

# PENGELOMPOKAN KUALITAS KERJA PEGAWAI MENGGUNAKAN ALGORITMA *K-MEANS++* DAN *COP-KMEANS* UNTUK MERENCANAKAN PROGRAM PEMELIHARAAN KESEHATAN PEGAWAI DI PT. PLN P2B JB DEPOK

<sup>1</sup>Chandra Muhammad Fikri, <sup>2</sup>Fenty Eka Muzayyana Agustin, <sup>3</sup>Fitri Mintarsih

<sup>1,2,3</sup>Teknik Informatika, Fakultas Sains dan Teknologi Universitas Islam Negeri Syarif Hidayatullah, Jakarta

<sup>1</sup>fferaf@gmail.com,

<sup>2</sup>fenty.eka@uinjkt.ac.id,

<sup>3</sup>fitri.mintarsih@uinjkt.ac.id

*Abstrak:* PT. PLN P2B JB Depok merupakan sebuah perusahaan yang bergerak dalam bidang jasa penyaluran dan pusat pengaturan beban listrik pada wilayah Jawa-Bali. Setiap divisi memiliki target dan harus selesai tepat waktu. Beban pekerjaan yang berat memaksa pegawai untuk bekerja lebih keras sehingga berakibat banyak pegawai yang sakit. Untuk mengurangi jumlah pegawai yang sakit PLN P2B JB Depok melakukan tindakan pencegahan setiap enam bulan sekali guna mencegah terjadinya penyakit yang sama pada semester berikutnya, tetapi jumlah penyakit masih tetap bahkan bertambah, karena itu dibutuhkan cara baru untuk merencanakan program kesehatan pegawai di semester berikutnya. Bu Fatma ingin menggabungkan database kesehatan pegawai dengan kualitas kerja pegawai agar mendapatkan penyakit dengan kelompok kualitas kerja yang berbeda-beda. Algoritma K-Means++ adalah algoritma yang mampu mengelompokkan data yang besar menjadi beberapa kelompok, tetapi K-Means++ tidak mampu mengelompokkan pegawai yang sama dalam satu kelompok kualitas kerja maka dari itu penggunaan COP-Kmeans dibutuhkan. Hasil dari perhitungan K-Means++ dan COP-Kmeans pada penelitian ini adalah 5 kelompok kualitas kerja dengan pegawai yang berbeda dan penyakit yang berbeda, dimana penyakit dan jumlah penyakit dalam kelompok tersebut akan digunakan sebagai bahan pertimbangan Bu Fatma untuk merencanakan program kesehatan pegawai di semester berikutnya.

*Keywords* – Algoritma K-Means++, Algoritma COP-Kmeans, database kesehatan pegawai.

*Abstract:* PT. PLN P2B JB Depok is a company engaged in the distribution of services and the center of electricity load management in the Java-Bali region. Each division has a target and must be completed on time. Heavy workloads force employees to work harder resulting in many ill employees. To reduce the number of employees who are sick PLN P2B JB Depok take precautions once every six months to prevent the occurrence of the same disease in the next semester, but the number of diseases still remain even increases, because it takes a new way to plan employee health programs in the next semester. Mrs. Fatma wants to combine employee health databases with the quality of employee work to get the disease with different work quality groups. The K-Means ++ algorithm is an algorithm capable of grouping large data into

multiple groups, but K-Means ++ is not able to group the same employees in a group of quality work hence the use of COP-Kmeans is required. The results of K-Means ++ and COP-Kmeans calculations in this study are 5 groups of quality work with different employees and different diseases, where the disease and the number of diseases in the group will be used as consideration for Mrs. Fatma to plan employee health programs in the next semester .

*Keywords:* K-Means Algorithm ++, COP-Kmeans Algorithm, employee health database.

## I. PENDAHULUAN

PT. PLN P2B JB Depok merupakan sebuah perusahaan yang bergerak dalam bidang jasa penyaluran dan pusat pengaturan beban listrik pada

wilayah Jawa-Bali, instalasi sistem transmisi tenaga listrik Jawa-Bali, mengelola transaksi tenaga listrik, mengelola sistem operasi tenaga listrik. Setiap divisi memiliki target dan harus selesai tepat waktu. Beban pekerjaan yang berat memaksa pegawai untuk bekerja lebih keras sehingga berakibat banyak pegawai yang sakit.

Untuk mengurangi jumlah pegawai yang sakit PLN P2B JB Depok melaksanakan program kesehatan setiap enam bulan sekali guna mencegah terjadinya penyakit yang sama pada semester berikutnya, karena jika dibiarkan dikhawatirkan pegawai yang bersangkutan akan dimutasi atau dipensiunkan dini. Contoh pada semester 1 tahun 2013 kasus cabut gigi mendominasi *database* kesehatan sebesar 0,0352% dari 5.706 kasus penyakit, maka akan dilakukan program kesehatan pada akhir semester yaitu diadakan seminar tentang “Kesadaran Menjaga Kesehatan Gigi”, tetapi program kesehatan yang telah dilaksanakan masih belum tepat dikarenakan pada semester berikutnya jumlah pegawai dengan penyakit yang telah ditentukan tidak ada perubahan, bahkan bertambah, Bu Fatma memiliki rencana menentukan program kesehatan tidak hanya menggunakan data kesehatan, tetapi menggunakan data kesehatan dan kualitas kerja pegawai untuk menentukan program kesehatan.

Menurut Lonnie [1] mengatakan bahwa terdapat hubungan yang erat antara kualitas kerja dengan kesehatan sehingga untuk menentukan program kesehatan bisa diambil menggunakan data kesehatan dan kualitas kerja tiap pegawai. Bu Fatma mengharapkan metode yang menggunakan data kesehatan dan kualitas kerja dapat memberikan hasil berupa pertimbangan program kesehatan yang harus didahulukan terlebih dahulu.

Pada saat ini PLN P2B JB Depok telah memiliki sebuah Aplikasi Sistem Manajemen

Kinerja Pegawai (SMKP) untuk menilai kualitas kerja pegawai berdasarkan jumlah *job description* dan tambahan pekerjaan yang telah diselesaikan dalam bentuk kelompok. Tetapi Aplikasi SMKP tidak mampu menilai kualitas kerja pegawai dari sisi kesehatan, contoh hasil akhir perhitungan kualitas kerja Juniza menurut Aplikasi SMKP adalah sangat baik tetapi pada saat kerja Juniza mengalami sakit gigi dan dihari yang lain sakit demam. Pegawai yang mengalami sakit pada saat kerja akan mempengaruhi kualitas kerja pegawai sehingga hasil penilaian aplikasi SMKP masih diragukan [1].

Berdasarkan wawancara penulis dengan Bu Fatma di PLN P2B JB Depok terdapat *database* kesehatan yang memiliki *record* sebanyak 5.722 baris. Pada *database* kesehatan tercatat sebanyak 1.919 pegawai sakit dari tanggal 5 Juli 2012 hingga 31 Mei 2013 atau semester 2 tahun 2012 berjumlah 16 *record* dan semester 1 tahun 2013 berjumlah 5.706 *record*.

Jumlah *record* yang terdapat dalam *database* kesehatan sangat banyak yakni 5.722 *record*, sehingga dibutuhkan sebuah cara untuk mengolah data yang berjumlah besar serta mampu menghasilkan penilaian kualitas kerja dalam bentuk kelompok masing-masing kualitas kerja. Menurut buku teknik Data Mining mampu menjawab masalah dari permasalahan cara untuk mengolah *database* kesehatan yang berjumlah 5.722 *record* menjadi informasi yang penting. Peneliti mencoba tiga metode data mining dengan kriteria mampu mengelompokkan data yang besar, cepat dan akurat yaitu *hierarchical clustering*, fuzzy c-means dan k-means. Berdasarkan jurnal [2] yang telah membandingkan metode K-Means dengan *hierarchical clustering*, Soumi dan Sanjay [3] membandingkan K-Means dengan Fuzzy C-Means, peneliti menganalisa bahwa kedua k-means

paling cocok digunakan untuk database kesehatan pegawai. Untuk menutupi kelemahan k-means peneliti menggunakan k-means++ dan cop-kmeans untuk mengelompokkan pegawai yang sama dalam kelompok yang sama.

## II. LANDASAN TEORI

### A. Konsep Dasar Data Mining

Istilah *data mining* sering disandingkan dengan jumlah data yang sangat besar yang akan dijadikan sebagai tempat menambang agar mendapatkan informasi baru pada data yang sangat besar. Istilah *data mining* bahkan memiliki nama lain seperti *knowledge mining from data*, *knowledge extraction*, *pattern analysis*, *data dredging*. *Knowledge mining from data* sangat tepat dibandingkan dengan *data mining* tetapi terlalu panjang untuk disebutkan. Jika dipersingkat menjadi *knowledge mining* sangat tidak tepat dikarenakan tidak mencerminkan menambang dengan jumlah data yang sangat besar. Menambang identik dengan proses menemukan emas pada tambang emas yang dalam, sehingga kata “data” dan “mining” sering dipakai untuk menyatakan *knowledge mining from data* dan hingga kini “*data mining*” menjadi lebih populer dibandingkan *knowledge mining from data*. *Data mining* sendiri adalah salah satu langkah dalam *knowledge mining from data* yang berfungsi untuk menggali informasi baru dari data yang telah dipersiapkan.

### B. Algoritma K-Means++

Algoritma *k-means* sering digunakan dalam teknik *clustering* yang bertujuan untuk meminimalkan jarak kuadrat yang telah dirata-ratakan antar titik dalam kluster yang sama. Tetapi algoritma k-means memiliki kelemahan yaitu tidak bisa memberikan akurasi yang tepat sekalipun

menggunakan perhitungan yang sederhana dan cepat. Menurut Jurnal *K-Means++ The Advantages of Careful Seeding* jika k-means ditambahkan dengan *randomized seeding technique* akan meningkatkan akurasi pada algoritma k-means.

Akurasi algoritma k-means sangat bergantung dengan nilai C (*centroid*) di awal perhitungan maka jika menggunakan nilai C yang berbeda akan memberikan hasil yang berbeda pula bahkan membutuhkan banyak iterasi (perulangan) untuk menentukan anggota suatu kluster jika nilai C tidak tepat. Dengan menambahkan rumus *randomized seeding technique* maka akan lebih menentukan nilai C pada saat awal perhitungan.

Setiap anggota memiliki peluang untuk terpilih menjadi *centroid* sehingga setiap anggota dihitung nilai peluang untuk terpilih dan yang paling mendekati adalah yang paling tepat.

Berikut rumus *randomized seeding technique*:

$$\frac{D(x')^2}{\sum_{x \in X} D(x')^2} \quad (1)$$

Keterangan :

$D(x')^2$  = Jarak *Euclidean Distance*

$\sum_{x \in X} D(x')^2$  = Jumlah Jarak *Euclidean Distance*

Rumus *randomized seeding technique* akan menghasilkan sebuah angka yang akan dijadikan patokan semakin jauh nilai objek maka semakin besar kemungkinan nilai objek akan menjadi nilai C berikutnya.

### C. Algoritma K-Means

Dalam statistik dan mesin pembelajaran, pengelompokan *K-Means* merupakan metode analisis kelompok yang mengarah pada pemartisian N objek pengamatan ke dalam K kelompok (*cluster*) dimana setiap objek pengamatan dimiliki oleh sebuah kelompok dengan *mean* (rata-rata) terdekat, dimana keduanya mencoba untuk menemukan pusat dari kelompok

dalam data sebanyak iterasi perbaikan yang dilakukan oleh kedua algoritma.

*K-Means* merupakan salah satu metode pengelompokan data *non hierarchy* (sekatan) yang berusaha mempartisi data yang ada ke dalam bentuk dua atau lebih kelompok. Metode ini mempartisi data kedalam kelompok sehingga data berkarakteristik sama di masukkan ke dalam satu kelompok yang sama dan data yang berkarakteristik berbeda di kelompokkan ke dalam kelompok yang lain.

Pengukuran jarak pada ruang jarak (*distance space*) Euclidean menggunakan formula:

$$D(X_1, X_2) = \sqrt{\sum_{j=1}^p |X_{2j} - X_{1j}|^2} \quad (2)$$

D adalah jarak antara data  $x_2$  dan  $x_1$  adalah nilai mutlak. Pengukuran jarak pada ruang jarak Manhattan menggunakan formula.

#### D. Algoritma COP-Kmeans

Data kesehatan yang digunakan pada penelitian ini memiliki sifat yaitu satu objek bisa sama dengan objek yang lainnya, pada *database* kesehatan yang dimaksud objek adalah nama pegawai. Pegawai memiliki kemungkinan sakit lebih dari satu kali karena itu dalam *database* kesehatan ada kemungkinan terdapat satu pegawai yang sama sakit lebih dari satu kali. Jika hanya menggunakan *k-means* tidak akan bisa mengelompokkan objek yang sama kedalam satu kelompok karena itu solusinya adalah algoritma *cop-kmeans*, dimana terdapat hubungan *must-link constraints* dan *cannot-link constraints* yang menyatakan hubungan antar objek. *Cop-kmeans* mengandalkan hubungan antar objek maka harus dibuat kumpulan hubungan antara satu objek dengan objek yang lainnya dengan kriteria *must-link constraints* dan *cannot-link constraints*.

#### E. Rapid Application Development

*Rapid Application Development* (RAD) adalah suatu pendekatan berorientasi objek terhadap pengembangan sistem yang mencakup suatu metode pengembangan serta perangkat-perangkat lunak. RAD bertujuan mempersingkat waktu yang biasanya diperlukan dalam siklus hidup pengembangan sistem tradisional antara perancangan dan penerapan suatu sistem informasi. Pada akhirnya, RAD sama-sama berusaha memenuhi syarat-syarat bisnis yang berubah secara cepat.

### III. METODOLOGI

Pada perancangan sistem informasi ini, penulis menggunakan metode pengembangan sistem *Rapid Application Development* (RAD) dengan *Unified Modelling Language* (UML) diagram untuk menggambarkan alur proses dan data pada sistem yang akan dirancang.

### IV. IMPLEMENTASI

#### A. Desain Model Algoritma K-Means++

Sebagaimana yang dijelaskan pada bab sebelumnya, perancangan pemodelan algoritma *K-Means++* dan *COP-Kmeans* di implementasikan pada pengelompokan kualitas kerja pegawai dalam aplikasi. Selanjutnya akan dijelaskan cara kerja algoritma yang di implementasikan dalam penelitian ini.

##### 1. Langkah Algoritma K-Means++

Llangkah awal pada *k-means++* adalah menentukan *centroid* pertama awal secara acak dari *database* kesehatan lalu untuk mendapatkan *centroid* berikutnya akan menggunakan metode *k-means++*.

*Centroid* 1 secara acak : No.1 (JUNIZA FIRDHA S, 20, 10, 20, 20).

Tabel 1. Perhitungan untuk mendapatkan *centroid* ke 2

| D1        | MIN       | D(x)/ $\sum$ D(x) | Cumulative  |
|-----------|-----------|-------------------|-------------|
| 0         | 0         | 0                 | 0           |
| 14,1      | 14,1      | 0,09              | 0,09        |
| <b>10</b> | <b>10</b> | <b>0,07</b>       | <b>0,16</b> |
| 22,3      | 22,3      | 0,15              | 0,31        |
| 14,1      | 14,1      | 0,09              | 0,40        |
| 34,6      | 34,6      | 0,23              | 0,63        |
| 14,1      | 14,1      | 0,09              | 0,72        |
| 14,1      | 14,1      | 0,09              | 0,81        |
| 14,1      | 14,1      | 0,09              | 0,91        |
| 14,1      | 14,1      | 0,09              | 1           |

Setelah menghitung jarak data dengan *centroid* 1 maka langkah selanjutnya adalah menentukan nilai acak untuk mendapatkan *centroid* ke dua. Nilai acak : 0,1. Maka akan kita dapatkan nilai *centroid* ke 2 adalah No 3 (20, 10, 30, 20).

Tabel 2. Perhitungan untuk mendapatkan *centroid* ke 3

| D1          | D3          | MIN         | D(x)/ $\sum$ D(x) | Cumulative |
|-------------|-------------|-------------|-------------------|------------|
| 0           | 10          | 0           | 0                 | 0          |
| 14,1        | 22,3        | 14,1        | 0,1               | 0,1        |
| 10          | 0           | 0           | 0                 | 0,1        |
| 22,3        | 14,1        | 14,1        | 0,1               | 0,2        |
| 14,1        | 22,3        | 14,1        | 0,1               | 0,3        |
| <b>34,6</b> | <b>26,4</b> | <b>26,4</b> | <b>0,2</b>        | <b>0,5</b> |
| 14,1        | 22,3        | 14,1        | 0,1               | 0,6        |
| 14,1        | 22,3        | 14,1        | 0,1               | 0,7        |
| 14,1        | 22,3        | 14,1        | 0,1               | 0,8        |
| 14,1        | 22,3        | 14,1        | 0,1               | 0,9        |

Setelah menghitung jarak data dengan *centroid* 2 maka langkah selanjutnya adalah menentukan nilai acak untuk mendapatkan *centroid* ke tiga. Nilai acak: 0,4. Maka, akan kita dapatkan nilai *centroid* ke 3 adalah No. 6, ID 22 (OKTAVIOLA PUTRI, 10, 20, 50, 30).

Tabel 3. Perhitungan untuk mendapatkan *centroid* ke 4

| D1   | D3          | D6          | MIN  | D(x)/ $\sum$ D(x) | Cumulative |
|------|-------------|-------------|------|-------------------|------------|
| 0,0  | 10,0        | 34,6        | 0,0  | 0,0               | 0,0        |
| 14,1 | 22,4        | 46,9        | 14,1 | 0,1               | 0,1        |
| 10,0 | 0,0         | 26,5        | 0,0  | 0,0               | 0,1        |
| 22,4 | 14,1        | 26,5        | 14,1 | 0,1               | 0,3        |
| 14,1 | 22,4        | 46,9        | 14,1 | 0,1               | 0,4        |
| 34,6 | 26,5        | 0,0         | 0,0  | 0,0               | 0,4        |
| 14,1 | 22,4        | 46,9        | 14,1 | 0,1               | 0,6        |
| 14,1 | 22,4        | 46,9        | 14,1 | 0,1               | 0,7        |
| 14,1 | <b>22,4</b> | <b>46,9</b> | 14,1 | <b>0,1</b>        | <b>0,9</b> |
| 14,1 | 22,4        | 46,9        | 14,1 | 0,1               | 1,0        |

Setelah menghitung jarak data dengan *centroid* 3 maka langkah selanjutnya adalah menentukan nilai acak untuk mendapatkan *centroid* ke empat. Nilai acak : 0,8. Maka akan kita dapatkan nilai *centroid* ke 4 adalah No. 9, ID 53 (SITI AISYAH USMAN, 20, 10, 10, 10).

Tabel 4. Perhitungan untuk mendapatkan *centroid* ke 5

| D1          | D3   | D6          | D10         | MIN  | D(x)/ $\sum$ D(x) | Cumulative |
|-------------|------|-------------|-------------|------|-------------------|------------|
| 0,0         | 10,0 | 34,6        | 14,1        | 0,0  | 0,0               | 0,0        |
| 14,1        | 22,4 | 46,9        | 0,0         | 0,0  | 0,0               | 0,0        |
| 10,0        | 0,0  | 26,5        | 22,4        | 0,0  | 0,0               | 0,0        |
| <b>22,4</b> | 14,1 | <b>26,5</b> | <b>30,0</b> | 14,1 | <b>1,0</b>        | <b>1,0</b> |
| 14,1        | 22,4 | 46,9        | 0,0         | 0,0  | 0,0               | 1,0        |
| 34,6        | 26,5 | 0,0         | 46,9        | 0,0  | 0,0               | 1,0        |
| 14,1        | 22,4 | 46,9        | 0,0         | 0,0  | 0,0               | 1,0        |
| 14,1        | 22,4 | 46,9        | 0,0         | 0,0  | 0,0               | 1,0        |
| 14,1        | 22,4 | 46,9        | 0,0         | 0,0  | 0,0               | 1,0        |
| 14,1        | 22,4 | 46,9        | 0,0         | 0,0  | 0,0               | 1,0        |

Setelah menghitung jarak data dengan *centroid* 4 maka langkah selanjutnya adalah menentukan nilai acak untuk mendapatkan *centroid* ke lima. Nilai acak : 0,01. Maka akan kita dapatkan nilai *centroid* ke 5 adalah No. 4, ID 20 (SURATMAN I, 20, 10, 40, 10).

### B. Desain Model Algoritma K-Means

Setelah mendapatkan *centroid* 1 hingga 5 maka langkah selanjutnya adalah mengelompokkan *instance* ke dalam kelompok kualitas kerja hingga terjadi *convergent*. *Convergent* pada sampel ini terjadi di iterasi ke 2. Berikut hasil perhitungan pada iterasi ke-2:

Tabel 5. Hasil Perhitungan *Centroid* Terdekat Untuk Setiap Objek di iterasi ke-2

| No | Nama                | C1     | C2     | C3     | C4     | C5     |
|----|---------------------|--------|--------|--------|--------|--------|
| 1  | JUNIZA FIRDHA S     | 36,056 | 10     | 34,641 | 12,122 | 22,361 |
| 2  | JUNIZA FIRDHA S     | 26,458 | 22,361 | 46,904 | 2,020  | 30     |
| 3  | MUHAMMAD BIN MIDI H | 42,426 | 0      | 26,458 | 20,454 | 14,142 |
| 4  | SURATMAN I          | 46,904 | 14,142 | 26,458 | 28,607 | 0      |
| 5  | PUTRI EKANDINI      | 26,458 | 22,361 | 46,904 | 2,020  | 30,000 |
| 6  | OKTAVIOLA PUTRI     | 62,450 | 26,458 | 0      | 45,085 | 26,458 |

| No | Nama              | C1     | C2     | C3     | C4    | C5 |
|----|-------------------|--------|--------|--------|-------|----|
| 7  | MOEKIT            | 26,458 | 22,361 | 46,904 | 2,020 | 30 |
| 8  | SAKINAH SIREGAR   | 26,458 | 22,361 | 46,904 | 2,020 | 30 |
| 9  | SITI AISYAH USMAN | 26,458 | 22,361 | 46,904 | 2,020 | 30 |
| 10 | SITI AISYAH USMAN | 26,458 | 22,361 | 46,904 | 2,020 | 30 |

C. Desain Model Algoritma COP-KMeans

Tentukan setiap pasangan yang telah ditentukan dengan rumus COP-KMeans. Berikut hasil dari perhitungan COP-KMeans:

Tabel 1. Hasil perhitungan COP-KMeans terhadap hasil K-Means

| No | Nama                 | C1 | C2 | C3 | C4 | C5 |
|----|----------------------|----|----|----|----|----|
| 1  | JUNIZA FIRDHA S      |    |    |    | OK |    |
| 2  | JUNIZA FIRDHA S      |    |    |    | OK |    |
| 3  | MUHAMMAD BIN MIDI. H |    | OK |    |    |    |
| 4  | SURATMAN I           |    |    |    |    | OK |
| 5  | PUTRI EKANDINI       |    |    |    | OK |    |
| 6  | OKTAVIOLA PUTRI      |    |    | OK |    |    |
| 7  | MOEKIT               |    |    |    | OK |    |
| 8  | SAKINAH SIREGAR      |    |    |    | OK |    |
| 9  | SITI AISYAH USMAN    |    |    |    | OK |    |
| 10 | SITI AISYAH USMAN    |    |    |    | OK |    |

D. Desain Model Algoritma Gold Standard

Berdasarkan rumus *gold standard* terdapat  $j$  dan  $k$  kedua inisial ini memiliki arti yang berbeda,  $j$  bernilai jumlah *class* pada *gold standard* sedangkan  $k$  bernilai jumlah dari cluster perhitungan K-Means++ dan COP-KMeans, sehingga  $j = 5$  dan  $k = 5$ . Langkah berikutnya adalah menghitung nilai tiap-tiap kelompok.

1. Class 0 Kelompok Sangat Baik

Pada tabel *gold standard* tidak terdapat kelompok sangat baik maka nilai dari  $\sum_0 \max_0 |\omega_0 \cap c_0|$  adalah 0.

2. Class 1 Kelompok Baik

| Hasil Perhitungan |          |      | Gold Standard |          |      |
|-------------------|----------|------|---------------|----------|------|
| No                | Kelompok | Nama | Class         | Kelompok | Nama |

| Hasil Perhitungan |          |                   | Gold Standard |          |                   |
|-------------------|----------|-------------------|---------------|----------|-------------------|
| No                | Kelompok | Nama              | Class         | Kelompok | Nama              |
| 4                 | Buruk    | PUTRI EKANDINI    | 1             | Baik     | PUTRI EKANDINI    |
| 7                 | Buruk    | SAKINAH SIREGAR   | 1             | Baik     | SAKINAH SIREGAR   |
| 8                 | Buruk    | SITI AISYAH USMAN | 1             | Baik     | SITI AISYAH USMAN |

3. Class 2 Biasa

| Hasil Perhitungan |              |            | Gold Standard |          |            |
|-------------------|--------------|------------|---------------|----------|------------|
| No                | Kelompok     | Nama       | Class         | Kelompok | Nama       |
| 3                 | Sangat Buruk | SURATMAN I | 2             | Biasa    | SURATMAN I |

4. Class 3 Buruk

| Hasil Perhitungan |          |                 | Gold Standard |          |                 |
|-------------------|----------|-----------------|---------------|----------|-----------------|
| No                | Kelompok | Nama            | Class         | Kelompok | Nama            |
| 5                 | Biasa    | OKTAVIOLA PUTRI | 3             | Buruk    | OKTAVIOLA PUTRI |

5. Class 4 Sangat Buruk

| Hasil Perhitungan |          |                      | Gold Standard |              |                      |
|-------------------|----------|----------------------|---------------|--------------|----------------------|
| No                | Kelompok | Nama                 | Class         | Kelompok     | Nama                 |
| 2                 | Baik     | MUHAMMAD BIN MIDI. H | 4             | Sangat Buruk | MUHAMMAD BIN MIDI. H |
| 1                 | Buruk    | JUNIZA FIRDHA S      | 4             | Sangat Buruk | JUNIZA FIRDHA S      |
| 6                 | Buruk    | MOEKIT               | 4             | Sangat Buruk | MOEKIT               |

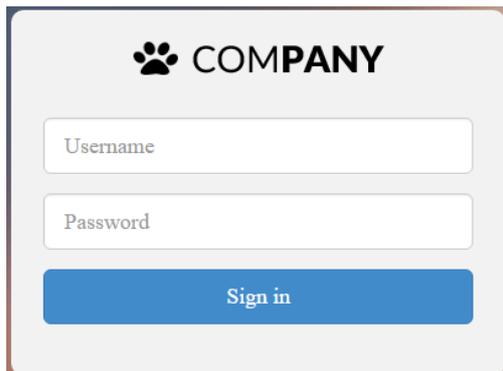
6. Substitusi Nilai Dari Setiap Kelompok

Setelah mendapatkan nilai dari tiap kelompok maka langkah selanjutnya adalah memasukkan nilai ke dalam rumus *purity* K-Means

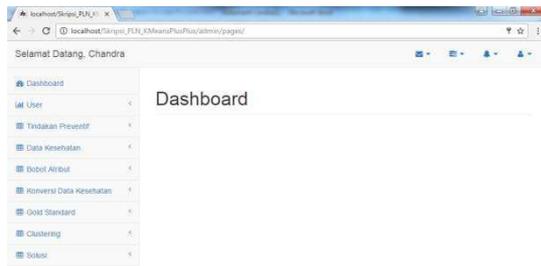
$$\begin{aligned}
 &= \frac{1}{N} \sum_k \max_j |\omega_k \cap c_j| \\
 &= \frac{1}{10} * ((\sum_0 \max_0 |\omega_0 \cap c_0|) + (\sum_1 \max_1 |\omega_1 \cap c_1|) + (\sum_2 \max_2 |\omega_2 \cap c_2|) \\
 &\quad + (\sum_3 \max_3 |\omega_3 \cap c_3|) + (\sum_4 \max_4 |\omega_4 \cap c_4|)) \\
 &= \frac{1}{10} * (0 + 0 + 0 + 0 + 0) \\
 &= \frac{1}{10} * (0) \\
 &= 0
 \end{aligned}$$

Menurut metode *purity* jika hasil mendekati 0 maka akurasi perhitungan K-Means++ dan COP-KMeans tidak tepat atau jika mendekati 1 maka akurasi perhitungan K-Means++ dan COP-KMeans tepat. Kesimpulan yang didapat dari perhitungan *purity* adalah perhitungan K-Means++ dan COP-KMeans tidak tepat karena bernilai 0.

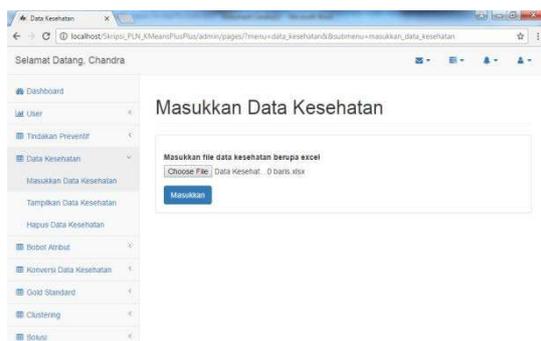
E. Desain Interface



Gambar 1. Login

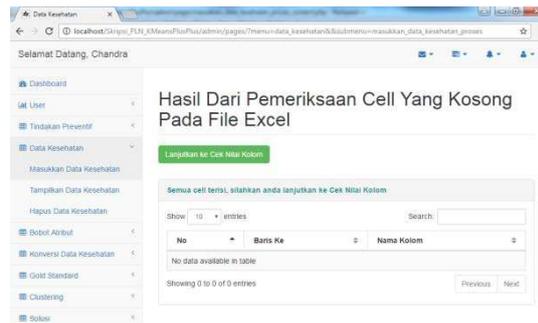


Gambar 2. Tampilan dashboard



Gambar 3. Masukkan database kesehatan berbentuk excel

Untuk memulai *clustering* dibutuhkan data masukkan yaitu data kesehatan pegawai melalui menu masukkan data kesehatan.

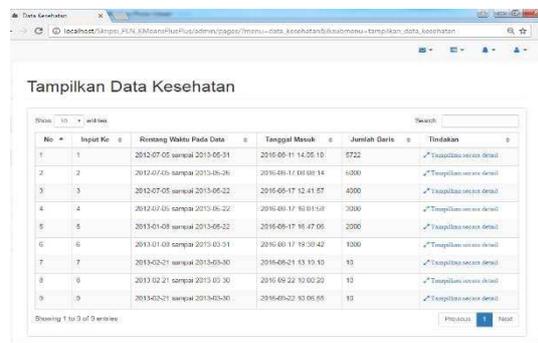


Gambar 4. Pemeriksaan cell yang kosong pada excel File yang telah diupload akan di periksa *cell* yang kosong, jika tidak ada maka akan lanjut ke tahap cek nilai pada *database*, jika iya aplikasi akan menolak *file* yang telah diupload.



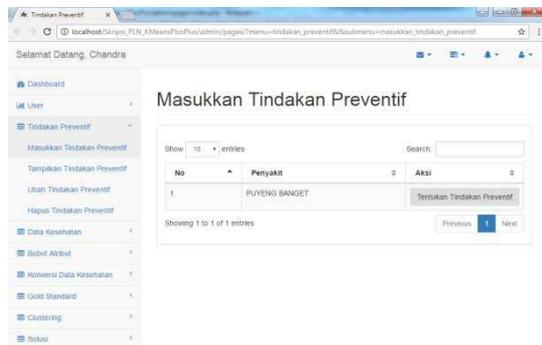
Gambar 5. Hasil cek nilai kolom

Jika tidak ada *cell* yang kosong maka akan dilanjutkan di tahap cek nilai kolom penyakit. Jika penyakit tidak ada di *database* atau telah ada tetapi belum diberi nilai maka akan muncul nama penyakit yang bermasalah dan harus diberi nilai jika data yang dimasukkan ingin di hitung. Jika penyakit ada dan telah diberi nilai maka hasil cek nilai kolom akan kosong.

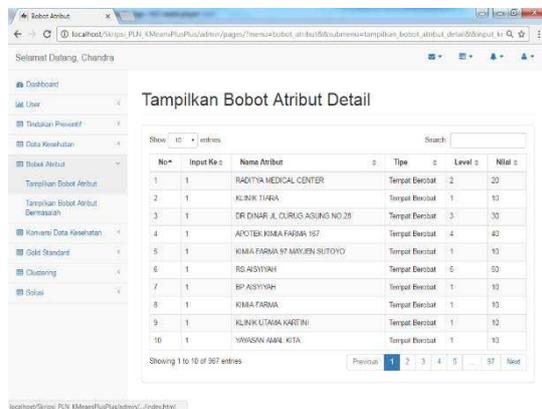


Gambar 6. Tampilkan data kesehatan yang telah dimasukkan

Setelah melewati proses cek nilai kolom penyakit maka data yang telah dimasukkan dapat dilihat pada menu tampilan data kesehatan.

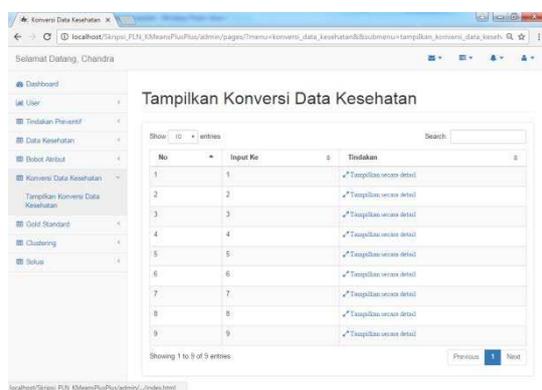


Gambar 7. Tentukan tindakan preventif jika belum ada di database

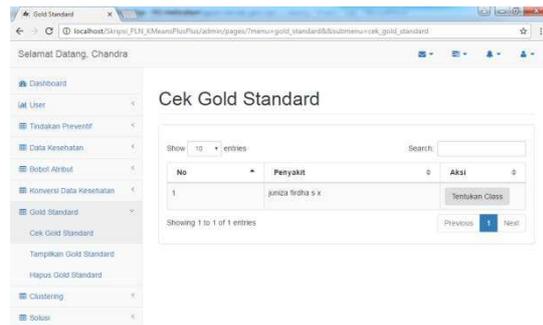


Gambar 8. Tampilan bobot atribut tiap-tiap atribut

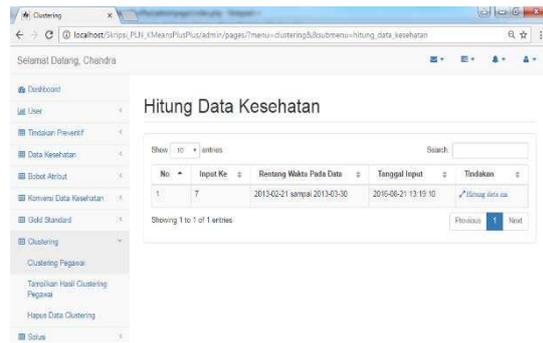
Setelah file telah diupload maka pada kolom rawat jalan, rawat inap, penyakit dan tempat berobat akan dirubah menjadi angka dan bisa dilihat pada Menu Bobot Atribut.



Gambar 9. Tampilan data kesehatan yang telah dirubah menjadi angka

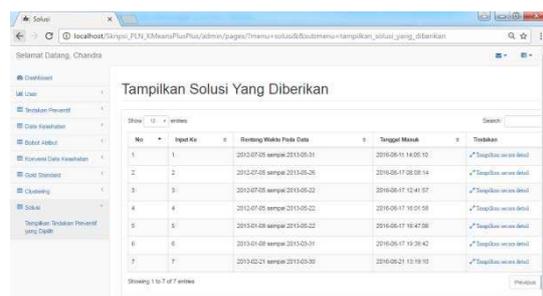


Gambar 10. Cek gold standard



Gambar 11. Hitung data kesehatan menggunakan K-Means++ dan COP-KMeans

Jika data yang telah dimasukkan tidak mengalami masalah maka ketika user klik Menu Clustering Pegawai akan muncul data yang telah dimasukkan user untuk dihitung hingga menghasilkan program kesehatan. Untuk mulai menghitung klik link hitung data ini, maka akan diproses oleh Aplikasi PPKP.



Gambar 12. Hasil perhitungan berupa solusi yang diberikan

## V. KESIMPULAN

Algoritma K-Means++ dan COP-KMeans dapat diimplementasikan pada database kesehatan pegawai dengan baik. Terbukti dari hasil perhitungan, kualitas kerja dapat dikelompokkan dalam 5 kelompok dan tidak ada pegawai yang

sama pada kelompok lain. Kecepatan perhitungan K-Means++ memang lebih lambat dibandingkan K-Means tetapi hasil K-Means++ lebih mendekati kebenaran dibandingkan K-Means. Tanpa COP-KMeans pegawai yang sama pasti berada di kelompok yang berbeda, sehingga jika tidak menggunakan COP-KMeans hasil penelitian akan ambigu.

Dan apakah penelitian ini telah menjawab rumusan masalah “Apakah hasil perhitungan dari K-Means++ dan COP-KMeans dapat memberikan perencanaan program pemeliharaan kesehatan pegawai yang sesuai dengan yang diinginkan oleh Bu Fatma?” jawaban dari rumusan masalah tersebut adalah iya, karena hasil akhir perhitungan K-Means++ dan COP-KMeans berupa 5 kelompok kualitas kerja dengan pegawai sebagai anggota kelompok dan penyakit yang diderita, pada tiap-tiap kelompok memiliki penyakit dan jumlah penderita yang beragam sehingga Bu Fatma dapat mempertimbangkan program kesehatan pegawai yang tepat untuk semester berikutnya.

## VI. SARAN

Aplikasi yang penulis buat tentu saja masih belum sempurna, masih banyak hal yang dapat dikembangkan guna membuat manfaat aplikasi menjadi lebih baik lagi untuk kedepannya. Oleh karena itu peneliti juga menyampaikan beberapa saran untuk pengembangan selanjutnya, yaitu:

1. Aplikasi tidak hanya mengolah *database* kesehatan pegawai tetapi bisa mengolah data yang menurut peneliti berikutnya cocok digunakan untuk metode K-Means++ dan COP-KMeans.
2. Pengamanan aplikasi perlu diperhatikan untuk pengembangan selanjutnya karena hasil dan sumber data dari aplikasi sangat rahasia

sehingga perlu adanya keamanan yang lebih pada aplikasi.

## REFERENSI

- [1] Lonnie Golden (2011). *The effects of working time on productivity and firm performance: a research synthesis paper*. GENEVA : Penn State University
- [2] Manvreet dan Usvir (2013). *Comparison Between K-Mean and Hierarchical Algorithm Using Query Redirection*. India: Department of CSE, Sri Guru Granth Sahib World University, Fatehgarh Sahib, Punjab.
- [3] Soumi dan Sanjay (2013). *Comparative Analysis of K-Means and Fuzzy C-Means Algorithms*. India: Department of Computer Science and Engineering, Amity University, Uttar Pradesh Noida.