

PENERAPAN K-MEANS CLUSTERING PADA DATA BANTUAN KEUANGAN UNTUK PEMBANGUNAN DESA PADA KABUPATEN KARAWANG

Ivan Rizwan ¹, Jajam Haerul Jaman ^{2*}

^{1,2,)} Universitas Singaperbangsa Karawang

Email : jajam.haeruljaman@staff.unsika.ac.id

Abstrak

K-means clustering adalah metode pengelompokan data yang digunakan untuk mengelompokkan data ke dalam beberapa kategori (atau "kelompok") berdasarkan kesamaan fitur. Dalam penerapannya pada data bantuan keuangan untuk pembangunan desa, k-means clustering dapat digunakan untuk mengelompokkan desa-desa berdasarkan tingkat kebutuhan bantuan keuangan yang berbeda. Dalam berbagai bidang, seperti bisnis, keamanan jaringan, sains data, dan pembelajaran mesin, clustering digunakan untuk mengelompokkan data menjadi grup yang homogen. Penerapan clustering dapat digunakan untuk menentukan jumlah cluster yang optimal dan mengelompokkan data ke dalam cluster yang sesuai. Ini dapat membantu dalam pengambilan keputusan dan meningkatkan efektivitas dan efisiensi dalam berbagai aplikasi. Dengan mengelompokkan desa-desa ke dalam beberapa kelompok yang berbeda, pemerintah dapat menyalurkan bantuan keuangan dengan lebih efektif dan efisien, sesuai dengan kebutuhan masing-masing kelompok.

Kata kunci: *k-means, clustering, dataset*

Abstract

K-means clustering is a data clustering method used to group data into categories (or "clusters") based on feature similarities. In its application to data on financial assistance for village development, k-means clustering can be used to group villages based on different levels of need for financial assistance. In various fields, such as business, network security, data science, and machine learning, clustering is used to group data into homogeneous groups. The application of clustering can be used to determine the optimal number of clusters and group data into appropriate clusters. This can help in decision making and increase effectiveness and efficiency in various applications. By grouping villages into different groups, the government can distribute financial assistance more effectively and efficiently, according to the needs of each group.

Keyword: *k-means, clustering, dataset*

*Penulis Korespondensi

Diterima : 08 Agustus 2023. Disetujui : 30 September 2023. Dipublikasikan : 30 September 2023

I. PENDAHULUAN

Pembangunan desa merupakan salah satu upaya untuk meningkatkan kualitas hidup masyarakat di desa. Bantuan keuangan merupakan salah satu bentuk dukungan yang dapat diberikan untuk mendukung pembangunan desa. Namun, dalam penyaluran bantuan keuangan, seringkali terdapat kesulitan dalam menentukan desa-desa mana yang membutuhkan bantuan keuangan lebih banyak. Oleh karena itu, diperlukan suatu metode yang dapat digunakan untuk mengelompokkan desa-desa berdasarkan tingkat kebutuhan bantuan keuangan.

Klustering adalah cara untuk mengklasifikasikan data mentah dengan baik dan mencari pola yang mungkin ada dalam dataset [1]. Metode ini mengelompokkan data ke dalam beberapa kategori (atau "kelompok") berdasarkan kesamaan fitur. Dalam penerapannya pada data bantuan keuangan untuk pembangunan desa, *k-means clustering* dapat digunakan untuk mengelompokkan desa-desa berdasarkan tingkat kebutuhan bantuan keuangan yang berbeda.

Proses ini adalah pengelompokan objek data ke dalam *cluster* yang tidak terkait sehingga data dalam *cluster* yang sama mirip, namun data yang termasuk dalam *cluster* yang berbeda berbeda. Kebutuhan untuk mengatur data yang tumbuh dengan cepat dan belajar informasi berharga dari data, yang membuat teknik *clustering* banyak diterapkan dalam banyak bidang aplikasi seperti kecerdasan buatan, biologi, manajemen hubungan pelanggan, kompresi data, ekstraksi data, pencarian informasi, pengolahan gambar, pembelajaran mesin, pemasaran, kedokteran, pengenalan pola, psikologi, *statistic* [2].

Penelitian ini bertujuan untuk mengevaluasi penerapan *k-means clustering*

pada data bantuan keuangan untuk pembangunan desa. Penelitian ini diharapkan dapat menghasilkan metode yang efektif dan efisien dalam mengelompokkan desa-desa berdasarkan tingkat kebutuhan bantuan keuangan, sehingga dapat membantu pemerintah dalam menyalurkan bantuan keuangan yang sesuai dengan kebutuhan masing-masing desa.

II. METODOLOGI PENELITIAN

Metode analisis yang digunakan dalam penelitian ini dikenal sebagai penemuan pengetahuan dalam basis data atau bisa disebut *Knowledge Discovery in Database* (KDD). KDD adalah komponen utama dalam proses mengubah data menjadi kumpulan informasi yang tidak ada celahnya. Informasi ini ditemukan dalam set data skala besar yang manfaat potensialnya sebelumnya tidak dipahami.

Pada tahun 1967, MacQueen pertama kali mengusulkan algoritma *k-means*, itu adalah salah satu algoritma pembelajaran *non-supervised* yang paling sederhana, yang digunakan untuk menyelesaikan masalah *cluster* yang terkenal [3].

Data Selection

Pada tahap ini, pengumpulan data dilakukan. Dataset yang digunakan pada penelitian kali ini merupakan data yang diambil dari website opendata.karawangkab.go.id yang diperoleh oleh Dinas Pemberdayaan Masyarakat dan Desa. Data yang terdapat pada dataset tersebut merupakan data bantuan keuangan untuk pembangunan desa pada kabupaten Karawang yang memiliki 10 atribut yaitu no, nama_kabupaten, kode_kecamatan, nama_kecamatan, add, dbh, dana_desa, bantuan_gubernur, satuan dan tahun. Tampilan dari dataset tersebut dapat dilihat pada Gambar 1 berikut.

KODE_KECAMATAN	NAMA_KECAMATAN	ADD	DBH	DANA_DESA	BANTUAN_GUBERNUR	SATUAN	TAHUN
3215010	PANGKALAN	4,117,856,000	2,060,743,200	8,865,817,000	1,040,000,000	RUPIAH	2020
3215011	TEGALWARU	4,334,363,000	2,396,283,300	9,762,887,000	1,170,000,000	RUPIAH	2020
3215020	CIAMPEL	3,821,932,000	1,752,904,600	9,052,964,000	910,000,000	RUPIAH	2020
3215031	TELUKJAMBE TIMUR	6,720,350,000	4,107,156,000	8,525,158,000	1,170,000,000	RUPIAH	2020
3215032	TELUKJAMBE BARAT	4,486,550,000	2,838,040,300	10,177,618,000	1,300,000,000	RUPIAH	2020
3215040	KLARI	9,350,492,000	4,352,900,600	13,080,444,000	1,690,000,000	RUPIAH	2020
3215050	CIKAMPEK	7,476,647,000	2,689,833,800	9,676,847,000	1,300,000,000	RUPIAH	2020
3215051	PURWASARI	4,536,870,000	2,169,499,400	8,949,749,000	1,040,000,000	RUPIAH	2020
3215060	TIRTAMULYA	4,894,280,000	2,436,316,300	10,255,173,000	1,300,000,000	RUPIAH	2020

Gambar 1. Tampilan Dataset

1. Data Preprocessing

Pada tahap ini, atribut yang tidak diperlukan akan dihapus agar

memudahkan proses pengolahan data. Atribut – atribut yang dihapus seperti no, nama_kabupaten, kode_kecamatan, satuan, dan tahun. Berikut hasilnya dapat dilihat pada tabel 1

Tabel 1. Hasil Preprocessing

NAMA_KECAMATAN	ADD	DBH	DANA_DESA	BANTUAN_GUBERNUR
PANGKALAN	4,117,856,000	2,060,743,200	8,865,817,000	1,040,000,000
TEGALWARU	4,334,363,000	2,396,283,300	9,762,887,000	1,170,000,000
CIAMPEL	3,821,932,000	1,752,904,600	9,052,964,000	910,000,000
TELUKJAMBE TIMUR	6,720,350,000	4,107,156,000	8,525,158,000	1,170,000,000
TELUKJAMBE BARAT	4,486,550,000	2,838,040,300	10,177,618,000	1,300,000,000
KLARI	9,350,492,000	4,352,900,600	13,080,444,000	1,690,000,000
CIKAMPEK	7,476,647,000	2,689,833,800	9,676,847,000	1,300,000,000
PURWASARI	4,536,870,000	2,169,499,400	8,949,749,000	1,040,000,000
TIRTAMULYA	4,894,280,000	2,436,316,300	10,255,173,000	1,300,000,000

Untuk penjelasan atribut – atribut yang terdapat pada dataset adalah sebagai berikut [4]

- Nama_Kecamatan : menyatakan nama kecamatan yang ada di Kabupaten Karawang sesuai ketentuan BPS merujuk pada Peraturan Kepala BPS Nomor 3 Tahun 2019 dengan tipe data teks
- ADD : menyatakan jumlah bantuan Anggaran Dana Desa dari Pemerintah Kabupaten Karawang dengan tipe data numerik
- DBH : menyatakan jumlah bantuan Dana Bagi Hasil dengan tipe data numerik
- Dana_Desa : menyatakan jumlah bantuan Dana Desa dari

Pemerintah Pusat dengan tipe data numerik

- Bantuan_Gubernur : menyatakan jumlah Bantuan Gubernur dari Pemerintah Provinsi Jawa Barat dengan tipe data numerik

2. Data Mining

Setelah data sudah melalui *preprocessing*. Maka saatnya data tersebut diolah dalam proses data mining dengan teknik *clustering* dan menggunakan metode *K-means* untuk mengolahnya. Pertama – tama dilakukan dulu pengecekan pada data apakah terdapat *missing value* atau nilai yang kosong, kemudian

memperbaikinya dengan mencari *mean/median* atau bahkan menghilangkannya. Kemudian data yang sudah berhasil di prediksi *missing valuenya* di transformasi agar bisa dilakukan proses *clustering* menggunakan tool R Studio. Terakhir memvisualisasikan *cluster – cluster plot* yang berhasil didapatkan menggunakan *library Cluster* yang ada pada *software* R Studio. Segala proses

pengolahan data ini sepenuhnya dilakukan menggunakan *tool* R Studio.

III. HASIL PENELITIAN

Penelitian ini sepenuhnya dilakukan menggunakan *tool* R Studio pada perangkat *Windows*. Pertama – tama lakukan terlebih dahulu penginstalan *package – package* yang dibutuhkan seperti yang terlihat pada gambar 2.

```
1 install.packages("tidyverse")
2 install.packages("factoextra")
3 install.packages("cluster")
4 library(tidyverse)
5 library(factoextra)
6 library(cluster)
```

Gambar 2. Daftar Package yang dibutuhkan

Kemudian proses import data dilakukan, disini peneliti melakukan *import data* menggunakan *command line* `data=read.delim("clipboard")` untuk

mengimportkan data secara otomatis dari blok – blok sel yang sudah di salin dari Excel dan tersimpan di *clipboard*. Berikut hasil dari data yang sudah diimport terdapat pada gambar 3.

data		30 obs. of 5 variables									
\$ NAMA_KECAMATAN	: chr	"PANGKALAN"	"TEGALWARU"	"CIAMPEL"	"TELUKJAMBE TIMUR"	...					
\$ ADD	: num	4.12e+09	4.33e+09	3.82e+09	6.72e+09	4.49e+09	...				
\$ DBH	: num	2.06e+09	2.40e+09	1.75e+09	4.11e+09	2.84e+09	...				
\$ DANA_DESA	: num	8.87e+09	9.76e+09	9.05e+09	8.53e+09	1.02e+10	...				
\$ BANTUAN_GUBERNUR	: num	1.04e+09	1.17e+09	9.10e+08	1.17e+09	1.30e+09	1.69e+09	1.30e+09	1.04e+09	1.04e+09	1.04e+09

Gambar 3. Data yang sudah berhasil di import

Kemudian peneliti melakukan pengecekan *missing value* atau NA. Berikut hasil pengecekannya terdapat pada gambar 4.

```

> data=read.delim("clipboard")
> view(data)
> dataset=data[,-1]
> row.names(dataset)=data[,1]
> view(dataset)
> summary(data)
  NAMA_KECAMATAN      ADD          DBH      DANA_DESA      BANTUAN_GUBERNUR
Length:30      Min.   :2.744e+09   Min.   :1.397e+09   Min.   :4.261e+09   Min.   :5.200e+08
Class :character  1st Qu.:4.487e+09   1st Qu.:2.367e+09   1st Qu.:9.758e+09   1st Qu.:1.170e+09
Mode  :character  Median :5.987e+09   Median :2.690e+09   Median :1.277e+10   Median :1.300e+09
                        Mean   :5.725e+09   Mean   :2.712e+09   Mean   :1.193e+10   Mean   :1.331e+09
                        3rd Qu.:6.720e+09   3rd Qu.:3.091e+09   3rd Qu.:1.403e+10   3rd Qu.:1.560e+09
                        Max.   :9.350e+09   Max.   :4.353e+09   Max.   :1.563e+10   Max.   :1.820e+09
                        NA's   :1         NA's   :1         NA's   :1         NA's   :1

```

Gambar 4. Hasil Pengecekan Missing Value

Berdasarkan tabel diatas, telah ditemukan adanya 4 *missing value* pada kolom dari row Karawang Barat yaitu empat buah kolom yang tidak memiliki nilai di dalamnya.

Maka proses selanjutnya adalah mengisi terlebih dahulu kolom – kolom yang kosong itu dengan cara mencari *mean*, hasil dapat dilihat pada gambar 5. Dilakukan proses penambahan nilai dengan menggunakan nilai rata – rata atau *mean* dari kolom – kolom pada baris lain secara satu persatu.

```

data$ADD=ifelse(is.na(data$ADD), mean(data$ADD,na.rm=TRUE),data$ADD)
data$DBH=ifelse(is.na(data$DBH), mean(data$DBH,na.rm=TRUE),data$DBH)
data$DANA_DESA=ifelse(is.na(data$DANA_DESA), mean(data$DANA_DESA,na.rm=TRUE),data$DANA_DESA)
data$BANTUAN_GUBERNUR=ifelse(is.na(data$BANTUAN_GUBERNUR), mean(data$BANTUAN_GUBERNUR,na.rm=TRUE),data$BANTUAN_GUBERNUR)

```

17	MAJALAYA	4405939000	2145476300	7801436000	910000000
18	KARAWANG TIMUR	2743540000	1397201500	4260866000	520000000
19	KARAWANG BARAT	5725366897	2712106710	11934646724	1331379310
20	RAWAMERTA	6193775000	3091129600	13760016000	1690000000
21	TEMPURAN	7416480000	3363663600	15454484000	1820000000
22	KUTAWALUYA	6040803000	2915816300	14006506000	1560000000

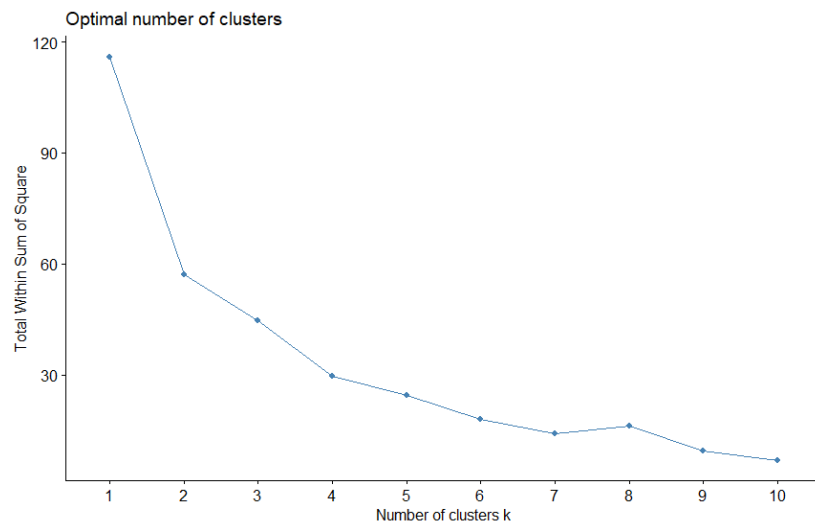
Gambar 5. Hasil Pengisian Nilai Kosong Menggunakan Mean

Selanjutnya menggunakan metode yang menjadikan *Sum of Square Error* (SSE) sebagai pertimbangan. Hasil *cluster* optimal ditentukan dengan menggunakan metode *elbow*, yang menunjukkan bahwa jumlah *cluster* yang optimal adalah 2, yang ditunjukkan oleh sudut siku pada gambar 6.

```

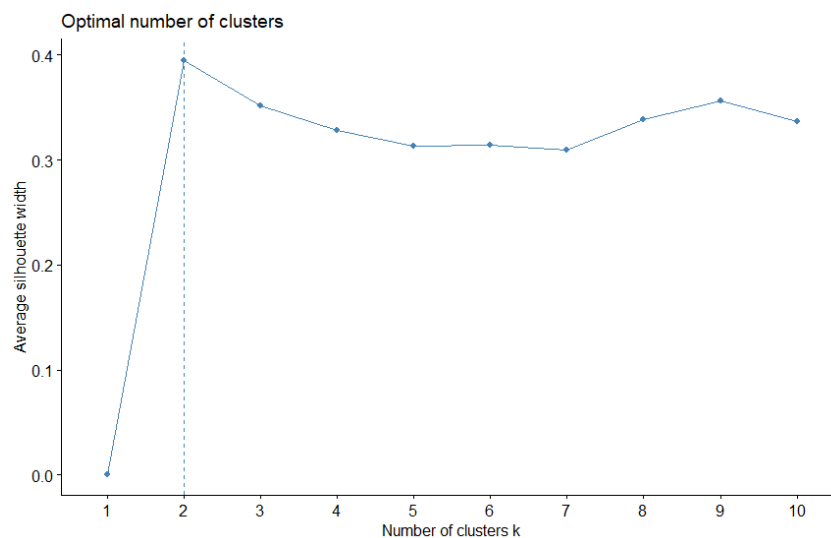
fviz_nbclust(datafix,kmeans,method="wss")
fviz_nbclust(datafix,kmeans,method="silhouette")

```



Gambar 6. Hasil Metode Elbow

Selanjutnya, untuk pertimbangan kedua. Peneliti mencoba menggunakan metode *silhouette* untuk mencari hasil *cluster* yang paling optimal, dan hasilnya sama seperti metode *elbow* tadi, dapat dilihat pada gambar 7.

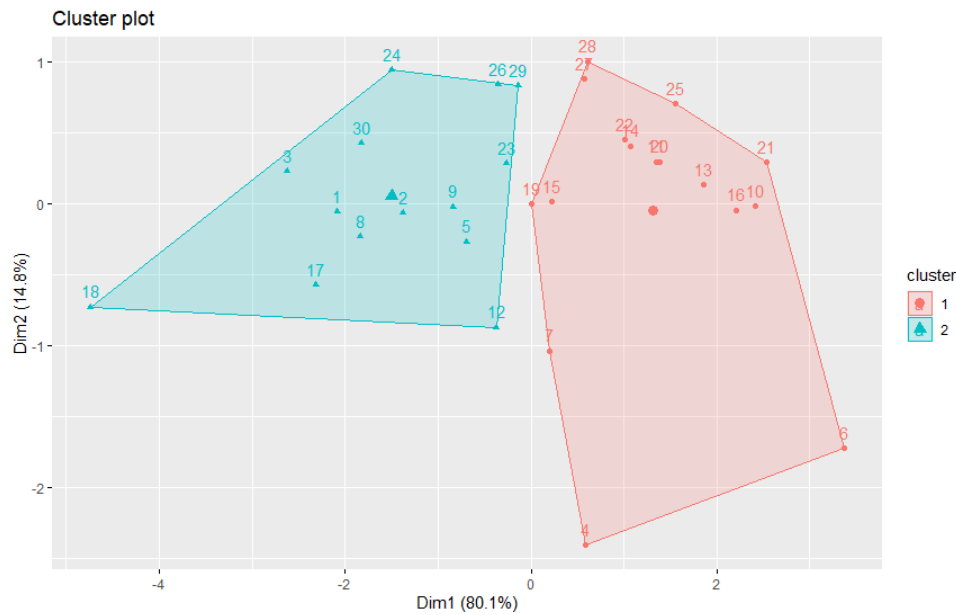


Gambar 7. Hasil Metode Silhouette

Kemudian setelah ditemukannya hasil *cluster* yang paling optimal menggunakan metode *elbow* & *silhouette* maka saatnya melakukan proses *plotting cluster*.

```
final=kmeans(datafix,2)
print(final)

fviz_cluster(final,data=datafix)
```



Gambar 8. Hasil Cluster Plotting

Setelah berhasil membagi data ke dalam 2 cluster, langkah selanjutnya adalah mengevaluasi dan menganalisis *cluster* untuk mengumpulkan dan menyimpulkan informasi. Berdasarkan hasil yang terdapat pada tabel 2.

Terlihat bahwa biaya bantuan keuangan yang dibutuhkan untuk pembangunan desa pada kecamatan yang terdapat pada *cluster* 1 lebih besar dari yang dibutuhkan oleh *cluster* 2.

Tabel 2. Hasil Pembagian Cluster

CLUSTER 1	CLUSTER 2
TELUKJAMBE TIMUR	PANGKALAN
KLARI	TEGALWARU
CIKAMPEK	CIAMPEL
JATISARI	TELUKJAMBE BARAT
BANYUSARI	TIRTAMULYA
CILAMAYA WETAN	PURWASARI
CILAMAYA KULON	KOTABARU
LEMAHABANG	MAJALAYA
TELAGASARI	KARAWANG TIMUR
KARAWANG BARAT	RENGASDENGKLOK
RAWAMERTA	JAYAKARTA
TEMPURAN	CILEBAR
KUTAWALUYA	BATUJAYA

PEDES	PAKISJAYA
CIBUAYA	
TIRTAJAYA	

IV. KESIMPULAN

K-Mean Clustering adalah metode pengelompokan data yang digunakan untuk mengidentifikasi pola dan mengelompokkan data berdasarkan kesamaan. Dalam konteks penerapan pada bantuan biaya keuangan pembangunan desa, metode ini dapat digunakan untuk mengelompokkan desa berdasarkan tingkat kebutuhan bantuan keuangan. Dengan menganalisis data yang tersedia, metode ini dapat digunakan untuk menentukan jumlah *cluster* yang optimal dan mengelompokkan desa ke dalam *cluster-cluster* yang sesuai. Hal ini dapat membantu dalam pengambilan keputusan dalam

pemberian bantuan keuangan dan memastikan bahwa dana yang tersedia diterapkan dengan efektif dan efisien. Sehingga nantinya jika ada permintaan dana dengan jumlah jauh lebih besar dari nilai clusternya, perlu dipertanyakan apa penyebabnya sebelum diturunkannya dana bantuan.

DAFTAR PUSTAKA

- Huang Z, "Extensions to the k-means algorithm for clustering large data sets with categorical values," *Data Mining and Knowledge Discovery*, Vol.2, pp:283–304, 1998.
- Sun Shibao, Qin Keyun,"Research on Modified k-means Data Cluster Algorithm"I. S. Jacobs and C. P. Bean, "Fine particles, thin films and exchange anisotropy," *Computer Engineering*, vol.33, No.13, pp.200– 201,July 2007.
- Sun Jigui, Liu Jie, Zhao Lianyu, "Clustering algorithms Research",*Journal of Software* ,Vol 19,No 1, pp.48-61,January 2008.
- Website Open Data Karawang, Dinas Pemberdayaan Masyarakat Dan Desa. <https://opendata.karawangkab.go.id/data-sets/dpmd/rekapitulasi-bantuan-keuangan-untuk-pembangunan-desamenurut-jenisnya-per-kecamatan-di-kabupaten-karawang-1638833298> diakses pada 22 Januari, 2023.
- Malik, D., & Rosalina, V. (2019). *TRADISIONAL BERBASIS ANDROID MENGGUNAKAN METODE*. 6(1), 12–19.
- Rianto, D. A., Assegaf, S., & Fernando, E. (2017). Perancangan Aplikasi Sistem Informasi Geografis (Sig) Lokasi Minimarket Di Kota Jambi Berbasis Android. *Jurnal Ilmiah Media SISFO*, 9(2), 295–304.
- Simargolang, M. Y., & Nasution, N. (2018). Aplikasi Pelayanan Jasa Laundry Berbasis WEB (Studi Kasus: Pelangi Laundry Kisaran). *Jurnal Teknologi Informasi*, 2(1), 9. <https://doi.org/10.36294/jurti.v2i1.402>
- Soraya, A., & Wahyudi, A. D. (2021). Rancang Bangun Aplikasi Penjualan Dimsum Berbasis Web (Studi Kasus: Kedai Dimsum Soraya). *Jurnal Teknologi Dan Sistem Informasi (JTISI)*, 2(4), 43–48. <http://jim.teknokrat.ac.id/index.php/JTISI>
- Taufik, A., & Ermawati. (2017). Perancangan Sistem Informasi Pemesanan Pentas Seni Berbasis Web Pada Sanggar Seni Getar Pakuan Bogor. *IJSE - Indonesian Journal on Software Engineering*, 3(2), 1–7. <http://ejournal.bsi.ac.id/ejurnal/index.php/ijse/article/view/2812/1836>