

VISUALISASI DATA PENGELOMPOKKAN KELULUSAN MAHASISWA DENGAN ALGORITMA CLUSTERING K-MEANS

Salsabila Adinda Hermawan¹, Dessy Ratna Sari Dewi Hartanto², Dewa Made Wiharta³, I Made Arsa Suyadnya⁴

^{1,2}Mahasiswa Program Studi Teknik Elektro, Fakultas Teknik, Universitas Udayana

^{3,4}Dosen Program Studi Teknik Elektro, Fakultas Teknik, Universitas Udayana

Jl. Raya Kampus Unud Jimbaran, Kec. Kuta Sel, Kabupaten Badung, Bali 80361

salsabilaadinda@student.unud.ac.id¹, dessyratanatan@student.unud.ac.id²,

wiharta@unud.ac.id³, arsa.suyadnya@unud.ac.id⁴

ABSTRAK

Implementasi teknologi di perguruan tinggi saat ini mengalami peningkatan pesat, salah satunya dengan menghasilkan kumpulan seperti data keuangan, data perpustakaan, dan data akademik. Penelitian ini memanfaatkan data kelulusan mahasiswa menggunakan metode *data mining*, yaitu algoritma *K-Means Clustering*. Penelitian ini dilakukan untuk membantu Fakultas Teknik Universitas Udayana dalam memahami pola kelulusan mahasiswa secara terperinci dengan mengelompokkan data kelulusan berdasarkan variabel asal mahasiswa (kabupaten atau kota) dan program studi di Fakultas Teknik Universitas Udayana, sehingga dapat memberikan informasi mengenai kelompok mahasiswa yang lulus tepat pada waktunya dan yang lulus tidak tepat pada waktunya. Fokus lain pada penelitian ialah desain visualisasi dari data hasil *clustering* dalam bentuk *chart* dan grafik yang dapat diakses menggunakan *dashboard*. *Dashboard* visualisasi data dibuat menggunakan *Google Looker Studio*, di mana data kelulusan mahasiswa yang telah melalui proses *clustering* divisualisasikan dalam bentuk *chart* dan grafik yang dapat memudahkan pihak Fakultas Teknik Universitas Udayana untuk memperoleh informasi dan wawasan baru yang berguna untuk meningkatkan performa akademik di Fakultas Teknik Universitas Udayana itu sendiri.

Kata Kunci: Data Akademik, Kelulusan Mahasiswa, Algoritma *K-Means*, *Clustering*, Visualisasi, *Clustering*, dan *Dashboard*.

ABSTRACT

The implementation of technology in universities is rapidly increasing, either of them are creating various groupings such as financial, library, and academic data. This research implemented the graduation data Udayana University Faculty of Engineering students by applying the data mining method, specifically the *K-Means Clustering* algorithm. The study aimed to assist Udayana University Faculty of Engineering in comprehending detailed graduation patterns by categorizing the data based on variables including students' origin (district or city) and the study programs in engineering faculty. This research also focused on the visualization design of clustered data in the form of charts and graphs. This data visualization dashboard was created using *Google Looker Studio*, which the clustered student graduation data is visualized in the form of charts and clustering graphs, which can make it easier for Udayana University Faculty of Engineering to achieve new information and useful insights to improve the academic performance of Udayana University Faculty of Engineering itself.

Keyword: Academic Data, Student Graduation, *K-Means* Algorithm, *Clustering*, Visualization, *Dashboard*.

1. PENDAHULUAN

Perkembangan teknologi telah memberikan dampak signifikan pada sektor

pendidikan, salah satunya dalam pengelolaan data di lingkungan Fakultas Teknik Universitas Udayana.

Perkembangan teknologi mendorong Fakultas Teknik Universitas Udayana untuk menghasilkan beragam jenis data, termasuk data kelulusan mahasiswa yang meningkat dari tahun ke tahun. Peningkatan terhadap jumlah data kelulusan mahasiswa ini memerlukan analisis mendalam guna memperoleh wawasan dan informasi yang relevan.

Tingkat kelulusan mahasiswa menjadi indikator penting dalam mengevaluasi keberhasilan akademis di perguruan tinggi. Keberhasilan Mahasiswa dapat diukur melalui kemampuannya menyelesaikan program studi sesuai dengan waktu yang ditetapkan. Dimana idealnya mahasiswa jenjang strata-1 menyelesaikan masa studinya dalam waktu 4 tahun. Namun, di Fakultas Teknik Universitas Udayana, seperti pada lingkungan akademik perguruan tinggi lainnya, terdapat tantangan dimana mahasiswa menghadapi kesulitan untuk menyelesaikan studinya sesuai batas waktu yang ditentukan [1].

Mengatasi masalah tersebut, dilakukan analisis dengan metode pengelompokan mahasiswa berdasarkan total Indeks Prestasi Kumulatif (IPK) akhir, Satuan Kredit Semester (SKS), dan lamanya masa studi mahasiswa. Langkah analisis ini penting bagi Fakultas Teknik Universitas Udayana guna mengevaluasi tingkat kelulusan mahasiswa mengacu pada asal mahasiswa dari Kabupaten atau Kota serta program studi. Analisis ini dapat memberikan manfaat signifikan bagi Fakultas Teknik Universitas Udayana dalam mendukung pengambilan strategi dan langkah-langkah yang memungkinkan universitas merancang kegiatan untuk mendukung kelulusan mahasiswa. Dengan memahami tren atau pola kelulusan, pihak Fakultas Teknik Universitas Udayana dapat menginisiasi program atau kegiatan bagi program studi atau kabupaten dan kota asal mahasiswa yang memerlukan perhatian khusus, dengan tujuan meningkatkan performa Fakultas Teknik Universitas Udayana.

Metode *K-Means Clustering* bertujuan mengelompokkan data kelulusan mahasiswa Strata 1 di Fakultas Teknik Universitas Udayana pada angkatan 2017 yang berasal dari Provinsi Bali. Metode *clustering* ini termasuk ke dalam teknik *multivariate* yang bertujuan mengkategorikan objek berdasarkan ciri-cirinya [2]. Data kelulusan mahasiswa dikelompokkan berdasarkan tingkat kelulusan tepat waktu, yakni kurang atau sama dengan 8 semester, dan kelulusan tidak tepat waktu, yakni lebih dari 8 semester. Analisa menggunakan *tools Google Collab* dan bahasa pemrograman Python dengan parameter yang dipertimbangkan melibatkan asal mahasiswa dari Kabupaten atau Kota serta program studi yang diambil.

Beberapa penelitian yang berhasil mengelompokkan data dengan implementasi metode *K-Means Clustering* diantaranya adalah: Penelitian oleh H. Priyatman dkk [2] menerapkan algoritma *K-Means Clustering* untuk mendapatkan prediksi terhadap waktu kelulusan mahasiswa, dengan tujuan memperoleh informasi tentang proyeksi waktu kelulusan mahasiswa melalui pengamatan pola kelompok mahasiswa berdasarkan pengukuran terhadap parameter indeks Prestasi Kumulatif (IPK) dan absensi. Hasil penelitian ini diharapkan bermanfaat bagi kampus dan mahasiswa dalam memprediksi kelulusan mahasiswa tepat waktu. Kampus dapat mengambil langkah-langkah yang dibutuhkan, seperti memberikan bimbingan, untuk mencegah kelulusan terlambat bagi mahasiswa. Penelitian yang dilakukan oleh M. R. Muttaqin & M. Defriani [3] mengimplementasikan metode *clustering* untuk mengelompokkan mahasiswa ke dalam kelompok-kelompok berdasarkan parameter nilai yang diperolehnya dari kurun waktu semester 1 sampai dengan semester 7. Tujuannya memberikan rekomendasi mengenai topik skripsi yang disesuaikan bagi setiap mahasiswa, sehingga pihak universitas dapat memberikan rekomendasi kepada mahasiswa dalam mengambil topik

skripsi. Diperoleh hasil analisis pengkelompok mahasiswa berdasarkan keahlian mereka, yang ditentukan oleh *cluster* dengan nilai tertinggi dari jumlah mata kuliah terbanyak dalam setiap keahlian. *Cluster* ini menjadi pedoman untuk merekomendasikan topik judul skripsi kepada mahasiswa.

Analisa data yang dilakukan memerlukan visualisasi *dashboard* guna membantu pihak Fakultas Teknik Universitas Udayana memahami hasil analisis *K-Means Clustering* dengan lebih efektif. Visualisasi ini melibatkan representasi data dalam bentuk *chart* dan *grafik* yang dapat diakses melalui sebuah *dashboard* yang dapat memberikan kemudahan dalam pengaksesan informasi. Melalui penggunaan representasi grafis dari berbagai jenis data (geografis, numerik, tekstual, dan jaringan), tujuan visualisasi informasi adalah untuk merangsang dan meningkatkan kemampuan kognitif [5].

Alat untuk membuat *dashboard* visualisasi penulis menggunakan *Looker Studio* yang sebelumnya dikenal juga sebagai *Google Data Studio* ini akan terintegrasikan dengan *Google Spreadsheet* yang memuat sumber data hasil *clustering*.

2. TINJAUAN PUSTAKA

2.1 Data Mining

Dalam *data mining*, kumpulan-kumpulan data yang berskala besar di klusterisasi untuk mencari pola dan informasi yang dapat berguna. Menemukan pola tersembunyi dalam kumpulan data besar adalah tujuan dari teknik penambangan data, yang terdiri dari penerapan berbagai metode berbeda untuk mencapai tujuan ini. Algoritma yang dikenal dengan Metode *K-Means clustering* yang digunakan dalam *data mining* tergolong sebagai contoh Machine Learning dengan kategori Unsupervised Learning. Metode ini memanfaatkan teknik *K-Means Clustering* untuk mempartisi data dengan tujuan membaginya menjadi beberapa kelompok.

2.2 Algoritma K-Means Clustering

Algoritma *K-Means Clustering* adalah teknik *clustering* non-hirarki, algoritma ini digunakan agar dapat mengelompokkan suatu set objek menjadi n kelompok sehingga variasi data dalam setiap kelompok mencapai tingkat yang minimal [4]. Melalui proses pembuatan kelompok (*cluster*) dari kumpulan data yang kelas atau grupnya tidak diketahui, tujuan analisis cluster adalah untuk menentukan bagian data mana yang termasuk dalam *cluster* mana. [2] Hal ini dicapai dengan terlebih dahulu membuat grup. Berikut ini adalah daftar langkah-langkah dalam *clustering* menggunakan metode *K-Means*, beserta beberapa contoh perhitungannya:

1. Menentukan banyaknya *cluster* (k) yang digunakan.
2. Tetapkan nilai pusat (*centroid*). Tahap awal, nilai *centroid* ditetapkan secara acak. Sementara untuk tahap iterasi dapat menggunakan rumus persamaan 1 berikut:

$$V_{ij} = \frac{1}{N_i} \sum_{k=0}^{N_i} Xkj \dots \dots \dots (1)$$

Keterangan:

V_{ij} = *Centroid* rata-rata *cluster* ke- i untuk variabel ke- j

N_i = Banyaknya anggota dari *cluster* ke- i

i, k = Indeks *cluster*

j = Indeks variabel

Xkj = nilai data ke- k variabel ke- j untuk *cluster* tersebut

3. Dalam menentukan rentang titik *centroid* dan masing-masing objek, digunakan metode *Euclidean Distance*. Metode ini menggambarkan rentang langsung interval dua titik dalam ruang *Euclidean*, dihitung dengan menggunakan rumus persamaan 2 berikut:

$$De = \sqrt{(x_i + s_i)^2 + (y_i + t_i)^2} \dots \dots \dots (2)$$

Keterangan:

De = Jarak *Euclidean*

i = Jumlah objek

(x, y) = Titik koordinat objek
 (s, t) = Titik koordinat *centroid*

4. Susunlah kelompok objek tersebut ke dalam kelompok-kelompok berdasarkan jarak dari pusat *centroid* yang membentuk masing-masing kelompok. Langkah 2 hingga langkah 4 harus diulangi secara iteratif hingga mencapai nilai *centroid* yang optimal.

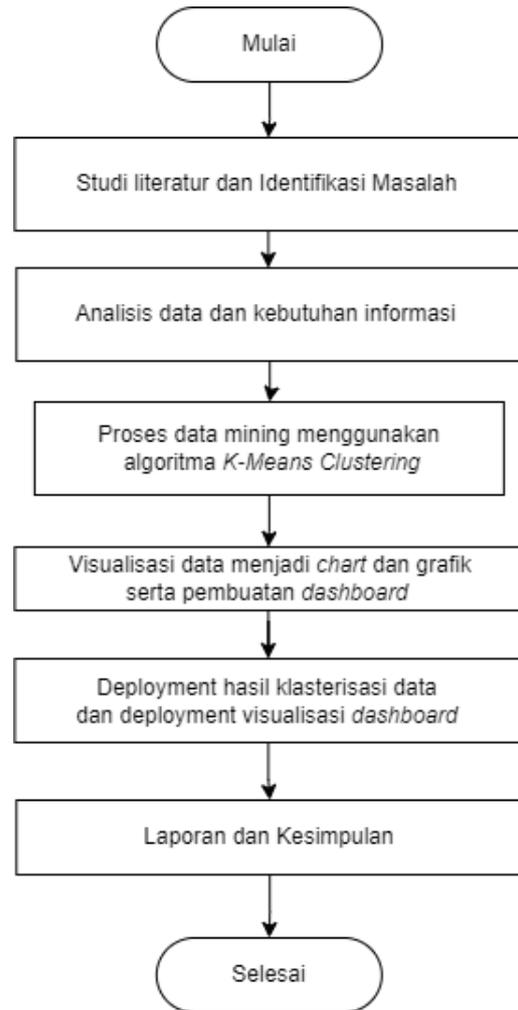
2.3 Visualisasi Data

Membuat representasi data yang menarik secara visual adalah tujuan visualisasi data, yang dicapai melalui pemanfaatan berbagai bentuk grafis seperti bagan, grafik, tabel, dan peta [7]. *Looker Studio* adalah alat gratis yang mengubah dan mengolah data ke dalam *dashboard* dan laporan yang informatif, mudah dibagikan, mudah dibaca dan mudah dipahami serta dapat disesuaikan sepenuhnya. Pada *Looker Studio* laporan atau *dashboard* yang dibuat dapat terhubung ke set data set, seperti database (termasuk BigQuery, MySQL), *Google Spreadsheet* hingga upload file CSV.

Pengguna tidak diharuskan memulai dengan halaman kosong karena alat ini menawarkan contoh format laporan dan galeri yang telah dikembangkan oleh komunitas [8].

3. METODE PENELITIAN

Unit Sumber Daya Informasi (USDI) Universitas Udayana Kampus Bukit Jimbaran yang berlokasi di Jalan Raya Kampus Unud di Bukit Jimbaran, Badung, Bali dijadikan sebagai lokasi penelitian. Periode pelaksanaan penelitian dimulai pada bulan Juni 2023 dan berlanjut hingga bulan Oktober 2023. Gambar 1 di bawah ini menggambarkan perkembangan penelitian yang dilakukan secara bertahap:



Gambar 1. Diagram Penelitian

Berikut penjelasan mengenai langkah pada penelitian, sebagai berikut:

Langkah 1. Studi Literatur dan Identifikasi Masalah

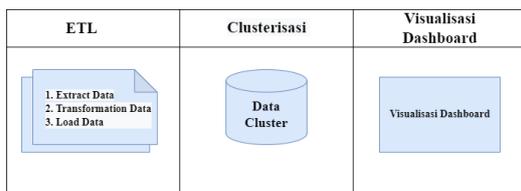
Tahap penelitian ini dimulai dengan pengumpulan data kelulusan mahasiswa Strata-1 Fakultas Teknik Universitas Udayana pada tahun angkatan 2017 di Provinsi Bali. Data ini bersumber dari Unit Sumber Daya Informasi (USDI) Universitas Udayana yang akan dimanfaatkan untuk mendapatkan *insight* baru dan membantu Fakultas Teknik Universitas Udayana untuk mengevaluasi tingkat kelulusan mahasiswa. Sebelum dianalisis, data tersebut diolah melalui proses ETL (*Extract, Transform, dan Load*) sehingga data dapat digunakan untuk dianalisis lebih lanjut.

Langkah 2. Analisis Data dan Kebutuhan Informasi

Analisis data dilakukan untuk mengetahui kelompok siswa mana yang lulus tepat waktu dan kelompok mana yang tidak lulus tepat waktu. Analisis ini bertujuan untuk membedakan kelompok siswa yang lulus tepat waktu dan kelompok siswa yang tidak lulus dengan mengidentifikasi ciri-ciri khas yang membedakan kelompok mahasiswa tersebut dengan mahasiswa yang lulus tepat waktu. Hal ini akan memungkinkan pemahaman yang lebih baik tentang faktor-faktor yang mempengaruhi tingkat kelulusan. Penelitian ini dilakukan dengan tujuan untuk membantu Fakultas Teknik Universitas Udayana dalam merumuskan kebijakan, strategi, ataupun program yang akan meningkatkan persentase mahasiswa yang lulus.

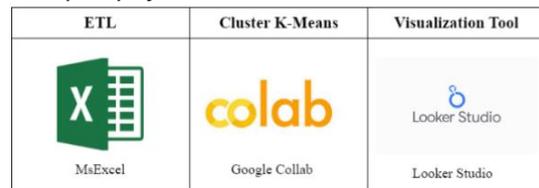
Langkah 3. Proses data mining menggunakan algoritma *K-Means Clustering*

Algoritme *K-Means Clustering*, yang dijelaskan dalam Arsitektur Logis, diterapkan pada proses *data mining* saat ini, dimana data dari akademik mahasiswa Fakultas Teknik Universitas Udayana diproses melalui proses ETL (*Extract, Transform, dan Loading*) untuk mendapatkan data kelulusan mahasiswa. Setelahnya, data tersebut dianalisis dengan tujuan mengidentifikasi kelompok mahasiswa lulus tepat waktu dan lulus tidak tepat waktu. Hasil analisis kemudian divisualisasikan melalui chart dan grafik yang terintegrasi dalam sebuah *dashboard* [9]. Gambar 2 merupakan gambar yang menjelaskan tahapan *logical architecture*.



Gambar 2. *Logical Architecture*

Tahap arsitektur fisik mencerminkan konfigurasi teknis yang diadaptasi dari arsitektur logis. Proses ini dimulai dengan proses ETL menggunakan *Microsoft Excel* untuk mengolah data. Kemudian, data dianalisis menggunakan metode *K-Means Clustering* melalui *Google Collab*. Langkah terakhir melibatkan visualisasi data dalam bentuk *chart* dan grafik yang ditampilkan di sebuah *dashboard* menggunakan *Looker Studio*. Skema perancangan arsitektur fisik ini dapat ditemukan dalam diagram berikut [9]. Gambar 3 menjelaskan mengenai tahapan *physical architecture*.



Gambar 3. *Physical Architecture*

Langkah 4. Visualisasi data menjadi *chart* dan grafik

Visualisasi yang akan ditampilkan terkait informasi *clustering* data kelulusan mahasiswa Fakultas Teknik Universitas Udayana angkatan 2017 asal bali akan dibuat menggunakan *Looker Studio*. Kelompok data hasil *clustering* pada langkah sebelumnya yang diproses melalui *Microsoft Excel* selanjutnya file yang tersimpan akan diekspor ke *Google Spreadsheet* dimana nantinya sumber data pada *Google Spreadsheet* tersebut akan diintegrasikan dengan *Looker Studio* untuk kemudian divisualisasikan ke representasi grafis yang disusun menjadi sebuah *dashboard* yang terpadu.

Langkah 5. *Deployment* hasil *clustering* data dan visualisasi dalam *dashboard*

Penerapan hasil *K-Means Clustering* data dan visualisasi pada *dashboard* dalam bentuk solusi yang dapat ditawarkan kepada Universitas Udayana mengacu pada analisis data kelulusan mahasiswa [9].

Langkah 6. Laporan dan Penarikan Kesimpulan

Tahap terakhir dari penelitian adalah pembuatan laporan yang termasuk mengenai pembahasan hasil serta penarikan kesimpulan atas penelitian yang telah dilakukan yakni visualisasi data pengelompokkan kelulusan mahasiswa dengan algoritma *K-Means Clustering*.

4. HASIL DAN PEMBAHASAN

4.1 Proses Pengumpulan Data

Penelitian ini menggunakan data kelulusan mahasiswa dari Fakultas Teknik

```
df_dataakademik
```

	nim	prodi	kabupaten	total_sks	ipk_akhir	lama_studi	tahun_lulus
0	1705511001	Sarjana Teknik Sipil	Kota Denpasar	151	3.80	4	2021
1	1705511003	Sarjana Teknik Sipil	Kota Denpasar	145	3.54	4	2021
2	1705511005	Sarjana Teknik Sipil	Jembrana	150	3.84	4	2021
3	1705511006	Sarjana Teknik Sipil	Buleleng	148	3.75	4	2021
4	1705511009	Sarjana Teknik Sipil	Badung	145	3.74	4	2021
...
316	1705552045	Sarjana Teknologi Informasi	Kota Denpasar	147	3.12	4	2022
317	1705552050	Sarjana Teknologi Informasi	Kota Denpasar	144	3.80	4	2022
318	1705552051	Sarjana Teknologi Informasi	Kota Denpasar	144	3.72	4	2021
319	1705552052	Sarjana Teknologi Informasi	Kota Denpasar	144	3.82	5	2022
320	1705552057	Sarjana Teknologi Informasi	Kota Denpasar	145	3.81	4	2022

321 rows x 7 columns

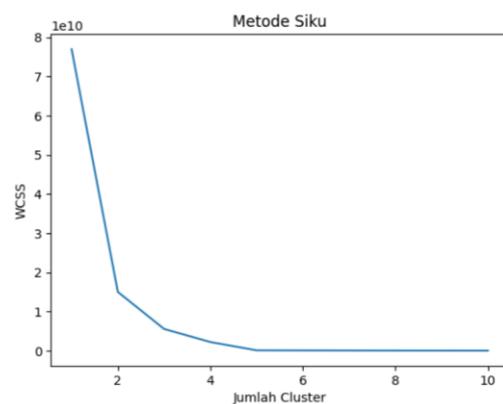
Gambar 4. Tampilan Tabel Data Akademik Mahasiswa di *Google Collab*

4.2 Proses Data Mining Menggunakan Algoritma *K-Means Clustering*.

4.2.1 Pencarian Jumlah Cluster

Penerapan *K-Means Clustering* ini dimulai dengan melakukan menentukan banyaknya *cluster* efektif dan penetapan nilai *cluster* efektif, yang merupakan tahap terpenting untuk bisa menghasilkan hasil *clustering* yang optimal. Dengan menggunakan metode *elbow*, seperti yang terilustrasikan pada Gambar 5, proses diinisialisasi dengan memperhitungkan dan mencari nilai optimal secara berulang dalam rentang dari 1 hingga 10.

Strata-1 Universitas Udayana pada tahun angkatan 2017 di Provinsi Bali yang telah melalui proses ETL (*Extract, Transform, Load*). Data tersebut diekspor ke dalam *Google Collab* untuk dilakukan analisis lebih lanjut dengan parameter yang digunakan meliputi nomor induk mahasiswa (NIM), program studi (prodi), kabupaten, total SKS, IPK akhir, lama studi, dan tahun kelulusan. Gambar 4 berikut menjelaskan rincian tampilan tabel data kelulusan mahasiswa menggunakan tools *Google Collab*.



Gambar 5. Jumlah Cluster (n)

Berdasarkan Gambar 5, dapat ditentukan bahwa jumlah *cluster* (n) yang optimal terhadap penerapan *K-Means*

Clustering adalah berjumlah 2 *cluster*. Penentuan dari nilai *centroid* masing-masing *cluster* akan digunakan untuk menentukan setiap data termasuk ke dalam *cluster* mana saja dengan memperhitungkan jarak dari tiap data terhadap *centroid* terpilihnya [4].

4.2.2 Penetapan Nilai Cluster dan Centroid Awal.

Titik pusat *cluster* (*centroid*) menjadi parameter kunci dalam menghitung jarak antara setiap data ke *cluster* yang bersangkutan (*distance*). Dalam proses ini penentuan nilai titik pusat *cluster* bisa ditentukan dengan metode *elbow* yang telah diterapkan. Kode Program 1 merupakan kode pemrograman untuk menentukan jumlah *cluster* (K).

Kode Program 1. Mengukur Jumlah Cluster (K)

```
# Menentukan Jumlah Cluster (K)
num_clusters = 2
# Mencari Titik Pusat Cluster
initial_centers =
X_normalized[[24, 26]]
```

Penelitian ini menetapkan jumlah *cluster* sebanyak dua, ditandai dengan variabel 'num_clusters = 2 yang bertujuan untuk menentukan jumlah kelompok data atau *cluster* yang akan terbentuk. Dimana masing-masing *cluster* teridentifikasi sebagai:

- C1 = Mahasiswa Lulus Tepat Waktu
- C2 = Mahasiswa Lulus Telat

Variabel 'initial_centers = X_normalized[[24, 26]]' memiliki peran penting dalam menetapkan titik awal *cluster* atau *centroid* dalam analisis *cluster*. Pada penilai ini indeks ke-24 dan ke-26 dari dataset X_normalized dipilih sebagai titik pusat *cluster*. Proses pemilihan ini merupakan langkah awal dalam pengelompokan data ke dalam *cluster* yang optimal, memungkinkan penentuan awal posisi *cluster* untuk analisis lebih lanjut.

4.2.3 Mengukur Jarak dari Setiap Data ke Pusat Cluster

Mengukur jarak dari setiap data ke titik pusat *cluster* langkah penting dalam proses *clustering* menggunakan metode algoritma *K-Means*. Tahap ini menghitung jarak dua kali, yaitu pada perhitungan pertama dalam iterasi ke-1 dan kedua dalam iterasi ke-2, menggunakan kode yang telah disediakan. Iterasi ke-1 bertujuan untuk menentukan *cluster* awal dengan cara memilih titik-titik pusat *cluster* secara acak atau dengan suatu pendekatan tertentu. Kode Program 2 merupakan code untuk mengimplementasikan proses iterasi ke-1 dan menampilkan hasil pusat *cluster* iterasi ke-1.

Kode Program 2. Mengukur Jarak dari Setiap Data ke Pusat Cluster

```
# Iterasi 1 - Menghitung Jarak
dari Setiap Data ke Pusat
Cluster
kmeans =
KMeans(n_clusters=num_clusters,
init=initial_centers, n_init=1,
random_state=42)
cluster_labels =
kmeans.fit_predict(X_normalized)
(kmeans.cluster_centers_)
```

Iterasi ke-1 dilakukan dengan menghitung jarak dari setiap data dalam dataset terhadap pusat *cluster*. Pengelompokan dilakukan menggunakan metode *K-Means Clustering* dengan dua *cluster* yang telah diperoleh sebelumnya. Penggunaan variabel 'initial_centers' sebagai titik awal pusat *cluster* memungkinkan perhitungan jarak yang lebih terfokus pada *cluster* yang sudah ada. Tabel 1 menampilkan hasil pusat *cluster* yang dihasilkan oleh proses iterasi ke-1.

Tabel 1. Hasil Iterasi ke-1

0	0.4336	0.59	0.0418	0.655	0.02
1	0.462	0.5524	0.889	0.6309	0.0456

Hasil dari iterasi ini menunjukkan pusat-pusat *cluster* yang ditemukan, diwakili oleh dua himpunan nilai numerik yang masing-masing menggambarkan posisi relatif dari

setiap fitur dalam dataset terhadap *cluster* yang terbentuk. Kode Program 3 merupakan *code* untuk mengimplementasikan proses iterasi ke-2 dan menampilkan hasil pusat *cluster* iterasi ke-2.

Kode Program 3. Menghitung Pusat *Cluster* yang Baru

```
# iterasi 2: Menghitung Pusat
Cluster yang Baru
kmeans.fit(X_normalized)
centroids_new =
kmeans.cluster_centers_
(kmeans.cluster_centers_)
```

Iterasi ke-2 dilakukan untuk memperbarui posisi titik-titik pusat *cluster* berdasarkan rata-rata dari anggota *cluster* saat ini, dan setiap titik data diperbarui ke *cluster* yang sesuai dengan pusat *cluster* terdekat yang baru dihitung. Iterasi ke-2 dilakukan tidak ada perubahan dalam posisi data. Tabel 2 menampilkan pusat *cluster* yang dihasilkan oleh proses iterasi ke-2.

Tabel 2. Hasil Iterasi ke-2

0	0.4336	0.59	0.0418	0.655	0.02
1	0.462	0.5524	0.889	0.6309	0.0456

Setelah menghitung iterasi ke-2, perlu adanya perbandingan jarak antara pusat *cluster* pada proses iterasi ke-1 dan proses iterasi ke-2 dalam algoritma *K-Means* adalah untuk mengevaluasi perubahan atau konvergensi *cluster* pada setiap iterasi. Hal ini penting karena proses *K-Means* bertujuan untuk menemukan pusat *cluster* yang optimal untuk grup data yang ada. Kode Program 4 merupakan *code* untuk mengukur jarak antara pusat *cluster* pada iterasi ke-1 dan pusat *cluster* iterasi ke-2 dengan menggunakan norma *Euclidean*

Kode Program 4. Menghitung Jarak Iterasi ke-1 dan Iterasi ke-2

```
# Menghitung jarak antara pusat
cluster pada iterasi 1 dan
iterasi 2
distance_between_centers =
np.linalg.norm(new_centers -
kmeans.cluster_centers_)
# Menampilkan hasil perbandingan
```

```
if distance_between_centers == 0:
    print("Titik centroid pada
iterasi ke-2 sama dengan iterasi
ke-1")
else:
    print("Titik centroid pada
iterasi ke-2 berbeda dari iterasi
ke-1")
# Menghitung jarak antara pusat
cluster pada iterasi 1 dan
iterasi 2
distance_between_centers =
np.linalg.norm(new_centers -
kmeans.cluster_centers_)
# Menampilkan hasil perbandingan
if distance_between_centers == 0:
    print("Titik centroid pada
iterasi ke-2 sama dengan iterasi
ke-1")
else:
    print("Titik centroid pada
iterasi ke-2 berbeda dari iterasi
ke-1")
```

Perhitungan jarak iterasi ke-1 dan iterasi ke-2 pada penelitian ini menggunakan perhitungan norma *Euclidean* dengan *numpy* untuk mengukur jarak antara pusat *cluster* pada kedua iterasi. Hasil dari perbandingan jarak iterasi ke-1 dengan jarak iterasi ke-2 ialah "Titik centroid iterasi ke-2 sama dengan iterasi ke-1". Jika jarak antara pusat *cluster* pada iterasi ke-1 dan iterasi ke-2 sama, ini menunjukkan bahwa posisi pusat *cluster* tidak mengalami perubahan antara kedua iterasi tersebut. Hal ini menandakan bahwa *cluster* telah mencapai konvergensi, artinya sudah tidak ada perubahan signifikan dalam posisi *cluster* lagi, dan algoritma sudah menemukan solusi yang stabil.

4.2.4 Hasil *K-Means Clustering*

Hasil *K-Means Clustering* pertama membagi program studi ke dalam kelompok berdasarkan tingkat kelulusan, sementara *K-Means Clustering* kedua mengelompokkan kabupaten dan kota berdasarkan tingkat kelulusan. Proses ini dilakukan untuk mempermudah mengetahui hasil *Clustering K-Means* mengacu pada anggota dari *Cluster* yang terbentuk. Tabel 3

menjelaskan mengenai rincian hasil *clustering* berdasarkan program studi yang dimuat ke dalam tabel.

Tabel 3. Hasil *Clustering* Berdasarkan Program Studi

Cluster 1		Cluster 2	
Lulus Tepat Waktu (4 tahun atau 8 semester)		Lulus Telat (lebih dari 4 tahun atau lebih dari 8 semester)	
Prodi	Total Mahasiswa	Prodi	Total Mahasiswa
Sarjana Teknik Sipil	89	Sarjana Teknik Sipil	2
Sarjana Teknologi Informasi	68	Sarjana Teknologi Informasi	9
Sarjana Arsitektur	61	Sarjana Arsitektur	4
Sarjana Mesin	40	Sarjana Mesin	4
Sarjana Teknik Elektro	38	Sarjana Teknik Elektro	6

Berdasarkan Tabel 3 data hasil *Clustering K-Means* terhadap kelulusan mahasiswa Fakultas Teknik Universitas Udayana tahun angkatan 2017, terlihat bahwa mayoritas mahasiswa cenderung lulus tepat pada waktunya, pada rentang durasi 4 tahun atau setara dengan 8 semester. Program studi sarjana teknik sipil memperlihatkan tingkat kelulusan tepat waktu yang paling tinggi dibandingkan program studi lainnya, sedangkan teknik elektro menunjukkan jumlah mahasiswa lulus tepat waktu yang relatif lebih rendah dibandingkan program studi lainnya. Tabel 4 menjelaskan mengenai rincian hasil *clustering* berdasarkan kabupaten atau kota yang dimuat ke dalam tabel.

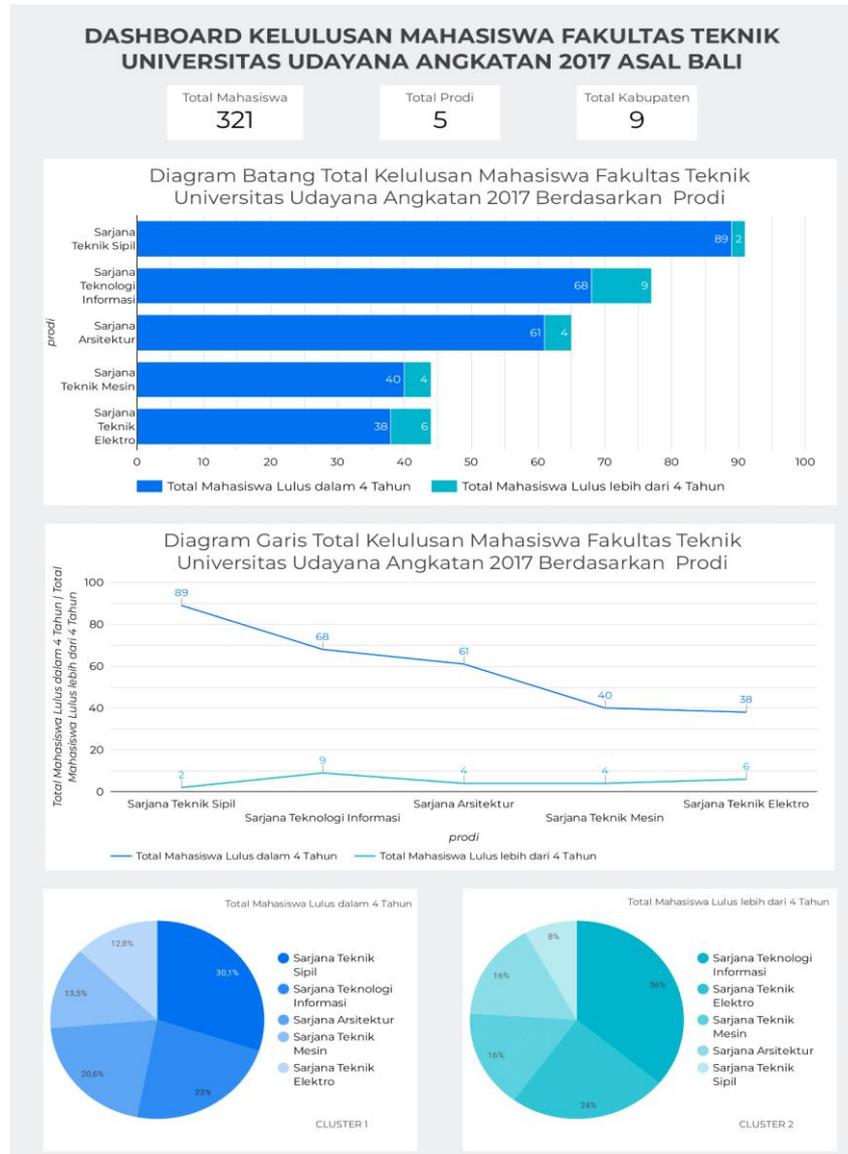
Tabel 4. Hasil *Clustering* Berdasarkan Kabupaten atau Kota

Cluster 1		Cluster 2	
Lulus Tepat Waktu (4 tahun atau 8 semester)		Lulus Telat (lebih dari 4 tahun atau lebih dari 8 semester)	
Kabupaten	Total Mahasiswa	Kabupaten	Total Mahasiswa
Kota Denpasar	161	Kota Denpasar	18
Badung	52	Badung	2
Gianyar	34	Gianyar	2
Tabanan	30	Tabanan	2
Buleleng	15	Buleleng	9
Jembrana	10	Jembrana	0
Karangasem	8	Karangasem	0
Klungkung	7	Klungkung	1

Berdasarkan Tabel 4 hasil analisis *Clustering K-Means* terhadap data kelulusan mahasiswa Fakultas Teknik Universitas Udayana berdasarkan wilayah asal yaitu kabupaten atau kota, terlihat bahwa sebagian besar mahasiswa dari wilayah Denpasar dan sekitarnya cenderung lulus tepat waktu pada rentang durasi 4 tahun atau setara dengan 8 semester. Wilayah Denpasar memiliki jumlah mahasiswa lulus tepat waktu yang signifikan dibandingkan wilayah lain, diikuti oleh Badung, Gianyar, dan Tabanan. Sementara itu, dalam *Cluster 2* yang menunjukkan mahasiswa yang lulus telat, wilayah Buleleng memiliki jumlah yang lebih signifikan dibandingkan dengan wilayah lainnya dalam kategori ini.

4.3 Visualisasi Dashboard

Visualisasi data bertujuan untuk mempresentasikan data menjadi informasi bentuk grafis yang mudah diterima pengguna. Visualisasi data ke dalam bentuk *dashboard* pada penelitian ini dibuat menggunakan *Looker Studio*. *Dashboard* dibagi menjadi 2 halaman dengan tampilan halamannya ditunjukkan pada Gambar 6 dan Gambar 7 berikut.

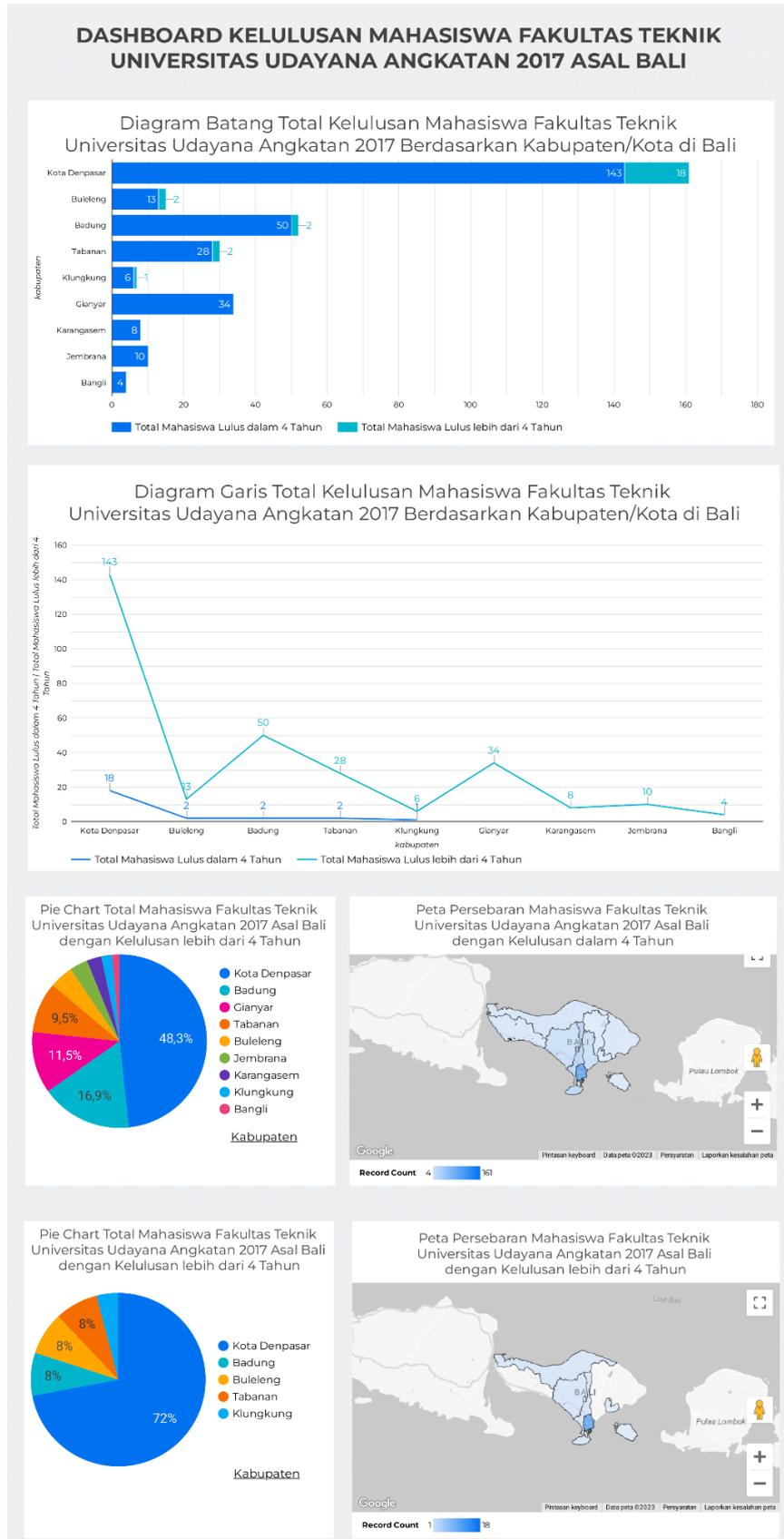


Gambar 6. Cluster Dashboard Halaman Pertama

Pada Gambar 6 tersebut ditampilkan halaman pertama dari dashboard yang dapat digunakan sebagai media visualisasi dan presentasi secara general. Informasi mengenai data yang dianalisa dalam penelitian, kategori yang diakses oleh pengguna beserta persebaran maupun persentasenya, jumlah aktivitas setiap page pada masing-masing kategori secara keseluruhan, dapat dilihat pada halaman pertama dashboard pada bagian scorecard.

Pada halaman pertama ditampilkan scorecard mulai dari total mahasiswa, total program studi hingga total kabupaten yang datanya digunakan pada dashboard ini yaitu

menggunakan data mahasiswa Fakultas Teknik Universitas Udayana angkatan 2017 asal bali. Dibawahnya dilanjutkan dengan tampilan diagram batang, diagram garis dan pie chart dari data kelulusan mahasiswa dengan dengan 2 cluster yang diuraikan mengacu pada asal program studinya. Dimana cluster 1 merupakan kelompok total mahasiswa yang lulus pada rentang durasi 4 tahun. Sedangkan cluster 2 merupakan kelompok total mahasiswa yang lulus lebih dari 4 tahun lamanya. Pada Gambar 7 berikut ditampilkan halaman ke-2 dari dashboard ini.



Gambar 7. Cluster Dashboard Halaman Kedua

Gambar 7 yang ditunjukkan di atas merupakan tampilan halaman kedua dari *dashboard* pada penelitian ini yang dimana pada halaman ini menunjukkan tampilan diagram batang, diagram garis, pie chart dan peta persebaran dari data kelulusan mahasiswa dengan dengan 2 *cluster* yang diuraikan berdasarkan kabupaten atau kota asal dari mahasiswa. Dimana 2 *cluster* disini sama dengan *cluster* pada halaman pertama.

Dashboard ini dibuat dengan ukuran 1200 piksel x 1800 piksel untuk halaman pertama sedangkan halaman kedua berukuran 1200 piksel x 2400 piksel.

4.4 Deployment hasil *clustering* data dan deployment *dashboard* visualisasi data

Pihak Fakultas Teknik Universitas Udayana dapat mengambil beberapa kegiatan-kegiatan seperti dengan mengadakan:

1. Bagi program studi dapat membuat program bimbingan studi bagi mahasiswa dengan mengadakan semester pendek yang dapat membantu meningkatkan prestasi belajar mahasiswa ataupun menyediakan program mentoring akademik oleh dosen, mahasiswa akhir, ataupun alumni yang dapat membantu mahasiswa dalam memahami materi, pengerjaan tugas, serta memberikan motivasi untuk meningkatkan semangat studi mahasiswa.
2. Bagi pihak Fakultas Teknik Universitas Udayana dapat melakukan kerja sama dengan pemerintah setempat melalui penyelenggaraan program-program pendidikan yang mendukung mahasiswa di wilayah tersebut, seperti pemberian beasiswa, pelatihan, atau bantuan akses internet.

5. KESIMPULAN

5.1 Kesimpulan

Berdasarkan analisis dari proses dan hasil penelitian, diperoleh kesimpulan

bahwa mayoritas mahasiswa Fakultas Teknik Universitas Udayana angkatan 2017 cenderung lulus tepat waktu, terutama dalam program studi Teknik Sipil yang memiliki tingkat kelulusan tertinggi, sementara program studi Teknik Elektro memiliki tingkat kelulusan yang relatif lebih rendah. Wilayah asal mahasiswa, seperti Denpasar, Badung, Gianyar, dan Tabanan, menunjukkan angka lulus tepat waktu yang signifikan, namun perlu perhatian khusus terhadap wilayah Buleleng yang memiliki jumlah mahasiswa yang lulus dengan keterlambatan. Selain itu, visualisasi data menggunakan *Looker Studio* dianggap sebagai solusi alternatif yang memudahkan pihak Fakultas Teknik Universitas Udayana dalam mengintegrasikan dan berbagi laporan dari berbagai sumber data, termasuk *Google Spreadsheet*, dengan memperhatikan keamanan informasi, untuk mendukung pengambilan keputusan yang berbasis pada data visualisasi.

6. DAFTAR PUSTAKA

- [1] A. Khaerunnisa. (2022). Analisis Tingkat Kelulusan Mahasiswa di Unisba dengan menggunakan Algoritma *K-Means Clustering*. *Jurnal Riset Matematika (JRM)*, 2(1), 73-82.
- [2] H. Priyatman, F. Sajid, D. Haldivany, "Clusterisasi Menggunakan Algoritma *K-Means Clustering* untuk Memprediksi Waktu Kelulusan Mahasiswa" *Jurnal Edukasi dan penelitian Informatika*, vol. 5, no. 1, pp. 62-66, 2019.
- [3] M. R. Muttaqin., M. Defriani. "Algoritma *K-Means* untuk Pengelompokan Topik Skripsi Mahasiswa". *ILKOM Jurnal Ilmiah*. Vol. 12, No. 2. pp. 1221-129.
- [4] K. G. F. F. Utomo, F. A. Bachtiar, N. Y. Setiawan, "Analisis Perilaku Pengguna pada PT Antar Surya Jaya : Harian Surya Surabaya menggunakan Metode *K-Means Clustering*", *Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer*, vol. 5, no. 11, pp. 4828-4835, 2020.
- [5] M. K. Hidayat, R. Fitriana, "Penerapan Sistem Intelijensia Bisnis Dan *K-Means*

- Clustering* Untuk Memantau Produksi Tanaman Obat”, *Jurnal teknologi Industri Pertanian*. vol 2, no 32, pp 204-219, 2022.
- [6] F. N. Hayati, M. Sofiani, Diana. Nurlaily. “Pemanfaatan *Google Data Studio* Untuk Visualisasi E-Rapor Siswa SMAN 2 Balikpapan”, *Jurnal Pengabdian Kepada Masyarakat*, vol 2, no 2, pp. 87-94, 2021.
- [7] Z. Nabila., dkk. “Analisis Data Mining Untuk *Clustering* Kasus Covid-19 di Provinsi Lampung dengan Algoritma *K-Means*”. *Jurnal Teknologi dan Sistem Informasi (JTSI)*. Vol. 2, No. 2, pp. 100-108. 2021.
- [8] D. Fernando. “Visualisasi Data Menggunakan *Google Data Studio*”. *Seminar Nasional Rekayasa Teknologi Informasi*, vol. 1, no. 10, pp. 71-77, 2018.
- [9] R. Akbar, M. Octaviany. “Perencanaan Visualisasi *Dashboard* dan *Clustering* dengan Menerapkan *Business Intelligence* pada Dinas DPMPTSP Kabupaten Dharmasraya”, *Jurnal Edukasi dan Penelitian Informatika*. vol. 7, no. 3, pp. 340-350, 2021.
- [10] A. Wicaksono, F. A. Bachtiar, N. Y. Setiawan. “Segmentasi Pelanggan Menggunakan *Fuzzy C-Means Clustering* berdasarkan RFM Model pada *E-Commerce*”, *Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer*, vol. 5, no. 4, pp 1351-1360, 2021.
- [11] T. A. Saputri, S. Muharni, A. Perdana, Sulistiyanto. “Pemanfaatan *Google Data Studio* Untuk Visualisasi Data Bagi Kepala Gudang UD Salim Abadi”, *Ilmu Komputer Untuk Masyarakat*, vol 2, no 2, pp. 67-72, 2021.
- [12] F. Pandensolang, F. Manoppo, A. Sumendap. “Implementasi *Business Intelligence* Untuk Analisa dan Visualisasi Perbandingan Perencanaan dan Realisasi Anggaran pada BNNP Sulawesi Utara” pp. 1-8.
- [13] A. T. Putri, R. Andreswari, Irfan Darmawan, “Analisis Perancangan *Business Intelligence* Dan *Dashboard* Pada Data Kuota Telekomunikasi Dengan Metode *Business Dimensional Lifecycle*”, *eProceedings of Engineering*. vol. 8, no. 5, pp 8964-8969, 2021.