

## ANALISIS DATA *TRACER STUDY* DENGAN MENGIDENTIFIKASI *OUTLIER* MENGUNAKAN TEKNIK *DATA MINING*

Oleh :

Dwi Welly Sukma Nirad<sup>1</sup>, Kridanto Surendro<sup>2</sup>

Sistem Informasi<sup>1,2</sup>

STMIK Indonesia Padang<sup>1</sup>, Institut Teknologi Bandung<sup>2</sup>

[dwiwelly@stmikindonesia.ac.id](mailto:dwiwelly@stmikindonesia.ac.id)

---

### Abstrak

Melakukan *drop out* ataupun permintaan pengunduran diri terhadap mahasiswa merupakan salah satu hal yang dihindari oleh perguruan tinggi karena dapat mempengaruhi rating dan penilaian terhadap kampus. Perguruan tinggi akan meminta mahasiswa mengundurkan diri berdasarkan beberapa ketentuan, salah satunya jika nilai mahasiswa terbilang sangat rendah. Mahasiswa dengan nilai rendah diasumsikan tidak akan meraih penghasilan yang tinggi di masa depan, atau disebut tidak sukses. Oleh karenanya penelitian ini mengangkat data *tracer study*, sehingga bisa terlihat apakah asumsi tersebut tepat. Alumni yang *drop out* atau mengundurkan diri tetapi mampu memperoleh penghasilan tinggi ditentukan sebagai standar outlier dalam penelitian ini. Data outlier diperoleh dengan menggunakan teknik data mining, yaitu teknik *association rule mining*. Teknik ini membantu menemukan rule yang tepat dalam menentukan mahasiswa outlier. Hasil penelitian menunjukkan validitas derajat mahasiswa outlier serta rekomendasi keputusan untuk perguruan tinggi dalam menangani mahasiswa yang teridentifikasi sebagai outlier.

*Kata-kata kunci:* *tracer study*, *outlier*, *association rule mining*

### Abstract

Doing *drop out* or request for resignation to students is one of the things avoided by the college because it can affect the rating and assessment of the campus. Colleges will ask students to resign based on some provisions, one of them if the value of students is very low. Students with low grades are assumed not to achieve a high income in the future, or called unsuccessful. Therefore this study raised the data *tracer study*, so it can be seen whether the assumption is appropriate. Alumni who *drop out* or resign but are able to earn high income are determined as outlier standards in this study. Outlier data is obtained using data mining techniques, namely *association rule mining* techniques. This technique helps to find the right rule in determining student outlier. The results showed the validity of outlier student degrees as well as the decision recommendation for the college in handling the students identified as outliers.

*Keywords:* *tracer study*, *outlier*, *association rule mining*

---

## 1. PENDAHULUAN

Pada dasarnya, perguruan tinggi terikat dengan elemen-elemen yang berada di lingkungannya, salah satunya dengan mahasiswa. Kebanyakan kasus yang menjadi ketakutan bagi mahasiswa adalah berhenti di tengah masa studi, baik melalui *drop out* atau permintaan pengunduran diri dari kampus. Pemberhentian mahasiswa dapat mempengaruhi kedua belah pihak. Bagi kampus dapat mempengaruhi nilai kredibilitasnya, sedangkan bagi mahasiswa dapat mempengaruhi kehidupan masa depan mereka. Padahal pemberhentian mahasiswa belum tentu merupakan keputusan yang tepat. Mahasiswa yang diberhentikan bisa saja mahasiswa yang akan sukses di masa depan. Sukses dalam penelitian ini

dideskripsikan sebagai orang yang memiliki penghasilan tinggi.

*Tracer study* adalah studi pelacakan terhadap alumni lulusan perguruan tinggi. *Tracer study* juga menyediakan informasi berharga mengenai hubungan antara pendidikan tinggi dan dunia kerja profesional, menilai relevansi pendidikan tinggi, informasi bagi pemangku kepentingan (*stakeholders*), dan kelengkapan persyaratan bagi akreditasi pendidikan tinggi (Roviati, et al., 2015). Pada penelitian ini, data *tracer study* dimanfaatkan untuk menganalisis karakteristik mahasiswa *outlier*.

*Outlier* merupakan data atau kelompok data yang bersifat menyimpang dari data lainnya, yang bisa juga mengacu sebagai abnormalitas, kesumbangan, penyimpangan, atau anomali (Aggarwal, 2013). Definisi

*outlier* pada setiap kasus akan berbeda karena *outlier* memiliki beberapa klasifikasi. Objek *outlier* dalam penelitian ini adalah mahasiswa, dengan definisi bahwa mahasiswa yang disebut *outlier* adalah mahasiswa yang berhenti di tengah masa studi (*drop out*) dan memiliki penghasilan tinggi pada masa berkarir.

Meskipun *outlier* berjumlah sedikit, namun keberadaan *outlier* dapat mempengaruhi proses pengambilan keputusan. Mahasiswa yang teridentifikasi sebagai *outlier* perlu dipertimbangkan untuk tidak langsung diberhentikan. Oleh karenanya perlu mengenali karakteristik mahasiswa tersebut. Melalui pengetahuan tentang karakteristik mahasiswa yang *outlier*, perguruan tinggi dapat melakukan tindakan tertentu sebelum benar-benar memberhentikan mahasiswa.

Salah satu cara untuk mengidentifikasi *outlier* dapat dilakukan dengan teknik *data mining*. *Data mining* akan dapat membantu menemukan hubungan yang tidak terduga dan untuk menyimpulkan data dalam cara baru yang bisa dipahami dan berguna bagi pemilik data (Hand, et al., 2001). Penelitian ini memanfaatkan salah satu teknik *data mining* dalam menemukan data *outlier*, yaitu *association rule mining*.

## 2. STUDI LITERATUR

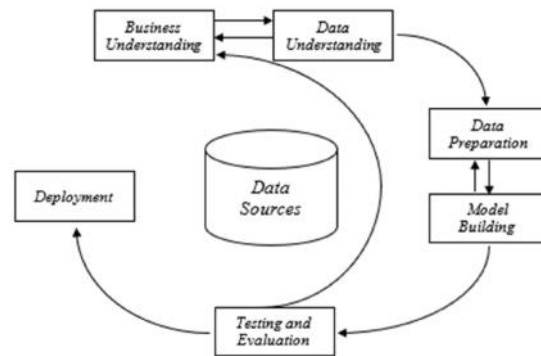
### a. *Data mining*

*Data mining* merupakan analisis dari pengamatan *data set* (biasanya berukuran besar), untuk menemukan hubungan yang tidak terduga dan untuk menyimpulkan data dalam cara baru yang dapat dipahami dan berguna bagi pemilik data (Hand, et al., 2001). Ditemukan juga pemahaman lain yang menyatakan bahwa *data mining* adalah proses dalam menemukan pola menarik dan *knowledge* dari sejumlah besar data (Han, et al., 2012). Sebagian orang menganggap *data mining* sama dengan *Knowledge Discovery from Data* (KDD). Namun lebih banyak yang sepakat mengatakan bahwa *data mining* dan KDD tidaklah sama, karena *data mining* merupakan salah satu langkah yang harus dilewati dalam proses KDD. Ilustrasinya dapat dilihat pada gambar 1.



Gambar 1 . Proses KDD dalam *database* (Foguem, et al., 2013)

Terdapat beberapa proses yang menjadi standar umum dalam pengerjaan *data mining*. Proses tersebut dibentuk oleh para peneliti dan praktisi sehingga menjadi panduan untuk mencapai best practice saat melakukan penambangan data. Salah satunya adalah proses *Cross-Industry Standard Process Data Mining* (CRISP-DM). Proses CRISP-DM ini telah diusulkan sejak tahun 1990-an. Proses tersebut digambarkan seperti gambar 2.



Gambar 2. Proses CRISP-DM (Olson & Delen, 2008)

Dari gambar 2, dapat dilihat bahwa proses CRISP-DM memiliki enam tahap. Keenam tahapan inilah yang akan membantu analisis *data mining* menjadi lebih optimal. Cakupan kerja enam tahap proses CRISP-DM menurut (Olson & Delen, 2008) yaitu:

#### 1. *Business understanding*

Tahap pemahaman bisnis mencakup penentuan tujuan bisnis, penilaian situasi saat ini, menentukan tujuan *data mining*, dan membangun perencanaan proyek.

#### 2. *Data understanding*

Tahap pemahaman data mencakup mengumpulkan data, mendeskripsikan data, menggali data, dan melakukan verifikasi kualitas data.

#### 3. *Data preparation*

Tahap persiapan data melingkupi pemilihan data, pembersihan, serta

menyesuaikan menjadi bentuk yang dibutuhkan, serta melakukan formatting bila dibutuhkan.

#### 4. *Model building*

Terdapat beberapa model atau teknik yang dapat digunakan dalam *data mining*. Setelah memahami data yang diperoleh secara mendalam, maka selanjutnya dapat dilakukan pemilihan model yang sesuai dengan data yang akan digunakan. Kesalahan dalam pemilihan model dapat mengurangi kualitas analisis data.

#### 5. *Testing and evaluation*

Model yang dipilih dan telah dibangun akan diuji serta dievaluasi pada tahap ini. Tahap ini akan menghasilkan identifikasi kebutuhan lain. Jika ternyata masih terdapat kebutuhan yang belum terpenuhi, maka fase akan diulang dari tahap awal, karena proses CRISP-DM adalah sebuah siklus yang berulang.

#### 6. *Deployment*

*Data mining* dapat digunakan untuk memverifikasi hipotesis sebelumnya, atau untuk menemukan *knowledge*. Jika telah mencapai tujuan, maka tahap selanjutnya adalah menerapkan model terpilih ke dalam proses bisnis. Model akan terus dipantau untuk kemungkinan terjadinya perubahan kondisi operasi, karena apapun yang terlihat benar saat ini belum tentu benar pada setahun mendatang.

### b. *Association rule mining*

*Association rules* adalah salah satu teknik *data mining* yang sederhana tetapi sangat berguna dalam pencarian pola pada *database* yang besar. Metodologi *Association rules*, atau analisis asosiasi adalah sebuah metodologi untuk mencari relasi (asosiasi) istimewa/menarik yang tersembunyi dalam himpunan data (atau *data set*) yang besar (Aprilla, et al., 2013).

Dalam menentukan *association rule* terdapat suatu interestingness measure (ukuran ketertarikan) yang didapatkan dari hasil pengolahan data dengan perhitungan tertentu. Ada dua ukuran, yaitu (Aprilla, et al., 2013):

- *Support*: bagian transaksi yang mengandung kedua X dan Y,

$$\text{support}(A) = \frac{\text{jumlah transaksi mengandung } A}{\text{total transaksi}}$$

- *Confidence*: seberapa sering item dalam Y muncul di transaksi yang mengandung X.

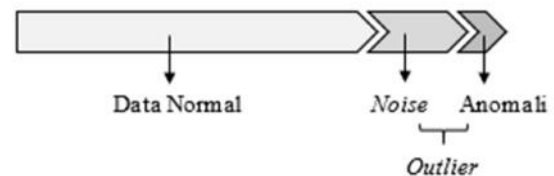
$$\text{confidence} = P(B|A) = \frac{\text{jumlah transaksi mengandung } A \text{ dan } B}{\text{jumlah transaksi mengandung } A}$$

Proses tersebut diturunkan menjadi langkah kerja sebagai berikut (Sharma & Verma, 2014):

1. Kumpulkan kandidat *itemset* sejumlah k, yang dihasilkan oleh 1-ekstensi dari besaran (k-1) -itemset, yang merupakan hasil dari iterasi sebelumnya.
2. Periksa nilai *support* untuk kandidat *k-itemsets* yang dihasilkan melalui *database*.
3. *Itemset* yang tidak mencapai nilai *support* minimum akan ditolak dan *itemset* yang tersisa disebut *large k-itemsets*.
4. Ulangi proses hingga tidak ada lagi *large itemset* dalam *database*.

### c. *Outlier*

Aggarwal (2013) menjelaskan, pada kebanyakan aplikasi, data dibuat dari satu atau lebih proses yang dapat merefleksikan aktivitas dalam sistem atau pengumpulan observasi tentang entitas. Ketika menghasilkan proses dalam cara yang tidak biasa, ini dapat membuat *outlier*. Meskipun begitu, *outlier* sering berisi informasi berguna tentang karakteristik sistem dan entitas yang tidak normal, yang mempengaruhi proses penghasilan data. Lebih lanjut, Aggarwal mengilustrasikan spektrum dari data normal ke *outlier* seperti gambar 3.



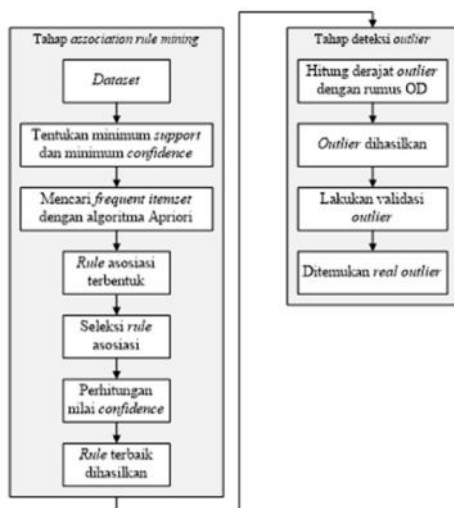
Gambar 3. Spektrum dari data normal ke *outlier* (Aggarwal, 2013)

*Noise* biasanya berkaitan dengan *outlier*. Ada pula yang menyebut bahwa *noise* merupakan *outlier* lemah. *Noise* dapat didefinisikan sebagai fenomena dalam data yang sebenarnya tidak menimbulkan interest untuk dianalisis, namun menghalangi analisis

data sehingga mengharuskan adanya tahap penghapusan *noise* sebelum analisis data ditampilkan (Singh & Upadhyaya, 2012). Umumnya *outlier* dikelompokkan ke dalam tiga kategori: global *outlier*, kontekstual (kondisional) *outlier*, dan kolektif *outlier* (Han, et al., 2012).

### 3. METODE PENELITIAN

Metode dalam penelitian ini dilakukan dengan mengkombinasi antara teknik *association rule mining* dengan teknik perhitungan derajat *outlier*. Teknik menghitung derajat *outlier* diadopsi dari penelitian terdahulu (Narita & Kitagawa, 2008). Perhitungan derajat *outlier* dapat menghasilkan jumlah data yang dinilai sebagai *outlier*. Maka, metode dalam penelitian ini dapat dipahami melalui gambar 4.



Gambar 4. Metode penelitian

Adapun langkah kerja pada penelitian ini dimulai dari pemilihan *data set*. *Data set* yang akan diteliti adalah *data set* yang tergolong sebagai big data. Selanjutnya dilakukan proses penggalian data (*data mining*). Pentingnya penggalian data dikarenakan kondisi umum data saat ini, yaitu kaya data namun miskin informasi. Proses penggalian data bertujuan untuk mendapatkan pola yang akan menghasilkan informasi berkualitas dan menemukan *knowledge* yang tersembunyi dari data tersebut. Dengan *knowledge* yang dimiliki,

keputusan yang diambil dapat menjadi tepat sasaran karena berdasarkan fakta dan kebutuhan organisasi. Untuk lebih jelasnya dapat dilihat pada gambar 5.

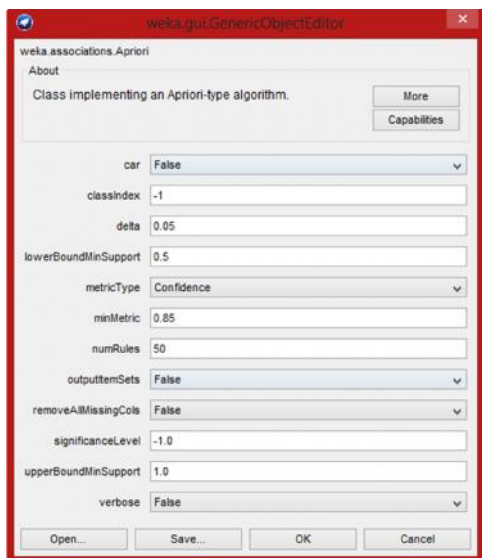


Gambar 5. Langkah kerja penelitian

### 4. HASIL PENELITIAN DAN PEMBAHASAN

Penelitian dilakukan dengan mengangkat data *tracer study* pada suatu kampus yang biasanya melakukan *drop out* terhadap mahasiswa apabila mahasiswa telah melampaui masa studi yang seharusnya atau nilai yang diperolehnya tidak mencapai batas standar minimum. Padahal jika dilihat secara lebih luas, banyak faktor yang harus dipertimbangkan sebelum memberhentikan mahasiswa dari kampus, karena bisa saja mahasiswa tersebut memiliki keunggulan lain yang belum terasah.

*Association rule mining* yang diterapkan dalam penelitian ini bertugas untuk menghasilkan *strong rule*. *Rule* yang muncul menunjukkan pola yang sering muncul pada data (*frequent pattern*). *Frequent pattern* ini berguna untuk memahami data secara mendalam. *Association rule mining* dilakukan dengan menggunakan aplikasi Weka. Dibutuhkan beberapa pengaturan untuk dapat menghasilkan *rule* terbaik. Jendela pengaturan pada Weka terkait teknik *association rule mining* yang menggunakan algoritma apriori dapat dilihat pada gambar 6.



Gambar 6. Jendela pengaturan teknik *association rule mining* di aplikasi Weka

Setelah *rule* terbentuk, dilakukan perhitungan derajat *outlier* dengan menggunakan program berbasis java. Berdasarkan ketentuannya, nilai derajat *outlier* harus diinput oleh *user* dengan nilai antara 0 dan 1. Pada penelitian ini, dimasukkan nilai 0,1 sebagai nilai derajat *outlier*. Maka, hasil yang diperoleh dari perhitungan derajat *outlier* ini dapat dilihat pada tabel di bawah ini. Karena jumlah datanya cukup banyak, maka hanya akan ditampilkan sebagian kecil saja. Dapat dilihat pada tabel 1.

Tabel 1. Hasil perhitungan derajat *outlier*

| Status Kelulusan | Keterangan        | Derajat outlier | Outlier |
|------------------|-------------------|-----------------|---------|
| Tidak lulus      | Mengundurkan Diri | 0,18            | True    |
| Tidak lulus      | Mengundurkan Diri | 0,13            | True    |
| Tidak lulus      | Mengundurkan Diri | 0,07            | False   |
| Lulus            | Lulus             | 0,07            | False   |
| Lulus            | Lulus             | 0               | False   |
| Lulus            | Lulus             | 0               | False   |
| Lulus            | Lulus             | 0               | False   |
| Tidak lulus      | Drop out          | 0,13            | True    |
| Tidak lulus      | Drop out          | 0,18            | True    |
| Tidak lulus      | Drop out          | 0,18            | True    |
| Tidak lulus      | Drop out          | 0,18            | True    |
| ...              | ...               | ...             | ...     |

Setelah mendapatkan hasil perhitungan derajat *outlier*, dari total data penelitian yang digunakan berjumlah 1.942 data, maka ditemukan data yang bersifat *outlier* sebanyak 259 data, atau sebesar 13% dari total data yang tersedia, yaitu 1.942 data. Tetapi, *outlier* yang diperoleh belum valid karena tidak semua data tersebut sesuai dengan definisi *outlier* terkait penelitian ini. Berdasarkan definisi, *outlier* yang dianggap valid adalah *outlier* dengan data yang berstatus tidak lulus. Sedangkan pada tabel data *outlier* yang ditampilkan di atas, masih ditemukan data mahasiswa dengan status kelulusan: lulus. Untuk itu dibutuhkan proses validasi temuan *outlier*.

Jumlah *outlier* yang ditemukan sebanyak 259 data diperoleh dari hasil perhitungan rumus derajat *outlier*. Adapun jumlah temuan *outlier* yang sesuai dengan realita yang mencapai 110 data merupakan jumlah temuan *outlier* yang telah di-filter sesuai dengan definisi *outlier* pada penelitian ini. Sedangkan jumlah *outlier* sebenarnya adalah jumlah perhitungan data *outlier* berdasarkan definisi *outlier* yang dihitung secara manual, yaitu sebanyak 138 data. Persentase akurasi penemuan *outlier* pada penelitian ini mencapai 80%. Dapat dilihat pada tabel 2.

Tabel 2. Data validasi temuan *outlier*

| Validasi  | Hasil perhitungan deteksi outlier |
|---|-----------------------------------|
| Jumlah <i>outlier</i> yang ditemukan                    | 259                               |
| Jumlah temuan <i>outlier</i> yang sesuai dengan realita | 110                               |
| Jumlah <i>outlier</i> sebenarnya                        | 138                               |
| Persentase akurasi                                      | 80%                               |

Dari 110 data yang dipastikan sebagai *outlier*, dapat diuraikan analisis pada masing-masing parameternya hingga menghasilkan fakta seperti uraian berikut.

1. Sebagian besar mahasiswa yang terdeteksi sebagai *outlier* memiliki IPK di bawah 2,75 dengan IQ lebih dari 120, namun lebih dari setengahnya memiliki

- psikologis (sikap kerja) yang hanya tergolong cukup baik.
2. Sebanyak 77 dari 110 data temuan *outlier* adalah mahasiswa yang kurang aktif dan tidak aktif dalam organisasi kampus selama masa perkuliahannya.
  3. Status kerja yang paling banyak data pada *outlier* adalah tidak bekerja. Namun persentase mahasiswa *outlier* yang memilih untuk berwirausaha lebih tinggi dibandingkan dengan persentase mahasiswa normal.
  4. Perbandingan mahasiswa yang bekerja sesuai dengan bidang kuliahnya pada data normal hampir berimbang. Sedangkan untuk data *outlier* didominasi oleh kategori tidak sesuai.
  5. Sebesar 67% dan 79% mahasiswa normal dan mahasiswa *outlier* tidak puas terhadap pekerjaannya saat ini.
  6. Persebaran data mahasiswa normal pada parameter jabatan dan status kerja tidak merata. Terdapat gap yang sangat tinggi antara status kerja: bekerja, dengan status kerja lainnya dimana mahasiswa yang berstatus bekerja mencapai 69%. Gap yang sama juga tampak pada parameter jabatan, dimana mahasiswa normal lebih banyak menjadi staf dibanding jabatan lainnya, yaitu 69% juga. Berbeda dengan mahasiswa *outlier*, persebaran mahasiswa untuk kedua parameter tersebut hampir merata.
  7. Gaji rata-rata mahasiswa lulusan kampus tersebut tahun 2006 ini berada pada kategori sedang, baik bagi mahasiswa normal maupun mahasiswa *outlier*. Namun untuk hitungan bonus, 73% mahasiswa *outlier* mendapat bonus sedang. Berbeda dengan mahasiswa normal yang 50%-nya justru tidak mendapatkan bonus dari tempat ia bekerja.
  2. Arahkan organisasi mahasiswa agar dapat menjadi wadah untuk melatih sikap kerja dan emosi mahasiswa.
  3. Berikan wawasan tentang sikap kerja baik dalam bentuk kuliah umum ataupun seminar.
  4. Jadikan keaktifan di organisasi kampus sebagai kewajiban bagi mahasiswa.
  5. Lakukan kontrol mendalam secara rutin terhadap kegiatan-kegiatan di organisasi kampus.
  6. Beri wawasan tentang peluang berwirausaha untuk masing-masing jurusan kuliah mahasiswa.
  7. Berikan dorongan untuk berani berwirausaha kepada mahasiswa sehingga mereka memiliki alternatif karir jika suatu saat mahasiswa (baik *outlier* maupun tidak) tidak diterima bekerja di perusahaan/instansi.
  8. Berikan pengetahuan sebelum penjurusan tentang masing-masing jurusan perkuliahan yang tersedia beserta peluang karir dan konsekuensi yang akan diterima baik selama masa kuliah maupun pada dunia kerja nantinya.
  9. Adakan tes minat dan bakat dengan materi dan metode pengujian yang lebih representatif sehingga dapat hasilnya dapat menjadi acuan penjurusan dengan lebih tepat.
  10. Latih kreativitas mahasiswa untuk berwirausaha dengan mengadakan kelas-kelas kreatif dan inovatif.

## 5. KESIMPULAN

Dari penelitian yang telah dilakukan, dapat diketahui bahwa pemanfaatan rumus *derajat outlier* pada teknik *association rule mining* mampu mendeteksi *outlier* dengan akurasi hingga 80%. Identifikasi data yang tepat akan memudahkan pemilihan teknik yang sesuai terhadap data penelitian sehingga kualitas informasi yang baik dapat tercapai. Namun perlu dilakukan analisis sebelum menentukan definisi *outlier*, karena perbedaan definisi *outlier* akan mempengaruhi hasil penelitian. Model yang diusulkan dalam penelitian ini dapat diterapkan pada kasus lain di luar bidang pendidikan untuk mengetahui kemampuan model terhadap berbagai bidang.

Sedangkan beberapa rekomendasi yang dapat disarankan untuk temuan fakta pada uraian sebelumnya yaitu sebagai berikut.

1. Kampus sebaiknya mempertimbangkan berbagai faktor atau parameter lain sebelum memutuskan untuk melakukan *drop out* terhadap mahasiswa, bukan hanya dari standar nilai tertentu.

**DAFTAR PUSTAKA**

- Aggarwal, C. C., 2013. *Outlier Analysis*. New York: Springer.
- Aprilla, D., Baskoro, D. A., Ambarwati, L. & Wicaksana, I. W. S., 2013. *Belajar Data Mining dengan RapidMiner*. Jakarta: Gramedia Pustaka Utama.
- Foguem, et al., 2013. Mining Association Rules for the Quality Improvement of the Production Process. *Expert Systems with Applications*, pp. 1034-1045.
- Hand, D., Mannila, H. & Smyth, P., 2001. *Principles of Data Mining*. Cambridge: MIT Press.
- Han, J., Kamber, M. & Pei, J., 2012. *Data Mining Concepts and Techniques*. USA: Elsevier.
- Narita, K. & Kitagawa, H., 2008. *Outlier Detection for Transaction Databases using Association Rules*. s.l., IEEE Computer Society, pp. 373-380.
- Olson, D. L. & Delen, D., 2008. *Advanced Data Mining Techniques*. Berlin: Springer.
- Roviati, E. et al., 2015. Tracer Study: Studi Rekam Jejak Alumni dan Respons Stakeholder Jurusan Tadris IPA-Biologi IAIN Syekh Nurjati Cirebon. *Scientiae Educatia*, 5(1).
- Sharma, N. & Verma, D. C. K., 2014. Association Rule Mining: An Overview. *International Journal of Computer Science and Communication*, pp. 10-15.
- Singh, K. & Upadhyaya, D. S., 2012. Outlier Detection: Applications And Techniques. *International Journal of Computer Science (IJCSI)*, 9(1), p. 307.