Deteksi Fraud Pada Akun Wifi Universitas Islam Indonesia Dengan Metode Principal Component Analysis

by Aku Siapa

Submission date: 24-Nov-2020 11:10PM (UTC+0700)
Submission ID: 1456151114
File name: paper-17523110-.docx (882.01K)
Word count: 2018
Character count: 13270

Deteksi Fraud Pada Akun Wifi Universitas Islam Indonesia Dengan Metode Principal Component Analysis

***Abstrak*—Fraud menjadi sebuah masalah yang dapat merugikan orang lain sehingga harus dilakukan tindakan. Fraud terjadi ketika pengguna membagikan akunnya dengan orang lain untuk mengakses wifi UIIConnect tanpa memikirkan celah keamanan yang dapat membahayakan data pengguna. Adanya fraud yang dilakukan dalam penggunaan akun wifi UIIConnect dapat dideteksi dengan menggunakan metode *principal component analysis*. *Principal component analysis* dapat melakukan pengelompokan akun-akun yang terindikasi melakukan fraud atau pun tidak. Tujuan dari penelitian ini adalah untuk mengembangkan model untuk mendeteksi fraud yang akan bermanfaat bagi Badan Sistem Informasi Universitas Islam Indonesia. Dalam proses mencapai tujuan dari penelitian, peneliti menggunakan lima langkah yaitu pengumpulan data, *pre-processing*, *labelling*, *clustering*, dan evaluasi. Dari penelitian ini telah berhasil dikembangkan model yang mampu mendeteksi pengguna yang terindikasi melakukan fraud. Akun yang terindikasi melakukan fraud akan dilakukan tindakan untuk keamanan data pengguna, sehingga pengguna harus berhati-hati dalam menggunakan akunnya.**

Keywords—fraud, principal component analysis, component, machine learning, clustering.

I. PENDAHULUAN

Universitas Islam Indonesia (UII) merupakan salah satu kampus yang memberikan fasilitas kepada mahasiswa, dosen, dan staf aktif di lingkungan UII untuk dapat mengakses wifi UIIConnect dengan menggunakan akun UII. UIIConnect saat ini telah terpasang lebih dari 700 *Access Points* di seluruh gedung UII. Total *bandwidth* yang disediakan UIIConnect mencapai 3.7 Gbps dan akses per user mencapai 125 Mbps [1]. Adanya fasilitas tersebut beberapa pengguna biasanya membagikan akunnya baik dengan teman atau orang terdekatnya untuk mengakses UIIConnect. Kondisi ini sangat berbahaya karena dapat dideteksi sebagai fraud.

Fraud merupakan penipuan yang dilakukan secara sengaja dengan tujuan untuk mendapatkan keuntungan pribadi yang dapat menyebabkan kerugian bagi orang lain [2]. Orang yang melakukan kejahatan ini biasanya disebut *fraudster*. Dalam kasus ini, ketika fraudster memiliki akun untuk mengakses wifi UIIConnect, akan sangat mungkin bagi *fraudster* untuk dapat mengkases platform lainnya dengan menggunakan akun tersebut. Hal ini dikarenakan seseorang cenderung *login* di berbagai macam platform dengan akun yang sama. Selain itu, *fraudster* yang mendapatkan akses akun orang lain bisa saja akan terjadi penipuan yang berujung pada masalah finansial seperti menargetkan akun bank untuk transfer dana ke akun sendiri atau akun *eCommerce* dan melakukan pembelian palsu.

Banyak aktivitas di kampus UII yang menggunakan wifi UIIConnect untuk mengakses internet setiap harinya. Aktivitas seperti kegiatan belajar mengajar atau aktivitas lain yang dilakukan oleh staf, dosen maupun mahasiswa pasti membutuhkan akses internet dengan menggunakan wifi UII. Terlebih lagi UII memberikan masing-masing akun yang dapat terhubung ke UIIConnect hingga 4 perangkat. Maka dari itu, sangat sulit mengidentifikasi akun yang melakukan fraud karena banyaknya pengguna yang terhubung dengan UIIConnect. Agar dapat mengatasi masalah akun yang melakukan fraud, maka perlu dilakukan analisis setiap lokasi dengan melihat *Acces Point* mana pengguna terhubung dan akses yang dilakukan pengguna dalam menggunakan wifi. Akan tetapi, Badan Sistem Informasi UII mengalami kesulitan untuk mengecek satu persatu akun yang melakukan fraud. Dari permasalahan tersebut, dapat ditarik kesimpulan bahwa Badan Sistem Informasi UII perlu memiliki suatu sistem untuk membantu dalam menganalisis akun yang melakukan fraud. Oleh karena itu, akan dibuat sistem untuk menghitung apakah akun yang menggunakan wifi UIIConnect melakukan fraud atau tidak. Deteksi tersebut berdasarkan lokasi yang didapat dari mac pada address acces point dan akses yang dilakukan pengguna. Selanjutnya sistem akan melakukan perhitungan terhadap dua faktor tersebut untuk mengidentifikasi apakah akun melakukan fraud.

Sistem akan dilengkapi dengan model *principal component analysis* untuk melakukan *clustering*. *Clustering* bertujuan untuk mengelompokkan akun-akun yang terindikasi melakukan fraud atau pun tidak. *Principal component analysis* (PCA) berguna untuk mengurangi dimensi permasalahan menjadi lebih sederhana dengan cara mengidentifikasi sebagian kecil komponen utama dan secara efektif merangkum sebagian besar variasi data [3]. PCA tetap menjaga *variance* sebanyak mungkin agar bisa menemukan variabel baru yang menggambarkan fungsi linier dari kumpulan data asli [3]. Jika terdapat data yang memuat nilai abnormal, karakteristik dari vektor akan sangat berpengaruh karena PCA sangat peka terhadap hal tersebut [4].

Diharapkan sistem yang dihasilkan dapat membantu Badan Sistem Informasi UII untuk mengetahui akun yang terdeteksi melakukan fraud. Selanjutnya Badan Sistem Informasi UII dapat melakukan tindakan terhadap akun yang terdeteksi melakukan fraud. Tindakan tersebut dapat berupa peringatan atau pemblokiran akun.

II. PENELITIAN TERKAIT

Penelitian yang dilakukan sebelumnya memiliki kasus dan cara yang berbeda dalam mendeteksi fraud. Viswanath melakukan penelitian tentang deteksi perilaku anomali pada jejaring sosial. Deteksi dilakukan dengan metode *unsupervised learning* yaitu *Principal Component Analisys* (PCA) yang digunakan untuk membedakan perilaku pengguna normal dan tidak normal. PCA memodelkan perilaku pengguna normal secara akurat dan mengidentifikasi anomali secara signifikan. Hasil evaluasi pendekatan yang dilakukan mencapai tingkat deteksi lebih dari 66% dan mencakup lebih dari 94% perilaku buruk dengan positive false kurang dari 0,3% [5]. Selain itu, Meng Bi juga melakukan penelitian tentang *anomaly detection* menggunakan metode PCA. Hasilnya PCA secara akurat

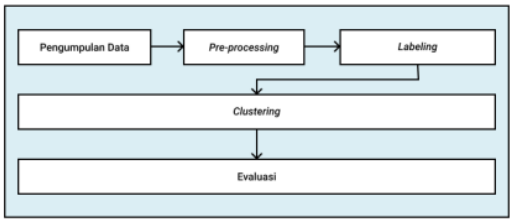
dapat menggambarkan perilaku pengguna normal dan anomali serta dapat meningkatkan efisiensi dan stabilitas [4].

Penelitian tentang deteksi anomali juga dilakukan oleh Paul. Penelitian ini menggunakan data jaringan aktivitas pengguna pada organisasi dan perusahaan yang disimpan sebagai *log*. *Log* ini akan digunakan sebagai fitur untuk melatih model dalam melakukan pengelompokan. Penelitian ini melakukan perbandingan metode *Gaussian Mixture Model* (GMM), *K-means* dan *Bayesian Gaussian Mixture Model* (BGMM). GMM menghasilkan *false positive* paling sedikit sebesar 0.33% sedangkan *K-means* 21.77% dan BGMM 5.67% [6].

Terdapat beberapa perbedaan dari penelitian-penelitian sebelumnya. Penelitian ini membahas mengenai anomaly pada penggunaan akun wifi UIIConnect atau terindikasi melakukan fraud. Fraud dapat diketahui berdasarkan faktor lokasi dan akses yang dilakukan pengguna. Sistem akan dilengkapi dengan model *principal component analysis* (PCA) untuk melakukan pengelompokan akun-akun yang terindikasi melakukan fraud atau pun tidak.

III. METODOLOGI PENELITIAN

Bagian ini menjelaskan metodologi penelitian yang digunakan untuk mengidentifikasi pengguna dibagi lima langkah yaitu pengumpulan data, *pre-processing*, *labelling*, *clustering*, dan evaluasi. Gambar 1 menunjukkan metodologi yang digunakan pada penelitian ini.



Gambar 1. Metodologi Penelitian

A. Pengumpulan Data

Pada langkah ini, dilakukan pengumpulan data internal yang diambil dari *database* Badan Sistem Informasi UII. Data tersebut berupa file csv yang akan digunakan sebagai data untuk kebutuhan sistem.

B. *Pre-processing*

Tahap *pre-processing* digunakan untuk mempersiapkan data sebelum dilakukan tahap selanjutnya. Berikut adalah langkah-langkah yang dilakukan pada tahap *pre-processing*:

- Mengambil beberapa kolom seperti *source address*, *source user* dan *application*
- Menghapus *source user* yang bernilai *null* atau NA dan menghitung jumlah freq applikasi
- Memberikan label kemiripan aplikasi
- Menghitung persentase kemiripan aplikasi
- Menghitung jumlah IP
- Menghitung jumlah Aplikasi

C. *Labelling*

Pada tahap labelling akan dilakukan pemberian label pengguna yang melewati batas *threshold*. Tahap ini digunakan untuk mengetahui pengguna yang terindikasi melakukan fraud.

D. *Clustering*

Bagian ini menjabarkan tentang proses *clustering* pengguna yang melakukan fraud dengan menggunakan metode *principal component analisys*. Proses ini bertujuan untuk mengelompokkan akun-akun yang terindikasi melakukan fraud atau pun tidak. Berikut adalah tahap-tahap yang dilakukan pada metode PCA:

1) Standardisasi Data

Standardisasi data bertujuan agar setiap variabel memiliki kontribusi yang sama. Berikut adalah hal-hal yang dilakukan pada standarisasi data:

- Menghitung rata-rata menggunakan persamaan:

$$\mu = \frac{1}{m}\sum_{i=1}^{m} x^{(i)} \quad (1)$$

- Menghitung *center*

$$x^{(i)} = x^{(i)} - \mu \quad (2)$$

- Menghitung kovarian matrix yang merupakan matrix M x M [7]

$$C = \frac{1}{M} X' X'^{T} \quad (3)$$

2) Menghitung Variance

Variance meruapakan sebaran data yang ditangkap oleh masing-masing *principal component*. Berikut adalah persamaan untuk menghitung *variance*:

$$\frac{1}{m}\sum_{i=1}^{m}\left(x^{(i)T}u\right)^2 = \frac{1}{m}\sum_{i=1}^{m} u^T x^{(i)} x^{(i)T} u = u^T\left(\sum_{i=1}^{m} x^{(i)} x^{(i)T}\right)u \quad (4)$$

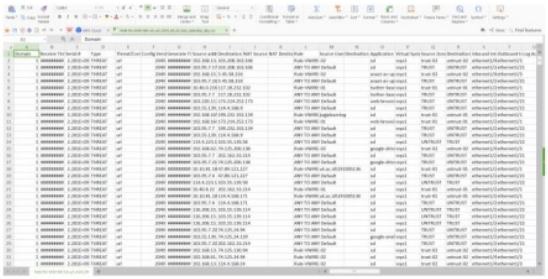
E. Evaluasi

Pada tahap evaluasi akan dilakukan pengukuran performa model. Tahap ini bertujuan untuk mengetahui apakah pemodelan sudah sesuai dengan yang diinginkan serta mengetahui sejauh mana pemodelan ini berhasil.

IV. HASIL

A. *Pengumpulan Data*

Data yang dibutuhkan dalam pembuatan model ini berupa hasil *record* setiap akun yang menggunakan wifi UII. Gambar 3 memuat data URL yang berisi *receive time*, *source user*, *source address*, *application* dan lain-lain.



Gambar 3. Data URL

B. Pre-processing

Setelah melakukan pengumpulan data, selanjutnya dilakukan pre-processing data. Proses yang dilakukan antara lain:

- Mengambil beberapa kolom seperti source address, source user dan application

	Source.User	Source.address	Application
1	NA	192.168.13.15	ssl
2	NA	103.95.7.17	ssl
3	NA	192.168.15.11	avast-av-update
4	NA	103.95.7.16	avast-av-update
5	NA	10.40.0.216	twitter-base
6	NA	103.95.7.7	twitter-base
7	NA	103.220.113.12	web-browsing
8	NA	103.55.139.35	ssl
9	jogjalearning	192.168.165.109	ssl
10	NA	192.168.164.254	web-browsing
11	NA	103.95.7.7	ssl
12	NA	103.55.139.35	ssl
13	NA	114.4.223.140	ssl
14	NA	192.168.62.89	google-drive-web
15	NA	103.95.7.7	ssl
16	NA	103.95.7.21	google-drive-web
17	uii.ac.id\19100	10.10.81.18	ssl
18	NA	103.95.7.4	ssl

Showing 1 to 19 of 800.000 entries. 3 total columns

Gambar 4. Hasil Pengambilan Kolom

- Menghapus source user yang bernilai null atau NA dan menghitung jumlah frequensi applikasi

	Source.User	Source.address	Application	Freq
1	uii.ac.id\	10.10.81.200	google-base	124
2	uii.ac.id\	10.10.81.200	ssl	114
3	uii.ac.id\	10.10.81.200	facebook-base	36
4	uii.ac.id\	10.10.81.200	facebook-video	106
5	uii.ac.id\	10.10.81.200	whatsapp-base	8
6	uii.ac.id\	10.10.81.200	web-browsing	6
7	uii.ac.id\	10.10.81.200	youtube-base	14
8	uii.ac.id\	192.168.30.11	web-browsing	2952
9	uii.ac.id\	10.10.41.214	ssl	87
10	uii.ac.id\	10.10.41.214	web-browsing	5
11	uii.ac.id\	10.10.41.214	windows-push-notifications	1
12	uii.ac.id\	10.10.81.252	web-browsing	38
13	uii.ac.id\	10.10.81.252	facebook-video	3
14	uii.ac.id\	10.10.81.252	facebook-base	39
15	uii.ac.id\	10.10.81.252	youtube-base	13
16	uii.ac.id\	10.10.81.252	google-base	81
17	uii.ac.id\	10.10.81.252	ssl	111
18	uii.ac.id\	10.10.81.252	whatsapp-base	8

Showing 1 to 19 of 1.216 entries. 4 total columns

Gambar 5. Hasil Source User dan Perhitungan Jumlah Frequensi Applikasi

- Memberikan label kemiripan aplikasi

	Source.User	Source.address	Application	Freq	Similar
1	uii.ac.id\	10.10.81.200	google-base	124	0
2	uii.ac.id\	10.10.81.200	ssl	114	0
3	uii.ac.id\	10.10.81.200	facebook-base	36	0
4	uii.ac.id\	10.10.81.200	facebook-video	106	0
5	uii.ac.id\	10.10.81.200	whatsapp-base	8	0
6	uii.ac.id\	10.10.81.200	web-browsing	6	0
7	uii.ac.id\	10.10.81.200	youtube-base	14	0
8	uii.ac.id\	192.168.30.11	web-browsing	2952	0
9	uii.ac.id\	10.10.41.214	ssl	87	0
10	uii.ac.id\	10.10.41.214	web-browsing	5	0
11	uii.ac.id\	10.10.41.214	windows-push-notifications	1	0
12	uii.ac.id\	10.10.81.252	web-browsing	38	0
13	uii.ac.id\	10.10.81.252	facebook-video	3	0
14	uii.ac.id\	10.10.81.252	facebook-base	39	0
15	uii.ac.id\	10.10.81.252	youtube-base	13	0
16	uii.ac.id\	10.10.81.252	google-base	81	0
17	uii.ac.id\	10.10.81.252	ssl	111	0
18	uii.ac.id\	10.10.81.252	whatsapp-base	8	0

Showing 1 to 19 of 1.216 entries. 5 total columns

Gambar 6. Hasil Pemberian Label

- Menghitung persentase kemiripan aplikasi

	Source.User	Similar0	Similar1
1		100.000000	0.00000
2	@uii.ac.id	100.000000	0.00000
3		14.285714	85.71429
4	@uii.ac.id	100.000000	0.00000
5		100.000000	0.00000
6		100.000000	0.00000
7		100.000000	0.00000
8		100.000000	0.00000
9		100.000000	0.00000
10		100.000000	0.00000
11		100.000000	0.00000
12		100.000000	0.00000
13		100.000000	0.00000
14		71.428571	28.57143
15	@mail.umy.ac.id	100.000000	0.00000
16		6.046512	93.95349
17		100.000000	0.00000
18		100.000000	0.00000

Showing 1 to 19 of 143 entries. 3 total columns

Gambar 7. Hasil Perhitungan Persentase

- Menghitung jumlah IP

	Source.User	freq
1		1
2	21@uii.ac.id	1
3		2
4	@uii.ac.id	1
5		1
6	3	1
7	2	1
8		1
9		1
10		1
11		1
12		1
13		1
14		2
15	@mail.umy.ac.id	1
16		23
17		1
18		1

Showing 1 to 19 of 143 entries. 2 total columns

Gambar 8. Hasil Perhitungan IP

- Menghitung jumlah Aplikasi

	Source.User	freq
1	[redacted]	122
2	[redacted]@uii.ac.id	1435
3	[redacted]	63
4	[redacted]uii.ac.id	26
5	[redacted]	1
6	[redacted]	156
7	[redacted]	30
8	[redacted]	2203
9	[redacted]	22
10	[redacted]	213
11	[redacted]	94
12	[redacted]	56
13	[redacted]	2
14	[redacted]	110
15	[redacted]@mail.umy.ac.id	10
16	[redacted]	9173
17	[redacted]	85
18	[redacted]	72

Showing 1 to 19 of 143 entries, 2 total columns

Gambar 9. Hasil Perhitungan Jumlah Aplikasi

Gambar 10 merupakan hasil akhir dari proses pre-processing yang menyajikan kolom akun yang digunakan, jumlah ip, jumlah aplikasi dan persentase kemiripan aplikasi.

	ipUser.freq	appUser.freq	similarAppUser.Similar0
[redacted]	1	122	100.000000
[redacted]uii.ac.id	1	1435	100.000000
[redacted]	2	63	14.285714
[redacted]uii.ac.id	1	26	100.000000
[redacted]	1	1	100.000000
[redacted]	1	156	100.000000
[redacted]	1	30	100.000000
[redacted]	1	2203	100.000000
[redacted]	1	22	100.000000
[redacted]	1	213	100.000000
[redacted]	1	94	100.000000
[redacted]	1	56	100.000000
[redacted]	1	2	100.000000
[redacted]	2	110	71.428571
[redacted]@mail.umy.ac.id	1	10	100.000000
[redacted]	23	9173	6.046512
[redacted]	1	85	100.000000
[redacted]	1	72	100.000000

Showing 1 to 19 of 143 entries, 3 total columns

Gambar 10. Hasil *Pre-processing*

C. Labelling

Dalam tahap labelling, masing-masing pengguna akan diberikan label antara 1 atau 0. Label 1 menunjukkan jika pengguna melewati batas *threshold* dan label 0 jika pengguna tidak melewati batas *threshold*. Gambar 11 menunjukkan tabel hasil labelling masing-masing pengguna

	ipUser.freq	appUser.freq	similarAppUser.Similar0	Hasil	Label
1	1	122	100.000000	100.00000	0
2	1	1435	100.000000	100.00000	0
3	2	63	14.285714	28.57143	0
4	1	26	100.000000	100.00000	0
5	1	1	100.000000	100.00000	0
6	1	156	100.000000	100.00000	0
7	1	30	100.000000	100.00000	0
8	1	2203	100.000000	100.00000	0
9	1	22	100.000000	100.00000	0
10	1	213	100.000000	100.00000	0
11	1	94	100.000000	100.00000	0
12	1	56	100.000000	100.00000	0
13	1	2	100.000000	100.00000	0
14	2	110	71.428571	142.85714	1
15	1	10	100.000000	100.00000	0
16	23	9173	6.046512	139.06977	1
17	1	85	100.000000	100.00000	0
18	1	72	100.000000	100.00000	0

Showing 1 to 19 of 143 entries, 5 total columns

Gambar 11. Hasil Pemberian Label

D. Clustering

Dalam clustering akan dilakukan tahap standarisasi data dan menghitung variance.

1) Standardisasi Data

Standardisasi data dilakukan dengan menggunakan skala sehingga data akan memiliki *impact* yang sama dan *comparable*.

a) Menghitung rata-rata

Gambar 12 merupakan hasil perhitungan mean masing-masing *variable*. Variabel tersebut berupa 'ipuser.freq' (1,336), 'appUser.freq' (367,9), 'similarAppUser.similar0' (92,135) dan 'Hasil' (99,76).

```
> summary(training)
  ipUser.freq      appUser.freq    similarAppUser.Similar0     Hasil           Label
 Min.   : 1.000   Min.   :   1.0   Min.   :  6.047         Min.   : 15.38   Length:113
 1st Qu.: 1.000   1st Qu.:  40.0   1st Qu.:100.000         1st Qu.:100.00   Class :character
 Median : 1.000   Median : 138.0   Median :100.000         Median :100.00   Mode  :character
 Mean   : 1.336   Mean   : 367.9   Mean   : 92.135         Mean   : 99.76
 3rd Qu.: 1.000   3rd Qu.: 392.0   3rd Qu.:100.000         3rd Qu.:100.00
 Max.   :23.000   Max.   :9173.0   Max.   :100.000         Max.   :155.56
```

Gambar 12. Hasil Perhitungan Rata-rata

b) Menghitung center

Selanjutnya pada Gambar 13 akan dilakukan *centering* untuk setiap *variable*. Variabel ipuser.freq menghasilkan nilai *centering* (1,336283), 'appUser.freq' (367,929204), 'similarAppUser.similar0' (92,135016) dan 'Hasil' (99,763717).

```
> pc$center
     ipUser.freq      appUser.freq similarAppUser.Similar0             Hasil
        1.336283         367.929204             92.135016         99.763717
```

Gambar 13. Hasil Perhitungan Center

c) Menghitung kovarian matrix

Pada gambar 14 merupakan hasil perhitungan kovarian matrix. Dari perhitungan tersebut, dapat dilihat kekuatan korelasi masing-masing *variable* dengan setiap *principal component*.

```
> pc$rotation
                              PC1         PC2         PC3         PC4
ipUser.freq             0.65208026 0.08829306 -0.07281259  0.7494625
appUser.freq            0.61625632 0.17659340 -0.47466905 -0.6031022
similarAppUser.Similar0 -0.43173798 0.55866990 -0.66432432  0.2452824
Hasil                   0.09285366 0.80554767  0.57276594 -0.1200430
```

Gambar 14. Hasil Perhitungan Kovarian Matrix

Gambar 15 merupakan hasil akhir dalam tahap standardisasi data.

```
> pc$x
          PC1         PC2         PC3         PC4
1  -0.42615026  0.1560631 -0.09801863  0.12817969
2   0.44849437  0.4066998 -0.77171027 -0.72779540
3   1.16646204 -5.6115693  0.02871986  0.07271441
4  -0.49009990  0.1377378 -0.04876166  0.19076431
5  -0.50675345  0.1329655 -0.03593433  0.20706239
6  -0.40350142  0.1625533 -0.11546381  0.10601430
7  -0.48743533  0.1385013 -0.05081404  0.18815662
8   0.96009153  0.5533022 -1.16576600 -1.22847238
9  -0.49276447  0.1369742 -0.04670929  0.19337200
10 -0.36553132  0.1734339 -0.14471013  0.06885468
11 -0.44480224  0.1507182 -0.08365201  0.14643354
12 -0.47011364  0.1434644 -0.06415447  0.17120662
13 -0.50608731  0.1331564 -0.03644742  0.20641047
14  0.70079226  1.5855269  2.27936949 -0.15134480
15 -0.50075818  0.1346835 -0.04055217  0.20119508
16 14.58832829  2.3061354 -1.20580428  0.74911546
17 -0.45079752  0.1490002 -0.07903417  0.15230085
19 -0.49409675  0.1365924 -0.04568310  0.19467585
20  0.52776528  0.4294155 -0.83276838 -0.80537426
21 -0.50342274  0.1339200 -0.03849980  0.20380278
22 -0.44413609  0.1509091 -0.08416511  0.14578162
23 -0.49476290  0.1364015 -0.04517001  0.19532777
24 -0.33555493  0.1820239 -0.16779933  0.03951814
26 -0.44546838  0.1505273 -0.08313892  0.14708546
27  0.05613665  0.2942664 -0.46949825 -0.34381267
28 -0.28292970  0.1971041 -0.20833371 -0.01198379
29 -0.40483371  0.1621715 -0.11443762  0.10731815
30 -0.11472487  0.2438718 -0.33404160 -0.17170496
```

Gambar 15. Hasil Standardisasi Data

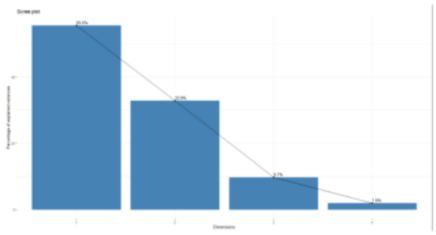
2) Menghitung Variance

Gambar 16 merupakan hasil dari perhitungan *variance* setiap *principal component*. Hasil variance yang didapat oleh 'PC1' (0.555), 'PC2' (0.3283), 'PC3' (0,09722) dan 'PC4' (0,01945) .

```
> summary(pc)
Importance of components:
                          PC1    PC2     PC3     PC4
Standard deviation     1.490 1.1460 0.62360 0.27890
Proportion of Variance 0.555 0.3283 0.09722 0.01945
Cumulative Proportion  0.555 0.8833 0.98055 1.00000
```

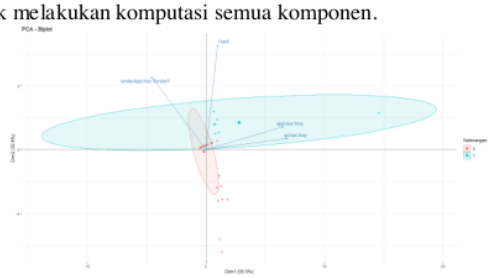
Gambar 16. Hasil Perhitungan *Variance*

Selanjutnya *variance* tersebut dapat dilakukan ploting untk melihat total *variance* yang dibawa oleh masing-masing *component*.



Gambar 17. Ploting *Variance*

Pada gambar 17 PC1 memiliki *variance* paling besar dengan nilai 55% dan PC4 dengan nilai terkecil 1,9%. Dari ploting tersebut dengan mengambil komponen PC1 dan PC2 maka akan didapatkan total *variance* dari seluruh data sebesar 88,3 % sehingga proses komputasi menjadi lebih baik karena tidak melakukan komputasi semua komponen.



Gambar 18. Hasil Biplot PC1 dan PC2

Selanjutnya pada gambar 18 dapat dilihat hasil biplot pada sumbu vertikal (Dim1) dan sumbu horizontal (Dim2). Kemudian pada biplot di atas juga divisualisasikan setiap variabel sebagai bentuk vektor. Dari plot tersebut, dapat dilihat arah vektor 'ipuser.freq' dan 'appUser.freq' cenderung horizontal seperti arah *principal component* yang kedua (Dim2). Hal ini mengindikasikan bahwa *variable* 'ipuser.freq' dan 'appUser.freq' lebih banyak dijelaskan atau diwakili oleh *principal component* yang kedua. Sebaliknya, arah vektor 'similarAppUser.similar0' dan 'Hasil' lebih mendekati arah *principal component* yang pertama (Dim1). Hal ini mengindikasikan jika informasi yang dibawa variabel 'similarAppUser.similar0' dan 'Hasil' lebih banyak diwakili oleh *principal component* yang pertama.

E. Evaluasi

Setelah mendapatkan *cluster* dalam pemodelan, selanjutnya dilakukan evaluasi untuk melihat apakah sudah sesuai dengan yang diinginkan. Tetapi pada tahap ini belum dapat dilakukan karena masih terdapat kekurangan yaitu faktor lokasi. Hal ini disebabkan *cross referencing* data URL dan data *access point* belum bisa dilakukan.

V. KESIMPULAN

Berdasarkan hasil dari pembuatan model yang dilakukan dalam penelitian ini, dapat disimpulkan bahwa penelitian ini telah berhasil melakukan clustering pengguna yang melaukan fraud dan tidak. PCA berhasil mendapatkan total variance sebesar 88,3% yang hanya menggunakan dua component principal. PCA berhasil mereduksi dimensi komponen sehingga proses komputasi menjadi lebih baik karena tidak melakukan komputasi semua komponen. Serta mendapatkan hasil *clustering* pengguna yang terindikasi melakukan fraud dan tidak.

Penelitian ini masih akan dilanjutkan dengan melakukan cross referencing data URL dan data Access Point. Dari data access point dapat diketahui lokasi setiap pengguna yang menggunakan wifi UIIConnect.

REFERENSI

[1] "Akses Internet," 2017. [Online]. Available: https://bsi.uii.ac.id/akses-internet/.

[2] Z. Zojaji, R. E. Atani, and A. H. Monadjemi, "A Survey of Credit Card Fraud Detection Techniques : Data and Technique Oriented Perspective," pp. 1–26, 2016.

[3] Y. Aït-sahalia and D. Xiu, "Using principal component analysis to estimate a high dimensional factor model with high-frequency data ☆," vol. 201, pp. 384–399, 2017.

[4] M. Bi, J. Xu, M. Wang, and F. Zhou, "Anomaly detection model of user behavior based on principal component analysis," *J. Ambient Intell. Humaniz. Comput.*, 2016.

[5] B. Viswanath *et al.*, "Towards Detecting Anomalous User Behavior in Online Social Networks," 2014.

[6] M. Paul and K. Medhe, "Using Machine Learning to Detect Anomalies in Internet Browsing Pattern of Users," 2019.

[7] P. N. Primandari and B. Hardiansyah, "Ekstraksi Fitur Menggunakan Principal Component Analisys (PCA)," 2018.

Deteksi Fraud Pada Akun Wifi Universitas Islam Indonesia Dengan Metode Principal Component Analysis

ORIGINALITY REPORT

6%	5%	2%	0%
SIMILARITY INDEX	INTERNET SOURCES	PUBLICATIONS	STUDENT PAPERS

PRIMARY SOURCES

1	doku.pub Internet Source	1%
2	dspace.uii.ac.id Internet Source	1%
3	Samundra Deep, Xi Zheng, Chandan Karmakar, Dongjin Yu, Leonard G. C. Hamey, Jiong Jin. "A Survey on Anomalous Behavior Detection for Elderly Care Using Dense-Sensing Networks", IEEE Communications Surveys & Tutorials, 2020 Publication	1%
4	Lutao Zheng, Guanjun Liu, Chungang Yan, Changjun Jiang, Mengchu Zhou, Maozhen Li. "Improved TrAdaBoost and Its Application to Transaction Fraud Detection", IEEE Transactions on Computational Social Systems, 2020 Publication	1%
5	Cong Zhang, Xiang Chen, Shuai Cao, Xu	

	Zhang, Xun Chen. "A Novel HD-sEMG Preprocessing Method Integrating Muscle Activation Heterogeneity Analysis and Kurtosis-Guided Filtering for High-Accuracy Joint Force Estimation", IEEE Transactions on Neural Systems and Rehabilitation Engineering, 2019 Publication	1%
6	123dok.com Internet Source	<1%
7	"Machine Learning, Image Processing, Network Security and Data Sciences", Springer Science and Business Media LLC, 2020 Publication	<1%
8	www.scribd.com Internet Source	<1%
9	id.123dok.com Internet Source	<1%
10	kikyputriani.wordpress.com Internet Source	<1%
11	www.irjet.net Internet Source	<1%

Exclude quotes	On	Exclude matches	Off
Exclude bibliography	On		