

Analisis Pola Keterkaitan Profil Dengan Tingkat Kehadiran Pegawai Menggunakan *Clustering K-Means*. Studi Kasus Pada Sekretariat BPJT

The Analysis of Association Rule Between Profile and Employee's Presence Using K-Means Clustering. Case Study BPJT Secretariat

Mujiono Sadikin¹ Tyas Atus Hapsari²

Fakultas Ilmu Komputer, Universitas Mercubuana Jakarta

Jl. Meruya Selatan No.1, Kembangan, Daerah Khusus Ibukota Jakarta

Email : mujiono.sadikin@mercubuana.ac.id¹, 41516120135@student.mercubuana.ac.id²

Abstrak

Berdasarkan amanat peraturan perundang undangan yang berlaku, Pemerintah wajib melaksanakan penegakan disiplin terhadap para pegawai, baik Pegawai Negeri Sipil (PNS) atau Pegawai non PNS. Salah satu aspek yang terkait dengan tingkat disiplin pegawai adalah kehadiran. Kehadiran ini berkenaan dengan tanggung jawab pegawai saat bekerja. Pegawai yang hadir tepat waktu dan tidak terlambat saat masuk kerja bisa dikatakan mempunyai sifat disiplin baik. Penelitian ini dilakukan untuk mengetahui pola keterkaitan kehadiran dan karakteristik pegawai di lingkungan Sekretariat Badan Pengatur Jalan Tol (BPJT). Dataset yang digunakan dalam penelitian ini adalah rekap kehadiran pegawai tahun 2017. Analisis pola dilakukan menggunakan teknik *clustering*. Hipotesa yang diajukan dalam penelitian ini adalah bahwa ada hubungan antara masa kerja dengan tingkat kehadiran pegawai. Hasil analisis menunjukkan bahwa terdapat beberapa kelompok pegawai yang perlu ditingkatkan kedisiplinannya dan kelompok pegawai yang sudah baik. Ciri-ciri pegawai yang kurang baik tingkat kehadirannya adalah laki-laki, sudah menikah, Non PNS, masa kerja 0-10 tahun, staf/ pelaksana, usia 31-40 tahun, tingkat kehadiran 70-80 %, dan ketidakhadiran lebih dari 20%. Sedangkan ciri-ciri kelompok pegawai yang memiliki kedisiplinan yang baik adalah laki-laki, sudah menikah, PNS, masa kerja 0-10 tahun, staf/ pelaksana, usia 31-40 tahun, tingkat kehadiran 90-100 %, dan ketidakhadiran lebih dari 0-5%. Dari hasil analisis dapat disimpulkan bahwa bahwa tidak ada hubungan antara masa kerja dan tingkat kedisiplinan kehadiran pegawai di lingkungan Sekretariat BPJT, karena ciri- ciri kelompok pegawai dengan kehadiran terbaik dan terburuk berada pada rentang masa kerja 0-10 tahun.

Kata Kunci: K-Means, *Clustering*, Data Mining, Kehadiran Pegawai, BPJT.

Abstract

Based on the mandate of the applicable laws and regulations, the Government is obliged to carry out discipline enforcement of employees, both Civil Servants (PNS) or non-PNS Employees. One aspect related to employee discipline is attendance. This presence relates to employee responsibilities at work. Employees who are present on time and does not late when arrive in the office is indicated to have good discipline. This study was conducted to determine the pattern of the relationship between the presence and characteristics of employees within the Secretariat of the Toll Road Regulatory Agency (BPJT). The dataset used in this study was a recap of employee attendance in 2017. Pattern analysis was performed by the using of clustering technique. The hypothesis proposed in this study is that there is a relationship between work tenure and employee attendance. The analysis results show that there are several groups of employees that need to be improved in discipline and a good group of employees as well. The characteristics of employees who are poor attendance rate are men, married, non PNS, tenure of 10-10 years, staff/ executive, age 31-40 years, attendance rate of 70-80%, and absenteeism of more than 20%. While the characteristics of a group of employees who have good discipline are male, married, civil servants, tenure of 10-10 years, staff/ executives, age 31-40 years, attendance rate of 90-100%, and absence of more than 0 -5%. From the results of the analysis it can be concluded that there is no relationship between the length of work and the level of discipline of employee attendance within the BPJT Secretariat, because the characteristics of the group of employees with the best and worst attendance are in the range of 0-10 years.

Keywords: K-Means, Clustering, Data Mining, Employee Presence, BPJT

1. Pendahuluan

Dari pengamatan secara umum terhadap rekap kehadiran pegawai di lingkungan Sekretariat BPJT terlihat bahwa ketertiban kehadiran pegawai masih perlu ditingkatkan. Hal ini ditunjukkan dengan ditemukannya pegawai yang datang melebihi jam kerja ataupun pulang sebelum waktunya. Tata tertib jam kerja yang berlaku adalah pukul 08.00 – 16.30 untuk Hari Senin-Kamis dan pukul 08.00-17.00 untuk Hari Jumat. Pegawai yang baru saja bekerja atau yang memiliki masa kerja sampai dengan 5 tahun pada umumnya memiliki rekap kehadiran baik sekali. Sedangkan pegawai yang memiliki masa kerja di atas 5 tahun terlihat lebih sering datang terlambat.

Untuk memastikan keterkaitan pola / profil data kehadiran dan karakteristik pegawai diperlukan eksplorasi lebih lanjut. Cara pengenalan pola informasi / pengetahuan dari sekumpulan data dikenal sebagai teknik data mining dan salah satu algoritmanya adalah klastering. Penerapan algoritma klastering untuk pengenalan profil data telah banyak digunakan. Di antara publikasi hasil penelitian terkait klastering antar lain: pengenalan pola kecelakaan lalu lintas (Rahmat C.T.I et al., 2017), pemprofilan calon mahasiswa baru (Ramadhani, 2014), pengenalan pola karakteristik siswa berprestasi (Sibuea & Safta, 2017), pemetaan pola hasil panen pada bidang pertanian (Ridlo, Defiyanti, & Primajaya, 2017), dan analisis pola informai produk dan penjualan produk pada minimarket (Ghofar & Kurniawan, 2018; Metisen & Sari, 2015).

Pada penelitian ini dilakukan pengujian atas hipotes apakah benar terdapat hubungan antara masa kerja dengan tingkat kehadiran pegawai di lingkungan Sekretariat BPJT. Hipotesa ini diambil sebagai dasar penelitian karena terdapat asumsi umum bahwa pegawai dengan masa kerja lebih lama, karena merasa lebih senior, berkinerja kehadiran kurang baik. Pengujian dilakukan dengan melakukan klasterisasi dataset kehadiran pegawai menggunakan teknik klastering k-means, memanfaatkan kakas rapid miner. Diharapkan dari hasil penelitian ini bisa menjadikan salah satu pertimbangan Kepala Sub Bagian Administrasi dan Kepegawaian dalam menentukan kebijakan atau strategi terkait peningkatan ketertiban pegawai.

Data pendukung dalam penelitian ini adalah rekap kehadiran tiap bulan. Di dalam rekap kehadiran ini terdapat atribut-atribut mengenai keterangan dan tautan bukti alasan sah untuk datang terlambat, izin, cuti ataupun dinas luar kota. Rekapitulasi kehadiran pegawai di lingkungan Sekretariat BPJT dan diunduh dari aplikasi kehadiran online kepegawaian bina marga. Pegawai PNS dan non PNS menggunakan aplikasi tersebut untuk rekap kehadiran setiap harinya. Alamat website untuk mengakses aplikasi online tersebut yaitu <http://www.kepegbima.com:8080/kehadiranbinamarga/>.

2. Studi Terkait

Klastering sebagai salah satu teknik penambangan informasi pada data telah banyak digunakan. Beberapa publikasi penelitian terkait analisis profile data dengan menggunakan teknik klastering disajikan pada bagian ini. Roni Setiawan dkk (Setiawan, 2016) meneliti profil calon mahasiswa LP3I Jakarta menggunakan k-means. Pada penelitiannya, penulis menguji klister dengan mencoba berbagai nilai k mulai k=2 sampai dengan k=5. Berdasarkan hasil uji coba jumlah cluster yang dapat mewakili profil atau karakteristik calon mahasiswa di Politeknik LP3I Jakarta adalah dengan k=4. Dengan k=4 dihasilkan tiga segmentasi calon mahasiswa berdasarkan rata-rata jumlah pembayaran uang registrasi. Penelitian ini menghasilkan pengetahuan mengenai karakteristik calon mahasiswa yang kemudian digunakan untuk menentukan strategi promosi.

Brilian Rahmat dkk (Rahmat C.T.I et al., 2017) menyajikan hasil penelitian tentang analisis data kecelakaan menggunakan k-means dengan aplikasi RapidMiner. Pada penelitiannya,

penulis menggunakan 500 set contoh data kecelakaan yang dikelompokkan menjadi 3 klastr. Hasil analisisnya menunjukkan pola keterkaitan antara frekuensi tingkat kecelakaan di tiap lokasi dan waktu-waktu rawan yang berpotensi terjadi kasus kecelakaan.

Penggunaan k-means untuk keperluan analisis yang lain ditunjukkan pada penelitian yang dipublikasikan di (Ghofar & Kurniawan, 2018). Makalah penulis ini menyajikan hasil analisis dataset pelanggan di UMS Store. Dengan menggunakan sampel data sebanyak 357, dilakukan pengelompokan atas 3 kelompok klaster. Dengan membandingkan nilai atribut dalam setiap kelompok dihasilkan kesimpulan bahwa kelompok pelanggan potensial berada dalam kluster ke 3. Kelompok inilah yang direkomendasikan sebagai target pemasaran UMS Store.

Dari literatur review yang telah dilakukan, terlihat bahwa K-Means fisibel untuk melakukan analisis kluster. Dengan pertimbangan tersebut, pengenalan pola kehadiran karyawan dalam penelitian ini juga didekati menggunakan algoritma K-means. Dari hasil klusterisasi kemudian dianalisis bagaimana profile karyawan yang bertetangga (mirip dalam satu klaster). Hipotesa yang ingin dibuktikan pada penelitian ini adalah apakah ada hubungan antara masa kerja dengan tingkat kedisiplinan kehadiran pegawai.

3. Material Dan Metode

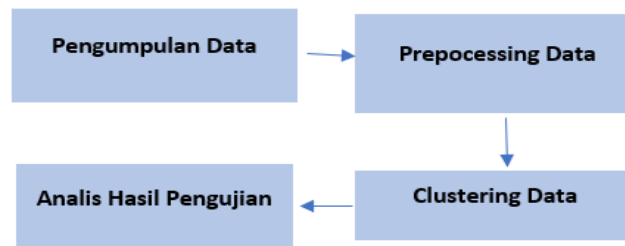
Penelitian ini dilakukan di Sekretariat Badan Pengatur Jalan Tol dengan melakukan analisis terhadap data rekap kehadiran 2017. Tujuan dari penelitian yang dilakukan ini adalah untuk mencari karakteristik kehadiran pegawai berdasarkan pola kehadiran. Hasil analisis selanjutnya digunakan sebagai pertimbangan strategi dalam meningkatkan kedisiplinan pegawai. Hasil penelitian ini juga diajukan sebagai bahan referensi untuk mengetahui kelompok karyawan mana yang memerlukan arahan lebih intensif. Pihak pimpinan BPJT mengharapkan agar kinerja pegawai meningkat dan mendukung pencapaian tujuan organisasi.

3.1. Dataset

Data kehadiran pegawai di Sekretariat Badan Pengatur Jalan Tol diambil dari rekap aplikasi kehadiran online. Setiap instant data mempunyai atribut yang terkait dengan informasi pribadi dan atribut yang terkait dengan kehadiran. Atribut yang terkait informasi pribadi adalah: *Jenis_Kelamin, Status_Perkawinan, Status_Kepegawaian, Masa_Kerja, Jabatan, dan Usia*. Sementara atribut terkait kehadiran adalah: *Per_Hadir, Per_Tidak_Hadir, Telat, Tidak Masuk, Cuti, Dinas Luar, dan Ijin tidak masuk*. Setiap pegawai memiliki file excel masing-masing untuk rekap kehadiran per tahun. Karenanya, kemudian dilakukan penggabungan data seluruh pegawai. Pada akhirnya diperoleh dataset sebanyak 1.036 instan data.

3.2. Tahapan Penelitian

Tahap pertama penelitian ini adalah pengumpulan data kehadiran masing-masing pegawai secara lengkap tiap bulan selama satu tahun. Tahap selanjutnya adalah *preprocessing*. Pada tahap *pre-processing* dilakukan penghapusan untuk instan-instan yang mengandung nilai attribute bias atau ambigu. Selanjutnya dilakukan normalisasi nilai attribute setiap instan. Pada data hasil pre-processing kemudian diterapkan klusterisasi menggunakan algoritma k-means untuk menguji beberapa kemungkinan pembagian kelompok. Setelah diperoleh pengklasteran yang paling sesuai dengan tujuan penelitian, maka langkah terakhir adalah analisa hasil pengujian. Analisis dilakukan terhadap masing-masing kelompok/ cluster yang ada. Tahapan penelitian terlihat seperti Gambar 1.



Gambar 1. Tahapan Penelitian

Normalisasi data pada tahap *preprocessing* proses menkodekan nilai atribut yang kontinyu menjadi kategorikal. Tabel 1 berikut menyajikan normalisasi nilai atribut instan yang diterapkan pada dataset.

Tabel 1. Pengkodean Nilai Attribute Dataset

| Atribut | Nilai Asli | Normalisasi | Atribut | Nilai Asli | Normalisasi |
|---------------------------|---------------|-----------------|--------------------------------|--------------------|-------------|
| Jenis_Kelamin | Perempuan | 0 | TLT_PSW *Telat | 0 menit | 0 |
| | Laki-laki | 1 | | >0, <=60 menit | 1 |
| Status_Perkawinan | Lajang | 0 | | >60, <120 menit | 2 |
| | Menikah | 1 | | >=120, < 180 menit | 3 |
| | Bercerai | 2 | | >=180, < 240 menit | 4 |
| Status_Kepegawaian | Non_PNS | 0 | >=240 menit | 5 | |
| | PNS | 1 | TM *Tidak Masuk | 0 hari | 0 |
| Masa_Kerja | 0 - 10 | 0 | | >0, <= 5 hari | 1 |
| | 11 - 20 | 1 | | >5, <= 10 hari | 2 |
| | 21 - 30 | 2 | | >10, <=15 hari | 3 |
| | > 30 | 3 | | > 15, <=20 hari | 4 |
| Jabatan | pelaksana | 0 | > 20 hari | 5 | |
| | non_eselon | 1 | C *Cuti | 0 hari | 0 |
| | fungsional | 2 | | >0, <= 5 hari | 1 |
| | eselon_4 | 3 | | >5, <= 10 hari | 2 |
| | eselon_3 | 4 | | >10, <=15 hari | 3 |
| eselon_2 | 5 | > 15, <=20 hari | | 4 | |
| Usia | 20 - 30 | 0 | > 20 hari | 5 | |
| | 31 - 40 | 1 | DL *Dinas Luar | 0 hari | 0 |
| | 41 - 50 | 2 | | >0, <= 5 hari | 1 |
| | 51 - 60 | 3 | | >5, <= 10 hari | 2 |
| Per_Hadir | < 70% | 0 | | >10, <=15 hari | 3 |
| | >= 70%, < 80% | 1 | | > 15, <=20 hari | 4 |
| | >= 80%, < 90% | 2 | > 20 hari | 5 | |

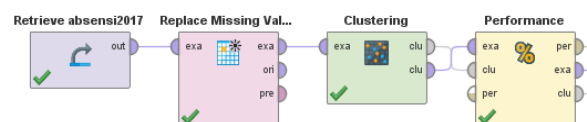
| Atribut | Nilai Asli | Normalisasi | Atribut | Nilai Asli | Normalisasi |
|-----------------|-----------------|-------------|--------------------------|-----------------|-------------|
| Per_Tidak_Hadir | >= 90%, <= 100% | 3 | ITM *Ijin tidak masuk | 0 hari | 0 |
| | <= 5% | 0 | | >0, <= 5 hari | 1 |
| | > 5%, < 10% | 1 | | >5, <= 10 hari | 2 |
| | >= 10%, < 15% | 2 | | >10, <=15 hari | 3 |
| | >= 15%, < 30% | 3 | | > 15, <=20 hari | 4 |
| | >= 20% | 4 | | > 20 hari | 5 |

3.3. Metode

Digunakan algoritma K-Means pada studi ini untuk menganalisis pola kehadiran karyawan BPJT dalam kaitannya dengan karakteristik karyawan. Algoritma K-Means merupakan teknik partisi untuk klastering, yang meskipun konvensional namun masih cukup powerful untuk berbagai keperluan analisis (Arora, Deepali, & Varshney, 2016). K-means *clustering* menggunakan jarak rerata (*mean*) antar objek sebagai indikator kesamaan di antara objek-objek. K-means membagi sejumlah objek ke dalam partisi-partisi berdasarkan kategori-kategori yang ada dengan melihat titik tengah yang diberikan. Pengelompokan objek dilihat dari jarak objek dengan titik tengah yang paling dekat. Setelah mengetahui titik tengah terdekat, objek tersebut akan diklasifikasikan sebagai anggota dari kategori tersebut.

Kakas yang digunakan pada penelitian ini adalah aplikasi *Rapid Miner*. Langkah-langkah dalam penyelesaian pengklasteran menggunakan algoritma k-mean Rapid Miner berturut-turut: *Import* dataset ke rapidminer studio, *Perancangan model clustering k-means*, dan *Pengujian dalam pembagian kelompok*.

Perancangan tahapan klastering k-means menggunakan Rapid Miner ditunjukkan seperti gambar 2. Rancangan terdiri dari dataset kehadiran 2017, *tool replace missing value*, *tool k-mean*, dan *tool performance*. Dalam perancangan *clustering* harus disilangkan ke *tool performance* untuk memperoleh hasil yang sesuai.



Gambar 2. Rancangan Proses *Clustering* Rapid Miner

Agar memperoleh hasil terbaik maka perlu dilakukan pengujian dalam pembagian kelompok objek, yaitu dengan mengubah-ubah nilai k. Kemudian dipilih kelompok yang paling mungkin menjawab pertanyaan apakah ada hubungan antara masa kerja dengan tingkat kedisiplinan kehadiran pegawai. Dengan prinsip utama adalah data dalam satu cluster memiliki tingkat kemiripan yang maksimum dan data antar cluster memiliki kemiripan yang minimum.

3.4. Evaluasi Cluster

Untuk melakukan evaluasi kualitas hasil *clustering* pada penelitian ini diukur digunakan *Silhouette Coefficient* sebagai parameter kinerja. *Silhouette Coefficient* merupakan index yang umum digunakan untuk melakukan evaluasi klasster, bisa diacu pada (et al., 2016; Romadon & Fitriana, 2019). *Silhouette Coefficient* merupakan indikator seberapa mirip suatu instan dengan instan lainnya dalam satu grup klaster dibandingkan dengan grup klaster yang lain (Rousseeuw, 1986). Fourmula untuk perhitungan *Silhouette Coefficient* diturunkan seperti persamaan (1), (2), dan (3).

Untuk setiap poin instan $i \in C_i$ (data point i di dalam cluster C_i), jarak rata-2 antara data poin ke i dengan semua poin yang lain adalah:

$$a(i) = \frac{1}{|C_i|-1} \sum_{j \in C_i, i \neq j} d(i, j) \quad (1)$$

Dan untuk poin instan $i \in C_i$, jarak rata-rata minimum ke poin yang lain dalam C

$$b(i) = \min_{i \neq k} \frac{1}{|C_k|} \sum_{j \in C_k} d(i, j) \quad (2)$$

Maka nilai *Silhouette Coefficient* poin i didefinisikan sebagai:

$$s(i) = \frac{b(i)-a(i)}{\max\{a(i), b(i)\}}, \text{ jika } |C_i| > 1 ; s(i) = 0, \text{ jika } |C_i| = 1 \quad (3)$$

4. Hasil Dan Pembahasan

Nilai *Silhouette Coefficient*

Dengan melakukan eksperimen *clustering* untuk $k=2$ s.d. $k=6$ diperoleh nilai SC seperti tabel 2. Untuk semua nilai k yang diujicobakan, diperoleh nilai SC yang tidak signifikan (semuanya di bawah 0.5). Namun demikian berdasarkan nilai SC, pilihan $k=6$ menghasilkan kualitas klaster yang paling baik.

Tabel 2. Analisis Nilai *Centroi*

| k | Rata-SC |
|---|---------------|
| 2 | 0.3732 |
| 3 | 0.3563 |
| 4 | 0.3474 |
| 5 | 0.3146 |
| 6 | 0.3798 |

Berdasarkan beberapa pengujian dengan menggunakan parameter k yang berbeda-beda diperoleh hasil yang paling baik adalah dengan $k=6$. Untuk pengujian menggunakan $k=2$ ataupun $k=4$, hasil cluster yang diperoleh masih kurang spesifik. Analisis nilai *centroid* diperlihatkan pada tabel *centroid* tabel 2. Urutan persentase kehadiran terbaik berdasarkan nilai yang diperoleh adalah *cluster 1*, *cluster 5*, *cluster 2*, *cluster 3*, *cluster 4* dan *cluster 0*. Untuk *cluster 1* memiliki karakteristik masa kerja 2 atau 21-30 tahun dan status kepegawaiannya sudah PNS. Cluster 5 memiliki karakteristik masa kerja 2 atau 21-30 tahun dan status kepegawaiannya PNS. *Cluster 2* memiliki karakteristik masa kerja 0 atau 0-10 tahun dan status kepegawaiannya Non PNS. *Cluster 3* memiliki karakteristik masa kerja 0 atau 0-10 tahun dan status kepegawaiannya Non PNS. *Cluster 4* memiliki karakteristik masa kerja 0 atau 0-10 tahun dan status kepegawaiannya Non PNS. Sedangkan *cluster 0* memiliki karakteristik masa kerja 0 atau 0-10 tahun dan status kepegawaiannya Non PNS.

Salah satu contoh analisis karakteristik pegawai pada klaster, diperlihatkan pada tabel 3 yang menyajikan ciri *cluster 0*. Warna kuning adalah frekuensi terbanyak. Pegawai pada *cluster 0*, sebanyak 327, menunjukkan bahwa pegawai pada klaster ini memiliki tingkat kedisiplinan yang cukup baik karena sebagian besar persentase kehadiran adalah **90-100%**, persentase tidak hadir adalah tidak lebih dari 5%, pernah telat ataupun pulang sebelum waktunya lebih dari 240 menit, tetapi tidak pernah membolos ataupun izin, dan memiliki masa kerja **0-10**. Pada *cluster 1* dengan total 415 item, dapat dilihat bahwa pegawai pada klaster ini memiliki tingkat kedisiplinan yang sangat baik karena sebagian besar persentase kehadiran adalah **90-100%**, persentase tidak

hadir tidak lebih dari 5%, tidak pernah telat ataupun pulang sebelum waktunya, tidak pernah bolos ataupun izin, dan memiliki masa kerja **0-10 tahun**.

Tabel 3. Analisis Nilai *Centroid*

| Attribute | cluster_0 | cluster_1 | cluster_2 | cluster_3 | cluster_4 | cluster_5 |
|--------------------|-----------|-----------|-----------|-----------|-----------|-----------|
| Jenis_Kelamin | 0.54 | 0.93 | 0.66 | 0.87 | 0.44 | 0.65 |
| Status_Perkawainan | 1.10 | 1.00 | 0.67 | 0.94 | 0.71 | 0.85 |
| Status_Kepegawaian | 1.00 | 1.00 | 0.42 | 0.31 | 0.47 | 0.30 |
| Masa_Kerja | 2.67 | 2.78 | 0.55 | 0.87 | 0.61 | 0.69 |
| Jabatan | 1.39 | 1.79 | 0.19 | 0.19 | 0.29 | 0.06 |
| Usia | 2.41 | 2.58 | 0.97 | 1.52 | 1.12 | 1.67 |
| Per_Hadir | 2.99 | 2.98 | 3.00 | 0.94 | 3.00 | 1.67 |
| Per_Tidak_Hadir | 0.15 | 0.10 | 0.11 | 3.50 | 0.17 | 2.69 |
| TLT + PSW | 4.60 | 0.43 | 0.51 | 0.83 | 4.46 | 4.54 |
| TM | 0.24 | 0.08 | 0.15 | 1.73 | 0.26 | 1.20 |

Tabel 4. Karakteristik Pegawai Pada *Cluster 0*

| Cluster 0 | | | | | |
|-----------------|------------|-----|-------------|-------------|-----|
| Jenis_Kelamin | Perempuan | 183 | TLT_PSW | 0 menit | 0 |
| | Laki-Laki | 144 | | <= 60 menit | 0 |
| Status_Perkawin | Lajang | 80 | < 120 menit | 0 | |
| | Menikah | 235 | < 180 | 50 | |
| | Bercerai | 12 | < 240 | 36 | |
| Status_Kepegaw | Non_PNS | 150 | >= 240 | 241 | |
| | PNS | 177 | = 0 Hari | 215 | |
| Masa_Kerja | 0 - 10 | 227 | TM | <=5 Hari | 112 |
| | 11 - 20 | 39 | | <10 Hari | 0 |
| | 21 - 30 | 35 | | <15 Hari | 0 |
| | 31 keatas | 26 | | <20 Hari | 0 |
| | | | | >=20 Hari | 0 |
| Jabatan | pelaksana | 284 | C | = 0 Hari | 324 |
| | non_eselon | 0 | | <=5 Hari | 3 |
| | Jafung | 9 | | <10 Hari | 0 |
| | eselon 4 | 34 | | <15 Hari | 0 |
| | eselon 3 | 0 | | <20 Hari | 0 |
| Usia | eselon 2 | 0 | DL | >=20 Hari | 0 |
| | 20-30 | 57 | | = 0 Hari | 111 |
| | 31-40 | 149 | | <=5 Hari | 142 |
| Per_Hadir | 41-50 | 52 | <10 Hari | 52 | |
| | 51-60 | 69 | <15 Hari | 17 | |
| | <70% | 0 | <20 Hari | 4 | |
| | <80% | 8 | >=20 Hari | 1 | |
| Per_Tidak_Hadir | <90% | 71 | ITM | = 0 Hari | 251 |
| | <=100% | 248 | | <=5 Hari | 76 |
| | <=5% | 183 | | <10 Hari | 0 |
| | <10% | 82 | | <15 Hari | 0 |
| | <15% | 49 | | <20 Hari | 0 |
| | <20% | 13 | >=20 Hari | 0 | |
| | >=20% | 0 | | | |

Cluster 2 dengan total 185 item, menunjukkan karakteristik pegawai yang memiliki tingkat kedisiplinan yang sangat baik karena sebagian besar persentase kehadiran adalah **90-100%**, persentase tidak hadir adalah tidak lebih dari 5%, tidak pernah telat ataupun pulang sebelum waktunya, tidak pernah bolos ataupun izin, dan memiliki masa kerja **21-30 tahun**. *Cluster 3* beranggotakan 109 pegawai dengan ciri-ciri memiliki tingkat kedisiplinan yang buruk karena sebagian besar persentase kehadiran adalah **kurang dari 70%**, persentase tidak hadir adalah lebih dari 20%, sering telat ataupun pulang sebelum waktunya, sering bolos, dan memiliki masa kerja **0-10 tahun**.

Secara keseluruhan hasil. Analisis menunjukkan bahwa kelompok pegawai dengan tingkat kehadiran **terbaik** ada di cluster 1 sedangkan pegawai dengan tingkat kehadiran **terburuk**

ada di cluster 3. Lebih rinci lagi ciri-ciri pegawai terbaik dan terburuk tersebut dapat dilihat pada tabel 4.

Tabel 4. Analisis Kehadiran Pegawai Terbaik dan Terburuk

| No | Ciri - Ciri Pegawai | |
|----|-------------------------------------|---|
| | Cluster 1 (Terbaik) | Cluster 3 (Terburuk) |
| 1. | Laki-laki | Laki-laki |
| 2. | Sudah menikah | Sudah menikah |
| 3. | PNS | Non PNS |
| 4. | Masa kerja 0-10 tahun | Masa kerja 0-10 tahun |
| 5. | Staf/ pelaksana | Staf/ pelaksana, |
| 6. | Usia 31-40 tahun | Usia 31-40 tahun |
| 7. | Tingkat kehadiran 90-100% per tahun | Tingkat kehadiran kurang dari 70% per tahun |
| 8. | Ketidakhadiran 0-5% per tahun | Ketidakhadiran lebih dari 20% per tahun |

5. Kesimpulan & Studi Lanjutan

Makalah ini menyajikan hasil studi anilis kinerja kehadiran pegawai pada Sekretariat BPJT dalam kaitannya dengan dengan karakteristik mereka. Digunakan algoritma *k-means clustering* sebagai kakas analisis. Hasil studi menunjukkan bahwa secara umum sebagian besar pegawai di Sekretariat BPJT memiliki tingkat kehadiran yang baik. Pegawai dengan kinerja kurang baik segingga memerlukan perhatian khusus, dari hasil analisis termasuk ke dalam *cluster* 3. Ciri-ciri pegawai dengan kinerja kurang baik tersebut adalah laki-laki, sudah menikah, Non-PNS, masa kerja 0-10 tahun, staf/ pelaksana, usia 31-40 tahun. Pegawai pada klaster ini mencapai tingkat kehadiran hanya 70-80 % per tahun, dan ketidakhadiran lebih dari 20% per tahun. Hasil penelitian juga menunjukkan bahwa tidak terdapat hubungan antara masa kerja dengan tingkat kehadiran pegawai di lingkungan Sekretariat BPJT karena ciri- ciri kelompok pegawai dengan tingkat kehadiran terbaik dan terburuk berada pada rentang masa kerja 0-10 tahun.

Pada studi berikutnya kami merencanakan untuk mengeksplorasi lebih jauh untuk mengetahui factor-faktor apa saja yang mempengaruhi kelompok pegawai di klaster 3 sehingga kinerja kehadiran mereka kurang baik. Beberapa kemungkinan factor yang akan dieksplorasi misalnya factor Pendidikan, factor peraturan ke-SDM-an, maupun factor keuangan. Diperlukan integrasi data dari bagian SDM dan keuangan untuk studi lanjutan ini.

Daftar Acuan

- Arora, P., Deepali, & Varshney, S. (2016). Analysis of K-Means and K-Medoids Algorithm for Big Data. *Physics Procedia*, Vol.78 pp. 507–512.
- Fitriah, D., Fahmi, H., ... Arymurthy, A. M. (2016). A Data Mining Based Approach for Determining the Potential Fishing Zones. *International Journal of Information and Education Technology*, Vol. 6(3), pp. 187–191.
- Ghofar, M. A., & Kurniawan, Y. I. (2018). Aplikasi Pengelompokan Pelanggan Pada Ums Store Menggunakan Algoritma K-Means. *Jurnal Teknologi dan Manajemen Informatika*, Vol. 4(1).
- Metisen, B. M., & Sari, H. L. (2015). Analisis clustering menggunakan metode K-Means dalam pengelompokan penjualan produk pada Swalayan Fadhila. *Jurnal Media Infotama*, Vol. 11(2), pp. 110–118.
- Rahmat C.T.I, B., Gafar, A. A., Fajriani, N., Ramdani, U., Uyun, F. R., P, Y. P., & Ransi, N. (2017). Implemetasi K-Means Clustering Pada Rapidminer Untuk Analisis Daerah Rawan

- Kecelakaan. In *Prosiding Seminar Nasional Riset Kuantitatif Terapan 2017* Vol. 1, pp. 58–62.
- Ramadhani, R. D. (2014). Data Mining Menggunakan Algoritma K-Means Clustering Untuk Menentukan Strategi Promosi Universitas Dian Nuswantoro. *Industrial Marketing Management*, Vol. 1(1), pp. 1–9.
- Ridlo, M. R., Defiyanti, S., & Primajaya, A. (2017). Implementasi Algoritme K-Means Untuk Pemetaan Produktivitas Panen Padi Di Kabupaten Karawang. In *Proceeding Citee 2017* (hal. 426–433). Yogyakarta.
- Romadon, Y. W., & Fitriana, D. (2019). The Comparative Study on Clustering Method Using Hospital Facility Data in Jakarta District and Surrounding Areas. *International Journal of Machine Learning and Computing*, Vol. 9(6).
- Rousseeuw, P. J. (1986). Silhouettes: a graphical aid to the interpretation and validation of cluster analysis. *Journal of Computational and Applied Mathematics*, Vol. 20 pp. 53–65.
- Setiawan, R. (2016). Penerapan Data Mining Menggunakan Algoritma K-Means Clustering Untuk Menentukan Strategi Promosi Mahasiswa Baru (Studi Kasus: Politeknik Lp3i Jakarta). *J. Lentera ICT*, Vol. 3(1) pp. 76–92.
- Sibuea, M. L., & Safta, A. (2017). Pemetaan Siswa Berprestasi Menggunakan Metode K-Means Clustering. *Jurnal Teknologi dan Sistem Informasi (Jurteks)*, Vol. 4(1) pp. 85–92.