

Implementasi *K-Means Clustering* Terhadap Mahasiswa yang Menerima Beasiswa Yayasan Pendidikan Battuta di Universitas Battuta Tahun 2020/2021 Studi Kasus Prodi Informatika

Baginda Harahap¹, Aripin Rambe²

Email : profesionalbaginda@gmail.com¹, arambe1903@gmail.com²

Abstrak

Setiap penerimaan mahasiswa baru di Universitas Battuta, Yayasan Pendidikan Battuta menyalurkan dana bantuan berbentuk beasiswa kepada calon Mahasiswa yang mendaftar di Universitas Battuta. Agar penyaluran ini tepat sasaran, maka dibutuhkan suatu metode yang tepat. Dalam penelitian ini digunakanlah *K-Means Clustering* untuk menyeleksi mahasiswa dalam menerima beasiswa tersebut, yaitu 50%, 60% dan 100%. Sehingga penelitian ini sangat cocok untuk digunakan dalam menyeleksi calon mahasiswa baru yang mendapatkan beasiswa dari Yayasan Pendidikan Battuta.

Kata Kunci: Beasiswa Yayasan, Mahasiswa, K-Means Clustering.

1. Pendahuluan

Universitas Battuta adalah salah satu perguruan tinggi yang ada di Kota Medan Provinsi Sumatera Utara yang dipayungi oleh Yayasan Pendidikan Battuta. Pada setiap tahunnya akan terjadi penerimaan mahasiswa baru, maka calon mahasiswa yang mendaftar di Universitas Battuta akan diberikan bantuan oleh Yayasan Pendidikan Battuta berbentuk beasiswa yaitu, 50% sampai tammat, 60% sampai tammat dan 100% sampai tammat. Beasiswa ini diberikan oleh Yayasan Pendidikan Battuta sebagai wujud pedulinya terhadap dunia pendidikan, masyarakat dan pengabdian kepada Bangsa Indonesia.

Pendidikan merupakan hal yang perlu diperhatikan pada saat ini, karna sebab melalui pendidikan akan membentuk priadi manusia yang baik menurut ukuran normatif. Menyadari keadaan saat ini pendidikan diberbagai perguruan tinggi boleh dikatakan sangat mahal, memang beasiswa saat ini sangat banyak baik melalui instansi pemerintah, industri dan bahkan pemerintah setempat turut

memberikan beasiswa kepada anak Bangsa Indonesia yang melanjutkan pendidikannya ke jenjang kulliah/perguruan tinggi, tapi kita lihat dari segi banyaknya beasiswa dan ketatnya seleksi mendapkatka beasiswa tersebut, masih banyak anak-anak kita yang tidak mendapatkan beasiswa itu, maka Yayasan Pendidikan Battuta turut untuk membantu dalam rangka memberikan bantuan berbentuk beasiswa kepada anak Bangsa Indonesia yang melanjutkan pendidikannya di Universitas Battuta.

Knowledge Discovery In Database (KDD) merupakan metode untuk memperoleh pengetahuan dari *database* yang ada. Dalam *database* terdapat tabel-tabel yang saling berhubungan/berelasi. Hasil pengetahuan yang diperoleh dalam proses tersebut dapat digunakan sebagai basis pengetahuan (*knowledge base*) untuk keperluan pengambilan keputusan. KDD berperan penting dalam proses data mining. Data Mining ialah proses ekstraksi data untuk di jadikan informasi yang tersembunyi dari *database* yang besar. Eko Prasetyo (2012) dalam

penelitian nya mengemukakan, ada empat kelompok pekerjaan yang berkenaan pada data mining, yang pertama adalah dengan model prediksi (*prediction modelling*), kemudian yang kedua adalah analisa kelompok (*Cluster analysis*), yang ketiga adalah deteksi anomali (*anomaly detection*) dan analisis asosiasi (*association analysis*). Analisa *cluster* ialah suatu teknik pendukung keputusan yang digunakan dengan tujuan mengumpulkan objek-objek menjadi kelompok yang multivariat berdasarkan karakteristik obyek tersebut.

K-Means merupakan salah satu algoritma *clustering* dengan metode partisi yang berbasis titik pusat (*centroid*).

Algoritma ini pertama kali diusulkan oleh MacQueen (1996) dan dikembangkan oleh Hartigan dan Wong tahun 1975. *K-Means* mempunyai kemampuan mengelompokkan data dalam jumlah yang cukup besar dengan waktu komputasi yang cepat dan efisien (Irwansyah dan Faisal, 2015).

K-Means Clustering adalah metode untuk mengkategorikan atau pengelompokan sekelompok objek sesuai dengan atribut yang sama atau karakteristik ke dalam sejumlah kgroups (jumlah bilangan bulat positif). Ini mendefinisikan sebuah *cluster* oleh massa yang yang mewakili mean dari *cluster*. Hal ini sesuai dengan permasalahan yang saya temukan di Universitas Battuta dalam rangka penerimaan mahasiswa baru yakni identifikasi beasiswa yang diberikan oleh Yayasan Pendidikan Battuta, dimana metode data mining dengan algoritma *K-Means clustering* cocok untuk mengumpulkan dan mengelompokkan mahasiswa yang menerima beasiswa yaitu 50% sampai tammat, 60% sampai tammat dan 100% sampai tammat.

Pada penelitian ini, di kemukakan suatu metode pengolahan data dengan mengelompokkan mahasiswa menggunakan algoritma *K-Means clustering* dari data mahasiswa yang mendaftar di Universitas Battuta. Hasil observasi dibandingkan dengan identifikasi menggunakan algoritma *K-Means clustering*, sehingga di harapkan beasiswa yang di salurkan kepada mahasiswa tidak salah sasaran.

Siswa-siswi yang berhak menerima potongan biaya administrasi sekolah merupakan siswa-siswi dengan status siswa yatim piatu, siswa yatim atau piatu, siswa kurang mampu, siswa kakak beradik dan siswa berprestasi. Sehingga dibutuhkan sebuah data mining *clustering* (Defiyanti et al., 2017). Akan tetapi tidak sampai tahap data mining *clustering* saja untuk dapat menentukan hasil yang lebih akurat dibutuhkan sebuah algoritma *K-Means* (Zeebaree et al., 2017).

SMK Swasta Jaya Krama beringin saat ini memiliki permasalahan dimana untuk menentukan siswa-siswi yang layak untuk mendapatkan potongan biaya administrasi sekolah dengan jumlah siswa 585 orang sehingga dibutuhkan sebuah *Clustering* data untuk menentukan siapa-siapa saja yang berhak mendapatkan potongan biaya administrasi sekolah (Basaran & Günes, 2016). Dalam hal ini data mining *clustering* mempunyai peranan dalam mengolah dan mengelompokkan data pada siswa-siswi SMK Swasta Jaya Krama Beringin yang berhak menerima potongan biaya administrasi sekolah (Alelyani et al., 2019).

Data Mining : Klasifikasi Menggunakan Algoritma C.45. Hasilnya Dengan menggunakan algoritma c.45 merupakan salah satu metode yang dapat digunakan untuk pohon keputusan, dan data mining merupakan salah satu proses dari KDD.

Penerapan Data Mining dengan Metode *Clustering* untuk Pengelompokan Data Pengirim Burung. Hasilnya Dengan menggunakan algoritma *clustering K-Means* dapat memberikan pemahaman yang realistis tentang semua masalah Mengelompokkan dalam hal ini burung, dimana setiap burung di *cluster*, dan dengan metode ini pencegahan flu burung dapat dilaksanakan dengan baik berdasarkan *cluster*.

Penerapan Data Mining dengan Algoritma Fp-Growth untuk Mendukung Strategi Promosi Pendidikan (Studi Kasus Kampus STMIK Triguna Dharma). Hasilnya Dengan menggunakan metode ini dapat dilihat dari hasil penelitian tersebut bahwa metode dalam pencarian Frequent Itemsset pohon keputusan menggunakan algoritma FP-Growth bekerja sangat baik dalam melakukan Frequent Itemsset dengan proses pembentukan FP-Tree dengan menghasilkan rule dari data sampel mahasiswa baru, memiliki tingkat akurasi yang tinggi, dan FP-Growth dapat diterapkan untuk mendukung strategi promosi pendidikan pada Perguruan Tinggi.

Identifying Dominant Amino Acid Pairs of Known Protein-Protein Interactions via *K-Means Clustering*. Hasilnya Metode *K-Means clustering* pada penelitian ini menerapkan statistik dan teknik komputasi untuk menemukan pola asam amino dan mengidentifikasi asam amino yang luar biasa. Dan hasilnya menggambarkan pola korelasi dari semua sampel asam amino, dan mengidentifikasi asam amino yang dominasi dari satu set kelompok memiliki nilai centroid tertinggi.

Unsupervised Patterned Fabric Defect Detection Using Texture Filtering and *K-Means clustering* [13]. Hasilnya dengan

menggunakan metode *K-Means clustering* dapat mengelompokkan cacat pada kain. dapat memberikan rekonstruksi yang lebih baik kualitas dalam hal kuantitas sedikit hologram tingkat kuantisasi; Namun kerugian serius untuk vector kuantisasi adalah kecepatan rendah perhitungan, menyebabkan ketidakstabilan hasil pengolahan hologram.

Penggunaan Algoritma Apriori dalam Menganalisa Prilaku Mahasiswa Dalam Memilih Matakuliah (STUDI KASUS : FKIP UPI “YPTK”). Hasilnya dengan menggunakan metode Apriori dapat digunakan untuk menganalisa pola prilaku mahasiswa dalam memilih matakuliah, dan pola yang diperoleh bisa dijadikan acuan.

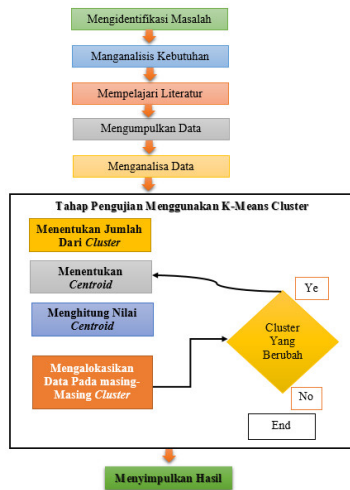
2. Metode Penelitian

Pada Metodologi penelitian ini dilakukan secara sistematis menjelaskan proses bagaimana kerangka kerja penelitian yang dilaksanakan. beberapa tahapan yang digambarkan dalam bentuk kerangka Gambar kemudian dapat di pahami mulai dari proses analisa kebutuhan sampai dengan hasil dari penelitian ini. Tahapan-tahapan dalam melakukan penelitian, yakni menganalisa kebutuhan data, proses pengumpulan data, analisa data menggunakan algoritma *K-Means*, kemudian dilakukan pengolahan data dengan RapidMiner9.10, dan hasil dari penelitian ini. Berdasarkan kerangka kerja nantinya akan dijabarkan langkah-langkah pada penelitian ini. yang akan di jelaskan pada Gambar 1.

2.1. Kerangka Kerja Penelitian

Tahapan pada poin ini yakni dimana kerangka kerja berfungsi untuk menggambarkan proses yang sistematis dalam penelitian ini, dari tahap menganalisa kebutuhan data hingga

menyimpulkan hasil. yang akan di jelaskan pada Gambar 1.



Gambar 1. Kerangka Kerja Penelitian

Pada Gambar 1 merupakan kerangka kerja penelitian dengan menggunakan metode *K-Means*. Adapun tahapan-tahapan proses dalam algoritma *clustering* menggunakan metode *K-Means* yakni sebagai berikut (Ong., 2013) :

- a. Memilih jumlah *cluster* *k*.
- b. Inialisasi *k* pusat *cluster* pada umumnya banyak cara dalam proses ini, Tetapi pilihan utama pada tahapan *cluster* adalah dengan cara random.
- c. Alokasikan semua data / objek ke *cluster* paling dekat.

Untuk melakukan proses pengolahan data pada titik tiap titik pusat *cluster* yakni dengan teori jarak Euclidean yang dirumuskan sebagai berikut :

$$D(i,j) = \sqrt{(X_{1i} - X_{1j})^2 + (X_{2i} - X_{2j})^2 + \dots + (X_{ki} - X_{kj})^2}$$

$D(i,j)$ = Jarak dari data ke *i* ke pusat *cluster* *j*

X_{ki} = Data ke *i* pada atribut data ke *k*

X_{kj} = Titik pusat ke *j* pada atribut ke *k*

d. Proses selanjut nya adalah menghitung ulang pusat *cluster* dengan keanggotaan *cluster* yang terbaru. rata-rata dari semua data/ objek dalam *cluster* merupakan pusat *cluster*. Sehingga mean bukanlah prioritas parameter yang di gunakan.

e. Menugaskan kembali setiap objek dengan menggunakan pusat *cluster* baru, apakah pusat *cluster* berubah hingga proses *clustering* selesai, ulang kembali proses “c” sampai di temukan nilai pada pusat *cluster* tidak ada berubah.

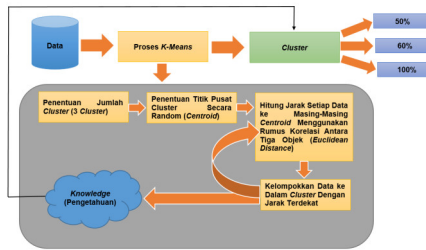
2.2. K-Means Clustering

K-Means clustering ialah metode yang termasuk pada *clustering* non-hirarki dimana setiap obyek yang masuk dalam kelompok adalah obyek-obyek yang sama dan berkorelasi. Data yang tergabung dalam kelompok mempunyai tingkat kemiripan yang lebih besar dan memiliki tingkat perbedaan yang besar pula dengan kelompok lainnya.

Pada dasarnya *clustering* adalah metode untuk mengkategorikan atau pengelompokan sekelompok objek sesuai dengan atribut yang sama atau karakteristik dengan data-data lainnya.

Clustering merupakan suatu metode pada data mining dimana proses kerja pada algoritma ini sifatnya tanpa arahan (*unsupervised*), artinya metode ini tidak lagi memerlukan lagi suatu training dan tanpa guru bahkan output tidak di perlukan. Pada data mining terdapat dua pembagian jenis metode *clustering* untuk proses pengelompokan data, yakni *hierarchical clustering* dan *non-hierarchical clustering*.

Adapun langkah-langkah pengolahan data dengan metode *K-Means Cluster* terlihat pada Gambar 2 :

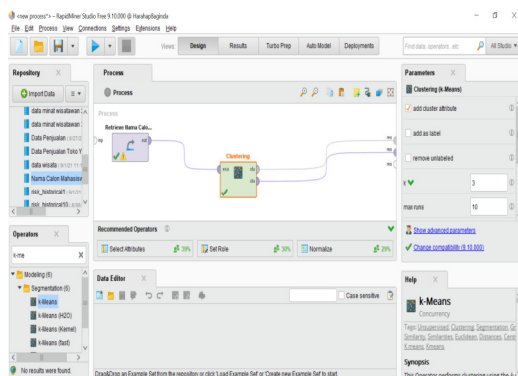


Gambar 2. Alur Kerja K-Means Clustering

Pada Gambar 2 merupakan proses algoritma *K-Means*, dimana tahap pertama adalah mengumpulkan data, seleksi data, data cleaning, dan data transformasi. Tahapan selanjutnya adalah proses pengolahan *K-Means* dengan menentukan jumlah *cluster* secara *random*, titik pusat *cluster*, menghitung jarak setiap data dan mengelompokkan data pada *cluster* sehingga menghasilkan *knowledge* yaitu *clustering*.

3. Hasil dan Pembahasan Implementasi RapidMiner 9.10.0

a. Design



Gambar 3. Bentuk Implementasi RapidMiner terhadap Data Calon Mahasiswa Prodi Informatika 2020/20201

Dari gambar 3 diatas kita dapat melihat kebutuhan operator yang akan kita gunakan untuk mengelola data calon

mahasiswa prodi informatika yang akan menerima beasiswa dari Yayasan Pendidikan Battuta. Yang pertama kita membutuhkan *Retrieve* yang berfungsi sebagai tempat data yang kita inputkan, pada hal ini peneliti memakai data calon mahasiswa baru universitas battuta prodi infomatika 2020/2021 file extensi penyimpanannya adalah excel. Yang kedua kita menggunakan operator *K-Means Clustering* yang berfungsi untuk mengelompokkan sesuai yang kita inginkan, hal ini peneliti mengatur *cluster* nya sebanyak 3 yaitu: 50%, 60% dan 100% kategori beasiswa yang akan diberikan kepada calon mahasiswa tersebut.

a. Cluster Model

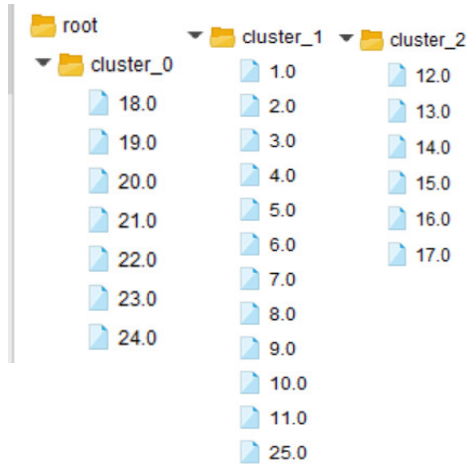
Cluster Model

```
Cluster 0: 7 items
Cluster 1: 12 items
Cluster 2: 6 items
Total number of items: 25
```

Gambar 4. Cluster Model

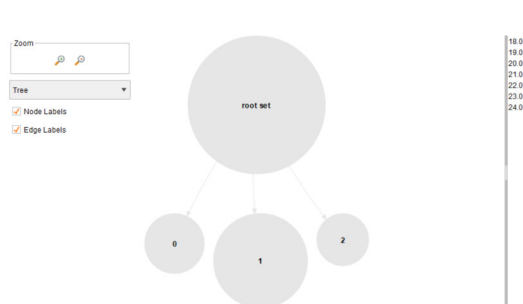
Pada gambar 4 ini dijelaskan bahwa jumlah data/items yang kita inputkan itu adalah 25 mahasiswa, *cluster* 0 itu adalah mahasiswa yang mendapatkan beasiswa 50% sampai tammat berjumlah 7 orang, *cluster* 1 itu adalah mahasiswa yang mendapatkan beasiswa 60% sampai tammat berjumlah 12 orang, dan *cluster* 2 itu adalah mahasiswa yang mendapatkan beasiswa 100% sampai tammat berjumlah 6 orang.

b. Folder View



Gambar 5. Penampilan masing-masing items di cluster

c. Graph



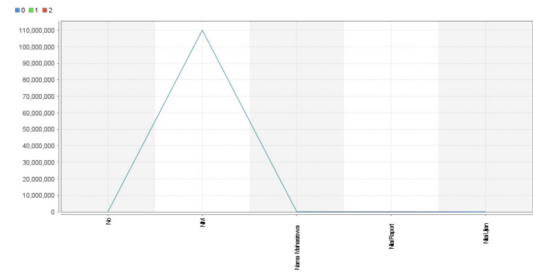
Gambar 6. Penampilan dari Graph

d. Centroid Table

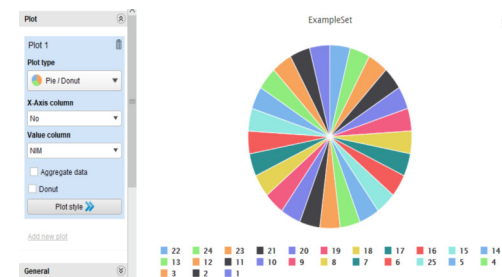
Attribute	cluster_0	cluster_1	cluster_2
No	21	7.583	14.500
NMI	110120037.714	110120006.667	110120020.833
Nama Mahasiswa	21	7.583	14.500
Nilai Raport	6.571	4.667	6.667
Nilai Ujian	7	4.167	5.167

Gambar 7. Penampilan Nilai Centroid Masing-Masing Cluster

e. Plot



Gambar 8. Penampilan dari Plot f. Visualizations



Gambar 9. Penampilan dari Visualizations

4. Kesimpulan

Berdasarkan hasil yang diperoleh dari penelitian ini dengan metode *K-Means clustering*. Hasilnya mendapatkan 3 kategori beasiswa Yayasan Pendidikan Battuta, yaitu 50% sampai tammat , 60% sampai tammat dan 100% sampai tammat. Hal ini dibuktikan berdasarkan perhitungan jarak terdekat berdasarkan penentuan nilai centroid secara random menggunakan rumus Euclidean Distance, pada jumlah mahasiswa prodi informatika dan nilai raport serta nilai ujian. Maka mahasiswa prodi informatika yang mendapatkan beasiswa 50% sampai tammat berjumlah 7 orang, mahasiswa prodi informatika yang mendapatkan beasiswa 60% sampai tammat berjumlah 12 orang, dan mahasiswa prodi informatika yang mendapatkan beasiswa 100% sampai tammat berjumlah 6 orang,

itulah hasil dari implementasi *K-Means Clustering* pada *software RapidMiner9.10.0*.

Tabel 1. Mahasiswa yang mendapatkan beasiswa dari Yayasan Pendidikan Battuta 50% sampai tammat.

Kode	Nama Mahasiswa
18	Azli Zamzami Zakiyya
19	Soeandi Malik Pratama
20	Bima Sakti
21	Fandy Ramadhan
22	Muhammad Agus Triwibowo
23	Agung Arif Syahputra
24	Delviana Hasibuan

Tabel 2. Mahasiswa yang mendapatkan beasiswa dari Yayasan Pendidikan Battuta 60% sampai tammat.

Kode	Nama Mahasiswa
1	Muhammad Zul Azuan
2	Siti Nur Alisyia
3	Bagas Syahputra
4	Bagus Hamdani Parapat
5	Salohot Harahap
6	Wahyu Saputra Tanjung
7	Windy Aswinta
8	Sri Uci Maylani
9	Indah Sartika
10	T. Ramatul Hayati
11	Mhd. Anif Aulia
25	Muhammad Alfin Pangaribuan

Tabel 3. Mahasiswa yang mendapatkan beasiswa dari Yayasan Pendidikan Battuta 100% sampai tammat.

Kode	Nama Mahasiswa
12	Dea Balqis Intan Pratiwi
13	Rihadatul Aisy Nasution
14	Muhammad Taufik Azhar Sihura
15	Andika Syahdewa
16	Diky Setiawan
17	Dewangga Putra

Daftar Pustaka

Sri.M., 2015. Penerapan Data Mining Dengan Metode *Clustering* Untuk Pengelompokan Data Pengiriman Burung. Prosiding Seminar Ilmiah Nasional Teknologi Komputer (SENATKOM 2015) Universitas Putra Indonesia YPTK Padang. Vol. 1, ISSN : 2460 – 4690.

Penangsang.O, Dimas.F.U.P dan Taufani.K., 2017. Optimal Placement and Sizing of Distributed Generation in Radial Distribution System Using *K-Means Clustering* Method. International Seminar on Intelligent Technology and Its Application. DOI: 978-1-5386-2708-2/17/\$31.00©/2017/IEEE.

Ngamsuriyaroj.S dan Thepsutum.K., 2017. Identifying Dominant Amino Acid Pairs of Known Protein-Protein Interactions via *K-Means Clustering*. IEEE 19th International Conference on High Performance Computing and Communications; IEEE 15th International Conference on Smart City; IEEE 3rd International Conference on Data Science and Systems. DOI:10.1109/HPCC-SmartCity-DSS.2017.37.

Shetty.K dan Kallimani.JS., 2017. Automatic Extractive Text Summarization using *K-Means Clustering*. International Conference on Electrical, Electronics, Communication, Computer and Optimization Techniques (ICEECCOT). DOI: 10.1109/ICEECCOT.2017.82846 27.

- Jamhur.A.I., 2016. Penerapan Data Mining Untuk Menganalisa Jumlah Pelanggan Aktif Dengan Menggunakan Algoritma C.45. Vol. 23, No. 2, Oktober 2016, Hal. 12-20 Copyright©2016 by LPPM UPI YPTK Padang. ISSN : 1412-5854.
- Aline.E.P dan Tedy.S., 2014. Penerapan Data Mining Untuk Klasifikasi Prediksi Penyakit ISPA (Infeksi Saluran Pernapasan Akut) Dengan Algoritma Decision Tree (ID3). Jurnal Sarjana Teknik Informatika Volume 2 Nomor 1, Februari 2014. e-ISSN: 2338-5197.
- Yuli.M., 2016. Data Mining : Klasifikasi Menggunakan Algoritma C4.5. Program Studi Pendidikan Informatika STKIP PGRI Sumbar. ISSN : 2407-0491.
- Sri.M., 2015. Penerapan Data Mining Dengan Metode *Clustering* Untuk Pengelompokan Data Pengiriman
- Defit.S., 2013. Penggunaan Algoritma Apriori Dalam Menganalisa Prilaku Mahasiswa Dalam Memilih Mata Kuliah (Studi Kasus : FKIP UPI "YPTK"). Jurnal Media Processor Vol.8, No.3, Oktober 2013.
- Hamdi.A.A, Sayed.M.S, Fouad.M.D dan Hadhoud.MM., 2018. Unsupervised Patterned Fabric Defect Detection Using Texture Filtering and *K-Means clustering*. International Conference on Innovative Trends in Computer Engineering (ITCE 2018), Aswan University, Egypt. DOI: 978-1-5386-0879-1/18/\$31.00 ©2018 IEEE.
- Ngamsuriyaroj.S dan Thepsutum.K., 2017. Identifying Dominant Amino Acid Pairs of Known Protein-Protein Interactions via *K-Means Clustering*. IEEE 19th International Conference on High Performance Computing and Communications; IEEE 15th International Conference on Smart City; IEEE 3rd International Conference on Data Science and Systems. DOI:10.1109/HPCC-SmartCity-DSS.2017.37.