No.	Jam Kerja P.	outlier	Umur	Kode Jenis.	Kode Keca.	Kode Pendid.	Jam Kerja P.
1	2368	false	44	1	19	5	48
2	1504	false	48	2	11	3	27
3	1198	false	44	2	21	5	24
4	1544	false	34	1	11	4	32
5	1712	false	42	2	21	9	40
6	1835	false	47	1	17	5	37
7	1712	false	48	2	8	7	48
8	1712	false	49	1	16	5	40
9	2368	false	30	1	18	4	48
10	344	false	31	2	20	3	7
11	1712	false	38	1	16	4	40
12	1835	false	35	2	18	8	37
13	2060	false	46	2	16	8	40
14	1712	false	33	1	12	8	40
15	1835	false	45	2	18	5	37
16	1712	false	25	1	16	5	48
17	1712	false	20	2	8	5	40

Gambar 3.13 Hasil Normalisasi Agregasi

Pada gambar 3.13 tersebut terdapat *outlier* atau salah satu teknik agregasi yang terdapat hasil *false* dan *true*. Nilai *false* adalah data yang normal, sedangkan nilai *true* adalah data unnormal maka harus dihapuskan. Pada tahap ini terdapat pengurangan item data, dari data awal berjumlah 3526 item menjadi 3516 item karena pada *outlier* yang bernilai *true* dihapuskan.

Gambar 3.14

Untuk penghapusan *outlier* yang bernilai *true* menggunakan menu *filter example* yang telah disediakan oleh *rapidminer* seperti yang terlihat pada gambar 3.14. Laporan hasil dari penghapusan *outlier* yang bernilai *true* dapat dilihat pada gambar 3.15 dibawah ini.

Row No.	id	Jam Kerja P.	cluster	outlier	Umur	Kode Jenis.	Kode Keca.	Kode Pendid.	Jam Kerja P.
1	1	2368	cluster_2	false	44	1	19	5	48
2	2	1504	cluster_2	false	48	2	11	3	27
3	3	1198	cluster_2	false	44	2	21	5	24
4	4	1544	cluster_2	false	34	1	11	4	32
5	5	1712	cluster_2	false	42	2	21	9	40
6	6	1835	cluster_2	false	47	1	17	5	37
7	7	1712	cluster_2	false	48	2	8	7	48
8	8	1712	cluster_2	false	49	1	16	5	40
9	9	2368	cluster_0	false	30	1	18	4	48
10	10	344	cluster_3	false	31	2	20	3	7
11	11	1712	cluster_2	false	38	1	16	4	40
12	12	1835	cluster_2	false	35	2	18	8	37
13	13	2060	cluster_2	false	46	2	16	8	40
14	14	1712	cluster_0	false	33	1	12	8	40
15	15	1835	cluster_2	false	45	2	18	5	37
16	16	1712	cluster_0	false	25	1	16	5	48

Gambar 3.15 Hasil penghapusan bernilai *true*

Disini juga dapat dilihat item data berkurang dari awal berjumlah 3526 item menjadi 3516 item karena *outlier* yang bernilai *true* telah dihapuskan.

4. Hasil dan Pembahasan

4.1 Hasil Analisis

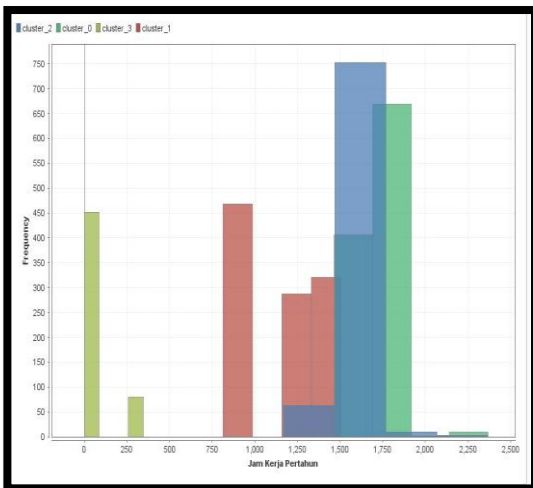
Berdasarkan penelitian yang telah dilakukan, telah didapatkan suatu pola seperti gambar 4.1, informasi dan pengetahuan baru dari proses data mining dalam menentukan dan pengkelompokkan jenis pengangguran yang ada di kabupaten lahat setelah melalui tahapan *knowledge discovery in database (KDD)* dari data penduduk tahun 2015, 2016, dan 2017. Dari penelitian ini telah didapat hasil:

1. Penduduk dengan rata-rata jam kerja pertahun $0 - \leq 600$ jam pertahun dapat disebut sebagai penganggur/tidak memiliki pekerjaan (*cluster 3*).
2. Penduduk dengan rata-rata jam kerja pertahun $\geq 601 - \leq 1300$, digolongkan sebagai setengah menganggur karena jam kerja penduduk tersebut jauh dibawah dari jam kerja normal (*cluster 1*).

3. Penduduk dengan rata-rata jam kerja pertahun ≥ 1301 - ≤ 2000 bisa disebut sebagai penganggur musiman, karena penduduk yang memiliki jam kerja ini terutama terdapat di sektor pertanian dan perikanan. Seperti pada musim hujan penyadap karet dan nelayan tidak dapat melakukan pekerjaan mereka dan terpaksa menganggur (*cluster 2*).
4. Sedangkan, penduduk dengan rata-rata jam kerja pertahun ≥ 2001 - ≤ 2500 disebut penduduk dengan rata-rata jam kerja normal (*cluster 0*).

4.2 Persentation

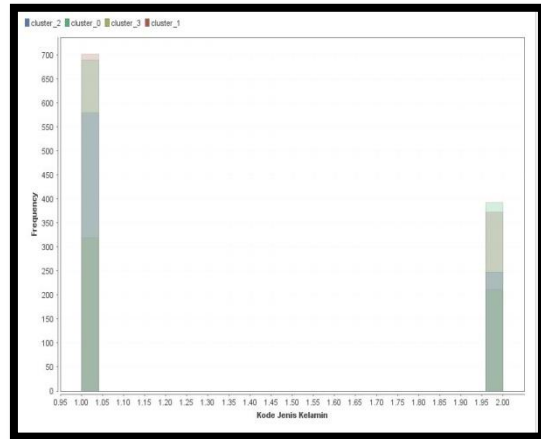
Ini merupakan tahapan terakhir, dimana pengetahuan yang telah ditemukan secara visual ditampilkan kepada user. Hasil pengelompokkan data dari rentan tahun 2015,2016, dan 2017 dengan total record 3516 seperti yang telah dijelaskan pada proses sebelumnya. Maka didapat hasil pengetahuan yang bisa dilihat pada gambar dibawah ini.



Gambar 4.2 Hasil keluaran akhir yang berbentuk diagram

Dari gambar 4.2 didapat hasil 3 *cluster*, pada *cluster* penganggur terdapat 530 item, pada *cluster* setengah menganggur 1075 item, pada *cluster* penganggur musiman 828 item, sedangkan pada *cluster* pekerja ada 1083 item.

- Balok biru= *cluster_2*
- Balok hijau = *cluster_0*
- Balok kuning = *cluster_3*
- Balok merah = *cluster_1*

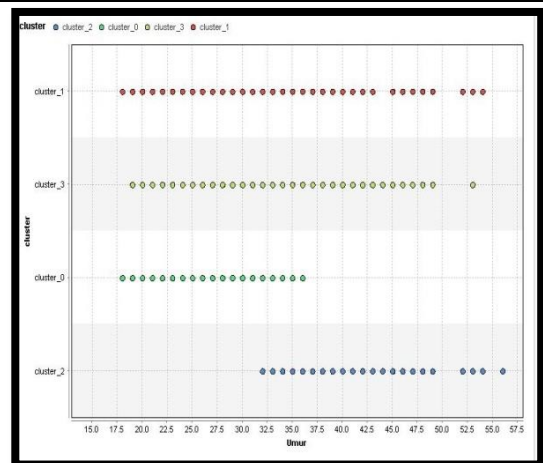


Gambar 4.3 Hasil jenkel

Pada gambar 4.3 adalah diagram dari jenis kelamin mana yang termasuk dalam sebuah *cluster*. Dengan mendapatkan hasil seperti dibawah ini:

Tabel 4.5 Tabel hasil jenkel

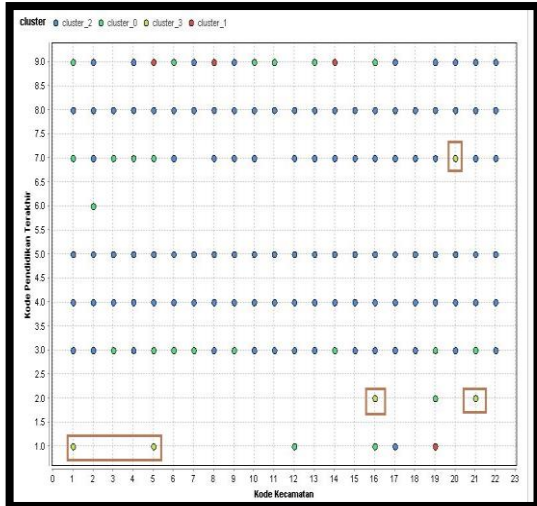
	Laki-laki	Perempuan
<i>Cluster 0</i>	685	398
<i>Cluster 1</i>	700	375
<i>Cluster 2</i>	520	248
<i>Cluster 3</i>	320	210



Gambar 4.4 Hasil pola umur

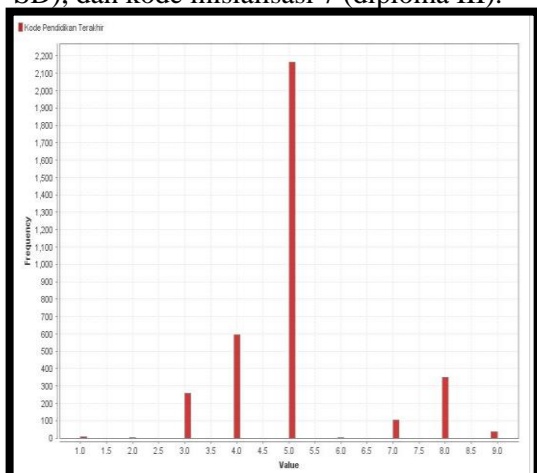
Dalam gambar 4.4 terlihat hasil pola umur yang diperoleh bahwa *cluster 0* dengan jumlah item tertinggi sebanyak 121 item dengan rentan umur 17-36 tahun, *cluster 1* dengan jumlah item tertinggi 162 item dengan rentan umur 17-54 tahun,

cluster 2 dengan jumlah item tertinggi 91 item dengan rentan umur 31-56 tahun, sedangkan pada cluster 3 dengan jumlah item 45 item dengan rentan umur 19-51 tahun.



Gambar 4.5 Hasil keluaran pola pendidikan akhir

Keluaran pola hasil *rapidminer* yang selanjutnya adalah pola pendidikan akhir seperti yang tertera pada gambar 4.5. Titik-titik berwarna kuning yang termasuk kedalam *cluster_3* yang artinya *cluster* penganggur/tidak memiliki pekerjaan. Bisa disimpulkan penganggur/tidak memiliki pekerjaan, pendidikan akhir mereka yaitu dengan kode inialisasi 1 (tidak menempuh pendidikan), kode inialisasi 2 (tidak tamat SD), dan kode inialisasi 7 (diploma III).



Gambar 4.6 Diagram hasil pendidikan akhir

Pada gambar 4.6 adalah diagram hasil pendidikan akhir apa saja yang ditempuh. Disini didapat informasi bahwa pendidikan dengan inialisasi 5 (SMA/SLTA) adalah pendidikan akhir yang paling banyak ditempuh oleh masyarakat kabupaten lahat. Dengan posisi kedua pendidikan terakhir apa saja yang ditempuh yaitu pendidikan akhir dengan inialisasi 4 (SLTP/SMA), selanjutnya posisi ketiga dengan pendidikan akhir dengan inialisasi 8 (S1/Strata1), posisi keempat dengan pendidikan akhir 3 (SD), sedangkan posisi kelima yaitu dengan pendidikan akhir 7 (DII/Diploma III), dan selanjutnya orang yang menempuh pendidikan akhir 9 (S2/Strata 2).

Dari hasil analisis diatas didapatkan penganggur/tidak memiliki pendidikan berjumlah 530 orang, dengan jumlah laki-laki 320 orang dan perempuan berjumlah 210 orang dengan rentan umur 20-50 tahun.

Pada klasifikasi setengah menganggur berjumlah 1075 orang dengan jumlah laki-laki 700 orang dan perempuan berjumlah 375 orang dengan rentan umur 20-42 tahun.

Penganggur musiman berjumlah 828 orang dengan jumlah laki-laki 520 orang dan perempuan berjumlah 248 orang dengan rentan umur 32-53 tahun.

Dan pekerja sebanyak 1083 orang dengan jumlah laki-laki 320 orang dan perempuan 210 orang dengan rentan umur 20-50 tahun.

Pada katagori pendidikan apa saja yang di tempuh penganggur/tidak memiliki pekerjaan adalah orang yang tidak menempuh pendidikan, tidak tamat SD, dan Diploma III. Hasil tersebut didapat dari data penduduk berjumlah 3516 orang.

5. Simpulan

Berdasarkan hasil penelitian dan pembahasan dari bab-bab sebelumnya, dapat ditarik kesimpulan sebagai berikut :

1. Penerapan *data mining* dengan teknik *clustering* dan algoritma *k-means* yang dilakukan yaitu untuk menghasilkan sebuah informasi tentang pengelompokkan jenis-jenis pengangguran menurut kriteria tertentu, jenis kelamin mana yang masuk dalam jenis-jenis pengangguran tersebut, serta rentan umur, dan apa saja jenis pengangguran yang ada.
2. Dari penelitian menggunakan teknik *clustering* ini didapat informasi, bahwa masih banyak penduduk dengan jam kerja jauh dibawah jam kerja normal atau bisa disebut juga sebagai setengah menganggur.

Wahyudi, Bambang. 2004, Manajemen Sumber Daya Manusia, Bandung : Sulita.

DAFTAR PUSTAKA

- Hermawati, Fajar Astuti. 2013, *Data Mining*, Yogyakarta : CV.Andi Offset.
- Indonesia, BPS. *Informasi Umum BPS*.
<https://lahatkab.bps.go.id/index.php/masterMenu/view/id/1#masterMenuTab1>. 15 Januari 2018
- Indonesia, BPS. *Visi dan misi BPS*.
<https://lahatkab.bps.go.id/index.php/masterMenu/view/id/1#masterMenuTab2>. 15 Januari 2018
- Indonesia, BPS. *Struktur Organisasi BPS*.
<https://lahatkab.bps.go.id/index.php/masterMenu/view/id/1#masterMenuTab3>. 15 Januari 2018
- Kadir, Abdul & Terra Ch.Triwahyuni., 2013, *Pengantar Teknologi Informasi Edisi Revisi*, Yogyakarta : Andi Yogyakarta.
- Sukirno, Sadono. 2017, *Makro Ekonomi Teori Pengantar*, Jakarta : PT.Raja Grafindo Persada.
- Vulandari, Retno Tri,S.Si, M.SI., 2017, *Data Mining Teori dan Aplikasi Rappidminer*, Yogyakarta : Gawa Media.