

Identifikasi Outlier Menggunakan Teknik Data Mining Clustering Untuk Analisis Data Tracer Study Pada Fakultas Ilmu Komputer Universitas Esa Unggul

Vrantika Br Samosir¹ Agung Mulyo Widodo², Nizirwan Anwar³ Binastya Anggara Sekti⁴, Nixon Erzed⁵

^{1,2,3,4,5}Program Studi Teknik Informatika Fakultas Ilmu Komputer Universitas Esa Unggul

E-mail: vrantikaborusamosir@student.esaunggul.ac.id¹, agung.mulyo@esaunggul.ac.id², nizirwan.anwar@esaunggul.ac.id³, anggara@esaunggul.ac.id, nixon@esaunggul.ac.id

ABSTRAK

Mengidentifikasi Outlier dalam analisis suatu data sangatlah penting. Terlebih data tersebut jenis data penting dalam suatu Universitas. Dengan mengidentifikasi Outlier pada suatu data tracer study juga penting, Karena Pihak Universitas dapat mengetahui kenapa data tersebut ada data pencilan atau data yang menyimpang. Mengidentifikasi Outlier menggunakan Teknik Data Mining Clustering Agoritma K-Means, dapat mengetahui informasi yang akurasi. Karena pengelolaha data nya menggunakan aplikasi Microsoft Exel serta untuk pengujian system dalam pengelolahannya dapat dilakukan di aplikasi Weka. Yang dimana hasilnya sudah langsung dapat diketahui, data tersebut akan di cluster sesuai dengan bidang pekerjaan yang sama dan posisi jabatan yang berkaitan. Yang dimana hasilnya dari Mengidentifikasi Outlier menggunakan Teknik Data Mining Algoritma K-Means, juga menambah informasi, bahwa lulusan Fakultas Ilmu Komputer Universitas Esa Unggul lebih banyak bekerja pada bidang pekerjaan Jasa Outsourching dengan Posisi Jabatan berkaitan dengan Quality Assurance Specialist.. Yang dimana presentasenya sebanyak 39%. Dan lulusan yang bekerja dibidang pekerjaan Industri Jasa Pengiriman Barang dengan Posisi Jabatan Proses Pengiriman Barang atau Kurir. Yang dimana presentase nya sebanyak 9%.

Kata kunci : Tracer Study, Miersoft Exel, Outlier, Weka, Algorima K-Means.

ABSTRACT

Identifying Outliers in the analysis of a data is very important. Moreover, the data is an important type of data in a university. Identifying outliers in a tracer study data is also important, because the university can find out why the data has outliers or distorted data. Identifying Outliers using the K-Means Algorithm Data Mining Clustering Technique, can find out accurate information. Because the data manager uses the Microsoft Exel application and for system testing the management can be done in the Weka application. Where the results can be known immediately, the data will be clustered according to the same field of work and related job positions. Which results from Identifying Outliers using the K-Means Algorithm Data Mining Technique, also adds information, that graduates of the Faculty of Computer Science, University of Esa Unggul work more in the field of Outsourcing Services work with Positions related to Quality Assurance Specialist.. Which percentage is 39 %. And graduates who work in the field of Freight Forwarding Services Industry with Positions in the Delivery of Goods Process or Courier. Which is where the percentage is 9%.

Keyword : Tracer Study, Microsoft Excel, Outlier, Weka, K-Means Algorithm.

1. PENDAHULUAN

Pemahaman kualitas diri dan penyesuaian terhadap dunia kerja menjadi hal mutlak dan mendasar yang harus dimiliki didalam sebuah universitas pendidikan tinggi diseluruh perguruan tinggi di Indonesia. Setiap universitas dituntut untuk memberikan pendidikan dan pelatihan yang di selenggarakan sesuai dengan perkembangan masa yang sejalan dan selaras tuntunan kompetensi yang dibutuhkan didalam dunia kerja.

Berkaitan dengan dunia kerja, *Tracer Study* melacak proses transisi data lulusan mahasiswa yang telah bekerja dan yang pernah bekerja. Dengan mengelompokkan data sesuai kemiripan data pada data bidang pekerjaan dan posisi jabatan lulusan yang berkaitan. Dan mengetahui apakah terdapat data *outlier* pada data atau tidak. Mengidentifikasi *Outlier* dalam analisis suatu data sangatlah penting. Terlebih data tersebut jenis data penting dalam suatu Universitas. Dengan mengidentifikasi *Outlier* pada suatu data *tracer study* juga penting, Karena Pihak Universitas dapat mengetahui kenapa data tersebut ada data pencilan atau data yang menyimpang. Mengidentifikasi *Outlier* menggunakan Teknik Data Mining Clustering Algoritma K-Means, dapat mengetahui informasi yang akurasi. Karena pengelolaha data nya menggunakan aplikasi Microsoft Exel serta untuk pengujian system dalam pengelolahannya dapat dilakukan di aplikasi Weka. Yang dimana hasilnya sudah langsung dapat diketahui, data tersebut akan di cluster sesuai dengan bidang pekerjaan yang sama dan posisi jabatan yang berkaitan.

2. LANDASAN TEORI

Pengertian Data

Data adalah suatu informasi mengenai suatu hal yang sudah sering terjadi dan berupa serangkaian angka,

fakta, gambar, tabel, grafik, kata, symbol, huruf dan lainnya yang mengekspresikan suatu pemikiran, kondisi, objek dan situasi (*Nuzulla Agustina*). Dataset adalah serangkaian dari fakta dan observasi yang dimana menjadi sebuah informasi jika dilakukannya analisis terhadap data tersebut. Hasil dari analisis tersebut akan menjadi informasi dan jika dikembangkan akan menjadi sebuah pengetahuan.

Pengertian Data Kualitatif

Data Kualitatif adalah data yang lebih berupa deskriptif . yaitu sebuah data yang tidak berbentuk angka. Biasanya data tersebut dibuat dengan menggunakan simbol, gambar ataupun bentuk verbal lain. Jenis data yang satu ini bisa didapatkan melalui isian Kuesioner dari *Tracer Study*, Observasi dari hasil *Tracer Study*, Studi Literatur, Wawancara dan lain sebagainya. Data kualitatif yang dibutuhkan dalam penelitian ini menggunakan Variabel yang dimana variabelnya terdiri; nama lulusan, tahun lulusan, program studi lulusan, nama perusahaan, nama bidang pekerjaan serta posisi jabatan lulusan bekerja.

Pengertian Data Mining

Data Mining adalah proses menemukan pola yang menarik dan pengetahuan dari data yang berjumlah besar (*Han dan Kamber (2011,36)*). Data mining adalah suatu pencarian dan analysis dari jumlah data yang sangat besar dan bertujuan untuk mencari arti dari pola dan aturan (*Linoff dan Berry (2011,p7)*). Lalu menurut *Connolly dan Begg, (2010)*, data mining adalah suatu proses ekstraksi atau penggalian data yang belum diketahui sebelumnya, namun dapat dipahami dan berguna dari database yang besar serta digunakan untuk membuat suatu keputusan bisnis yang sangat penting. Dan menurut *Vercellis (2009, p77)*, data mining merupakan aktivitas yang menggambarkan sebuah

proses analysis yang terjadi secara iteratif pada database yang besar, yang bertujuan mengekstrak informasi dan *konowledge* yang akurat dan berpotensi berguna untuk *konowledge workers* yang berhubungan dengan pengambilan keputusan dan pemecahan masalah.

Data mining clustering adalah salah satu teknik dalam data mining yang digunakan pada penelitian ini untuk mengelompokkan data ke dalam kelompok-kelompok yang serupa berdasarkan kesamaan karakteristik. Tujuan utama dari data mining clustering adalah menemukan kemiripan data bidang pekerjaan lulusan serta posisi jabatan lulusan yang dijadikan bermacam macam *cluster*.

Pengertian Clustering K-Means

Menurut *Han dan Kamber (2011)*, Clustering adalah proses pengelompokkan kumpulan data menjadi beberapa kelompok sehingga objek di dalam satu kelompok memiliki banyak kesamaan dan memiliki banyak perbedaan dengan objek dikelompok lain. Perbedaan dan persamaannya biasanya berdasarkan nilai atribut dari objek tersebut dan dapat juga berupa perhitungan jarak.

Algoritma K-Means adalah salah satu metode Data Mining yang bersifat tanpa arahan (*unsupervised*). Ada dua jenis data clustering yang sering dipergunakan dalam proses pengelompokkan data yaitu *hierarchical (hirarki)* data clustering dan *non hierarchical (non hirarki)* data clustering. K-Means merupakan salah satu metode data clustering *non hirarki* yang berusaha mempartisi data yang ada ke dalam bentuk satu atau lebih *cluster*/kelompok. dan memudahkan menemukan cluster sesuai kemiripan pada data dalam penelitian ini.

Pengertian Elbow

Dalam konteks clustering, "*elbow*" merujuk pada metode visual yang digunakan untuk membantu menentukan jumlah optimal kelompok atau klaster

dalam data. Metode ini didasarkan pada analisis varian dalam *clustering*. Berikut ini adalah rumus manual menghitung jumlah *cluster* optimal.

$$W(S, C) = \sum_{K=1}^K \sum_{i \in S_k} \|y_i - C_k\|^2$$

Keterangan :

S = K-cluster yang telah terbentuk

y_i = data y pada indeks ke I

C_k = rata rata K-cluster pada nilai k
 $k=1,2,..K$)

Untuk mendapatkan perbandingannya adalah dengan menghitung *SSE (Sum Of Square Error)* dari masing – masing nilai cluster. Semakin besar jumlah cluster K maka nilai *SSE* akan semakin kecil. Jika nilai cluster pertama dengan nilai cluster kedua memberikan sudut dalam grafik atau nilainya mengalami penurunan paling besar maka nilai *cluster* tersebut adalah nilai yang terbaik.

Pengertian Tracer Study

Tracer Study adalah Studi pelacakan jejak lulusan/alumni yang bertujuan untuk mengetahui *outcome* pendidikan dalam bentuk transisi dari dunia pendidikan tinggi ke dunia kerja, output pendidikan yaitu penilaian diri terhadap penguasaan dan perolehan kompetensi, proses pendidikan berupa evaluasi proses pembelajaran dan kontribusi pendidikan tinggi terhadap perolehan kompetensi serta input pendidikan berupa penggalan lebih lanjut terhadap informasi Fakultas Ilmu Komputer lulusan di Universitas Esa Unggul. Kemudian dari informasi yang diperoleh dapat diambil berbagai kebijakan dan Tindakan yang manfaat bagi para lulusan dan pengembangan Lembaga tersebut.

Pengertian Weka

WEKA (Waikato Environment for Knowledge Analysis) adalah perangkat lunak yang menggunakan *algoritma machine learning* untuk melakukan beberapa proses yang berkaitan dengan

system temu kembali informasi atau data mining. *WEKA* merupakan *API Java* yang menyediakan *API* untuk pengolahan dalam data mining yang berbasis *open source (GPL)* dan *berengine JAVA*. *Weka* dikembangkan oleh Universitas *Waikato* di *selandia Baru* dan juga merupakan perangkat lunak gratis yang tersedia dibawah *GNU (General Public License)*.

Pengertian Outlier

Data adalah suatu informasi mengenai suatu hal yang sudah sering terjadi dan berupa serangkaian angka, fakta, gambar, tabel, grafik, kata, symbol, huruf dan lainnya yang mengekspresikan suatu pemikiran, kondisi, objek dan situasi (*Nuzulla Agustina*). *Datasset* adalah serangkaian dari fakta dan observasi yang dimana menjadi sebuah informasi jika dilakukan nya analisis terhadap data tersebut. Hasil dari analisis tersebut akan menjadi informasi dan jika dikembangkan akan menjadi sebuah pengetahuan.

3. METODOLOGI

Metode penelitian penulis ini dilakukan dengan metode kualitatif adalah yang dimana data berupa deskriptif, metode deskriptif dapat diartikan sebagai prosedur atau cara memecahkan masalah penelitian dengan memaparkan keadaan objek yang diteliti sebagaimana adanya. Berdasarkan fakta – fakta yang actual pada saat sekarang. yaitu sebuah data yang tidak berbentuk angka contohnya program studi lulusan, jenis pekerjaan atau posisi jabatan bekerja. Yang diberi kode angka. Jenis data yang satu ini bisa didapatkan melalui isian *Quesioner* dari *Tracer Study*, Observasi dari hasil *Tracer Study* dan Wawancara.

Pengertian Data Cleanning

Data *Cleanning* adalah proses Langkah awal melakukan pengolahan

data, yang dimana, proses data *cleanning* ini adalah proses pembersihan pada data. Yang dimana data lulusan yang tidak lengkap atau kurang lengkap. Atau tahap memperbaiki data yang tidak sesuai atau salah ketik seperti huruf kurang atau kelebihan huruf pada kata ataupun kalimat. Atau disebut mengatur data yang kurang rapi dalam data yang sesuai. Proses data *cleaning* ini memudahkan dalam pemberian *code* setiap program studi lulusan, bidang pekerjaan lulusan dan posisi jabatan lulusan bekerja. Sesuai dengan kemiripan data pada lulusan.

Pengertian Data Transformasi

Pada tahap ini, data akan di transformasikan ke bentuk yang sesuai untuk *preprocessing* data mining menggunakan Teknik Data Mining *Clustering*. Yang dimana tahap ini adalah proses pengkodean data pada setiap pernyataan pernyataan yang dimana pernyataan nya tersebut dari program studi lulusan, bidang pekerjaan lulusan dan posisi jabatan lulusan bekerja di dalam suatu perusahaan. Yang nantinya hasil pengkodean terhadap data pernyataan tersebut adalah memudahkan penulis dalam pengolahan data secara manual dan melalui *exel*, sesuai dengan kemiripan data pada lulusan.

Metode Pengujian sistem

Tahap ini adalah Langkah yang dipilih dalam metode pengolahan data dalam aplikasi *Weka*. Yang dimana dalam proses *clustering* nya data lulusan dengan menggunakan *algoritma K-Means*. Yang dimana lebih memudahkan penulis.

4. HASIL DAN PEMBAHASAN

Dalam penelitian ini, database lulusan yang didapatkan dari Pihak Biro Konseling dan *Tracer Study* Universitas Esa Unggul. Yang dimana jumlah data hasil proses Data *Cleanning* sebanyak 620 data lulusan. Yang dimana *variable* yang dipakai adalah nama lulusan, tahun

lulusan program studi lulusan, nama perusahaan, bidang pekerjaan lulusan serta posisi jabatan lulusan bekerja.

Selanjutnya akan dilakukan transformasi pada data, yang dimana hasil tranformasi data tersebut adalah hasil pengkodean terhadap data. Yang dimana data yang dipakai dalam tranformasi data(pengkodean data) tersebut meliputi; nama lulusan, bidang pekerjaan lulusan serta posisi jabatan lulusan bekerja.

Dataset hasil transformasi data tersebut kemudian akan dilakukan clustering data. Baik *clustering* data dilakukan pada Microsoft excel maupun dari pengujian *system* yaitu aplikasi *Weka*. Dataset hasil *clustering* Microsoft Exel tersebut akan di konversi terlebih menjadi format csv sehingga dapat memudahkan proses *clustering Weka* serta mengidentifikasi *outlier* dalam *Weka*. Data uji dalam penelitian ini berupa *ARFF*.

Penelitian ini akan mengevaluasi *performance algoritma* dari *Machine Learning (ML)*. khususnya dalam penelitian ini menggunakan *unsupervised*. Maka acuan dalam penelitian ini menggunakan *attribute* dan memilih *InterquartileRange* dalam mengidentifikasi *Outlier* dalam Hasil *Clustering* data menggunakan *algoritma K-Means*. Langkah awal sebelum melakukan perhitungan *cluster* pada yaitu menemukan jumlah cluster optimal pada jumlah 620 data lulusan.

Hasil Data Transformasi Data

Hasil dari Transformasi data ini adalah, pemberian kode setiap *variable*. Yang dimana *variabel* dalam penelitian ini adalah meliputi; nama lulusan, bidang pekerjaan lulusan dan posisi jabatan lulusan bekerja. Manfaat dalam pemberian kode ini adalah memudahkan dalam *pengclustering* data lulusan sesuai kemiripan data yang sama. Yang dimana *variable* tersebut dijadikan *atribut*. Sehingga memudahkan dalam pengujian *system* di aplikasi *Weka*.

No	Keterangan Program Studi
1	Teknik Informatika
2	Sistem Informasi

Tabel 4 1 Keterangan Pengkodean Program Studi

No	Keterangan Program Studi
1	Industri Komputer, Pemasangan Jaringan dan konsultasi computer.
2	Bidang Teknologi Keuangan (<i>FinTech</i>)
3	Bursa Kerja Online
4	<i>Call Center Management</i>
5	<i>Digital Printing Supplier</i>
6	<i>Event Organizer/MC & Entertainment</i>
7	<i>Fitness Center</i>
8	<i>Foundation/Yayasan</i>
9	Asuransi
10	Industri Retail
11	Jasa konsultan pengembangan Teknologi Informasi
12	<i>Jasa Outsourcing</i>
13	Lembaga Konsultasi Psikologi
14	Multimedia
15	Otomotif
16	Pelayanan Kesehatan
17	Pendidikan & Kursus
18	Photo Studio
19	<i>Produsen & Distributor Produk</i>
20	Industri Pelayanan Transportasi
21	<i>Website Design</i>

22	Industri Jasa Pengiriman barangan
23	Bidang Bisnis <i>Investasi</i>
24	Bidang <i>Manufaktur</i>
25	Pengelolaan jalan tol ruas Tangerang
26	<i>Business Directory & Online Advertising</i>
27	<i>Desain Interior</i>
28	<i>IT & Security System</i>
29	Industri Cakram Optik
30	Industri Listrik dan Energi Indonesia
31	Jasa Pelayanan Penyewaan
32	Jasa Akuntansi & Management Pajak
33	Jasa Pelayanan Penyiaran Informasi
34	Jasa Layanan Penanganan Data
35	Industri Minyak & Gas
36	Industri Office Furniture
37	Jasa Pelayanan Penyiaran Informasi
38	industri kuliner
39	<i>Importir</i> bahan kimia
40	Industri Pertanian dan Perkebunan

Tabel 4 2 Keterangan Pengkodean Bidang Pekerjaan Lulusan

No	Keterangan Posisi / Jabatan Pekerjaan Lulusan
1	Administrator Jaringan Transportasi
2	Analisis Aplikasi Kesehatan
3	<i>Customer Support Specialist</i>
4	Desainer Grafis

5	<i>Distribution Application Developer</i>
6	<i>Konsultan Public Relation</i>
7	Konsultan Teknologi Informasi
8	Manager Pemasaran
9	Mekanik
10	Multimedia Developer
11	<i>Network Engineer</i>
12	Pengembang Perangkat Lunak Asuransi
13	<i>Quality Assurance Specialist</i>
14	<i>Retail Systems Analyst</i>
15	Staff Nutrition Center
16	Teknisi
17	Tenaga Pengajar
18	Spesialis Basis Data
19	<i>Supervisor Call Center</i>
20	Administrator Jaringan
21	Pemasaran dan Komunikasi
22	Konsultan Public Relation
23	<i>Sales Marketing/ Sales Executive</i>
24	<i>Software Engineer/Developer</i>
25	<i>UI/UX Designer/Developer</i>
26	Kurir
27	<i>Penetration Tester</i>
28	Staff Administrasi Pemasaran
29	Bidang operator Jalan Tol Tangerang
30	<i>Visualisasi Ruang 3D</i>
31	visitor management
32	pembuatan <i>chip poliester</i>

33	<i>Project Planning Engineer</i>
34	<i>Announcer</i>
35	<i>Cybersecurity Specialist</i>
36	Administrator Jaringan Yayasan
37	Instruktur Kelas Fitness
38	<i>MC (Master of Ceremony)</i>
39	<i>Perancang Sensor dan Internet of Things (IoT)</i>

Tabel 4 3 Keterangan Pengkodean Posisi/Jabatan Lulusan Bekerja

Setelah diberikan kode setiap masing masing dari *variable* tersebut. maka hasil dari transformasi pada data 620 data sebagai berikut;

Nama	BP	PJ
HAIKAL SHIDDIQ	16	2
YUSTIKA RINI	12	17
RAMA PRATAMA	2	24
EDWIN ADRYANTO ARAFAT	10	14
NOVRIANTA ZUHRY SEMBIRING	12	13
EDO PRAKOSO	10	14
NUGIE RAMADHANEGAR	10	14
NAFIZA SAMDITYA PUTRA	2	24
OKA IRVIAN IRMANTO S	4	19
ILHAM TRI KURNIAWAN	2	24
FIQRI ZARKASIH	10	14
RANDY SWANDY	9	16
WINDO YOFENA	12	13
AZHAR PRABUDI	16	2
SITI ZULIATUL FAIDAH	9	16
YANITA	15	9
HANIF NURYANTO	9	16
ADITIA SAPUTRA	12	13
NURAENI	2	24
KRISBIANTO	12	13
MUHAMAD ANGGA GUSTIAWAN	12	13
MUSTAJAB	2	24
ALLAND CHANDRA KESUMA	15	9
ANUGERAH CAHYO ADHI	19	5
ALFATH DIIN ALYA SUNNII	19	5
RIANSA RACHMAN	19	8
AHMAD AJI RIMANDANANG	9	12
RAHMAD DARMAWAN	12	13
MUHAMMAD ZIDANE	2	24
WARIMAN	2	24
QOIDATUN NIKMAH	9	16

Gambar 4 1 Hasil Transformasi Data

Elbow

Untuk menghitung jumlah *cluster* optimal pada data dengan rumus metode

elbow. Berikut rumus menghitung jumlah cluster optimal pada data;

$$WCSS = \sum_{i=1}^k \sum_{x \in C_i} \|x - \mu_i\|^2$$

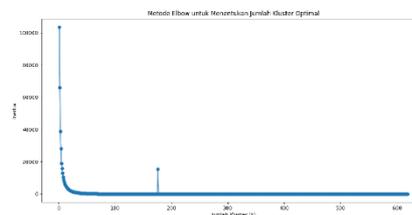
Di mana:

k adalah jumlah klaster yang diinginkan.

C_i adalah klaster ke-i

x adalah data poin dalam klaster C_i

μ_i adalah pusat (*centroid*) klaster C_i



Gmbar 4. 2 Hasil Cluster Optimal

Hasil dari grafik diatas menjelaskan hasil perhitungan menentukan jumlah *cluster* optimal pada 620 data lulusan sebanyak 5 cluster optimal. Yang dimana pada hasil iterasi ke 5 terjadi nya penurunan *signifikasi* pada hasil perhitungan *clustering* data.

Cluster	BP	PS
C1	3	3
C2	19	5
C3	2	24
C4	1	24
C5	10	14

Tabel 4 . 4 Hasil 5 Cluster Optimal

Keterangan ;

BP : Bidang Pekerjaan

PS : Posisi Jabatan

Algoritma K-Means

Algoritma K-Means adalah metode atau cara melakukannya hasil *clustering* pada 620 data sesuai dengan kemiripan data. Yang dimana, dalam clustering data tersebut hanya menggunakan atribut – atribut sebagai berikut; nama lulusan, bidang pekerjaan dan posisi jabatan lulusan bekerja. Berikut dibawah ini adalah cara menghitung jarak antara

dengan data lulusan. Dari hasil pemilihan nilai k secara acak pada hasil *cluster* yang ditentukan *elbow*. Dan dibawah ini adalah cara melakukan *clustering* manual dengan terlebih dahulu mencari nilai *centroidnya*. Hasil nilai *centroid* dimana yang jarak nya paling tersekat itu menentukan kategori *cluster* berapa data lulusan tersebut.

Example ; Data Lulusan

ABDILAH Hidayatulloh	23	27
----------------------	----	----

$$C1 = \sqrt{((23-3)^2 + (27-3)^2)}$$

$$C1 = \sqrt{((20)^2 + (24)^2)}$$

$$C1 = \sqrt{400 + 576}$$

$$C1 = 31.2409987$$

$$C2 = \sqrt{((23-19)^2 + (27-5)^2)}$$

$$C2 = \sqrt{((4)^2 + (22)^2)}$$

$$C2 = \sqrt{16 + 484}$$

$$C2 = 22,36067977$$

$$C3 = \sqrt{((23-2)^2 + (27-24)^2)}$$

$$C3 = \sqrt{((21)^2 + (3)^2)}$$

$$C3 = \sqrt{441 + 9}$$

$$C3 = 21,21320344$$

$$C4 = \sqrt{((23-1)^2 + (27-24)^2)}$$

$$C4 = \sqrt{((22)^2 + (3)^2)}$$

$$C4 = \sqrt{484 + 9}$$

$$C4 = 22,20360331$$

$$C5 = \sqrt{((23-10)^2 + (27-24)^2)}$$

$$C5 = \sqrt{((13)^2 + (3)^2)}$$

$$C5 = \sqrt{169 + 9}$$

$$C5 = 18,38477631$$

Maka Hasil Iterasi Pertama dari Hasil Pengelolaan Clustering Pertama adalah sebagai berikut;

Nama	BP	P1	P2	C1	C2	C3	C4	C5	JARAK TERDEKAT	CLUSTER
ABDILAH Hidayatulloh	23	27	12.2409987	22.36067977	21.21320344	22.20360331	18.38477631	18.38477631	18.38477631	1
ABDUL HAMID	30	16	29.8666481	15.5554919	29.12044	30.083113	20.097513	25.5582482	25.5582482	2
Abdul Lathif Gofar	1	24	21.0950231	26.17250469	0	0	19.453624	0	19.453624	4
ABDUL RAHMAN	29	8	16.925046	0	23.345231	24.081391	19.665588	0	19.665588	2
ABDUL RAHMAT	2	24	23.037796	25.4950737	0	0	12.8042403	0	12.8042403	4
ADEL MOHAMMAD	0	11	9.0528544	18.2647261	11	11.445258	8.26011259	2.54231125	8.26011259	5
ABU AMAR SATRIAWAN	17	17	19.7988991	13.1051206	16.502491	17.464430	7.6137711	2.61577311	7.6137711	5
ACHMAD FANUZI	0	0	0	26.1245155	21.021799	21.0950231	13.838408	0	13.838408	4
Achmad Fandiyan Fery Sugandemara	7	07	34.2344851	34.7002498	13.927838	14.178121	23.1944071	13.9283861	13.9283861	5
ACHMAD FITYAN ALDYAN	32	7	29.2745623	13.13284644	34.481373	35.153339	23.0877928	13.13284644	13.13284644	2
ADE SILAKMAN	31	7	39.2704621	13.13284644	34.481373	35.153339	23.0877928	13.13284644	13.13284644	2
ADHITYA IMAM WIGARSONO	27	4	34.0328143	8.26237748	32.015621	31.802419	19.723828	8.0623775	8.0623775	2
ADI WIDANTONO	19	20	25.345231	15	17.464249	18.439089	10.166538	20.826338	20.826338	5
ADOLAN WICAKO	9	12	19.166538	12.2055361	13.891444	14.422202	2.2060798	2.2060798	2.2060798	5
ADINDA DHANA CHAEDIR	2	24	21.037796	25.4950737	0	0	12.8042403	0	12.8042403	4
ADITIA SAPUTRA	12	13	13.453624	10.8321081	14.886099	15.516340	2.2060798	2.2060798	2.2060798	5
ADITYA EKA YULIARSO	12	13	13.453624	10.8321081	14.886099	15.516340	2.2060798	2.2060798	2.2060798	5
ADITYA NARA NABIL	12	13	13.453624	10.8321081	14.886099	15.516340	2.2060798	2.2060798	2.2060798	5
ADITYA RIZKY GUSTIAN	28	27	34.8544469	23.7072085	26.172105	27.1461151	22.20360331	22.20360331	22.20360331	5
Ahmad Supriyo Anggra	19	28	29.6814462	21	17.464249	18.439089	16.643112	16.643112	16.643112	5
AGUNG FALCHON	22	26	29.8328678	23.2123444	20.097513	21.0950231	14.075042	14.075042	14.075042	5
AGUNG AERAB (WANDOPERA MUSA SA)	12	13	13.453624	10.8321081	14.886099	15.516340	2.2060798	2.2060798	2.2060798	5
AGUNG TORONG	10	14	13.0362402	10.7212096	13.892428	14.075042	0	0	0	5
AGUS PRYANTO	16	2	13.0362404	4.24264087	30.07681	26.621054	13.8164279	4.24264087	4.24264087	2
AGUS SALIM	11	7	8.8421710	8.44621121	19.231884	19.723828	7.0726791	7.0726791	7.0726791	5
AMANDA ADIS TANZILA	26	24	24.9491761	18.3558408	14	15.1241908	10.8620948	0	10.8620948	5
AMAD AI RIMANDIANANG	9	12	19.166538	12.2055362	13.891444	14.422202	2.2060798	2.2060798	2.2060798	5
AMAM DIMPATY	19	5	16.3441053	0	18.404098	20.172021	12.1719211	0	12.1719211	4
AMAM DUNALDI	19	8	16.7820446	3	23.345231	24.081391	10.8166538	0	10.8166538	2
AMAM FAZI	19	8	16.7820446	3	23.345231	24.081391	10.8166538	0	10.8166538	2

Gambar 4 3 Hasil Iterasi Pertama Clustering Data

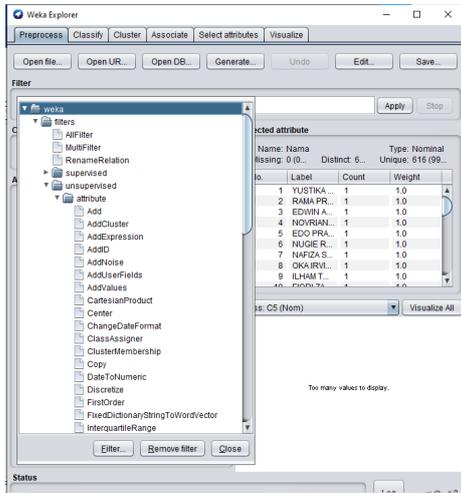
Dari hasil iterasi pertama tersebut, terlihat pada Gambar 4.3 dalam pernyataan jarak terdekat. Jarak terdekat ialah nilai terendah dari 5 nilai *centroid cluster* tersebut. Dan ini lah hasil perhitungan jarak tiap data ke tiap pusat *cluster* nya.

Selanjutnya, melakukan hasil perhitungan Iterasi Kedua. Yang dimana hasil perhitungan sedikit berbeda. Yang dimana hasil pengkodean terhadap bidang pekerjaan dan posisi / jabatan pekerjaan melakukan perhitungan dengan nilai *centroid* C1, C2, C3, C4 dan C5.

C1	3,3	4,15
C2	21,69189189	7,389189189
C3	3,935483871	25,74193548
C4	1	21,5
C5	15,34615385	17,09935897

Tabel 4 5 Keterangan Nilai Centroid Cluster Iterasi Kedua

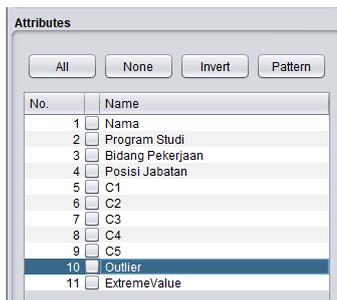
Pada Tabel 4.5 diatas adalah hasil perhitungan jumlah *cluster* yang terdapat di iterasi pertama dibagi dengan jumlahnya. Yang dimana dipilih adalah jarak terdekatnya ke pusat *cluster*. Lalu dilakukan perhitungan *clustering* lagi pada Iterasi Kedua ini. Dengan 620 data lulusan tersebut.



Gambar 4 13 Pemilihan Identifikasi Outlier



Gambar 4 14 Metode Pemilihan Outlier



Gambar 4 15 Hasil Outlier

Selected attribute			
Name:	Outlier	Type:	Nominal
Missing:	0 (0%)	Distinct:	1
Unique:	0 (0%)	Count:	620
Weight:	0	Weight:	620.0
1	no	620	620.0
2	yes	0	0.0

Gambar 4 16 Keterangan Hasil Identifikasi Outlier

No.	Nama	Program Studi	Bidang Pekerjaan	Posisi Jabatan	C1	C2	C3	C4	C5	Outlier	ExtremeValue
1	YUSTI	Teknik Informatika			17.0	4.06	13.6	15.2	11.9	17.8	no
2	RAMA	Teknik Informatika			24.0	14.9	25.3	19.5	1.66	25.0	no
3	EDWIN	Teknik Informatika			14.0	2.35	12.6	11.6	13.0	21.2	no
4	NOVI	Teknik Informatika			13.0						no
5	EDO	Teknik Informatika			14.0						no
6	NUJUL	Teknik Informatika			14.0	2.35	12.6	11.6	13.0	21.2	no
7	NAFZA	Teknik Informatika			24.0	14.9	25.3	19.5	1.66	25.0	no
8	OKAL	Teknik Informatika			19.0	10.1	25.4	14.4	6.18	24.2	no
9	ILHA	Teknik Informatika			24.0	14.9	25.3	19.5	1.66	25.0	no
10	FIQRI	Teknik Informatika			14.0	2.35	12.6	11.6	13.0	21.2	no
11	RANI	Teknik Informatika			19.0	4.36	14.7	12.8	10.8	20.9	no
12	WINDY	Teknik Informatika			13.0	0.12	10.5	12.1	14.8	20.3	no
13	SITI	Teknik Informatika			19.0	4.36	14.7	12.8	10.8	20.9	no
14	HANI	Teknik Informatika			19.0	4.36	14.7	12.8	10.8	20.9	no
15	ADITI	Teknik Informatika			13.0	0.12	10.5	12.1	14.8	20.3	no
16	NUR	Teknik Informatika			24.0	14.9	25.3	19.5	1.66	25.0	no
17	KRS	Teknik Informatika			13.0	0.12	10.5	12.1	14.8	20.3	no
18	MUK	Teknik Informatika			13.0	0.12	10.5	12.1	14.8	20.3	no
19	MUST	Teknik Informatika			24.0	14.9	25.3	19.5	1.66	25.0	no
20	ANHA	Teknik Informatika			12.0	3.84	12.3	8.63	14.3	23.2	no
21	RAH	Teknik Informatika			13.0	0.12	10.5	12.1	14.8	20.3	no
22	MUHA	Teknik Informatika			24.0	14.9	25.3	19.5	1.66	25.0	no
23	WARU	Teknik Informatika			24.0	14.9	25.3	19.5	1.66	25.0	no
24	QOUD	Teknik Informatika			19.0	4.36	14.7	12.8	10.8	20.9	no
25	MUHA	Teknik Informatika			24.0	14.9	25.3	19.5	1.66	25.0	no
26	MUR	Teknik Informatika			19.0	4.36	14.7	12.8	10.8	20.9	no
27	SOHY	Teknik Informatika			17.0	4.06	13.6	15.2	11.9	17.8	no
28	DANA	Teknik Informatika			14.0	2.35	12.6	11.6	13.0	21.2	no
29	MUHA	Teknik Informatika			24.0	14.9	25.3	19.5	1.66	25.0	no

Gambar 4 17 Hasil Identifikasi Outlier

Dari Gambar 4.13 sampai Gambar 4.17 adalah langkah langkah dalam proses mengidentifikasi *Outlier* setelah melakukan *Clustering* pada 620 data tersebut. Coba perhatikan Gambar 4.15 adalah *atribut* yang sudah diunsupevised dan memilih *InterQuatrtil* yang dimana tahapan mengidentifikasi *Outlier* dalam Pengujian Sistem diaplikasi *Weka*. Dan Coba Perhatikan Gambar 4.16, terlihat kita bisa mengetahui berapa banyak *Outlier* dalam hasil *Clustering* data tersebut. yang dimana opsi Yes menyatakan adanya data *Outlier* dan opsi Tidak menyatakan tidak adanya data yang *Outlier*.



Gambar 4 18 Hasil Grafik dari Clustering Data di Weka

Pada Gambar 4.18 terlihat jelas adalah hasil grafik dari proses dilakukannya *clustering* data dan hasil identifikasi *outlier*. Yang dimana sebanyak 5 cluster pada *attribute* nama lulusan, bidang pekerjaan lulusan serta posisi/jabatan lulusan bekerja.

5. KESIMPULAN

1. Penelitian ini dilakukan dengan cara pengolahan data yang digunakan pada data lulusan tahun 2016 – 2020, sebanyak 620 data. Data tersebut di kumpulkan di Biro Konseling dan Tracer Study (BKTS) Universitas Esa Unggul. Yang dimana dalam pengolahan data secara manual di aplikasi Microsoft Exel dan dengan

metode Elbow, dapat ditentukan berapa cluster yang di dikelompokkan data lulusan. Sehingga memudahkan penulis dalam pengelolaan data tersebut. Pengelolaan dilakukan dengan menghitung jarak ke tiap data ke titik pusat data cluster nya. Sehingga ada Namanya centroid (C1,C2,C3,C4 dan C5). Pengeolahan data juga dilakukan dengan rumus, eucoldian.

2. Hasil dari penelitian ini dari 2 aplikasi yaitu Microsoft Exel dan aplikasi Weka dapat disimpulkan ;
 - Cluster 1 (C1) = sebanyak 105 lulusan yang bekerja di bidang pekerjaan Produsen & Distributor Produk Pemasaran dan Komunikasi dengan Posisi Jabatan berkaitan dengan Pemasaran dan Komunikasi. Yang dimana presentasinya sebanyak 17%.
 - Cluster 2 C2 = sebanyak 53 lulusan yang bekerja dibidang pekerjaan Industri Jasa Pengiriman Barang dengan Posisi Jabatan yang berkaitan Proses Pengiriman Barang atau Kurir. Yang dimana presentase nya sebanyak 9%.
 - Cluster 3 C3 = sebanyak 240 lulusan yang bekerja dibidang pekerjaan Jasa Outsourcing dengan Posisi Jabatan yang berkaitan dengan Quality Assurance Specialist. Yang dimana presentase nya sebanyak 39%.
 - Cluster 4 C4 = sebanyak 128 lulusan yang bekerja di bidang pekerjaan Jasa Asuransi dengan Posisi Jabatan berkaitan dengan Pengembangan Perangkat Lunak Asuransi. 21%.
 - Cluster 5 C5 = sebanyak 94 lulusan yang bekerja di bidang pekerjaan Pendidikan dan Tenaga Kursus dengan posisi jabatan yang berkaitan dengan Tenaga Pengajar. Yang dimana presentase nya sebanyak 15%.

Berisi mengenai simpulan dari hasil penelitian dan saran.

(Kosong 1 spasi tunggal 10 pt: antara KESIMPULAN dan DAFTAR PUSTAKA)

DAFTAR PUSTAKA

- Laeli Nur Azizah “Pengertian Data, Jenis Data serta contohnya, sumber data, metode pengumpulan data. 2021”.
- Dwi Welly Sukma Nirad, Kridanto Surendro “Analisis Data Tracer Study dengan Mengidentifikasi Outlier menggunakan Teknik Data Mining, 2 Agustus 2018”.
- Arief Abdurrakhman,S.T.,M.T. (Kasubdit Pengembangan Kewirausahaan dan Karir ITS), 05 Desember 2021. “PENGERTIAN TRACER STUDY”
- Silvia Febrianti , Munawir , Liza Fitria, Universitas Samudra. 02 Oktober 2021 “PENERAPAN METODE K – MEANS CLUSTERING TERHADAP ALUMNI BERDASARKAN KUESIONER TRACER STUDY APPLICATION OF METHOD K – MEANS CLUSTERING
- Brilliananda Widhi Nugraha, Ali Mahmudi, Febriana Santi Wahyuni, 02 September 2021 “PENERAPAN METODE K-MEANS UNTUK PENGELOMPOKAN TINGKAT KEPUASAN PENGGUNA LULUSAN PADA TRACER STUDY PUSAT KARIR ITN MALANG”
- Linda Tanti, Binus University, 29 Oktober 2021.” Metode Data Mining Clustering”.
- DR. DERWIN SUHARTONO, S.KOM., M.T.I, Binus University, “Weka: Software untuk Memahami Konsep Data Mining”.
- Rony Setiawan, 30 Oktober 2021.”PENGERTIAN DATA MINING, TUJUAN DATA

- MINING, METODE DATA MINING DAN PENERAPAN DATA MINING”.
- Dwi Welly Sukma Nirad , Kridanto Surendro, Agustus 2018 "Analisis Tracer Study Dengan Mengidentifikasi Outlier Menggunakan Teknik Data Mining"
- DR. DERWIN SUHARTONO, S.KOM., M.T.I, Binus University, “Weka: Software untuk Memahami Konsep Data Mining”.
- Rony Setiawan, 30 Oktober 2021.”PENGERTIAN DATA MINING, TUJUAN DATA MINING, METODE DATA MINING DAN PENERAPAN DATA MINING”.
- Akhmad Choeruddin Wakhid, 23 November 2017, “Implementasi Data Mining Menggunakan Algoritma K- Means Clustering Untuk mengetahui pola pemilihan program studi mahasiswa baru IAI SALATIGA”
- Brilliananda Widhi Nugraha, Ali Mahmudi, Febriana Santi Wahyuni, 2 September 2021 “Brilliananda Widhi Nugraha, Ali Mahmudi, Febriana Santi Wahyuni”
- Irham Sodik, 07 Agustus 2017, “SISTEM TRACER ALUMNI PADA SMK NEGERI 1 UNDANAWU MENGGUNAKAN METODE K-MEANS CLUSTERING”
- Irwanto, 12650064, 2016 “PENERAPAN DATA MINING UNTUK MENGETAHUI POLA PEMILIHAN PROGRAM STUDI BARU UIN SUNAN KALIJAGA MENGGUNAKAN ALGORITMA K-MEANS”.
- LUIGI AJENG PRATIWI, 109091000053, 2013 “IMPLEMENTASI ALGORITMA DALAM DATA MINING UNTUK PEMINATAN JURUSAN BAGI SISWA KELAS X SMA NEGERI 29 JAKARTA”.
- Septi Anggarawati, 22 September 2022, “PENERAPAN ALGORITMA K-MEANS CLUSTERING UNTUK MENENTUKAN LINIEARITAS PEKERJAAN ALUMNI BERDASARKAN TRACER STUDY”.