

IMPLEMENTASI *UNSUPERVISED LEARNING* PADA NILAI JASMANI KESAMAPTAAN SEKOLAH POLISI NEGARA DENGAN METODE *CLUSTERING ANALYSIS*

Zaenal Muttaqin¹, Donny Fernando², Selvia Sulastriani³

Program Studi Sistem Informasi, Fakultas Teknologi Informasi, Universitas Serang Raya
Jl. Raya Serang-Cilegon Km. 5 Kp. Drangong Kelurahan Taktakan

E-mail: *d.zaey.vu@gmail.com¹, mr.donny2008@gmail.com², selviasulastriani27@gmail.com³

Abstrak - Agglomerative adalah salah satu dari metode *Hierarchical Analysis Clustering* (HAC) yang merupakan bagian dari konsep *Unsupervised Learning* dengan tujuan untuk mengelompokkan data-data dari hasil perhitungan terhadap variabel yang ada. Dalam metode *Agglomerative* terdapat proses perhitungan jarak seperti *Manhattan* dan *Euclidean* untuk mengukur kesamaan (*similarity*) antar objek. Metode hierarki yang digunakan dalam penelitian ini yaitu metode *Single Linkage*, *Average Linkage* dan *Complete Linkage*. Pengelompokan Nilai Kesamaptan diperlukan untuk dapat dilakukan penugasan sesuai dengan kelompok nilai yang telah ditentukan. Setiap nilai pada variabel di analisa melalui perpaduan perhitungan metode jarak dengan setiap metode hierarki. Diperoleh hasil bahwa metode perhitungan jarak Manhattan dengan metode *Single Linkage* lebih variatif (heterogen) sebanyak 17 klaster, sedangkan untuk *Average Linkage* kombinasi dengan perhitungan jarak Manhattan menghasilkan 24 klaster. Analisa terakhir diperoleh dari metode *Complete Linkage* kombinasi dengan perhitungan jarak Euclidean menghasilkan 24 klaster.

Kata Kunci: *Agglomerative, Clustering, Euclidean, Manhattan, Nilai Kesamaptan,*

I. PENDAHULUAN

Pada era digitalisasi saat ini semakin banyak aktifitas yang dilakukan oleh setiap organisasi baik institusi pemerintahan maupun swasta disegala bidang dengan memanfaatkan Teknologi Informasi (TI). Setiap sistem informasi yang digunakan secara otomatis akan berkaitan dengan penyimpanan suatu pangkalan data yang dikenal dengan database. Data yang tersimpan dalam database diperoleh dari transaksi proses bisnis yang disimpan dengan baik dan merupakan salah satu asset bagi organisasi tersebut.

Perkembangan data saat ini semakin meningkat dan tidak dapat dihindari. Berdasarkan pada perkembangan zaman, teknologi database pun berubah dan meningkat sesuai dengan kebutuhannya. Data sendiri pun terbagi menjadi dua, yaitu data terstruktur (*structured*) dan data tidak terstruktur (*unstructured*). Data terstruktur merupakan data yang dapat diproses, disimpan, diambil dalam format tetap dan biasanya memiliki informasi yang lengkap. Sedangkan data tidak terstruktur tidak memiliki bentuk dan struktur khusus dan memiliki sifat kurang lengkap (Adinugroho, 2018). Secara sifat data terbagi menjadi dua jenis, yaitu data kuantitatif dan data kualitatif. Data kualitatif adalah data yang disajikan dalam bentuk verbal (lisan/tulisan), bukan dalam bentuk angka dan tidak terukur, sedangkan data kuantitatif adalah data yang disajikan dalam bentuk

angka dan terukur yang informasi dapat diperoleh dalam bentuk angka (Lubis, 2015).

Sekolah Polisi Negara (SPN) merupakan lembaga pendidikan yang melaksanakan dan mengemban tugas dalam pembentukan personil kepolisian yang ditugaskan untuk memelihara keamanan dan ketertiban umum di Negara Kesatuan Republik Indonesia. Salah satu SPN yang ada di Indonesia salah satunya adalah SPN yang berlokasi di Kampung Tapos, Desa Karung Kambing, Kecamatan Mandalawangi, Kabupaten Pandeglang, Provinsi Banten, kode pos 42261.

Salah satu kegiatan pembentukan personil kepolisian bagian kesamaptan adalah kegiatan ujian jasmani yang mencakup lari 250 meter, pull up, push up, sit up dan shuttle run. Data hasil nilai yang sudah dikelompokkan akan diketahui kelompok siswa mana yang memiliki nilai ujian fisik lebih besar dibandingkan dengan siswa yang nilai ujian fisiknya lebih kecil. Siswa dengan nilai ujian fisik yang besar akan mendapatkan binaan khusus untuk melatih kekuatan fisiknya, dimana kemungkinan besar akan dimasukan ke Satuan Brimob, sedangkan untuk kelompok siswa yang ujian fisik kesamaptaannya lebih kecil akan diposisikan masuk kedalam kesatuan Polda ataupun Polres.

Saat ini proses pengolahan nilai siswa di SPN masih dilakukan secara manual dalam bentuk form penilaian yang ditulis oleh tim penilai. Selanjutnya form penilaian tersebut di kumpulkan dan di input secara terkomputerisasi dengan menggunakan aplikasi Microsoft Excel. Data yang telah di input

tersebut selanjutnya direkapitulasi dan dilakukan sortir data secara *ascending* untuk melihat dan mengukur nilai ujian fisik dari yang tertinggi sampai dengan yang terendah. Setelah mendapatkan hasil sort data selanjutnya tim pengambil keputusan akan mengelompokkan siswa-siswa yang akan ditugaskan ke Brimob, Polda atau Polres secara manual oleh pihak pimpinan.

Hal ini akan mengakibatkan terjadinya penilaian secara subjektif oleh pihak pimpinan terhadap peserta ujian dan mengakibatkan terjadinya kesenggangan antar peserta yang memiliki kedekatan dengan pihak internal SPN. Selain itu dengan dilakukannya pengambilan keputusan pengelompokan oleh pimpinan akan memakan waktu yang lama dalam pengambilan keputusannya yang berakibat kepada ketidakefektifan kerja.

Berdasarkan hal tersebut diatas, maka perlu dilakukan suatu pengelompokan nilai berdasarkan data-data yang dengan menerapkan *Unsupervised Learning* dengan metode *Hierarchical Clustering*. Pada dasarnya *Hierarchical Clustering* terbagi menjadi 2 model, yaitu *Agglomerative Hierarchical Clustering* (HAC) dan *Divisive Hierarchical Clustering* (DHC). Adapun proses perhitungan jarak (*metrics*) yang digunakan yaitu *Manhattan* (*Manhattan Block*) dan *Euclidean*.

Tujuan dari penelitian yang dilakukan adalah :

1. Bagaimana mengelola suatu data dari hasil tes jasmani kesamaptaaan di Sekolah Polisi Negara untuk dapat dikelompokkan berdasarkan pada nilai kesamaptaaan jasmani dengan menggunakan metode *Agglomerative Hierarchical Clustering* pada Sekolah Polisi Negara (SPN) Mandalawangi
2. Bagaimana pengambilan keputusan pembinaan khusus bagi siswa dari hasil pengelompokan nilai jasmani dengan nilai tertinggi .

II. TINJAUAN PUSTAKA

Kajian Pustaka merupakan gambaran dari teori dasar yang relevan, yang sesuai dengan bidang atau topik yang dijadikan bahan penelitian. Istilah kajian pustaka dapat diartikan secara langsung dari literature review, namun bagian ini tidak hanya meninjau pada bagian permukaan saja melainkan lebih masuk kedalam, hal itu dilakukan agar bisa melihat secara luas isi dari penelitian sebelumnya bisa melakukan evaluasi dari isi pustaka yang digunakan.

Fungsi dari kajian Pustaka itu sendiri yaitu untuk meningkatkan pemahaman dan memiliki akses terhadap database informasi ilmiah yang berhubungan dengan topik yang sedang diteliti. Dapat juga memberikan landasan teori terhadap penelitian sehingga dapat menunjukkan posisi penelitian dibandingkan dengan penelitian yang sudah dilakukan oleh peneliti sebelumnya, untuk

mengidentifikasi informasi, ide, teknik, metode, yang berhubungan dengan topik penelitian.

Penelitian yang akan dilaksanakan dan dikaji berdasarkan penelitian-penelitian sebelumnya diantaranya :

Anang Juni Pangestu Teknologi informasi-Teknik Informatika dari Universitas Nusantara PGRI Kediri dengan judul jurnal "Pembuatan kelompok belajar siswa Sekolah Dasar menggunakan metode *Agglomerative Hierarchical Clustering*" dalam jurnalnya menjelaskan bahwa Metode *Agglomerative Hierarchical Clustering* dapat menentukan anggota kelompok berdasarkan nilai jarak dari setiap siswa, jadi kelompok yang terbentuk memilikinilai rata-rata yang hampir sama. Peneliti lain Widyanti, wawan lakso noyuly saptomo, yustina wahyuretnoutami Program studi sistem informasi, STMK Sinar Nusantara Surakarta Jurnal Ilmiah Sinus (JIS) vol : 18, no : 1, ISSN print : 1693-1173, dalam jurnalnya yang berjudul "Penerapan *Agglomerative Hierarchical Clustering* Untuk Segmentasi Pelanggan" menjelaskan bahwa *Agglomerative Hierarchical Clustering* dapat digunakan untuk pengelompokan pelanggan dan dikaitkan dengan strategi pemasaran sehingga terciptanya sistem *Agglomerative Hierarchical Clustering* untuk *clustering customer* dan membantu perusahaan dalam menentukan strategi pemasaran yang tepat sesuai dengan jenis *customer* terhadap profitabilitas perusahaan.

Febri Liantoni, Laila Cahyani Institut Teknologi Adhi Tama Surabaya, Indonesia Jurnal ilmiah edutic / Vol.3, No.2 P-ISSN 2407-4489, E-ISSN 2528-7303 dalam jurnalnya yang berjudul "Pemanfaatan *Hierarchical Clustering* untuk pengelompokan daun berdasarkan fitur moment invariant" menjelaskan bahwa Uji coba dengan menggunakan metode centroid linkage clustering didapatkan nilai akurasi clustering sebesar 87%, sedangkan dengan menggunakan metode k-means didapatkan nilai akurasi *clustering* sebesar 81%. Kinerja metode clustering *linkage clustering* lebih baik dibandingkan k-means. Selain itu Penggunaan metode *seven moment invariant* dapat dibuktikan bahwa perubahan posisi, rotasi dan skala tidak begitu berpengaruh pada nilai fitur *invariant* yang dihasilkan.

1. Data Mining

Data mining adalah proses mengetahui sesuatu yang bermanfaat dengan kesesuaian baru, cara dan gaya yang ada dengan cara pemilihan data yang mempunyai ukuran besar. Data tersebut disimpan didalam gudang menggunakan teknologi sosialisasi cara serta statistik dan teknik metematik. Analisis pemantauan data base dalam memilih hubungan yang tidak diduga juga untuk merangkum data dengan pola atau dengan metode baru yang dapat dipahami, bermanfaat untuk pemilik data. (Larose : 2005)

Data mining memiliki fungsi yang begitu penting dalam mendapatkan informasi, yang berhasil dan menambah pengetahuan kegunaannya. Data mining memiliki empat fungsi dasar, yaitu :

- a. *Prediksi (Prediction)*
Peroses untuk menentukan cara dari data dengan menggunkan beberapa variabel untuk memprediksi variable lainnya yang tidak diketahui jenis atau nilainya.
- b. *Deskripsi (Description)*
Peroses untuk menentukan suatu cara untuk menentukan suatu cirri khusus dari data yang terdapat didalam suatu database.
- c. *Klasifikasi (Classification)*
Proses untuk menentukan fungsi agar dapat menggambarkan kelompok atau konsep dari suatu data. Peroses yang digunakan untuk menggambarkan data tersebut merupakan hal penting serta bisa memprediksi kecenderungan data pada masa mendatang.
- d. *Asosiasi (Association)*
Peroses yang dipakai untuk menentukan suatu hubungan yang terdapat pada nilai atribut dari sekumpulan data.

2. Klustering (*Clustering*)

Clustering adalah peroses pengelompokan objek ke dalam class dengan mengetahui kesamaan objek. Sebuah cluster adalah kumpulan data yang mempunyai persamaan satu dengan yang lain, dalam cluster yang mempunyai kesamaan dan memiliki perbedaan dengan cluster yang lain. *Clustering* merupakan pembelajaran yang tidak dapat ditebak karena tidak dibutuhkan target output atau yang sering disebut *unsupervised learning*.

Clustering dibagi menjadi 2 metode yaitu: *hierarchical* dan *non-hierarchical*. *Hierarchical* adalah metode pengelompokan data yang diawali dengan mengelompokkan beberapa objek yang memiliki kesamaan paling dekat, kemudian dilanjutkan kepada objek lain yang memiliki jarak yang lebihdekat, sehingga cluster akan membentuk hirarki dari yang paling mirip dan yang tidak mirip.

3. *Hierarchical Agglomerative Clustering (HAC)*

HAC (Hierarchical Agglomerative Clustering) adalah metode pengelompokan data yang menggabungkan dua buah klaster yang mempunyai kemiripan. Pada metode ini jumlah klaster tidak ditentukan secara spesifik. Pada beberapa kasus penentuan jumlah klaster juga sering dilakukan dengan menggabungkan metode lain didalamnya. *HAC* dapat diartikan juga suatu algoritma klaster yang dapat digunakan untuk mengklaster dokumen. *HAC* sering dipakai jika belum adanya informasi jumlah kelompok, pengelompokan bersifat *divisive (top to down)*, yang bermakna satu klaster sama tapi menjadi sebuah klaster, atau bersifat *agglomerative (bottom up)*, yang bermakna sebuah data menjadi sebuah klaster. Teknik hirarki (*hierarchical methode*), yaitu teknik klaster yang berbentuk

kontruksi hirarki atau berdasarkan tingkatan tertentu seperti struktur pohon. Penjelasan diatas dapat disimpulkan peroses pengelompokan dilakukan secara tingkatan atau bertahap.

Teknik pengelompokan hirarki (*Single Linkage, Complete Linkage, Average Linkage*), dibedakan berdasarkan konsep jarak antar kelompok. Penentuan jarak antar kelompok tersebut yaitu :

a. Metode *Single Linkage*

Untuk pengelompokan data berdasarkan pendekatan ini digunakan persamaan sebagai berikut:

$$\min_{j \in J} (dik, djk)$$

Berdasarkan data jarak yang sudah dihitung dan memperlakukan setiap objek data sebagai single klaster, pilih jarak dua kelompok terkecil dengan mengelompokan masing-masing dua buah data yaitu :

$$d(i,j)k = \min_{j \in J} (dik, djk)$$

b. Metode *Complete Linkage*

Untuk mengelompokan data berdasarkan pendekatan complete linkage, digunakan persamaan sebagai berikut:

$$\max_{j \in J} (dik, djk)$$

Berdasarkan data jarak yang sudah dihitung dan memperlakukan setiap objek data sebagai single klaster, pilih jarak dua kelompok terbesar dengan mengelompokan masing-masing dua buah data yaitu:

$$d(i,j)k = \max_{j \in J} (dik, djk)$$

c. Metode *Average Linkage*

Untuk mengelompokan data berdasarkan pendekatan average linkage, digunakan persamaan sebagai berikut :

$$d(i,j)k = \text{average}_{j \in J} (dik, djk)$$

Berdasarkan data jarak yang sudah dihitung dan memperlakukan setiap objek data sebagai single klaster, pilih jarak dua kelompok terbesar dengan mengelompokan masing-masing dua buah.

Pada metode *HAC (Hierarchical Agglomerative Clustering)* terdapat beberapa metode didalamnya. Pada penelitian ini digunakan metode *Single Linkage*, karena lebih cocok digunakan dalam mengelompokan siswa yang memiliki fisik lebih dari pada siswa yang lainnya. Metode *Single Linkage* yaitu mengukur jarak terdekat antara data nilai siswa satu dengan siswa lainnya. Jadi akan lebih mudah dan lebih akurat dalam mengelompokan.

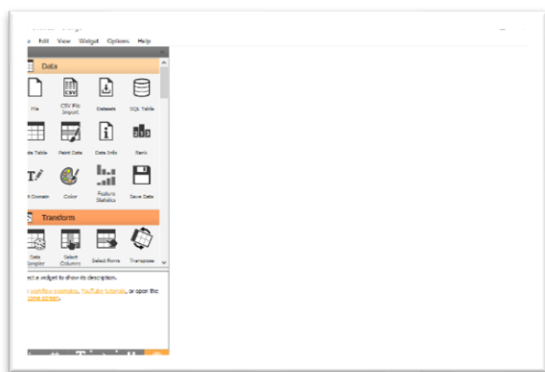
Selain itu pada penelitian ini juga sudah mencoba melakukan pengelompokan dengan metode *Complete Linkage*, karena metode *Complete Linkage* itu digunakan untuk mengukur jarak yang terbesar. Dapat juga diartikan cara mengukur jarak data dari

yang terjauh, maka akan lebih lama dalam mengelompokkan nilai siswanya. Oleh karena itu pada penelitian ini digunakan metode *Single Linkage*, *Average Linkage* dan *Complete Linkage*.

4. Aplikasi Data Mining ORANGE

Data mining hadir sebagai bidang yang berusaha mengolah data, yang saat ini semakin banyak dihasilkan dan tersimpan, menjadi suatu pengetahuan yang bermanfaat serta berharga. Implementasi dari teknik-teknik data mining umumnya dilakukan dengan menggunakan bahasa pemrograman tertentu yang menuntut *coding skill* seseorang.

Orange atau dikenal juga dengan Orange Data Mining merupakan perangkat lunak open source untuk melakukan proses data mining atau data analytic melalui konsep *visual programming*. Aplikasi Orange digunakan untuk menganalisa data melalui 4 (empat) tahapan yaitu pemanggilan file data, menampilkan tabel, menentukan metode pengukuran jarak, dan menampilkan hasil analisa.



Gambar 1. Aplikasi Orange User Interface

III. METODE PENELITIAN

Ada beberapa tahap penelitian yang dilakukan adalah: studi literatur, observasi, analisis/ implementasi metode *Unsupervised Learning*, *Data Visualization* dan kesimpulan. Pada tahap studi literatur dilakukan studi kepustakaan yaitu proses mengumpulkan bahan referensi mengenai data mining, *unsupervised learning*, *clustering* dan Hierarchical Agglomerative Clustering (HAC) dari berbagai buku, jurnal, artikel, dan beberapa referensi lainnya.

Selanjutnya adalah observasi, yaitu merupakan suatu teknik yang digunakan untuk mengumpulkan data dengan cara melakukan pengamatan secara langsung ke tempat yang dijadikan objek penelitian. Adapun permasalahan yang akan dijadikan bahan penelitian di Sekolah Polisi Negara. Hal ini dilakukan agar mendapatkan informasi, data yang dibutuhkan dalam penelitian mengetahui proses apa saja yang ada pada instansi tersebut.

Setelah observasi dilakukan analisis data hasil ujian jasmani dengan menggunakan metode Unsupervised Learning Clustering Analysis. Setelah

dianalisis maka grafik yang dihasilkan akan ditampilkan dalam bentuk Visualisasi Data. Berdasarkan dari hasil pengolahan dan visualisasi data tersebut petugas penilai dapat melakukan kesimpulan dan keputusan kepada siswa kesamptaan.

A. Jenis Data

Dalam penelitian ini menggunakan jenis data primer. Data yang diperoleh langsung dari pihak yang bersangkutan dalam penelitian ini, yaitu bagian Urminpres Subbagremin. Yakni data hasil nilai kesemptaan jasmani tahun 2019/2020. Adapun data sekunder diperoleh dari beberapa literatur berupa buku ataupun jurnal dengan jenis penelitian yang serupa untuk mendukung penelitian ini dengan berbagai teori.

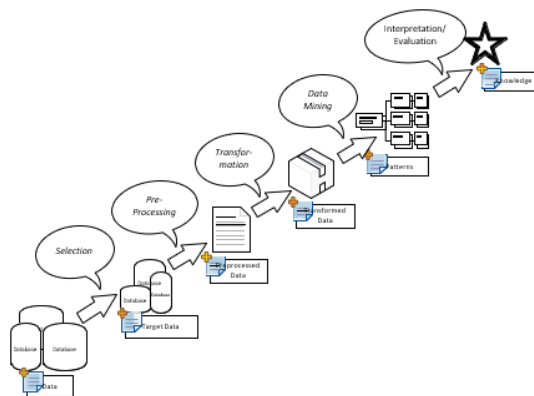
B. Dataset

Hasil ujian kesemptaan jasmani siswa tahun angkatan 2019/2020 pada Sekolah Polisi Negara sebagai berikut :

Tabel 1. Data Peserta Kesamptaan

No	Nama	lari 250 mtr	pull-up	push-up	Sit-up	shuttle run
1	Dicki andhika setyawan	96	70	100	100	78
2	Muhamad alfiyaser	87	88	100	100	68
3	Romdhani	79	100	100	100	70
4	Fajar rafli	75	100	100	100	82
5	Noviandi ananda	74	100	100	100	82
6	Muhammad bama	83	70	100	100	66
7	Arman suwondo	80	58	100	100	90
8	Muhamad adrianto	83	58	100	100	78
9	Yudi pratama	87	46	100	100	70
10	Muhamad saeful milah	72	70	100	100	100
11	Muhamad sufiyan	75	76	100	97	78
12	Panji galuh sadewo	81	82	100	100	40
...						
100	David jovani sembinging	68	58	77	85	66

C. Knowledge Discovery in Databases (KDD)



Gambar 2. Knowledge Discovery in Databases

Berdasarkan gambar di atas, proses penemuan pengetahuan tersebut merupakan urutan berulang dari langkah-langkah sebagai berikut:

a. *Data Cleaning and Integration*

Tren yang populer di industri informasi adalah melakukan pembersihan data dan penggabungan data sebagai langkah preprocessing, di mana data yang dihasilkan disimpan dalam data warehouse

b. *Data Selection and Transformation*

sebelum proses seleksi data, khususnya dalam kasus warehousing (pergudangan data). Data reduction (pengurangan data) juga dapat dilakukan untuk mendapatkan representasi yang lebih kecil dari data asli tanpa mengorbankan integritasnya.

c. *Data Mining*

Data mining (penambangan data) merupakan proses penting di mana metode cerdas diterapkan untuk mengekstrak pola data.

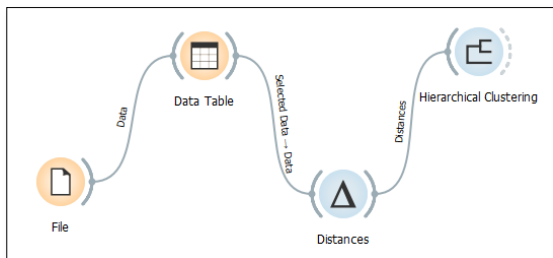
d. *Evaluation and Presentation*

Pattern Evaluation (Evaluasi pola) berfungsi untuk mengidentifikasi pola-pola yang benar-benar menarik yang mewakili pengetahuan berdasarkan ukuran-ukuran yang menarik. Sedangkan Knowledge presentation (Presentasi pengetahuan) digunakan untuk memvisualisasikan dan merepresentasikan pengetahuan yang ditambang kepada pengguna.

IV. HASIL DAN PEMBAHASAN

Proses klusterisasi hierarki terhadap data yang tersedia menggunakan kombinasi antara aplikasi ms. Excel dan Orange Data Mining V.3.32. Ms. Excel digunakan untuk membuat daftar nama-nama dari setiap peserta Kesamaptaan sebanyak 100 peserta.

Aplikasi Orange digunakan untuk menganalisa data melalui 4 (empat) tahapan yaitu pemanggilan file data, menampilkan tabel, menentukan metode pengukuran jarak, dan menampilkan hasil analisa (gambar 2).



Gambar 3. Tahapan Analisa Data

Data yang akan dianalisa di panggil dari ms. Excel dalam format *xlsx dan dibaca pada bagian File. Dalam menampilkan kumpulan data yang telah di panggil di bagian File disediakan fasilitas Data Table. Proses selanjutnya yaitu menentukan metode

pengukuran jarak terhadap variabel-variabel yang tersedia.

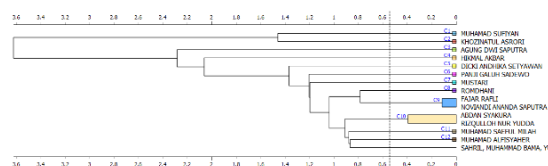
Dalam proses analisa data digunakan dua metode perhitungan jarak, yaitu Manhattan dan Cosine. Setiap hasil dari perhitungan jarak dilakukan analisa metode hierarki single linkage, complete linkage dan average linkage. Standarisasi yang digunakan untuk menghitung perbandingan antara perhitungan jarak Manhattan dan Euclidean berikut dengan metode hierarkinya adalah Annotation “Nama”, Pruning “Max Depth 10” dan Selection “Height Ratio 15%” seperti yang ditunjukkan pada gambar 3.

	NAMA	NILAI LARI	NILAI PULL-UP	NILAI PUSH-UP	NILAI SIT-UP	NILAI SHUTTLE RUN
1	DICKI ANDHUK...	95	70	100	100	79
2	MUHAMMAD AL...	87	58	100	100	68
3	ROMDHANI	79	100	100	100	70
4	FALAH RAFU	75	100	100	100	82
5	NOVIANDI ANL...	74	100	100	100	82
6	MUHAMMAD ...	83	70	100	100	66
7	ANKHAN SUWID...	80	58	100	100	90
8	MUHAMMAD AD...	83	58	100	100	78
9	YUDI PRATAMA	87	46	100	100	70
10	MUHAMMAD SA...	72	70	100	100	100
11	MUHAMMAD SIL...	75	76	100	100	97
12	PANI GALUH S...	81	82	100	100	40
13	YOSSI ADIVANS...	74	58	100	100	87
14	DEDE RIANTO	73	70	100	100	78
15	DIMAS AGUS R...	72	70	100	100	82
16	ANDONI JOHA...	80	39	100	100	80
17	MARDANI	76	46	100	100	87
18	AULIA RAHMAN	75	68	100	100	68
19	RONALDI RAH...	75	64	100	100	70
20	AHMAD ZAINU...	74	70	100	100	66
21	PURNAMA PUT...	73	58	100	100	82
22	KHORRA AZIZ	72	58	100	100	80

Gambar 4. Analisa Data

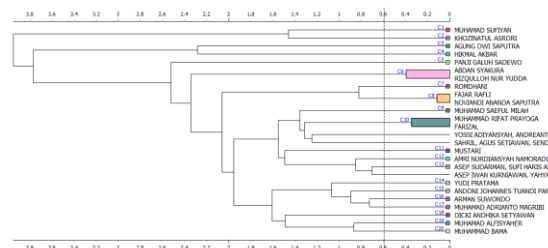
A. Implementasi *Manhattan Metrics Distance*

Dalam mengimplementasikan perhitungan jarak dengan metode Manhattan maka diperoleh hasil seperti pada gambar dibawah berikut.



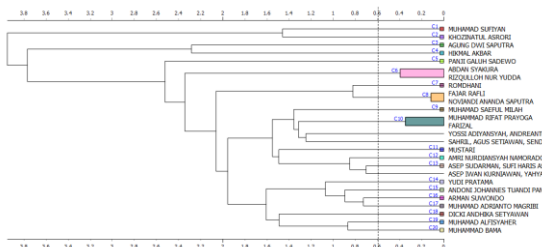
Gambar 5. Dendrogram Manhattan – Single Linkage

Perhitungan jarak Manhattan dengan metode hierarki Single Linkage diperoleh hasil sebanyak 17 kluster (gambar 5).



Gambar 6. Dendrogram Manhattan – Average Linkage

Sedangkan untuk metode hierarki Average Linkage diperoleh hasil sebanyak 24 kluster (gambar 6).

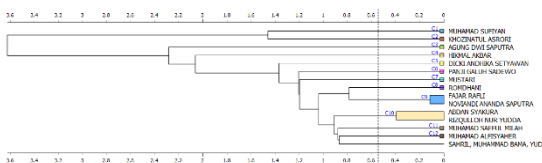


Gambar 7. Dendrogram Manhattan – Complete Linkage

Hasil dari metode hierarki Complete Linkage diperoleh hasil sebanyak 20 kluster (gambar 7).

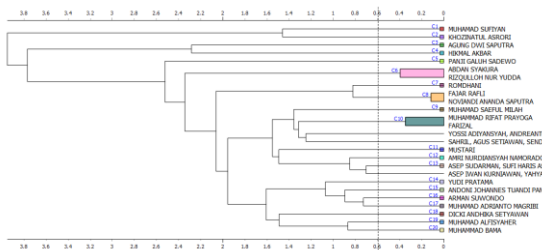
B. Implementasi Euclidean Metrics Distance

Sedangkan dalam mengimplementasikan perhitungan jarak dengan metode Euclidean maka diperoleh hasil seperti pada gambar dibawah berikut.



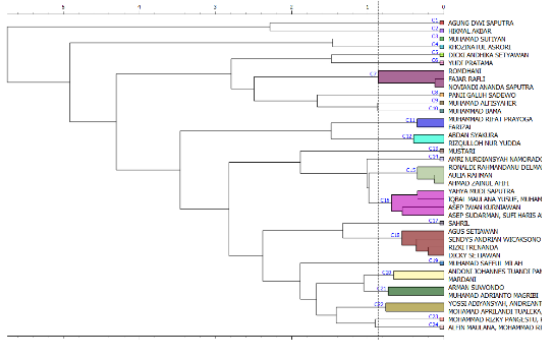
Gambar 8. Dendrogram Euclidean – Single Linkage

Perhitungan jarak Euclidean dengan metode hierarki Single Linkage diperoleh hasil sebanyak 12 kluster (gambar 8).



Gambar 9. Dendrogram Euclidean – Average Linkage

Sedangkan untuk metode hierarki Average Linkage diperoleh hasil sebanyak 20 kluster (gambar 9).



Gambar 10. Dendrogram Euclidean – Complete Linkage

Hasil dari metode hierarki Complete Linkage diperoleh hasil sebanyak 24 kluster (gambar 10).

V. KESIMPULAN

Berdasarkan hasil dari perhitungan maka dapat disimpulkan sebagai berikut :

1. Hasil dari metode hierarki Single Linkage dengan metode pengukuran jarak Manhattan lebih variatif dan kompleks dalam mengelompokkan Samapta berdasarkan nilai yang diperoleh yaitu 17 kluster atau selisih 4 kluster lebih banyak dibandingkan pengukuran jarak Euclidean.
2. Hasil dari metode hierarki Average Linkage dengan metode pengukuran jarak Manhattan lebih variatif dan kompleks dalam mengelompokkan Samapta berdasarkan nilai yang diperoleh yaitu 24 kluster atau selisih 4 kluster lebih banyak dibandingkan pengukuran jarak Euclidean.
3. Hasil dari metode hierarki Complete Linkage dengan metode pengukuran jarak Euclidean lebih variatif dan kompleks dalam mengelompokkan Samapta berdasarkan nilai yang diperoleh yaitu 24 kluster atau selisih 4 kluster lebih banyak dibandingkan pengukuran jarak Manhattan.

DAFTAR PUSTAKA

Adinugroho S, Arum Sari Y, (2018), *Implementasi Data Mining Menggunakan WeKa*, UB Press Malang.

Afrita Lubis A, dkk, *Kajian Metode Pengklasteran Hierarki Dengan Berbagai Pengukuran Jarak*. *E-Jurnal Statistika Sigma Hu Rhu*

Aggarwal, C. Charu, (2015), *Data Mining – Textbook, Springer International Publisher*. ISBN 978-3-319-14142-8 (eBook)

Irwansyah E, Faisal M, (2015). *Advanced Clustering – Teori dan Aplikasi*, Deepublish Publisher, Yogyakarta

Nishom M, (2019), *Perbandingan Akurasi Euclidean Distance, Minkowski Distance dan Manhattan Distance Pada Algoritma K-Means Clustering Berbasis Chi-Square*. *Jurnal Informatika : Jurnal Pengembangan IT (JPIT)*. DOI: 10.30591/jpit.v4i1.1253

Praman Setia, dkk (2018), *Data Mining Dengan R – Konsep Serta Implementasi*, Penerbit IN-Media, Bogor

Suherman E, dkk, (2019), *Agglomerative Hierarchical Clustering Dengan Berbagai Pengukuran Jarak Dalam Mengkluster Daerah Berdasarkan Tingkat Kemiskinan*, *Prosiding Annual Research Seminar (ARS)*. ISBN : 978-979-587-846-9.