



Digital Receipt

This receipt acknowledges that **Turnitin** received your paper. Below you will find the receipt information regarding your submission.

The first page of your submissions is displayed below.

Submission author: **Ridowati Gunawan**
Assignment title: **Periksa similarity**
Submission title: **Deteksi Outlier Menggunakan Algor...**
File name: **Deteksi_Outlier_Menggunakan_Alg...**
File size: **550.09K**
Page count: **6**
Word count: **4,781**
Character count: **16,350**
Submission date: **23-Jun-2020 02:50PM (UTC+0700)**
Submission ID: **1348473887**

Konferensi Nasional Sistem Informasi 2014, STMK Diponegara Makassar, 27 Februari – 01 Maret 2014

KNSI2014-252
DETEKSI OUTLIER MENGGUNAKAN ALGORITMA LOCAL OUTLIER FACTOR (STUDI KASUS DATA AKADEMIK MAHASISWA UNIVERSITAS ABC)

Daniel Tomi Raharjo¹, Ridowati Gunawan²

^{1,2}Program Studi Teknik Informatika, Fakultas Sains dan Teknologi, Universitas Sanata Dharma, Universitas Sanata Dharma, Paingan Maguwarjo Depok Sleman Yogyakarta
¹dans.winer@gmail.com, ²rido@asd.ac.id

Abstrak

Pada saat melakukan pengamatan terhadap sekumpulan data, seringkali diinginkan untuk mendapatkan data-data yang tidak biasa. Data yang tidak biasa dari kumpulan data yang ada disebut dengan outlier. Proses untuk melakukan deteksi outlier dilakukan dengan menggunakan berbagai pendekatan. Salah satunya adalah dengan menggunakan pendekatan density based, dengan algoritma local outlier factor. Pada penelitian ini digunakan data akademik mahasiswa program studi Teknik Informatika Universitas ABC untuk angkatan 2007-2008 dari mulai test masuk sampai dengan prestasi akademik mahasiswa sampai dengan semester 4. Secara intuitif sering kali dapat dengan cepat diperoleh kesimpulan bahwa mahasiswa dengan nilai test yang baik akan diperoleh index prestasi tiap semester yang baik pula. Algoritma local outlier factor dapat digunakan untuk memperoleh mahasiswa-mahasiswa yang berbeda dari yang secara intuitif dapat disimpulkan. Hal ini sangatlah penting bagi dosen pembimbing akademik atau pengampu program studi untuk memberikan pemangku secara khusus secara cepat, sehingga tidak membawa pengaruh yang buruk seperti mahasiswa harus menundukkan diri atau terkena program khusus. Dengan menggunakan algoritma local outlier factor ditemukan beberapa mahasiswa yang ternyata outlier. Pencapaian prestasi akademik yang tidak sesuai dengan hasil test masuk maupun yang tidak sesuai dengan tingkat pencapaian rata-rata dari mahasiswa lainnya dapat diperoleh dengan menerapkan algoritma local outlier factor ini. Pengembangan perangkat lunak yang dilakukan dapat secara fleksibel untuk menemukan atribut-atribut yang akan dijadikan faktor penentu outlier.

Kata kunci : outlier, local outlier factor, data akademik.

1. Pendahuluan

1.1 Latar Belakang Masalah

Teknik untuk mengeni outlier dikenal dengan istilah Deteksi Outlier, dalam deteksi outlier terdapat 3 pendekatan yang secara umum sering dipakai yaitu Clustering Based, Distance Based dan Density Based. Dari ketiga pendekatan tersebut muncul beragam algoritma yang digunakan untuk mencari outlier dari sekumpulan data. Akan tetapi, tidak semua algoritma tersebut cocok digunakan untuk menganalisis sebuah outlier dari kelompok data, apalagi jika data tersebut memiliki dimensi yang tinggi atau memiliki atribut yang banyak. Salah satu algoritma yang ada adalah algoritma LOF (Local Outlier Factor) algoritma ini digunakan dalam pendekatan Density Based, algoritma ini tidak secara eksplisit menyatakan bahwa suatu obyek adalah sebuah outlier, namun algoritma memberikan bobot / derajat outlier terhadap suatu obyek nilai.

Dalam perkembangan teknik data mining juga digunakan untuk mencari berbagai bidang, mulai dari ekonomi, bisnis dan juga dalam bidang pendidikan, banyak sekali penelitian dilakukan dalam bidang pendidikan. Salah satu fungsi data mining dalam dunia pendidikan adalah untuk mengembangkan sebuah metode yang dapat menemukan keunikan dari sebuah data yang berasal dari sistem pendidikan tersebut, dan menggunakan metode tersebut untuk lebih memahami siswa, sehingga dapat mengembangkan sebuah sistem yang sesuai [2].

Universitas ABC adalah sebuah institusi pendidikan yang memiliki banyak data, salah satu data yang dimiliki adalah data akademik mahasiswa. Data akademik yang akan diteliti adalah data penerimaan mahasiswa baru (PMB) dan data nilai per semester dari setiap mahasiswa. Data PMB

KNSI 2014 1254

Deteksi Outlier Menggunakan Algoritma Local Outlier Factor

by Gunawan Ridowati

Submission date: 23-Jun-2020 02:50PM (UTC+0700)

Submission ID: 1348473887

File name: Deteksi_Outlier_Menggunakan_Algoritma_Local_Outlier_Factor.pdf (550.09K)

Word count: 4781

Character count: 16350

KNSI2014-252

DETEKSI OUTLIER MENGGUNAKAN ALGORITMA LOCAL OUTLIER FACTOR (STUDI KASUS DATA AKADEMIK MAHASISWA UNIVERSITAS ABC)

Daniel Tomi Raharjo¹, Ridowati Gunawan²

^{1,2} Program Studi Teknik Informatika, Fakultas Sains dan Teknologi, Universitas Sanata Dharma
Universitas Sanata Dharma, Paingan Maguwoharjo Depok Sleman Yogyakarta
¹ dans.winner@gmail.com, ² rido@usd.ac.id

Abstrak

Pada saat melakukan pengamatan terhadap sekumpulan data, seringkali diinginkan untuk mendapatkan data-data yang tidak biasa. Data yang tidak biasa dari kumpulan data yang ada disebut dengan *outlier*. Proses untuk melakukan deteksi *outlier* dilakukan dengan menggunakan berbagai pendekatan. Salah satunya adalah dengan menggunakan pendekatan *density based*, dengan algoritma *local outlier factor*. Pada penelitian ini digunakan data akademik mahasiswa program studi Teknik Informatika Universitas ABC untuk angkatan 2007-2008 dari mulai test masuk sampai dengan prestasi akademik mahasiswa sampai dengan semester 4. Secara intuitif sering kali dapat dengan cepat diperoleh kesimpulan bahwa mahasiswa dengan nilai test yang baik akan diperoleh index prestasi tiap semester yang baik pula. Algoritma *local outlier factor* dapat digunakan untuk memperoleh mahasiswa-mahasiswa yang 'berbeda' dari yang secara intuitif dapat disimpulkan. Hal ini sangatlah penting bagi dosen pembimbing akademik atau pengelola program studi untuk dapat memberikan penanganan secara khusus secara cepat, sehingga tidak membawa pengaruh yang buruk seperti mahasiswa harus mengundurkan diri atau terkena program khusus. Dengan menggunakan algoritma *local outlier factor* ditemukan beberapa mahasiswa yang ternyata *outlier*. Pencapaian prestasi akademik yang tidak sesuai dengan hasil test masuk maupun yang tidak sesuai dengan tingkat pencapaian rata-rata dari mahasiswa lainnya dapat diperoleh dengan menerapkan algoritma *local outlier factor* ini. Pengembangan perangkat lunak yang dilakukan dapat secara fleksibel untuk memasukkan atribut-atribut yang akan dijadikan faktor penentu *outlier*.

Kata kunci : *outlier*, *local outlier factor*, data akademik.

I. Pendahuluan

I.1 Latar Belakang Masalah

Teknik untuk mengenali *outlier* dikenal dengan istilah Deteksi Outlier, dalam deteksi *outlier* terdapat 3 pendekatan yang secara umum sering dipakai yaitu: *Clustering Based*, *Distance Based* dan *Density Based*. Dari ketiga pendekatan tersebut muncul beragam algoritma yang digunakan untuk mencari *outlier* dari sekumpulan data. Akan tetapi tidak semua algoritma tersebut cocok digunakan untuk menganalisis sebuah *outlier* dari kelompok data, apalagi jika data tersebut memiliki dimensi yang tinggi atau memiliki atribut yang banyak. Salah satu algoritma yang ada adalah algoritma LOF (*Local Outlier Factor*) algoritma ini digunakan dalam pendekatan *Density Based*, algoritma ini tidak secara eksplisit menyatakan bahwa suatu obyek adalah

sebuah *outlier*, namun algoritma memberikan bobot / derajat *outlier* terhadap suatu obyek nilai.

Dalam perkembangannya teknik data mining juga digunakan untuk meneliti berbagai bidang, mulai dari ekonomi, bisnis dan juga dalam bidang pendidikan, banyak sekali penelitian dilakukan dalam bidang pendidikan. Salah satu fungsi data mining dalam dunia pendidikan adalah untuk mengembangkan sebuah metode yang dapat menemukan keunikan dari sebuah data yang berasal dari sistem pendidikan tersebut, dan menggunakan metode tersebut untuk lebih memahami siswa, sehingga dapat mengembangkan sebuah sistem yang sesuai [2].

Universitas ABC adalah sebuah institusi pendidikan yang memiliki banyak data, salah satu data yang dimiliki adalah data akademik mahasiswa. Data akademik yang akan diteliti adalah data penerimaan mahasiswa baru (PMB) dan data nilai per semester dari setiap mahasiswa. Data PMB

terdiri dari data diri calon mahasiswa beserta dengan hasil test yang diikuti oleh calon mahasiswa tersebut. Dari hasil test tersebut akan ditentukan apakah calon mahasiswa tersebut diterima atau tidak menjadi mahasiswa pada Universitas ABC. Selain data PMB, data tiap semester juga disimpan.

Setiap akhir semester IV setiap Program Studi di Universitas ABC melakukan evaluasi pembelajaran terhadap hasil akademik yang diperoleh oleh mahasiswa. Proses ini dikenal dengan nama evaluasi sisip program. Evaluasi ini untuk mengetahui kemampuan setiap mahasiswa, dan untuk memutuskan apakah mahasiswa tersebut harus dipertahankan di prodi tersebut, atau harus dikeluarkan (DO). Dalam memutuskan hasil evaluasi tersebut maka seorang pembimbing akademik ataupun ketua program studi harus memperhatikan riwayat akademik setiap mahasiswa, yaitu berupa nilai test masuk dan juga nilai akademik dari semester 1 sampai semester 4, mahasiswa yang memiliki nilai test masuk tinggi maka akan memiliki potensi prestasi / nilai akademik yang tinggi, begitu pula sebaliknya mahasiswa yang memiliki nilai test masuk yang rendah maka akan memiliki potensi prestasi / nilai akademik yang rendah juga. Namun prediksi tersebut tidak selalu tepat, karena bisa saja mahasiswa yang memiliki nilai test masuk tinggi namun ternyata prestasinya biasa saja atau rendah, begitu pula sebaliknya. Sehingga dapat dikatakan mahasiswa yang memiliki pola tidak umum atau dikenal dengan istilah *outlier*. Untuk menghindari mahasiswa yang terkena proses evaluasi sisip program yang berakhir pada DO, maka perlu diambil tindakan dini dengan memperhatikan data mahasiswa yang berpotensi menjadi *outlier*, sehingga memudahkan pembimbing akademik maupun ketua program studi dalam mengambil keputusan.

Untuk itu penulis melakukan penelitian menggunakan algoritma *Local Outlier Factor* untuk mendeteksi *outlier* dari nilai akademik mahasiswa Universitas ABC, mulai dari nilai per semester mahasiswa dengan membandingkan nilai test masuk mahasiswa tersebut. Nilai test masuk yang akan dibandingkan adalah nilai Tes Potensi akademik yaitu meliputi komponen penalaran verbal, kemampuan numerik, penalaran mekanik, hubungan ruang dan Bahasa Inggris.

1.2 Tujuan

Tujuan Penelitian ini adalah menerapkan algoritma *Local Outlier Factor* ke dalam sebuah sistem untuk mendeteksi *outlier* dari data akademik mahasiswa.

1.3 Rumusan Masalah

1. Bagaimana algoritma *Local Outlier Factor* dapat mendeteksi *outlier* dari data nilai akademik mahasiswa?
2. Apakah algoritma *Local Outlier Factor* dapat mendeteksi *Outlier* data nilai akademik mahasiswa?

1.4 Metodologi Penelitian

Metodologi yang digunakan untuk menyelesaikan permasalahan adalah menggunakan metode KDD (*Knowledge Discovery in Database*), yang dikemukakan oleh Jiawei Han dan Kamber [4].

2. Dasar Teori

2.1 Outlier

Secara lebih sederhana *outlier* adalah data yang terlalu berbeda atau tidak konsisten dengan satu set data. *Outlier* dapat disebabkan oleh kesalahan eksekusi atau juga oleh pengaturan program secara default [3].

2.2 Metode Density-Based

Metode *density-based* tidak secara eksplisit mengklasifikasikan sebuah obyek adalah *outlier* atau bukan, akan tetapi lebih kepada pemberian nilai kepada obyek sebagai derajat kekuatan obyek tersebut. Ukuran derajat kekuatan ini adalah *local outlier factor (LOF)*. Pendekatan untuk pencarian *outlier* ini hanya membutuhkan sebuah parameter yaitu *MinPts*. *Minpts* adalah jumlah tetangga terdekat yang digunakan untuk mendefinisikan *Local Neighborhood* suatu obyek.

2.2.1 Algoritma Local Outlier Factor (LOF)

Secara sederhana algoritma *LOF* dideskripsikan sebagai berikut [3] :

1. Menghitung *k-distance* dari obyek p .
Tujuan dari perhitungan *k-distance* ini adalah untuk menentukan tetangga dari p , secara sederhana *k-distance* dari sebuah obyek p , adalah jarak maksimal dari P terhadap tetangga terdekatnya.
2. Menemukan *k-distance neighborhood* dari p .
k-distance neighborhood suatu objek p dinotasikan $N_k\text{-distance}(p)$, atau $N_k(p)$ dimana berisi setiap objek dengan jarak tidak lebih besar dari *kdistance* (p).
3. Menghitung *reachability distance* dari obyek, *Reachability distance* dari suatu obyek p terhadap obyek o adalah $\text{distance}(p, o)$ atau $k\text{distance}(o)$, dengan membandingkan keduanya dan dicari nilai yang maksimum sehingga, $\text{reach-dist}(p, o) = \max\{k\text{distance}(o), \text{distance}(p, o)\}$.
4. Menghitung kepadatan lokal dari obyek p .
Kepadatan lokal / *local reachability density* dari p di definisikan seperti pada persamaan (1) :

$$Ird_{MinPts}(P) = \frac{\sum_{o \in N_{MinPts}(P)} reach-dist_{MinPts}(P, o)}{|N_{MinPts}(P)|}$$

5. Menghitung dan membuat peringkat LOF dari setiap obyek. *Local outlier factor* dari P didefinisikan seperti pada persamaan (2):

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

dalam algoritma LOF sebuah obyek dikatakan sebagai outlier apabila memiliki nilai LOF yang tinggi atau menjauhi 1, sedangkan obyek yang memiliki nilai LOF rendah atau mendekati 1 maka obyek tersebut tidak dapat dikategorikan sebagai outlier. Nilai LOF yang tinggi mengindikasikan bahwa obyek tersebut memiliki kepadatan yang rendah terhadap lingkungannya sehingga berpotensi menjadi sebuah outlier.

3. Pengujian

Pengujian dilakukan untuk memastikan apakah sistem yang telah dibangun, dapat mendeteksi sebuah outlier dari data akademik mahasiswa Universitas ABC. Dalam pengujian ini data yang digunakan adalah data akademik mahasiswa Teknik Informatika angkatan 2007-2008, yang meliputi nilai tes masuk dan nilai akademik yaitu indeks prestasi semester (IPS) semester 1-4. Data yang diolah adalah data hasil pembuatan gudang data yang telah dilakukan oleh Rosa dkk [6] pada penelitian sebelumnya.

3.1 Hasil Pengujian Deteksi Outlier

Pengujian terhadap data dilakukan untuk menentukan apakah muncul data *outlier*. Beberapa pengujian yang dilakukan adalah :

1. Pengujian terhadap mahasiswa yang masuk dengan menggunakan jalur test angkatan 2007-2008. Dengan nilai *minPts*=10 dan batas *outlier* 1.4. Hasil pengujian dapat dilihat pada tabel 1.

Tabel 1. Hasil Pengujian Deteksi Outlier Jalur Test

Se- me- ster	No. Urut Mahasi- swa	LOF	ips1	ips2	ips3	ips4	final
1	28	1.9	1.2	-	-	-	1.12
	28	1.84	-	1.65	-	-	1.12
3	28	1.89	-	-	1.57	-	1.12
	74	1.75	-	-	0.19	-	2.56
	50	1.51	-	-	0.59	-	2.56

4	28	1.91	-	-	-	1.68	1.12
	67	1.48	-	-	-	0	2.04
	68	1.43	-	-	-	0	2.00

2. Pengujian deteksi outlier untuk jalur prestasi dengan menggunakan nilai *minPts*=20, dan batas outlier yang digunakan adalah 1,5. Jalur prestasi berdasarkan data nilai raport. Untuk itu perlu dilakukan proses normalisasi untuk nilai tersebut. Hasil normalisasi nilai raport diperlihatkan pada kolom final. Hasil pengujian dapat dilihat pada tabel 2.

Tabel 2. Hasil Pengujian Deteksi Outlier Jalur Prestasi.

Se- me- ster	No. Urut Mahasi- swa	LOF	ips1	ips2	ips3	ips4	final
1	76	3.49	.85	-	-	-	2.85
	124	2.21	1.45	-	-	-	2.90
	15	1.88	3.55	-	-	-	2.16
	79	1.61	1.85	-	-	-	2.98
	94	1.53	1.90	-	-	-	2.77
2	17	3.46	-	0.32	-	-	2.72
	124	3.15	-	0.44	-	-	2.90
	94	1.73	-	1.35	-	-	2.77
	107	1.75	-	3.92	-	-	3.33
3	17	4.68	-	-	0.69	-	2.72
	107	1.64	-	-	3.76	-	3.33
	3	1.51	-	-	3.78	-	2.78
4	88	2.72	-	-	-	0.94	2.89
	94	2.31	-	-	-	1.20	2.77
	122	1.97	-	-	-	1.47	3.05
	87	1.58	-	-	-	1.83	2.80

3. Pengujian deteksi outlier untuk jalur prestasi dan jalur test dapat dilihat pada tabel 3.

Tabel 3. Hasil Pengujian Deteksi Outlier Mahasiswa Jalur Prestasi Dan Jalur Test.

Se- me- ster	No. Urut Mahasi- swa	LOF	ips1	ips2	ips3	ips4	Final
1	76	2.37	0.85	-	-	-	2.85
	28	2.17	1.72	-	-	-	1.12
	26	1.74	1.44	-	-	-	2.96
	124	1.67	1.45	-	-	-	2.90
	27	1.62	4.00	-	-	-	2.32
	36	1.62	1.40	-	-	-	2.40
2	17	3.17	-	0.32	-	-	2.72
	124	2.31	-	0.44	-	-	2.90
	28	2.09	-	1.65	-	-	1.12
	107	1.83	-	3.92	-	-	3.33
	68	1.7	-	1.00	-	-	2.00

	5						
	74	1.64	-	1.04	-	2.56	
3	74	3.03	-	-	0.19	2.56	
	17	2.64	-	-	0.69	2.72	
	50	2.54	-	-	0.59	2.56	
	28	2.17	-	-	1.53	1.12	
	46	1.89	-	-	0.31	2.76	
	67	1.78	-	-	1.28	2.04	
	107	1.72	-	-	3.76	3.33	
4	67	2.35	-	-	-	2.04	
	68	2.35	-	-	-	2.00	
	47	2.2	L	-	-	0.05	2.40
	46	2.13	L	L	L	0.31	2.76
	28	1.96	-	-	-	1.68	1.12
	72	1.8	-	-	-	0.82	2.96
	39	1.72	-	-	-	0.83	2.44
	88	1.7	L	-	-	0.94	2.56
	74	1.65	-	-	-	0.92	2.56

3.2 Analisis Hasil Pengujian

Analisis dilakukan terhadap setiap hasil pengujian yang telah dijelaskan pada 3.1.

1. Dari hasil pengujian terhadap data untuk mahasiswa dengan jalur test sebagai berikut :
 - a. Pada semester 1 terdapat satu mahasiswa yang berpotensi menjadi *outlier* yaitu mahasiswa no 28, dimana derajat *outliernya* cukup tinggi yaitu 1.9, mahasiswa ini memiliki nilai **20** yang rendah, serta nilai masuk yang **14** rendah juga.
 - b. Pada semester 2 obyek yang menjadi *outlier* tetap mahasiswa no 28, karena mahasiswa tersebut masih memiliki nilai **11** ips yang rendah.
 - c. Pada semester 3 terdapat 3 orang mahasiswa yang berpotensi *outlier*, mahasiswa pertama masih sama dengan semester 1 dan 2 yaitu mahasiswa no 28, mahasiswa selanjutnya yaitu mahasiswa dengan no 74 dan 50. Mahasiswa no 74 memiliki nilai ips yang sangat rendah yaitu 0.19 dan memiliki nilai masuk yang tinggi, mahasiswa no 50 juga memiliki nilai ips yang sangat rendah yaitu 0.59 dan memiliki nilai tes **8** masuk yang tidak terlalu tinggi ataupun tidak terlalu rendah.
 - d. Pada semester 4 terdapat 3 orang mahasiswa yang memiliki potensi sebagai *outlier*, mahasiswa pertama masih sama dengan semester **14** sebelumnya yaitu mahasiswa dengan no 28, kemudian mahasiswa 67 dan 68. Mahasiswa no 67 dan 68 memiliki nilai ips **0**, dan nilai test masuk tidak terlalu tinggi atau tidak terlalu rendah.

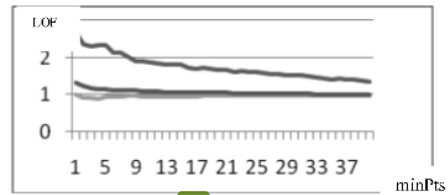
2. Dari pengujian data akademik mahasiswa jalur prestasi didapat hasil sebagai berikut :
 - a. Pada semester 1 terdapat 5 mahasiswa yang memiliki potensi sebagai *outlier* yaitu mahasiswa 76, 124, 15 79 dan 94. Mahasiswa-mahasiswa tersebut memiliki nilai ips yang rendah dan memiliki nilai final yang berada di tengah-tengah, tidak terlalu tinggi dan tidak terlalu rendah
 - b. Pada semester 2 terdapat 4 mahasiswa yang memiliki potensi sebagai *outlier* yaitu mahasiswa no 17, 124, 94 dan 107. Mahasiswa 94 dan **107** adalah mahasiswa yang memiliki nilai ips rendah dan nilai final yang berada di tengah-tengah, mahasiswa ini sebelumnya juga menjadi *outlier* pada semester 1, mahasiswa no 17 yang sebelumnya tidak berpotensi menjadi *outlier*, kini memiliki potensi **3** menjadi *outlier*, hal itu dikarenakan ips dari mahasiswa tersebut pada semester 2 sangat rendah yaitu 0.32. mahasiswa yang selanjutnya yaitu mahasiswa dengan no 107, berbeda dengan mahasiswa sebelumnya, mahasiswa ini berpotensi menjadi *outlier* karena memiliki nilai IPS yang sangat tinggi yaitu 3.92, serta nilai final yang juga tinggi.
 - c. Pada semester 3 terdapat 3 mahasiswa yang memiliki potensi sebagai *outlier*, mahasiswa no 17 menjadi *outlier* karena memiliki nilai ips yang rendah sedangkan mahasiswa no 107 dan no 3 menjadi *outlier* karena memiliki nilai ips yang cenderung lebih tinggi dari yang lainnya.
 - d. Pada semester **16** terdapat 4 mahasiswa yang berpotensi menjadi *outlier* mahasiswa tersebut adalah mahasiswa no 88, 94, 122 dan 87. Keempat mahasiswa tersebut memiliki nilai ips yang tergolong rendah dan nilai final yang sedang.
3. Dari pengujian data akademik mahasiswa untuk seluruh jalur yaitu jalur prestasi dan jalur test didapat hasil sebagai berikut :
 - a. Pada semester 1 terdapat 6 buah mahasiswa yang berpotensi menjadi *outlier* mahasiswa tersebut yaitu mahasiswa no 76, 28, 26, 124, 27 dan 36. Mahasiswa 76, 28 dan 124 adalah mahasiswa yang juga berpotensi menjadi *outlier* pada pengujian sebelumnya yaitu pengujian dengan memisahkan jalur prestasi dan jalur test. Sedangkan mahasiswa 28, 27 dan 36, pada pengujian sebelumnya yaitu pengujian mahasiswa 2007-2008 jalur test, tidak memiliki nilai LOF yang tinggi, namun pada pengujian dengan menggabungkan kedua data, mahasiswa tersebut memiliki nilai LOF yang cenderung tinggi. Mahasiswa 76, 28, 26 27 dan 36 berpotensi menjadi *outlier* karena memiliki nilai ips yang cenderung rendah daripada yang

- lainnya dan mahasiswa no 27 menjadi outlier karena memiliki nilai i_4 yang cenderung tinggi dari yang lainnya.
- Pada semester 2 diperoleh 6 mahasiswa yang berpotensi menjadi outlier karena memiliki nilai LOF yang cenderung tinggi yaitu mahasiswa no 17, 124, 28, 107, 68, 74. Pada pengujian jalur test dan jalur prestasi di semester 2 mahasiswa no 17, 124, 107 dan mahasiswa 28, 17 juga berpotensi sebagai outlier. Sedangkan mahasiswa no 68 pada pengujian jalur test semester 2 tidak memiliki nilai LOF yang cenderung tinggi, namun pada pengujian kali nilai LOF nya menjadi cenderung tinggi.
 - Pada semester 3 diperoleh 7 mahasiswa yang berpotensi menjadi outlier, mahasiswa tersebut yaitu mahasiswa no 74, 17, 50, 28, 46, 67 dan 107. Mahasiswa 74, 17, 50, 28, 67 dan 107 adalah mahasiswa yang memang memiliki potensi sebagai outlier pada pengujian jalur prestasi dan jalur test, sementara mahasiswa no 46 sebelumnya tidak memiliki nilai LOF yang tinggi dan tidak berpotensi sebagai outlier.
 - Pada semester 4 terdapat 9 mahasiswa yang memiliki potensi sebagai outlier mahasiswa tersebut yaitu mahasiswa no 67, 68, 47, 46, 28, 72, 39, 88, 74. Mahasiswa yang berpotensi menjadi outlier pada semester ini adalah mahasiswa yang nilai ips nya rendah.
- Dalam pengujian diatas tiap semester menghasilkan hasil yang berbeda-beda.
 - Saat data pengujian digabung, data hasil outlier menjadi berbeda dengan data pengujian saat dipisah. sebagian besar hasil outlier nya sama, namun terkadang terdapat mahasiswa yang tadinya tidak memiliki nilai LOF yang tinggi, namun pada saat digabung memiliki nilai LOF yang cenderung tinggi. Hal itu terjadi karena pertambahan jumlah data dan perbedaan penggunaan atribut.
 - Dari hasil review dan analisa oleh ketua program studi, hasil deteksi outlier diatas memang memiliki makna sebagai outlier, sehingga menurut ketua program studi data hasil pengujian dapat diterima sebagai outlier.

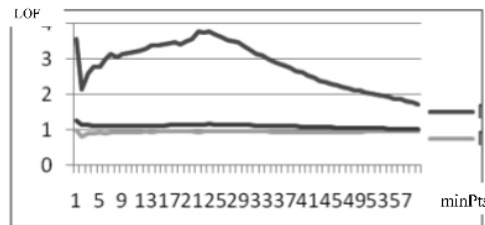
3.3 Pengujian Efek Perubahan Nilai Atribut $minPts$ Terhadap Nilai LOF

Algoritma LOF dipengaruhi oleh satu nilai atribut penting yang $minPts$.

- Pengujian untuk data mahasiswa angkatan 2007-2008 semester 1 jalur test terhadap jumlah $minPts$. dapat dilihat pada gambar 1.

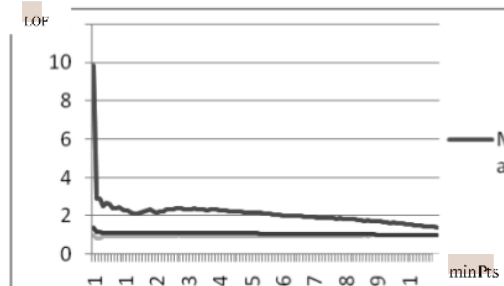


Gambar 1. Grafik perubahan $minPts$ untuk jalur test



Gambar 2. Grafik perubahan $minPts$ untuk jalur prestasi

- Pengujian Data Mahasiswa angkatan 2007-2008 semester 1 jalur prestasi dan jalur test.



3.4 Kesimpulan Hasil Pengujian Efek Nilai Atribut $minPts$ Terhadap Nilai LOF

Dari hasil percobaan dengan melakukan perubahan nilai $minPts$ maka dapat disimpulkan bahwa :

- Pada pengujian pertama yaitu dengan data mahasiswa 2007-2008 jalur test semester 1, nilai rata-rata LOF (mean) cenderung $minPts$ stabil pada saat $minPts$ 5-20.
- Pada pengujian kedua yaitu dengan data mahasiswa 2007-2008 jalur Prestasi, nilai rata-rata LOF (mean) cenderung stabil pada kisaran $minPts$ 8-38.
- Pada pengujian ketiga yaitu dengan data Mahasiswa 2007-2008 Jalur Prestasi dan Regular, nilai rata-rata LOF(mean) cenderung stabil pada kisaran $minPts > 7$.
- Dalam hal ini dapat disimpulkan bahwa nilai LOF sangat bergantung pada kepadatan jarak

obyek data dalam suatu jangkauan $minPts$, ketika suatu cluster memiliki kepadatan yang tinggi maka nilai LOF akan cenderung kecil, dan ketika suatu cluster memiliki kepadatan yang rendah maka nilai LOF akan cenderung lebih besar

4. Kesimpulan

Dari hasil penelitian ini, dengan menggunakan algoritma *Local Outlier factor* untuk melakukan deteksi outlier terhadap data akademik mahasiswa dapat ditarik kesimpulan sebagai berikut :

1. Algoritma *Local Outlier Factor* dapat mendeteksi outlier dari data akademik mahasiswa
2. Nilai atribut $minPts$ sangat mempengaruhi nilai derajat LOF dari suatu obyek data
3. Algoritma LOF tidak menyatakan secara eksplisit bahwa suatu obyek data merupakan sebuah outlier, penentuan obyek data mana saja yang akan menjadi sebuah outlier, ditentukan oleh pengguna dengan melihat besaran nilai derajat outliernya.

Daftar Pustaka:

- [1] Agymang, M., & Ezeife, C. I. (n.d.), *LSC-Mine: Algorithm for Mining Local Outliers*.
- [2] Baker, R. S., & Yacof, K., 2009, *The State of Educational Data Mining in 2009: A Review and Future Visions*
- [3] Breunig, M. M., Kriegel, H.-P., Ng, R. T., & Sander, J., 2000, *LOF: Identifying Density-Based Local Outliers. ACM 2000*.
- [4] Han, J., & Kamber, M., 2006, *Data Mining Concept and Techniques*, San Fransisco: Elsevier.
- [5] Rosa, P. H., Gunawan, R., & Wijono, S. H. (2013). The Development of Academic Data Warehouse as a Basic for Decision Making A Case Study at XYZ University. *International Conference an Enterprise Information Systems and Application*.

Deteksi Outlier Menggunakan Algoritma Local Outlier Factor

ORIGINALITY REPORT

13%

SIMILARITY INDEX

1%

INTERNET SOURCES

8%

PUBLICATIONS

6%

STUDENT PAPERS

PRIMARY SOURCES

- 1** Fankhänel, Beate (Prof. Dr. Eberhard Müller, Prof. Dr. Günther Frischat, Prof. Dr. Wolfram Beier and TU Bergakademie Freiberg, Werkstoffwissenschaft und Werkstofftechnologie). "Funktionale langfaserverstärkte Glasmatrixkomposite", Technische Universität Bergakademie Freiberg Universitätsbibliothek "Georgius Agricola", 2009.
Publication 2%
- 2** Submitted to Yonsei University
Student Paper 2%
- 3** Submitted to Swinburne University of Technology
Student Paper 1%
- 4** Submitted to The University of Manchester
Student Paper 1%
- 5** www.ece.neu.edu
Internet Source 1%

Kreupl, Franz. "Coulomb-Blockade bei

6

Raumtemperatur in selbstorganisierten Arrays von Pt-Clustern", Publikationsserver der Universität Regensburg, 2000.

Publication

1%

7

Polklas, Thomas. "Entwicklung eines numerischen Verfahrens zur strömungsmechanischen Auslegung des Abströmgehäuses einer Niederdruck-Dampfturbine", DuEPublico: University of Duisburg-Essen Publications Online, 2004.

Publication

1%

8

Dorothee Peters, Wilko Heitkoetter, Rasmus Voelker, Axel Moeller et al. "Validation of an open source high voltage grid model for AC load flow calculations in a delimited region", IET Generation, Transmission & Distribution, 2020

Publication

1%

9

Lautemann, Sven-Eric. "Schemaevolution in objektorientierten Datenbanksystemen auf der Basis von Versionierungskonzepten", Publikationsserver der Goethe-Universität Frankfurt am Main, 2003.

Publication

<1%

10

Submitted to University of New England

Student Paper

<1%

11

Tiwari, Sushma. "Situation of Landless Madheshi Dalit: A Case Study of Lalparsa VDC

<1%

of Parsa District", International Journal of Social Sciences and Management, 2014.

Publication

12

www.drumpool.com

Internet Source

<1%

13

www.morningstar.com

Internet Source

<1%

14

Submitted to National University of Ireland,
Galway

Student Paper

<1%

15

Schröder, Kathrin(Spohn, Tilman). "Eine kompakte Sonde für Temperatur- und Wärmeleitfähigkeitsmessungen in den Geowissenschaften", Münster University, Germany, Document Server, 2006.

Publication

<1%

16

Lindner, Guido. "Algorithmenauswahl im KDD-Prozess", Universität Karlsruhe, 2005.

Publication

<1%

17

Submitted to University of Arizona

Student Paper

<1%

18

Erich Novak, Harald Pfeiffer. "Coin Tossing Algorithms for Integral Equations and Tractability", Monte Carlo Methods and Applications, 2004

Publication

<1%

19

Jain, Brijnesh Johannes. "Structural Neural Learning Machines", Technische Universität Berlin, 2006.

Publication

20

Pantle, Ursa. "Asymptotic properties of estimators for random fields induced by stationary germ-grain models", Universität Ulm. Fakultät für Mathematik und Wirtschaftswissenschaften, 2006.

Publication

21

Submitted to Universidade do Porto

Student Paper

22

Simon, Bernd-Paul. "Ein Annotationsschema für deutsche Sätze", Technische Universität Berlin, 2006.

Publication

23

Chirstoph Flamm, Ivo L. Hofacker, Peter F. Stadler. " RNA The Computational Biology of RNA Secondary Structures ", Advances in Complex Systems, 2011

Publication

24

Scheinast, Werner (and Technische Universität Dresden, Mathematik und Naturwissenschaften, Physik, Institut für Kern-und Teilchenphysik, Institut für Kern- und Hadronenphysik des FZ Rossendorf). "Schwellennahe Erzeugung von Kaonen und Antikaonen in Proton-Kern-

Stößen", Saechsische Landesbibliothek- Staats- und Universitaetsbibliothek Dresden, 2004.

Publication

25

Submitted to University of Melbourne

Student Paper

<1%

26

Matzke, Dirk. "Mehrdimensionale Multiressourcenplanung mit Constraintlösern", Technische Universität Berlin, 2006.

Publication

<1%

27

Diana Gordon. "Chapter 36 Panel Discussion: "Empirical Versus Formal Methods"", Springer Nature, 2001

Publication

<1%

28

Marlene Almeida de Ataíde, Jayson Vaz Guimarães. "JUVENTUDE(s), VIOLÊNCIA URBANA E PERIFERIA EM SÃO PAULO: A CIDADE E SUAS CONTRADIÇÕES", Revista da Anpege, 2012

Publication

<1%

Exclude quotes On

Exclude matches < 5 words

Exclude bibliography On