

Mengidentifikasi Kelompok Sektor Perkebunan di Indonesia Berdasarkan Produktivitas Hasil Bumi

Hesti Pratiwi¹⁾, Ari Purno Wahyu Wibowo²⁾

^{1),2)} Program Studi SI Teknik Informatika, Fakultas Teknik, Universitas Widyatama
email: hesti.1402@widyatama.ac.id

Abstract

Known to be the agricultural country, Indonesia offers plenty of potential in terms of agricultural productivity, mainly soil-related production. In addition to that, of all sectors, a plantation sector also helps to sustain national economy. Therefore, the development from the agricultural sector, in this case refers to plantation, needs to be monitored by the associated institution in an attempt for optimizing the output or production of the aforementioned sector. This study is aimed at providing the classification of information on the agricultural sector area based on its productivity and the crop production. A grouping of these plantation sectors was developed by data mining method that implemented K-Means algorithm, which then visualized on the Business Intelligence Tableau application. From that classification, the study obtained three clusters from the plantation sectors, such as "great" that meant above the target, "good" that referred to inability to reach the target yet still above average, and "underperformed" which meant the plantation sectors in these clusters needed to be properly looked at because of its underperforming productivity.

Keywords: Clustering, Kmeans, Tableau, Indonesian Plantation

Abstrak

Dikenal dengan julukan negara agraris, Indonesia memiliki banyak potensi hasil bumi dari pertanahan, tak terkecuali dari sektor perkebunan yang memegang peran penting dari keseluruhan perekonomian nasional. Oleh karena itu, perkembangan produktivitas hasil bumi dari sektor pertanian, dalam hal ini perkebunan perlu dipantau supaya bisa menjadi perhatian instansi terkait dalam mengoptimalkan daerah penghasil tersebut. Penelitian ini bertujuan untuk menyediakan informasi klasifikasi terhadap daerah sektor pertanian berdasarkan produktivitas serta produksi hasil bumi. Pengelompokan sektor perkebunan ini dikembangkan dengan metode *data mining* dengan mengimplementasikan algoritma K-Means yang divisualisasikan pada aplikasi *Business Intelligence* Tableau. Dari hasil klasifikasi tersebut akan di dapat tiga *cluster* sektor perkebunan yang diantaranya adalah "great" yang berarti di atas target, "good" yang berarti belum mencapai target namun masih di atas rata-rata, dan "underperformed" yang berarti sektor perkebunan dalam *cluster* ini perlu perhatian khusus karena produktivitasnya di bawah performa.

Kata kunci: Clustering, Kmeans, Tableau, Perkebunan Indonesia.

1. Pendahuluan

Pandemic Covid-19 telah memberikan dampak bagi semua sektor dalam kehidupan, tak terkecuali adalah sektor perkebunan yang merupakan penunjang pada sektor ekonomi negara. Banyak negara yang mengalami hal serupa, maka sebagai negara agraris, Indonesia diharapkan dapat memanfaatkan peluang dalam sektor perkebunan dengan meningkatkan angka ekspor ke berbagai negara di dunia. Status negara agraris yang disandang oleh Indonesia menunjukkan sektor pertanian mempunyai potensi dan dapat menjadi andalan nasional untuk menjalankan dan memperkuat ketahanan pangan nasional [1].

Indonesia termasuk ke dalam tiga besar negara di dunia dengan penghasil hasil bumi

dari perkebunan terbesar. Untuk itu, informasi mengenai klasifikasi dari sektor perkebunan di Indonesia dibutuhkan sebagai alat pantau instansi terkait dalam mengoptimalkan produksi hasil buminya.

Informasi klasifikasi yang dimaksud adalah mengelompokkan perkebunan di Indonesia berdasarkan produktivitas pada masing-masing daerahnya dengan tujuan untuk mengetahui bagaimana keadaan produktivitas hasil bumi pada masing-masing daerah di Indonesia, sehingga instansi terkait dapat melakukan pemantauan dan memberikan strategi khusus pada daerah-daerah yang produktivitasnya di bawah performa.

Clustering adalah proses pengelompokan data dari sekumpulan data yang tidak teratur dengan memperhatikan kemiripan antar satu data dengan yang lainnya. Penelitian ini menggunakan metode k-means di mana algoritma ini merupakan salah satu algoritma yang paling banyak diminati dalam metode clustering dan termasuk ke dalam algoritma yang mudah diimplementasikan. Terlepas dari hal itu, algoritma k-means ini sendiri tentu memiliki kelemahan diantaranya memiliki sifat yang sensitif dalam menentukan jumlah cluster (k) awal dan solusi akhir menyatu pada local minima.

Penelitian ini terbagi menjadi beberapa tahapan proses dalam melakukan klasifikasi kelompok perkebunan pada masing-masing daerah dan proses visualisasi. Pada proses visualisasi juga ada perhitungan *growth year on year* dari produktivitas untuk masing-masing daerah yaitu perbandingan pertumbuhan tahun berjalan dengan tahun sebelumnya.

2. Tinjauan Pustaka dan Metode

A. Sektor Perkebunan di Indonesia

Perkebunan adalah salah satu sektor andalan dari pendapatan nasional Indonesia dan penyumbang devisa terbesar negara. Secara status pengusahaannya, perkebunan di Indonesia dibagi ke dalam tiga:

- 1) Perkebunan Rakyat, yaitu usaha perkebunan yang dikelola oleh perorangan dan tidak berbadan hukum.
- 2) Perkebunan Besar Negara, yaitu usaha perkebunan di bawah pengelolaan negara dalam hal ini pemerintah.
- 3) Perkebunan Besar Swasta, yaitu usaha perkebunan di bawah pengelolaan perusahaan di luar pemerintah dan berbadan hukum.

B. Data Mining

Data Mining merupakan suatu proses yang menguraikan pencarian atau penemuan informasi menarik di dalam suatu *database* atau *raw data* dengan memakai metode atau teknik tertentu. Proses yang terjadi di dalam *data mining* ini sendiri menggunakan beberapa metode diantaranya teknik statistik, konsep matematika, kecerdasan buatan, dan *machine learning* dalam menemukan dan menguraikan informasi yang berguna untuk pengetahuan terkait dengan berbagai *database* berkapasitas besar.

Output dari proses *data mining* ini nantinya mampu dijadikan sebagai bekal untuk mengidentifikasi dan menentukan langkah di masa depan. Beberapa teknik yang populer digunakan pada proses *data mining* diantaranya adalah asosiasi, estimasi, prediksi, *clustering* serta klasifikasi. Pelaksanaan *data mining* dalam penanganan dan pengolahan data pada sebuah kasus, akan sangat membantu dalam meningkatkan daya tarik digital serta dapat menghasilkan informasi yang lebih bermakna.

C. Algoritma K-Means

K-means adalah algoritma *clustering* yang sangat sederhana yang membagi kelompok data menjadi beberapa *cluster k*. *K-means* merupakan salah satu algoritma pelatihan *unsupervised*, awal mula algoritma ini dipublikasikan oleh Stuart Lloyd di tahun 1984 yang hingga saat ini merupakan algoritma yang banyak digunakan. Alasan algoritma ini banyak digunakan adalah karena *k-means* cukup mudah untuk diimplementasikan, relatif cepat, dan

mudah dimodifikasi. Algoritma *Clustering K-Means* dapat dengan mudah diimplementasikan dengan Algoritma Genetika.

Tahapan proses pada algoritma *K-means* dapat dijelaskan sebagai berikut (Larose, 2005; Prasetyo, 2012; Khotimah, 2014; Parlina et al., 2018; Primatha, 2018):

1. Tahap pertama adalah menentukan jumlah *cluster* (*k*) atau kelompok pada suatu *dataset*.
2. Selanjutnya tentukan nilai pusat (*centroid*). Penentuan nilai awal *centroid* dilakukan secara acak. Setelah itu ketika tahap perulangan, rumus perhitungan yang dipakai seperti persamaan di bawah ini:

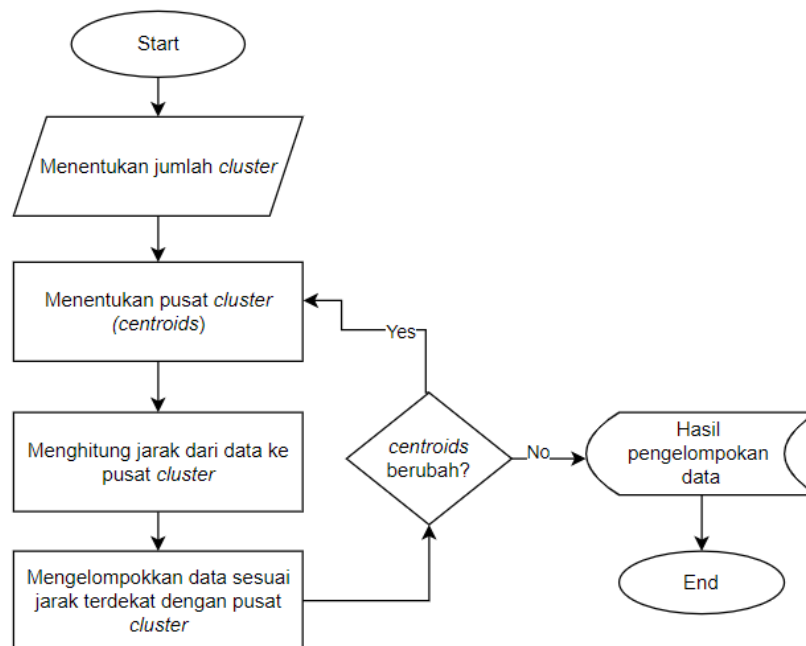
$$V_{ij} = \frac{1}{N_i} + \sum_{k=0}^{N_i} X_{kj} \quad (1)$$

Keterangan:

V_{ij}	:	<i>Centroid</i> / nilai rata-rata <i>cluster</i> indeks ke- <i>i</i> pada <i>variable</i> indeks ke- <i>j</i>
N_i	:	Jumlah data <i>cluster</i> indeks ke- <i>i</i>
i, k	:	Indeks atau penanda pada <i>cluster</i>
j	:	Indeks atau penanda pada <i>variable</i>
X_{kj}	:	Nilai data indeks <i>k variable</i> indeks ke- <i>j</i> pada <i>cluster</i> tersebut

3. Untuk setiap *row data*, hitung jarak minimum data tersebut dengan *centroid*. Terdapat beberapa metode perhitungan yang dapat digunakan dalam mengukur jarak minimum data ke *centroid*, diantaranya adalah metode *Euclidean*, *Manhattan / City Block*, serta *Minkowsky*.
4. Mengelompokkan objek data berdasarkan hitungan jarak ke *centroid* dengan jarak hitungan minimum terdekat.
5. Kembali ke tahapan nomor 2 hingga tahapan nomor 4, lakukan perulangan sampai didapat *centroid* yang bernilai optimal.

Dari langkah-langkah di atas, dapat divisualisasikan dalam bentuk *flow* proses algoritma *k-means* ini bekerja adalah sebagai berikut:



Gambar 1 Flowchart Algoritma K-Means

Gambar di bawah ini merupakan ilustrasi dari proses yang dilakukan dan didapatkan oleh algoritma *k-means*:



Gambar 2 Ilustrasi Proses *K-Means Clustering*

Dari gambar tersebut, dapat disimpulkan bahwa tujuan pada penelitian ini adalah untuk mengklasifikasikan *raw data* menjadi kelompok-kelompok data yang bisa dianggap sebagai informasi yang lebih dapat diolah menjadi sesuatu yang lebih bermanfaat.

D. Tableau

Tableau adalah aplikasi *business intelligence* yang digunakan dalam memvisualisasikan data dan mengolah data menjadi sebuah informasi yang lebih interaktif. Beberapa kelebihan dari fitur tableau adalah sebagai berikut:

1. Memiliki kecepatan analisis. Pengguna tidak diwajibkan untuk mahir dalam Bahasa pemrograman tertentu atau ahli *coding*.
2. Tidak bergantung pada aplikasi lain. Memudahkan pengguna supaya tidak perlu untuk mengolah data di banyak aplikasi.
3. Menyajikan dalam bentuk visualisasi yang interaktif. Tableau menyediakan berbagai macam fitur yang bisa menghasilkan visualisasi yang menarik dan interaktif.
4. Dapat berkolaborasi secara *real-time*. Artinya aplikasi ini memungkinkan pengguna untuk menggunakan berbagai sumber data dalam satu waktu yang sama.
5. Mengimpor berbagai ukuran dan *range data*, semua ukuran dan *range data* bisa diolah di Tableau.

E. ETL (Extract Transform Load)

ETL merupakan akronim dari kata *Extract*, *Transform*, dan *Load*. ETL merupakan proses integrasi data yang memadukan data dari berbagai sumber ke dalam satu repositori yang konsisten dan dimasukkan ke dalam *database* atau sistem lain. Pada awal munculnya ETL di tahun 1970, ETL diperkenalkan dengan tujuan untuk mengumpulkan atau mengintegrasikan proses pemuatan data ke dalam superkomputer yang selanjutnya akan dianalisis lebih dalam lagi. Sejak akhir 1980-an hingga pertengahan tahun 2000, ETL merangkak menjadi proses primer yang diperuntukkan untuk membangun gudang data sehingga bisa mendukung aplikasi *Business Intelligence* (BI).

Tiga tahapan proses yang digambarkan pada ETL adalah sebagai berikut:

1) *Extract*

Ini merupakan tahapan pertama dari proses yang melibatkan ekstraksi data dari sumber data yang sesuai. Sumber data yang digunakan bisa berbentuk *flat file* dengan ekstensi seperti *csv*, *xls*, dan atau *txt*.

2) *Transform*

Pada tahapan ini dilakukan proses cleansing sesuai dengan target skema yang sudah disediakan. Proses-proses tersebut bisa berupa normalisasi data, pengecekan duplikasi, filtering data, pengurutan serta pengelompokan data.

3) *Load*.

Dan terakhir adalah tahapan load data. Pada proses ini, data yang sudah melalui proses transformasi sebelumnya akan disimpan ke dalam wadah seperti table yang sudah disiapkan pada data *warehouse*.

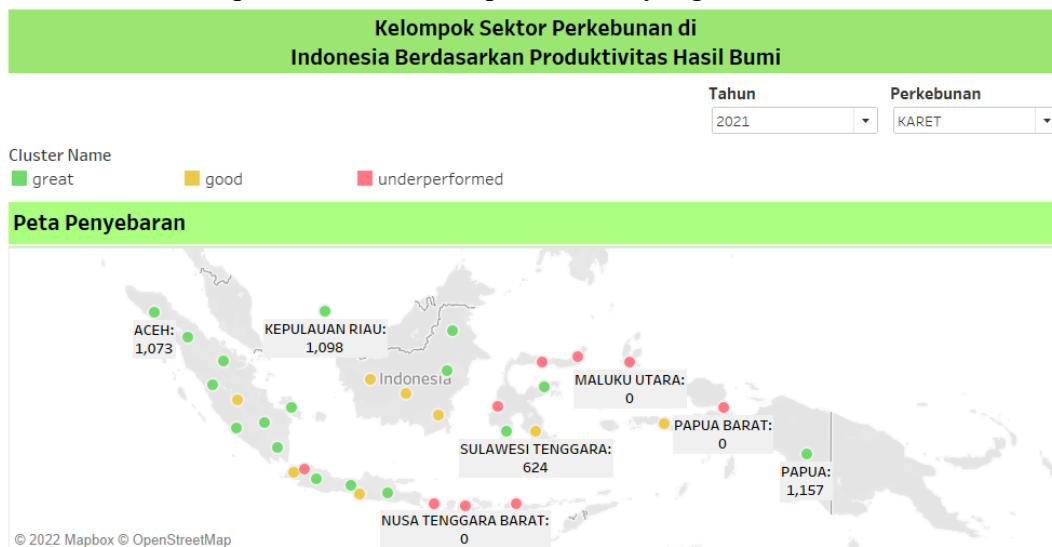
3. Hasil dan Pembahasan

Proses pengklasifikasian sektor perkebunan di daerah-daerah Indonesia pada penelitian ini dikelompokkan berdasarkan *variable* produktivitas pembentuk kelompok. *Variable* tersebut tersedia datanya dalam basis *yearly* atau tahunan.

Tahapan proses pengujian algoritma *K-Means* adalah seperti berikut:

- 1) Membuat *cluster* untuk pengelompokkan data, diantaranya:
 - a. *Cluster great*, untuk kelompok daerah yang memiliki produktivitas bagus.
 - b. *Cluster good*, untuk kelompok daerah yang memiliki produktivitas cukup baik.
 - c. *Cluster underperformed*, untuk kelompok daerah yang memiliki produktivitas di bawah target.
- 2) Menetapkan *centroids* (nilai pusat) bagi masing-masing *cluster*.
- 3) Proses pengelompokkan data ke dalam tiga *cluster* menggunakan algoritma *k-means*.
- 4) Visualisasi menggunakan *tools* Tableau untuk menampilkan hasil dari pengelompokkan sektor perkebunan pada daerah-daerah di Indonesia.

Pada gambar 3 dan gambar 4 berikut ini adalah gambaran dari hasil pengelompokan data perkebunan pada daerah-daerah di Indonesia yang divisualisasikan dalam bentuk peta penyebaran serta bentuk tabel yang berisi informasi tentang masing-masing daerah berada ke dalam *cluster* apa untuk tahun dan perkebunan yang terfilter.



Gambar 3 Informasi Peta Sebaran Cluster

Tabel Produktivitas

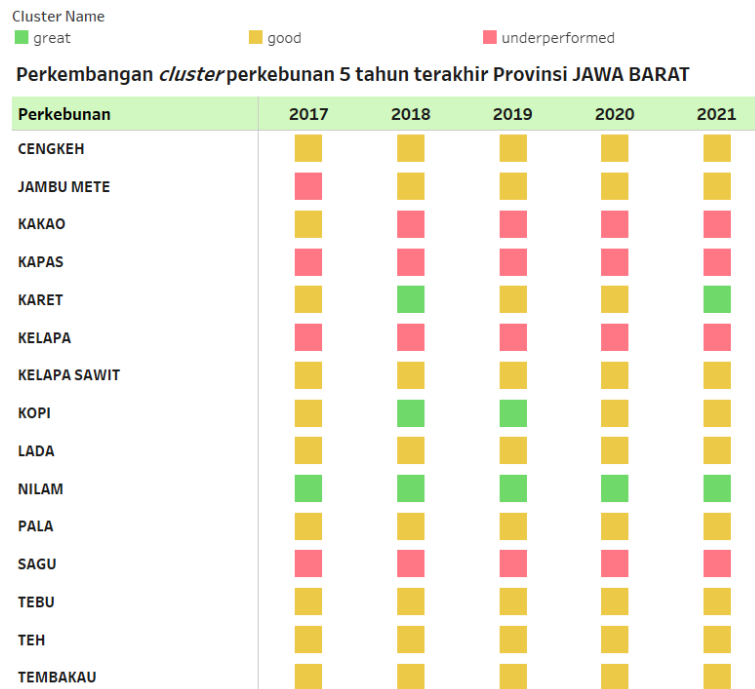
Provinsi	Produktivitas	Cluster
JAWA TIMUR	1,267	●
JAWA TENGAH	1,214	●
BENGKULU	1,199	●
KALIMANTAN TIMUR	1,195	●
KEPULAUAN BANGKA BELITUNG	1,176	●
PAPUA	1,157	●
SUMATERA SELATAN	1,154	●
SULAWESI SELATAN	1,144	●
SUMATERA BARAT	1,125	●

Tabel Produktivitas

Provinsi	Produktivitas	Cluster
SUMATERA UTARA	1,047	●
JAWA BARAT	1,032	●
LAMPUNG	990	●
KALIMANTAN SELATAN	960	●
BANTEN	957	●
KALIMANTAN BARAT	916	●
JAMBI	876	●
KALIMANTAN TENGAH	808	●
DI YOGYAKARTA	650	●
SULAWESI TENGGARA	624	●
MALUKU	619	●
BALI	286	●
SULAWESI UTARA	0	●
SULAWESI BARAT	0	●
PAPUA BARAT	0	●
NUSA TENGGARA TIMUR	0	●
NUSA TENGGARA BARAT	0	●
MALUKU UTARA	0	●
GORONTALO	0	●
DKI JAKARTA	0	●

Gambar 4 Visualisasi Hasil Clustering

Selanjutnya untuk gambar 5 adalah hasil visualisasi dari kelompok produktivitas semua perkebunan selama lima tahun terakhir untuk masing-masing daerah terpilih. Dengan kata lain, tampilan ini merupakan hasil lebih detail dari visualisasi pada gambar 3 dan gambar 4.



Gambar 5 Informasi produktivitas 5 tahun terakhir

Tabel Produktivitas *Year on Year*

Provinsi	2020	2021	YoY (%)
BALI	0	286	28600.00%
KALIMANTAN UTARA	433	1,121	158.89%
SULAWESI SELATAN	653	1,144	75.19%
SULAWESI TENGGARA	358	624	74.30%
MALUKU	374	619	65.51%
PAPUA	880	1,157	31.48%
SULAWESI TENGAH	860	1,068	24.19%
ACEH	967	1,073	10.96%
RIAU	992	1,066	7.46%
JAWA BARAT	968	1,032	6.61%
KEPULAUAN RIAU	1,040	1,098	5.58%
BANTEN	912	957	4.93%
DI YOGYAKARTA	629	650	3.34%
LAMPUNG	963	990	2.80%
SUMATERA BARAT	1,103	1,125	1.99%
SUMATERA UTARA	1,030	1,047	1.65%
JAMBI	862	876	1.62%
JAWA TENGAH	1,196	1,214	1.51%
JAWA TIMUR	1,253	1,267	1.12%
KEPULAUAN BANGKA BELITUNG	1,167	1,176	0.77%
KALIMANTAN TIMUR	1,189	1,195	0.50%
KALIMANTAN SELATAN	956	960	0.42%
KALIMANTAN BARAT	913	916	0.33%
BENGGKULU	1,196	1,199	0.25%
KALIMANTAN TENGAH	806	808	0.25%
SUMATERA SELATAN	1,153	1,154	0.09%
SULAWESI UTARA	0	0	0.00%
SULAWESI BARAT	0	0	0.00%
PAPUA BARAT	0	0	0.00%
NUSA TENGGARA TIMUR	0	0	0.00%
NUSA TENGGARA BARAT	0	0	0.00%
MALUKU UTARA	0	0	0.00%
GORONTALO	0	0	0.00%
DKI JAKARTA	0	0	0.00%

Gambar 6 Informasi *Growth* produktivitas

Gambar di atas ini menampilkan informasi dari *growth* produktivitas pada masing-masing daerah dalam basis *yearly* untuk tahun dan perkebunan terpilih. Di mana untuk mendapatkan *growth year on year* adalah sebagai berikut:

$$YoY = \frac{P_n - (P_{n-1})}{(P_{n-1})} \times 100\% \tag{2}$$

Keterangan:

P_n : Produktivitas di tahun terpilih

P_{n-1} : Produktivitas di tahun sebelumnya

Perhitungan nilai *year on year* pada proses visualisasi ini bertujuan untuk rekapitulasi prosentase *growth* produktivitas dari masing-masing daerah setiap tahunnya.

4. Kesimpulan

Berdasarkan hasil evaluasi terhadap percobaan pembentukan *clustering* pada data sektor perkebunan di Indonesia maka didapat kesimpulan sebagai berikut:

- Penentuan nilai *centroids* (nilai pusat) awal data sangat berpengaruh terhadap keberhasilan pengelompokkan data menggunakan algoritma *k-means*.
- Proses berapa kali melakukan iterasi juga berpengaruh terhadap kesesuaian hasil *clustering data*.
- Algoritma *K-Means Clustering* dapat digunakan sebagai dasar untuk memilih data potensial dari data yang memiliki dimensi tinggi.

Pada penelitian ini menunjukkan bahwa ada 3 klasifikasi *cluster* dari masing-masing jenis perkebunan untuk daerah-daerah di Indonesia. Ketiga *cluster* tersebut adalah *Great*, *Good*, dan *Underperformed*.

Daftar Pustaka

- [1] Darmawan, Didit; Genua, Veronika; Kristianto, Sonny; Murdaningsih; Hutubessy, Josina. *Tanaman Perkebunan Prospektif Indonesia*. Pasuruan: CV. Penerbit Qiara Media.
- [2] Han J, Kamber M, dan Jian P. *Data Mining*. 2012. *Concepts and techniques*. USA: Elsevier Inc.
- [3] Y. Mardi. 2017. *Data Mining: Klasifikasi Menggunakan Algoritma C4.5*, J. Edik Inform., vol. 2, no. 2, pp. 213–219.
- [4] Tan, Pang-Ning; Steinbach, Michael dan Kumar, Vipin. 2005. *Intoruction into Data Mining*. Boston: Addison-Wesley Longman Publishing Co., Inc.
- [5] Florin Gorunescu. 2011. *Data Mining Concepts, Models and Techniques*. Springer-Verlag Berlin Heidelberg
- [6] P. Vassiliadis, A. Simitsis, and E. Baikousi, A Taxonomy of ETL Activities, in *Proceedings of the ACM Twelfth International Workshop on Data Warehousing and OLAP*, New York, NY, USA, 2009, pp. 25–32.
- [7] Ismi D. P., Panchoo, & S., Mrinto. (2016). K-Means Clustering Based Filter Feature Selection on High Dimensional Data. *International Journal of Advances in Intelligent Informatics*. 38-45, <http://dx.doi.org/10.12928/ijain.v2il.54>.
- [8] Metisen B. M., & Sari H. L. (2016). Analisis Clustering Menggunakan Metode K-Means dalam Pengelompokan Penjualan Produk pada Swalayan Fadhila. *Jurnal Media Infotama*, 11(2). 1858-2680.
- [9] Yunita (2018). Penerapan Data Mining Menggunakan Algoritma K-Means Clustering Pada Penerimaan Mahasiswa Baru (Studi Kasus: Universitas Islam Indragiri). *Jurnal Sistemasi*, 7 (September), 238–249. <https://doi.org/10.32520/stmsi.v7i3.388>.
- [10] Abdullah H. A, (2018). Improve the Performance of K-Means by Using Genetic Algorithm for Classification Heart Attack. *International Journal of Electrcal and Computer Engineering (IJECE)*, 2088-8708. <https://doi.org/10.11591/ijece.v8i2.pp1256-1261>.