

# Clustering Data Remunerasi PNS Menggunakan Metode K-Means Clustering Dan Local Outlier Factor

Made Pasek Agus Ariawan<sup>1</sup>, Nyoman Putra Sastra<sup>2</sup>, I Made Sudarma<sup>3</sup>

Submission: 23-11-2019, Accepted: 25-12-2019

**Abstract**— Remuneration is a reward for services performed by employees. Remuneration is given to employees who have good performance. Remuneration can be given by an agency if it has implemented the Financial Management of Public Service Bodies (PK-BLU). The problem with remuneration is that the validation carried out by the direct supervisor of the employee concerned is still doubtful. Based on this, we need a system that can detect outlier data from the remuneration data of Civil Servants and classify the data using data mining techniques. The algorithm for finding outlier data is the Local Outlier Factor (LOF) algorithm and the algorithm that can be used to do clustering is a k-means clustering algorithm. The K-means algorithm has problems in determining the optimal number of clusters. Problems with the K-means method can be solved using the elbow method. Determination of this method is seen from the Sum Square Error (SSE) graph of several cluster numbers. This study aims to classify the remuneration data of civil servants by using the k-means clustering method with improvisation at the pre-processing stage and determining the optimal number of clusters. Local Outlier Factor method with a MinPts value of 150 can detect the most outlier data with 162 data outliers detected or 22.98%. The optimal number of clusters with the elbow method is 4 clusters with a Silhouette value of 0.542, Dunn of 0.040 and Purity of 0.89.

**Kata Kunci**— Clustering, K-Means, LOF, Outlier

**Intisari**— Remunerasi adalah imbalan atas jasa yang dilakukan oleh pegawai. Pemberian remunerasi diberikan kepada pegawai yang memiliki kinerja yang baik. Remunerasi dapat diberikan oleh suatu instansi apabila sudah menerapkan Pengelolaan Keuangan Badan layanan Umum (PK-BLU). Permasalahan dari pemberian remunerasi adalah validasi yang dilakukan oleh atasan langsung pegawai yang bersangkutan masih diragukan validitasnya. Berdasarkan hal tersebut, diperlukan suatu sistem yang dapat mendeteksi data bersifat outlier dari data remunerasi Pegawai Negeri Sipil dan mengelompokkan data tersebut menggunakan teknik data mining. Algoritma untuk mencari data outlier adalah algoritma Local Outlier Factor (LOF) dan algoritma yang dapat digunakan untuk melakukan clustering (pengelompokan) adalah algoritma k-means clustering. Algoritma K-means memiliki permasalahan dalam penentuan jumlah cluster yang terbaik. Permasalahan metode K-means ini dapat diselesaikan dengan menggunakan metode elbow. Penentuan metode ini dilihat dari grafik Sum Square Error (SSE) dari beberapa jumlah cluster. Penelitian ini bertujuan untuk mengelompokkan data remunerasi pegawai

negeri sipil dengan menggunakan metode k-means clustering dengan improvisasi pada tahap pre-processing dan penentuan jumlah cluster optimal. Metode Local Outlier Factor dengan nilai MinPts 150 dapat mendeteksi data outlier paling banyak dengan jumlah data terdeteksi outlier sebanyak 162 data atau sebesar 22,98%. Jumlah cluster optimal dengan metode elbow berjumlah 4 cluster dengan nilai Silhouette sebesar 0,542, Dunn sebesar 0,040 dan Purity sebesar 0,89.

**Kata Kunci**— Clustering, K-Means, LOF, Outlier

## I. PENDAHULUAN

Remunerasi adalah imbalan kerja berupa gaji, honor, tunjangan tetap, insentif, bonus atas prestasi, pesangon, dan/dana pensiun. Remunerasi merupakan penghasilan yang bisa diperoleh dan dibelanjakan oleh pegawai atau karyawan (*take home pay*) atas hasil pekerjaan yang telah dilakukan. Adanya remunerasi akan menghilangkan anggapan bahwa tidak ada korelasi positif antara kinerja dengan penghasilan. Ini berarti pegawai/karyawan yang berkinerja baik akan memiliki penghasilan yang tidak sama. Pemerintah telah berupaya mengubah sistem penggajian agar menjadi lebih baik melalui sistem remunerasi. Remunerasi terdiri dari gaji pokok yang ditambah tunjangan – tunjangan yang bersumber dari rupiah murni dan tunjangan lain yang bersumber dari Pendapatan Negara Bukan Pajak.

Berdasarkan Peraturan Menteri Keuangan Republik Indonesia Nomor 176 /PMK.05/2017 Remunerasi dapat diberikan oleh suatu instansi apabila sudah menerapkan Pengelolaan Keuangan Badan layanan Umum (PK-BLU). BLU adalah instansi di lingkungan Pemerintah yang dibentuk untuk memberikan pelayanan kepada masyarakat berupa penyediaan barang dan/atau jasa yang dijual tanpa mengutamakan mencari keuntungan dan dalam melakukan kegiatannya didasarkan pada prinsip efisiensi dan produktivitas. Instansi-instansi ini meliputi pendidikan, kesehatan, pengelola kawasan, pengelola dana dan barang jasa lainnya.

Jumlah penyelenggara pendidikan tinggi di Indonesia telah mengalami peningkatan dalam kurun beberapa tahun terakhir. Tercatat terdapat sebanyak 4.622 perguruan tinggi di Indonesia baik perguruan tinggi negeri maupun swasta [1] dengan jumlah perguruan tinggi negeri sebanyak 418 perguruan tinggi dan yang telah menerapkan BLU sebanyak 63 perguruan tinggi. Selain itu berdasarkan data dari kementerian keuangan terdapat 203 instansi yang telah menerapkan BLU.

Pada tahun 2012 Universitas Udayana telah ditetapkan sebagai Instansi Pemerintah yang menerapkan Pengelolaan Keuangan Badan layanan Umum (PK-BLU) sesuai Keputusan Menteri Keuangan Republik Indonesia Nomor 441/KMK.05/2011, Tertanggal 27 Desember 2011. Sebagai

<sup>1</sup>Mahasiswa, Program Magister Teknik Elektro, Jl. PB.Sudirman, Denpasar, Bali, Indonesia. Tlp: 081999921941 e-mail: [pasekagusariawan@gmail.com](mailto:pasekagusariawan@gmail.com)

<sup>2, 3</sup>Dosen, Jurusan Teknik Elektro dan Komputer Fakultas Teknik Universitas Udayana, Jln. Jalan Kampus Bukit Jimbaran 80361 INDONESIA; e-mail: [putra.sastra@unud.ac.id](mailto:putra.sastra@unud.ac.id), [imasudarma@gmail.com](mailto:imasudarma@gmail.com)



Perguruan Tinggi yang sudah berbentuk BLU, Universitas Udayana bisa melakukan Reformasi Birokrasi, sebagai prasyarat untuk mengajukan Tunjangan Remunerasi Universitas Udayana.

Di Universitas Udayana terdapat dua tahapan dalam pemberian remunerasi pegawai negeri sipil, yakni remunerasi yang dibayarkan setiap bulan dan remunerasi yang dibayarkan setiap semesternya. Remunerasi tiap bulan yang diterima oleh PNS sebesar 30% dari keseluruhan remunerasi yang didapat dan sisanya sebesar 70% dibayarkan setiap semester. Terdapat tiga parameter dalam pemberian remunerasi yaitu kehadiran, sikap dan perilaku, dan capaian kinerja dari pegawai selama enam bulan. Dengan adanya pemberian remunerasi bagi PNS tentunya perlu dilakukan evaluasi terkait pemberian remunerasi ini dengan cara mengelompokkan data pegawai dan melakukan analisis *outlier* pada data tersebut.

Permasalahan yang muncul dari pemberian remunerasi adalah validasi yang dilakukan oleh atasan langsung pegawai yang bersangkutan masih diragukan validitasnya. Berdasarkan hal tersebut, diperlukan suatu sistem yang dapat mendeteksi data bersifat *outlier* dari data remunerasi Pegawai Negeri Sipil dan mengelompokkan data tersebut menggunakan teknik data mining. *Clustering* merupakan teknik pengelompokan data yang sering digunakan. Salah satu algoritma yang digunakan dalam metode *clustering* adalah algoritma *K-Means* [2]. Algoritma *K-Means* adalah metode *clustering* non hierarki yang memiliki kelebihan pada waktu komputasi yang relatif cepat [3].

Prinsip kerja dari algoritma *k-means* adalah membagi data kedalam kelompok data yang memiliki kemiripan sifat data. Tetapi, algoritma *k-means* memiliki kekurangan yaitu sangat sensitive dalam menentukan jumlah partisi awal *cluster* [4]. Selain itu, dalam implementasinya algoritma *k-means* memiliki beberapa kelemahan seperti terdapat *outlier*, nilai *sum square error* besar dan nilai akurasi yang relative kecil [5]. *Outlier* adalah data yang memiliki karakteristik menyimpang dari data lainnya. Deteksi *outlier* penting karena dapat memberikan informasi berharga. Deteksi data - data menyimpang ini disebut *Outlier Mining* [6].

Merliana, Ernawati dan Santoso [7] menentukan jumlah *cluster* optimal pada algoritma *k-means* dengan menggunakan metode elbow, penentuan jumlah *cluster* pada metode ini dilihat dari grafik penurunan nilai *sum of square error* dari beberapa jumlah *cluster*.

Bhatt, Dhakar dan Chaurasia [8] menggunakan metode LOF mendeteksi *outlier* untuk meningkatkan hasil evaluasi *k-means clustering*. Selain itu, Penelitian yang dilakukan oleh Nugraha [9] juga menggunakan metode LOF untuk mendeteksi *outlier* pada *clustering* data siswa.

Berdasarkan penelitian yang telah dilakukan maka penelitian ini bertujuan untuk mengelompokkan data remunerasi pegawai negeri sipil dengan menggunakan metode *k-means clustering* dengan improvisasi pada tahap *pre-processing* dan penentuan jumlah *cluster* optimal.

## II. METODOLOGI PENELITIAN

*Clustering* merupakan pengelompokan data kedalam data tertentu, *clustering* digunakan untuk mengelompokkan data dengan data lainnya berdasarkan kemiripan pada objek data

[10]. Dalam perancangan penelitian ini akan dilakukan dalam beberapa tahapan antara lain:

### A. Pengumpulan data

Sumber data dalam penelitian ini adalah data sekunder, sedangkan proses pengumpulan data dilakukan di Universitas Udayana dengan memanfaatkan data remunerasi tenaga kependidikan di Universitas Udayana tahun 2017. Data terbagi menjadi dua data yaitu data remunerasi 70% untuk data sasaran kerja tenaga kependidikan dan data remunerasi 30% untuk data kehadiran dan sikap & perilaku.

### B. Seleksi data

Data yang diperoleh kemudian diseleksi dan dipilih data yang sesuai untuk dilakukan perhitungan. Data yang dibuang adalah data-data yang kurang relevan dengan penelitian. Dari data spreadsheet yang diperoleh, variabel yang digunakan untuk proses perhitungan adalah poin hitung kinerja, nilai kehadiran, dan nilai perilaku.

### C. Deteksi Outlier

*Outlier* adalah nilai-nilai yang tidak normal jauh dari sebagian besar nilai lain didalam dataset. *outlier* dapat memberikan wawasan yang berharga, *outlier* juga mempengaruhi pengujian. Penghapusan *outlier* adalah topik kontroversial, tetapi sebagian besar analisis sangat sensitif terhadap *outlier* karena dapat mempengaruhi hasil. [11].

Alur metode *local outlier factor* adalah sebagai berikut:

- Menghitung jarak antara pengamatan dan tetangga terdekat  $k$ .
- Menghitung tetangga terdekat yang tidak lebih dari nilai  $k - distance$ .
- Menghitung *Reachability Distance*.
- Menghitung *local reachability density* (kepadatan *local*) dari setiap objek. Dengan persamaan berikut :

$$lrd_{MinPts}(p) = \frac{o \in \sum_{N_{MinPts}(p)} reachdist_{MinPts}(p,o)}{|N_{MinPts}(p)|} \dots\dots(1)$$

Keterangan :

$lrd_{MinPts}(p)$  : *local reachability density* dari objek  $p$ .

$reachdist(p)$  : *reachability distance* dari objek  $p$  ke objek  $o$

$N_{MinPts}(p)$  : jumlah tetangga  $p$  dalam suatu *minPts*

- Menghitung LOF untuk setiap objek data dengan persamaan berikut :

$$LOF_{MinPts}(p) = \frac{o \in \sum_{N_{MinPts}(p)} \frac{lrd_{MinPts}(o)}{lrd_{MinPts}(p)}}{|N_{MinPts}(p)|} \dots\dots\dots(2)$$

Keterangan :

$LOF_{MinPts}(p)$  : derajat *outlier* dari objek  $p$

$lrd_{MinPts}(o)$  : *local reachability distance* dari objek  $p$ .

$reachdist(p)$  : *reachability distance* dari objek  $p$  ke objek  $o$

$N_{MinPts}(p)$  : jumlah tetangga  $p$  dalam suatu *minPts*

### D. Normalisasi data

Dalam teknik normalisasi data ini, transformasi linear dilakukan pada data asli. Nilai minimum dan maksimum dari data diambil dan setiap nilai diganti sesuai dengan rumus 3[12]. Data dirubah sesuai dengan kebutuhan penilitan. Pada penelitian ini data akan diubah kedalam rentang nilai 0-1.

$$v' = \frac{v - \min_A}{\max_A - \min_A} \dots \dots \dots (3)$$

**E. Proses Clustering**

Tahapan pengelompokan data remunerasi tenaga kependidikan dengan menggunakan metode *K-means*. Hasil yang diharapkan adalah *cluster* data remunerasi tenaga kependidikan dengan penentuan *cluster* sebanyak *K* menggunakan metode elbow, setelah menentukan jumlah *cluster* tentukan asumsi titik pusat *cluster* penentuan titik pusat *cluster* dapat dilakukan secara acak kemudian dilanjutkan menghitung jarak objek ke centroid pada penelitian ini proses yang digunakan akan menggunakan perhitungan jarak dengan metode *Euclidian distance*, data yang telah dihitung dengan dengan metode *Euclidian distance* akan dikelompokan berdasarkan jarak minimum.

Alur *Clustering* data dengan algoritma *k-means* adalah sebagai [13]:

- 1) Pilih titik *k* secara acak sebagai pusat *cluster*.
- 2) Tetapkan objek ke pusat kluster terdekat menurut fungsi jarak Euclidean.
- 3) Hitung centroid atau rata-rata dari semua objek di setiap *cluster*.
- 4) Ulangi langkah 2, 3 dan 4 hingga poin yang sama ditetapkan untuk setiap *cluster* secara berurutan.

**F. Evaluasi**

Tahapan evaluasi hasil *clustering* akan dilakukan uji performa dilakukan dengan menggunakan perhitungan *Dunn Index*, *Silhouette Index* pengukuran akurasi dengan *purity measure*.

*Silhouette index* (SI) digunakan untuk validasi *cluster* karena teknik ini adalah salah satu teknik yang terkenal. Indeks ini menjadi salah satu pengukuran kinerja terbaik yang mampu menunjukkan objek mana yang ditempatkan dengan baik di dalam *cluster* mereka[14].

*Dunn Index* digunakan untuk menentukan seberapa berbeda antara satu *cluster* dan *cluster* lainnya. *Dunn Index* dihitung dengan mengukur rasio jarak terbesar antar cluster dengan jarak terkecil di dalam cluster. Semakin tinggi nilai *Dunn Index* maka semakin baik cluster yang terbentuk [15].

Untuk mengukur hasil *clustering* dapat digunakan nilai *purity* dari suatu *cluster*. *Purity* (kemurnian) suatu *cluster* dipresentasikan sebagai anggota *cluster* yang paling banyak sesuai (cocok) di suatu kelas. Suatu *cluster* dinilai baik apabila nilai *purity* mendekati 1 dan buruk bila *purity* mendekati 0[16].

**III. HASIL DAN EVALUASI**

**A. Data Cluster.**

Data pada penelitian ini berupa spreadsheet data remunerasi 70% dan 30% dalam bentuk xls. Data input termasuk kategori data terstruktur. Data terstruktur direpresentasikan dalam skema yang jelas sehingga mudah untuk dianalisa maupun diintegrasikan dengan data terstruktur lainnya. Data

terstruktur biasanya disimpan dengan skema yang terdefinisi sehingga mudah untuk dilakukan query, dianalisa, dan diintegrasikan dengan data terstruktur lainnya. Berbeda dengan data tidak terstruktur, yang secara alami susah untuk dilakukan query, dianalisa, maupun diintegrasikan dengan sumber data lain. Tabel I dan Tabel II merupakan contoh input data pada penelitian ini. Komponen yang ada pada data input yang digunakan adalah Poin total kinerja, nilai kehadiran dan nilai perilaku.

TABEL I  
DATA REMUNERASI 70%

Poin Hitung Kinerja	Poin Hitung Penunjang	Poin Pakai Penunjang	Jumlah SK Penunjang	Poin Total Kinerja
372,9	1,375	1,375	3 / 3	374,3
9				6
126,2	9,025	7,375	18 / 24	133,6
3				0
130,2	15,166	12,141	18 / 29	142,4
8				2
168,9	12,100	9,075	18 / 29	178,0
7				5
73,26	3,733	3,733	7 / 7	76,99
0				3
134,0	1,650	1,650	6 / 6	135,7
9				4
84,64	5,383	5,383	13 / 13	90,02
6				9
75,75	2,475	2,475	9 / 9	78,22
1				6
110,3	13,791	11,316	18 / 27	121,6
5				6

TABEL II  
DATA REMUNERASI 70%

Nilai Kehadiran	Poin Kehadiran	Nilai Perilaku	Poin Perilaku	Total Poin
100,00	6,00	100,00	6,00	12,00
100,00	6,00	88,14	6,00	12,00
11,11	0,67	99,44	6,00	6,67
100,00	6,00	85,56	6,00	12,00
94,44	5,67	86,64	6,00	11,67
100,00	6,00	90,88	6,00	12,00
100,00	6,00	100,00	6,00	12,00
94,42	5,67	86,81	6,00	11,67
91,67	5,50	86,04	6,00	11,50
83,33	5,00	88,02	6,00	11,00
94,67	5,68	90,44	6,00	11,68
37,73	2,26	88,92	6,00	8,26

**B. Seleksi Data**

Proses seleksi data bertujuan untuk memilih variabel yang digunakan dalam proses *clustering*. Data yang tidak digunakan/dibuang adalah data tenaga kependidikan yang sudah berstatus pensiun. Pemilihan variabel dilakukan dengan



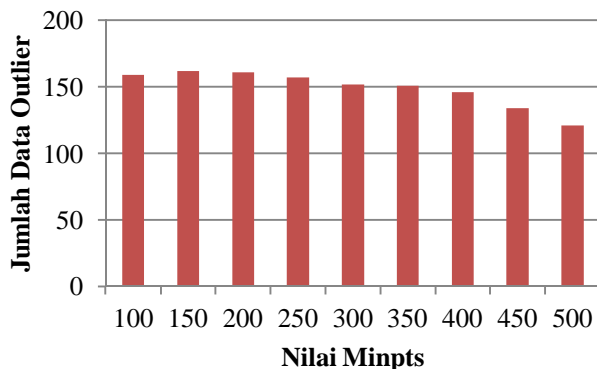
memilih variabel yang digunakan sebagai dasar pembayaran remunerasi. Dari variabel data yang ada pada Tabel I, variabel yang digunakan adalah nip, poin pakai kinerja. Sementara variabel yang digunakan pada tabel remun 30% adalah nip, nilai kehadiran dan nilai perilaku. Tabel III adalah tabel setelah melakukan pemilihan variabel dan integrasi kedua tabel tersebut.

TABEL III  
TABEL PEMILAHAN VARIABEL.

NIP	Perilaku	Kehadiran	Kinerja
'195403071980031004'	93,67	100	155,372
'195612311982111001'	81,6936	74,4136	125,395
'195710131980031003'	95,094	74,232	94,036
'195712311979011005'	89,8282	87,0991	152,6205
'195809051981031003'	87,895	90,9625	168
'195812311983032009'	86,8811	95,8778	156,09
'195907271984031001'	67,365	64,2213	84
'195909241986032003'	87,855	75,335	168
'195910051981031004'	90,895	62,1825	115,354
'195911251982112001'	78,99	82,625	141,6575

### C. Deteksi Outlier

Pada Gambar 1 ditampilkan grafik hasil pengujian deteksi outlier pada data remunerasi tenaga kependidikan tahun 2017. Pengujian terhadap deteksi outlier dengan menggunakan metode *local outlier factor* dilakukan dengan mengubah masukan parameter MinPts dengan nilai masukan adalah 100, 150, 200, 250, 300, 350, 400, 450, 500 dan berdasarkan pengujian, didapatkan bahwa parameter MinPts 150 merupakan parameter MinPts yang paling banyak mendeteksi data outlier sebanyak 162 data.



Gambar 1: Grafik hasil pengujian deteksi outlier

Tabel IV Menampilkan jumlah data yang terdeteksi sebagai outlier dengan nilai parameter MinPts yang berubah-ubah dan nilai batas outlier sebesar 1.4. Hasil deteksi tertinggi pada nilai MinPts 150 dengan presentase data terdeteksi outlier sebesar 22.98%. Dari hasil pengujian perubahan nilai MinPts, didapatkan bahwa semakin bertambahnya nilai MinPts maka data outlier yang ditemukan akan semakin berkurang. Hal ini menandakan bahwa semakin tinggi nilai MinPts sangat mempengaruhi nilai LOF dari suatu objek. Semakin tinggi

nilai MinPts berarti semakin luas pula kepadatan objek tersebut.

TABEL IV  
TABEL PEMILAHAN VARIABEL.

Minpts	Data Terdeteksi		Persentase	
	Outlier	Normal	Outlier	Normal
100	159	546	22,55%	77,45%
150	162	543	22,98%	77,02%
200	161	544	22,84%	77,16%
250	157	548	22,27%	77,73%
300	152	553	21,56%	78,44%
350	151	554	21,42%	78,58%
400	146	559	20,71%	79,29%
450	134	571	19,01%	80,99%
500	121	584	17,16%	82,84%

### D. Normalisasi Data

Pada tahapan normalisasi data, data diubah menjadi rentang 0-1 dengan menggunakan metode min-max. Tabel V adalah potongan tabel data sebelum dinormalisasi. Tabel VI adalah potongan tabel hasil normalisasi menggunakan metode normalisasi pada variabel perilaku, kehadiran, dan kinerja.

TABEL V  
TABEL DATA SEBELUM DINORMALISASI

No	NIP	Perilaku	Kehadiran	Kinerja
1	'195403071980031004'	93,67	100	155,372
2	'195612311982111001'	81,6936	74,4136	125,395
3	'195710131980031003'	95,094	74,232	94,036
4	'195712311979011005'	89,8282	87,0991	152,6205
5	'195809051981031003'	87,895	90,9625	168
6	'195812311983032009'	86,8811	95,8778	156,09
7	'195907271984031001'	67,365	64,2213	84
8	'195909241986032003'	87,855	75,335	168
9	'195910051981031004'	90,895	62,1825	115,354
10	'195911251982112001'	78,99	82,625	141,6575

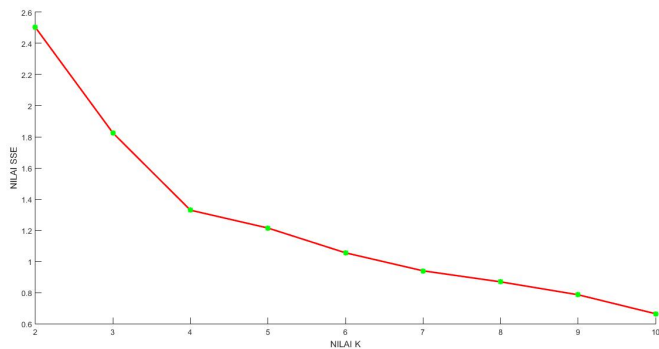
TABEL VI  
TABEL DATA SETELAH DINORMALISASI

No	NIP	Perilaku	Kehadiran	Kinerja
1	'195403071980031004'	0,9398	1	0,9185
2	'195612311982111001'	0,7572	0,708	0,725
3	'195710131980031003'	0,9615	0,7059	0,5226
4	'195712311979011005'	0,8812	0,8528	0,9007
5	'195809051981031003'	0,8517	0,8969	1
6	'195812311983032009'	0,8363	0,953	0,9231
7	'195907271984031001'	0,5387	0,5917	0,4579

8	'195909241986032003'	0,8511	0,7185	1
9	'195910051981031004'	0,8975	0,5684	0,6602
10	'195911251982112001'	0,7159	0,8017	0,83

**E. Penentuan Jumlah Cluster Optimum.**

Penentuan jumlah *cluster* optimal dengan menggunakan metode *elbow*. Gambar 2 adalah hasil dari proses metode *elbow*. Penurunan terjadi secara drastis dan diikuti grafik yang stabil pada titik 4.



Gambar 2: Grafik metode *elbow*

Tabel VII. Menunjukkan penurunan nilai SSE yang besar dari 2-3 dan 3-4, tetapi dari titik 4 dan selanjutnya terjadi penurunan yang stabil dan membentuk sudut siku pada titik 4 seperti yang ditunjukkan pada Gambar 2, sehingga nilai K yang digunakan adalah titik 4.

TABEL VII  
HASIL UJI METODE *ELBOW*.

K	SSE	Selisih
2	2,502471	0
3	1,824653	0,677818
4	1,328894	0,495759
5	1,214211	0,114683
6	1,055075	0,159136
7	0,93957	0,115505
8	0,868629	0,070942
9	0,785851	0,082777
10	0,663806	0,122045

**F. Clustering Menggunakan metode K-Means**

Proses hasil *clustering* menghasilkan tabel yang disimpan dalam *database* Hasil *cluster* dalam bentuk tabel yang dapat dilihat pada Tabel VIII.

TABEL VIII  
HASIL *CLUSTERING*.

Perilaku	Kehadiran	Kinerja	Jarak	Cluster
0,9398	1	0,9185	0,0811	2

0,8812	0,8528	0,9007	0,0814	4
0,8517	0,8969	1	0,0442	3
0,8363	0,9530	0,9231	0,0306	1
0,9460	0,7746	0,9548	0,1027	4
0,8550	0,9334	0,8762	0,0356	1
0,8908	0,9588	0,9272	0,0430	1
0,8618	0,7614	0,9409	0,0654	4
0,9618	0,9248	0,9351	0,0363	2
0,8563	0,9694	0,9350	0,0414	1
0,8595	0,9483	0,9586	0,0360	3
0,9354	0,8791	1	0,0682	2
0,9082	0,9150	1	0,0586	2
0,8753	0,9368	1	0,0219	3

Setelah itu data dipisahkan berdasarkan jumlah tiap-tiap *cluster*nya. *Cluster* 1 memiliki jumlah anggota 90, *cluster* 2 memiliki jumlah anggota 105, *cluster* 3 memiliki jumlah anggota 134 dan *cluster* 4 memiliki jumlah anggota 214.

**G. Uji Performa Cluster**

Metode *silhouette* dan *dunn index* mencari nilai tertinggi. Nilai *silhouette* tinggi menunjukkan data ditempatkan pada *cluster* yang tepat. Nilai *dunn* tinggi menunjukkan bahwa *cluster* berbeda dari *cluster* lainnya. Hasil metode ini ditunjukkan pada Tabel IX.

TABEL IX  
HASIL UJI DUNN DAN SILHOUTTE .

k	Silhouette	Dunn
2	0,512103	0,003121
3	0,528518	0,024445
4	0,542554	0,040361
5	0,489245	0,02124
6	0,457715	0,015977
7	0,449444	0,029326
8	0,435976	0,026804
9	0,448652	0,014316
10	0,480519	0,025445

Dilihat dari hasil pada Tabel IX, Nilai k=4 memiliki nilai *Silhouette* dan *Dunn* index tinggi. Hal ini sesuai dengan metode *elbow* yang digunakan untuk menentukan nilai k=4 merupakan nilai k yang optimal.

Sedangkan untuk Evaluasi *cluster* menggunakan *Purity* adalah sebagai berikut :

TABEL X  
HASIL UJI *PURITY*

Cluster	Jumlah	Baik	Sangat Baik	Purity	Purity (%)
---------	--------	------	-------------	--------	------------



Cluster 1	134	37	97	0.72	72%
Cluster 2	90	3	87	0.97	97%
Cluster 3	214	0	214	1	100%
Cluster 4	105	13	92	0.88	88%
Total	543				
Purity				0.89	89%

Tabel X menunjukkan hasil perhitungan yang diperoleh dengan nilai *purity* sebesar 0,89 atau 89%. Hasil yang diperoleh ini menunjukkan nilai akurasi dengan uji *purity* cukup bagus karena nilai *purity* mendekati 1.

#### H. Pelabelan Cluster

Setelah dilakukan proses *clustering*, data dikembalikan ke bentuk sebelum dinormalisasi agar lebih memudahkan dalam proses analisa hasil *cluster* tersebut.

Kriteria yang digunakan pada Metode SAW ini berjumlah 3 yaitu perilaku, kehadiran, dan kinerja, dalam Metode SAW kriteria tersebut diberikan bobot sesuai tingkat kepentingan kriteria tersebut.

Penentuan bobot kriteria perilaku dan kehadiran berdasarkan insentif kinerja minimal (gaji) diberikan sebesar 30% dari total nilai remunerasi per jabatan. Bobot kriteria kinerja ditentukan berdasarkan Insentif Kinerja Lebih diberikan sebesar 70% dari total nilai remunerasi per jabatan.

Terdapat 4 alternatif yang akan dilakukan proses perhitungan dengan Metode SAW. Penentuan alternatif berdasarkan jumlah *cluster* optimal yang telah ditentukan pada proses *clustering*. Nilai dari kriteria merupakan rata-rata dari nilai perilaku, kehadiran, dan kinerja pada tiap-tiap *cluster*. Alternatif dan nilai kriteria ditunjukkan pada Tabel XI.

TABEL XI  
ALTERNATIF DAN NILAI KRITERIA.

Alternatif	Kriteria		
	Avg(Perilaku)	Avg(Kehadiran)	Avg(Kinerja)
Cluster1	88,3334	94,3280	154,3033
Cluster2	94,4764	94,4718	163,0378
Cluster3	88,0889	94,7874	166,9786
Cluster4	88,1004	84,1811	163,3812

Tabel XII. Merupakan hasil dari perankingan *cluster*. Dari hasil yang ditampilkan pada Tabel XII. *Cluster* 3 merupakan *cluster* terbaik dengan nilai 0.9899, diikuti dengan *cluster* 2 dengan nilai 0.9830, *cluster* 4 dengan nilai 0.9580 dan *cluster* terburuk adalah *cluster* 1 dengan nilai 0.9364.

TABEL XII  
HASIL DARI PERANKINGAN.

Cluster	Saw	Rank
1	0,9364	4
2	0,9830	2
3	0,9899	1
4	0,9580	3

*Cluster* 1 merupakan *cluster* terburuk diantara *cluster* lainnya. Hal ini dikarenakan pada *cluster* 1 nilai rata-rata kinerja pada *cluster* ini terendah di antara *cluster* lain. Rentang nilai kinerja dari anggota *cluster* ini juga terendah diantara *cluster* lain.

*Cluster* 2 merupakan *cluster* terbaik kedua dengan rata-rata nilai kinerja terbesar kedua dari *cluster* lain, meskipun rata-rata nilai perilaku lebih besar dari *cluster* 4. Hal ini dikarenakan pembobotan kriteria kinerja lebih tinggi dari pada kriteria lainnya sehingga kriteria kinerja lebih menentukan perankingan dari pada kriteria lainnya.

*Cluster* 3 merupakan *cluster* terbaik dengan rata-rata nilai kinerja dan kehadiran yang tertinggi diantara *cluster* lain.

*Cluster* 4 merupakan *cluster* peringkat ketiga, dilihat dari rata-rata nilai kinerja pada *cluster* 4 lebih baik dari *cluster* 2, tetapi rata-rata nilai perilaku dan kehadiran jauh lebih kecil dibandingkan *cluster* 2 sehingga hasil setelah dilakukan perankingan *cluster* 4 lebih buruk dari *cluster* 2.

#### I. Visualisasi cluster dengan Box and whisker.

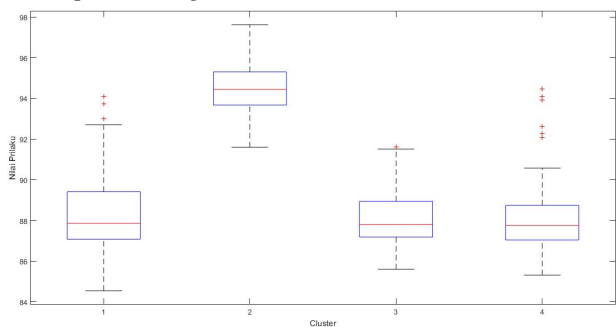
Grafik *Box and whisker* digunakan untuk menunjukkan rentang nilai variabel pada setiap *clusternya*. Grafik *Box and whisker* untuk variabel perilaku ditampilkan pada Gambar 3 Tabel XIII menampilkan nilai Min, Q1, median, Q3 dan nilai Max dari variabel perilaku pada setiap *clusternya*.

TABEL XIII  
VISUALISASI CLUSTER VARIABEL PERILAKU

	cluster 1	cluster 2	cluster 3	cluster 4
Min	84,545	91,595	85,598	85,312
Q1	87,079	93,670	87,180	87,044
Median	87,873	94,474	87,818	87,761
Q3	89,426	95,298	88,935	88,785
Max	94,085	97,618	91,620	94,465

Berdasarkan Gambar 3 analisis untuk variabel perilaku adalah sebagai berikut.

- 1) *Box Cluster* 2 berada lebih tinggi dari *box* lainnya menandakan nilai perilaku dari *cluster* 2 lebih tinggi dari *cluster* lainnya.
- 2) Bagian atas *box cluster* 1, *cluster* 3 dan *cluster* 4 lebih panjang daripada bagian bawahnya yang mengartikan nilai dibawah *median* lebih seragam dan nilai diatas *median* lebih beragam.
- 3) *Cluster* 1 memiliki *box* dan *whisker* yang lebih panjang dari *cluster* lainnya yang mengartikan bahwa nilai pada *cluster* 1 lebih beragam dari *cluster* lainnya



Gambar 3: Box and whisker prilaku

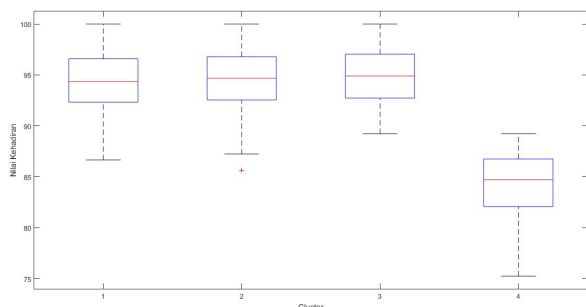
Grafik *Box and whisker* untuk variabel kehadiran ditampilkan pada Gambar 4. Tabel XIV menampilkan nilai Min, Q1, median, Q3 dan nilai Max dari variabel kehadiran pada setiap *cluster*nya.

TABEL XIV  
 VISUALISASI *CLUSTER* VARIABEL KEHADIRAN.

	cluster 1	cluster 2	cluster 3	cluster 4
Min	86,651	85,618	89,227	75,239
Q1	92,331	92,545	92,725	82,141
Median	94,394	94,711	94,896	84,709
Q3	96,626	96,780	97,036	86,819
Max	100,000	100,000	100,000	89,223

Berdasarkan Gambar 4, analisis untuk variabel kehadiran adalah sebagai berikut.

1. *Box cluster 4* lebih dibawah dari *cluster* lainnya yang menandakan bahwa nilai kehadiran pada *cluster 4* lebih rendah dari *cluster* lainnya
2. *Cluster 1* dan *cluster 2* memiliki *box* dan *whisker* yang hampir sama hal ini menandakan persebaran dari data pada *cluster 1* dan *cluster 2* hampir sama.
3. *Cluster 3* memiliki *box* dan *whisker* yang paling pendek dari *cluster* lain hal ini menandakan *cluster 3* memiliki nilai yang lebih seragam dari *cluster* lainnya.



Gambar 4: box and whisker kehadiran

Grafik *Box and whisker* untuk variabel kinerja ditampilkan pada Gambar 5. Tabel XV. menampilkan nilai Min, Q1,

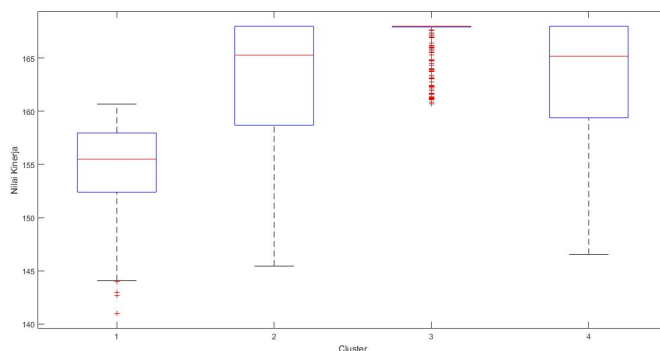
median, Q3 dan nilai Max dari variabel kinerja pada setiap *cluster*nya.

TABEL XV  
 VISUALISASI *CLUSTER* VARIABEL KINERJA.

	Cluster 1	Cluster 2	Cluster 3	Cluster 4
Min	140,975	145,444	160,709	146,552
Q1	152,392	158,684	167,924	159,463
Median	155,584	165,490	168,000	165,178
Q3	157,979	168,000	168,000	168,000
Max	160,671	168,000	168,000	168,000

Berdasarkan Gambar 5. analisis untuk variabel kinerja adalah sebagai berikut.

1. *Cluster 3* memiliki *box* terpendek yang menandakan nilai pada *cluster 3* paling seragam diantara *cluster* lainnya. *Cluster 3* memiliki nilai seragam dari nilai *median*, *quartile* tiga sampai nilai *maximum*
2. *Cluster 1* dibawah dari *cluster* lainnya yang mengartikan nilai dari kinerja pada *cluster 1* lebih rendah dari *cluster* lainnya.
3. *Cluster 2* dan *cluster 4* memiliki *box* yang lebih panjang pada bagian bawah yang mengartikan bahwa karakteristik data kinerja pada *cluster 2* dan *cluster 4* sama yaitu nilai dibawah *median* lebih beragam dan nilai diatas *median* lebih seragam.



Gambar 5: box and whisker kinerja

#### IV. KESIMPULAN

Kesimpulan yang dapat ditarik dari penelitian ini adalah Faktor yang cukup berpengaruh terhadap hasil deteksi *outlier* pada penelitian ini adalah penentuan *MinPts*. Berdasarkan beberapa percobaan yang dilakukan, Metode *Local Outlier Factor* dengan nilai *MinPts* 150 dapat mendeteksi data *outlier* paling banyak dengan jumlah data terdeteksi *outlier* sebanyak 162 data atau sebesar 22,98% dari total data. Metode *Elbow*, jumlah *cluster* yang digunakan pada pada penelitian ini berjumlah 4 *cluster* dengan nilai *Silhouette* sebesar 0,542, *Dunn* sebesar 0,040 dan *purity* sebesar 89%.

#### REFERENSI



- [1] Kemenristekdikti, "forlap.ristekdikti.go.id," diakses pada tanggal 3 april 2018, 2018. [Daring]. Tersedia pada: <https://forlap.ristekdikti.go.id/perguruan tinggi/homegraphpt> pada tanggal 3 april 2018.
- [2] N. G. Yudiarta, M. Sudarma, dan W. G. Ariastina, "Pengelompokan Berita Pada Unstructured Textual Data," *Maj. Ilm. Teknol. Elektro*, vol. 17, no. 3, hal. 339–344, 2018.
- [3] M. R. Ridlo, S. Defiyanti, dan A. Primajaya, "Implementasi Algoritma K-Means Untuk Pemetaan Produktivitas Panen Padi Di Kabupaten Karawang," in *CITEE 2017*, 2017, hal. 426–433.
- [4] widiarina, "Algoritma Cluster Dinamik Untuk Optimasi Cluster Pada Algoritma K-Means Dalam Pemetaan Nasabah Potensial Algoritma Cluster Dinamik Untuk Optimasi Cluster Pada Algoritma K-Means Dalam," *Tesis Magister Ilmu Komputer, Nusa Mandiri*, vol. 1, no. 1, hal. 33–36, 2013.
- [5] M. Nishom dan M. Y. Fathoni, "Implementasi Pendekatan Rule-Of-Thumb untuk Optimasi Algoritma K-Means Clustering," *J. Inform. J. Pengemb. IT*, vol. 3, no. 2, hal. 237–241, 2018.
- [6] K. G. Sharma, Y. Singh, dan A. K. Srivastava, "Variance on Factor," in *IMPACT*, 2017, hal. 101–103.
- [7] N. P. E. Merliana, E. Ernawati, dan A. J. Santoso, "ANALISA Penentuan Jumlah Cluster Terbaik Pada Metode K-Means Clustering," in *Prosiding seminar nasional multi disiplin ilmu*, 2015, hal. 978–979.
- [8] V. Bhatt, M. Dhakar, dan B. K. Chaurasia, "Filtered Clustering Based on Local Outlier Factor in Data Mining," vol. 9, no. 5, hal. 275–282, 2016.
- [9] N. Idham, "Penerapan Outlier Analysis Sebagai Salah Satu Rekomendasi Kelompok Belajar Terhadap Siswa Kelas 6 Di Sdn Pagelaran II Program Studi Teknik Informatika," Universitas Komputer Indonesia, 2017.
- [10] Z. Li *et al.*, "File yang so pernah download," *Jutei*, vol. 2, no. 2, hal. 23–32, 2018.
- [11] N. B. Hartono, "Analisis Outlier Dan Heteroskedastisitas Dengan Menggunakan Regresi Robust Weight Least Square," Universitas Negeri Semarang, 2016.
- [12] B. Santoso, I. Cholissodin, dan B. D. Setiawan, "Optimasi K-Means untuk Clustering Kinerja Akademik Dosen Menggunakan Algoritma Genetika," *J. Pengemb. Teknol. Inf. dan Ilmu Komput.*, vol. 1, no. 12, hal. 1652–1659, 2017.
- [13] G. Ngurah, W. Paramartha, D. E. Ratnawati, dan A. W. Widodo, "Analisis Perbandingan Metode K-Means Dengan Improved Semi-Supervised K-Means Pada Data Indeks Pembangunan Manusia (IPM)," vol. 1, no. 9, hal. 813–824, 2017.
- [14] A. R. Mamat, F. S. Mohamed, dan M. A. Mohamed, "Silhouette index for determining optimal k-means clustering on images in different color models," *106 Int. J. Eng. Technol.*, vol. 7, no. 2, hal. 105–109, 2018.
- [15] A. D. Savitri, F. A. Bachtar, dan N. Y. Setiawan, "Segmentasi Pelanggan Menggunakan Metode K-Means Clustering Berdasarkan Model RFM Pada Klinik Kecantikan ( Studi Kasus : Belle Crown Malang )," *J. Pengemb. Teknol. Inf. dan Ilmu Komput.*, vol. 2, no. 9, hal. 2957–2966, 2018.
- [16] Z. Arifin, S. Santosa, dan M. A. Soeleman, "Klasterisasi Genre Cerpen Kompas Menggunakan Agglomerative Hierarchical Clustering- Single Linkage," *J. Teknol. Inf.*, vol. 13, no. 2, hal. 92–100, 2017.