

## Penerapan Algoritma K-Means Untuk Clustering Big Five Personality

**Patrick Lay**

Matana University

**Ary Budi Warsito**

Matana University

Jl. CBD Barat Kav, Rw 1, Curug Sangereng, Kelapa Dua, Tangerang Regency, Banten 15810

Korespondensi penulis: [patrick.lay@student.matanauniversity.ac.id](mailto:patrick.lay@student.matanauniversity.ac.id)

**Abstract.** Personality is the combination of characteristics or qualities that form an individual's distinctive character. Every individual on earth has different personalities depending on their biological and environmental factors. To find out about one's personality, we can take a personality test which is a tool or method used to assess human personality. One of the most popular approaches to personality nowadays is the Big Five theory of personality, where this theory suggests that human personality is composed of five basic traits: Openness, Conscientiousness, Extroversion, Agreeableness and Neuroticism. While doing research, we found how some people tend not to have knowledge about their personality types. Therefore, we started this research with the aim of determining the user's personality, strengths and weaknesses of their personality in order to help them develop and improve themselves. This research used the PCA method and K-Means algorithm to aid in the clustering process.

**Keyword :** big five personality, clustering, k-means, pca

**Abstrak.** Kepribadian adalah kombinasi dari karakteristik dan kualitas yang membentuk karakter seorang individu. Setiap individu yang ada di bumi memiliki kepribadian yang berbeda tergantung pada faktor biologis dan lingkungan. Untuk mencari tahu tentang kepribadian seseorang, dapat dilakukan sebuah tes kepribadian yang merupakan alat atau metode yang digunakan untuk menilai kepribadian manusia. Salah satu pendekatan kepribadian yang paling populer saat ini adalah teori Big Five, dimana teori ini mengatakan bahwa kepribadian manusia terdiri dari lima sifat utama, diantaranya: Terbuka terhadap hal-hal baru, sifat berhati-hati, ekstraversi, mudah setuju atau sepakat dan neurotisme. Saat melakukan penelitian ini, kami menemukan bahwa beberapa orang cenderung untuk tidak memiliki pengetahuan mengenai jenis kepribadian yang dimilikinya. Untuk itu penulis memulai penelitian ini dengan tujuan untuk menentukan kepribadian pengguna, dimana nantinya penelitian ini diharapkan dapat membantu mereka memahami kepribadian, kelebihan serta kelemahan dari kepribadian mereka supaya dapat membantu dalam mengembangkan dan memperbaiki dirinya. Penelitian ini menggunakan metode PCA dan algoritma *K-Means* untuk membantu proses klasterisasinya.

**Kata kunci :** big five personality, klasterisasi, k-means, pca

### PENDAHULUAN

Kepribadian adalah keseluruhan sikap, perasaan, ekspresi dan tempramen, ciri-ciri khas dan perilaku seseorang. Sikap, perasaan, ekspresi dan tempramen itu akan terwujud dalam tindakan seseorang jika dihadapkan pada situasi tertentu. Kepribadian menjadi satu hal yang sangat berpengaruh dalam keberhasilan dan kegagalan seseorang dalam kehidupan, baik dalam bidang pekerjaan, maupun tindakannya dalam lingkungan masyarakat (Yuliasari & Hermawan, 2017).

Setiap individu di muka bumi ini memiliki kepribadian yang beragam, dan untuk mengetahui kepribadian itu dibutuhkanlah sebuah alat berupa tes kepribadian. Tes kepribadian merupakan tes yang dipergunakan untuk mengetahui perbedaan diantara setiap kepribadian, dan kepribadian itu sifatnya individual, yang artinya tidak seorang pun mempunyai kepribadian yang sama antara satu dengan yang lain (Cronbach, 1970).

Belakangan ini tes kepribadian dinilai sangat penting dan menjadi salah satu syarat untuk melamar pekerjaan. Untuk itu banyak tes kepribadian telah dikembangkan, diantaranya tes projektif yang mengukur kepribadian seseorang dengan menggunakan grafis maupun tes objektif yang mengandalkan kejujuran pengisinya yang terdiri atas beberapa pertanyaan atau pernyataan (Mudrika, 2011).

Salah satu tes kepribadian objektif yang dinilai akurat adalah Big Five Personality Test. Tes kepribadian ini merupakan tes kepribadian yang mengukur 5 (lima) buah kepribadian utama yang dimiliki oleh setiap manusianya, antara lain: Openness, Conscientiousness, Extroversion, Agreeableness dan Neuroticism (OCEAN).

Namun meskipun tes kepribadian dianggap penting, ada juga sebagian masyarakat yang masih belum memahami tes kepribadian itu sendiri, dan bahkan kepribadian yang dimilikinya. Maka dari itu penulis memulai penelitian ini dengan tujuan untuk menentukan kepribadian pengguna, dimana nantinya penelitian ini diharapkan dapat membantu mereka memahami kepribadian, kelebihan serta kelemahan dari kepribadian mereka supaya dapat membantu dalam mengembangkan dan memperbaiki dirinya.

Penelitian ini menggunakan metode PCA dan algoritma K-Means untuk membantu proses klusterisasinya

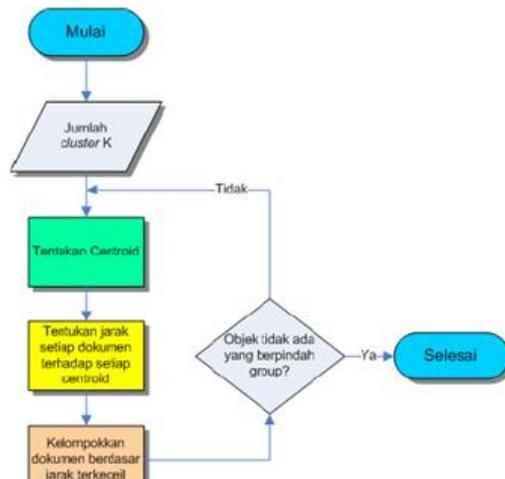
## **METODE PENELITIAN**

Pada penelitian ini data yang digunakan berupa file CSV yang memuat hasil kuesioner dari 1.015.342 orang responden. Data ini diambil dari situs Kaggle (<https://www.kaggle.com/datasets/tunguz/big-five-personality-test>).

Penelitian ini menggunakan jenis algoritma machine learning *Unsupervised Learning* dengan metode Klusterisasi (*Clustering*). Clustering adalah proses partisi satu set objek data ke dalam sebuah himpunan bagian yang disebut *cluster*. Di dalam satu *cluster*, setiap objek memiliki karakteristik yang mirip satu sama lain dan berbeda dengan *cluster* yang lain (Irwansyah, 2017).

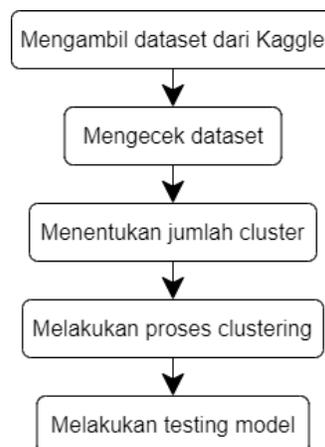
Metode yang digunakan dalam proses klusterisasi adalah *Principal Component Analysis* (PCA), yakni salah satu metode pada machine learning yang berfungsi untuk menyimpan informasi dan mereduksi dimensi data. Dimensi-dimensi tersebut direduksi dengan membentuk variabel-variabel baru yang disebut *Principal Components*, yang merupakan kombinasi linier dari variabel-variabel lama (Ramadhan, 2022).

Algoritma yang digunakan dalam penelitian ini adalah algoritma K-Means yang berfungsi untuk mempartisi data yang ada ke dalam *cluster*, sehingga data yang memiliki karakteristik sama akan dikelompokkan ke dalam sebuah *cluster* yang sama (Merliana, 2015).



**Gambar 1.** Flowchart K-Means (Siswanto, 2018)

Penelitian dilakukan dengan menggunakan tools Google Colaboratory dengan bahasa pemrograman Python.



**Gambar 2.** Diagram alir penelitian

Alur pengerjaan penelitian ini dapat dilihat pada Gambar 2. Penelitian akan dimulai dengan mengambil dataset dari Kaggle, mengecek dataset, menentukan jumlah cluster, melakukan proses clustering, dan melakukan testing model.

## HASIL DAN PEMBAHASAN

Dalam penelitian ini, terdapat lima tahap yang penulis lakukan untuk membantu pengguna dalam menentukan kepribadiannya.

Tahap pertama adalah mengambil dataset dari Kaggle, dimana pada penelitian ini kami mengambil dataset yang terdiri dari 1.015.342 jawaban responden terhadap kuesioner yang berhubungan dengan *Big Five Personality Test* ini.

Tahap kedua adalah mengecek dataset yang dapat dilihat pada Gambar 3, dimana pada tahapan ini kami mengecek kembali data yang digunakan, mencari data yang memiliki *missing value*, kemudian mengeliminasi.

```
print('Is there any missing value? ', data.isnull().values.any())
print('How many missing values? ', data.isnull().values.sum())
data.dropna(inplace=True)
print('Number of participants after eliminating missing values: ', len(data))

Is there any missing value? True
How many missing values? 89227
Number of participants after eliminating missing values: 1013401
```

**Gambar 3.** Mengecek dataset

Tahap ketiga adalah menentukan jumlah cluster, dimana pada penelitian ini penulis menggunakan 5 buah cluster untuk mengelompokkan 5 buah kepribadian yang ada pada *Big Five Personality*, antara lain: *Extroversion* (Ekstrovert), *Neurotic* (Cenderung memiliki emosi negatif), *Agreeable* (Mudah setuju atau sepakat), *Conscientious* (Sifat berhati-hati), dan *Openness* (Terbuka terhadap hal-hal baru).

```
# I define 5 clusters and fit my model
kmeans = KMeans(n_clusters=5)
k_fit = kmeans.fit(df_model)
```

**Gambar 4.** Menentukan jumlah cluster

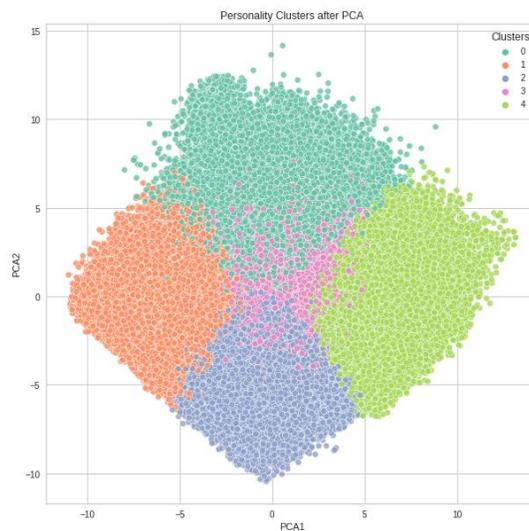
Tahap keempat adalah melakukan proses *clustering*, dimana pada Gambar 5 ditunjukkan jumlah dari setiap cluster. Cluster 0 menunjukkan cluster *Extroversion*, cluster 1 menunjukkan cluster *Neurotic*, cluster 2 menunjukkan cluster *Agreeable*, cluster 3 menunjukkan cluster *Conscientious*, dan cluster 4 menunjukkan cluster *Openness*.

```
[ ] df_model.Clusters.value_counts()

2    226854
3    213104
1    209167
4    201269
0    163087
Name: Clusters, dtype: int64
```

**Gambar 5.** Proses *clustering*

Di tahap proses *clustering* ini, kami juga menggunakan PCA untuk membantu memvisualisasikan grafik *clustering* yang dapat dilihat pada Gambar 6.



**Gambar 6.** Visualisasi grafik *clustering*

Tahap kelima adalah melakukan testing model, dimana penulis memasukkan sampel data yang ingin diketahui jenis kepribadiannya pada model yang telah dibuat seperti pada Gambar 7. Berdasarkan data yang dimasukkan, model akan menampilkan jenis kepribadian yang sesuai dengan klusternya. Gambar 8 menunjukkan bahwa hasil yang dikeluarkan oleh model sudah akurat dan Gambar 9 menunjukkan hasil *testing* dalam bentuk grafik.

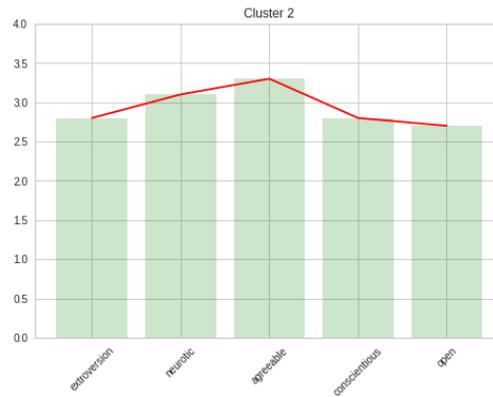
```
my_data = pd.read_excel('/content/drive/My Drive/UAS-ML/IPIP-FFH-data-8Nov2018/my-personality.xlsx')
my_data

   ext1  ext2  ext3  ext4  ext5  ext6  ext7  ext8  ext9  ext10  est1  est2  est3  est4  est5  est6  est7
0    1.0  2.0  3.0  1.0  2.0  3.0  4.0  5.0  4.0  3.0  2.0  3.0  4.0  5.0  1.0  2.0  3.0
Warning: Total number of columns (58) exceeds max_columns (20) limiting to first (20) columns.
```

**Gambar 7.** Sampel data untuk testing

| Sum of my question groups |              |          |           |               |      |
|---------------------------|--------------|----------|-----------|---------------|------|
|                           | extroversion | neurotic | agreeable | conscientious | open |
| 0                         | 2.8          | 3.1      | 3.3       | 2.8           | 2.7  |

**Gambar 8.** Hasil testing



**Gambar 9.** Visualisasi hasil testing

## KESIMPULAN

Penerapan algoritma K-Means untuk *clustering Big Five Personality* dapat memberikan hasil yang baik dalam mengelompokkan objek dan menentukan kepribadian pengguna.

Kedepannya implementasi ini diharapkan dapat melakukan testing pada sampel data yang lebih banyak untuk di tes tingkat keakuratannya, sehingga dapat mendapatkan hasil kepribadian yang dapat membantu pengguna dengan lebih akurat.

## DAFTAR REFERENSI

- Cronbach, Lee J. (1970). *Essential of Psychological Testing*. New York: Harper and Row.
- Irwansyah, E. (2017). CLUSTERING. 17 Desember 2022, dari <https://socs.binus.ac.id/2017/03/09/clustering>.
- Maryanto. (2021). Analisis Clustering Menggunakan PCA & K-Means. Diakses pada 17 Desember 2022, dari <https://rpubs.com/mere1803/719475>.
- Merliana, N. P. E. (2015). Perbandingan Metode K-Means Dengan Fuzzy C-Means Untuk Analisa Karakteristik Mahasiswa Berdasarkan Kunjungan Ke Perpustakaan (Studi Kasus Sekolah Tinggi Agama Hindu Negeri Tampung Penyang Palangka Raya). S2 thesis, UAJY. Diakses pada 17 Desember 2022, dari <http://e-journal.uajy.ac.id/id/eprint/8035>.
- Mudrika, N. (2011). Membaca Kepribadian Menggunakan Tes MBTI (Myer Briggs Type Indicator). Diakses pada 17 Desember 2022, dari <https://nafismudrika.wordpress.com/2011/02/18/membaca-kepribadian-menggunakan-tes-mbti-myer-briggs-type-indicator>.
- Ramadhan, Ashari. (2022). Analisis PCA: Sederhanakan Data dengan Reduksi Dimensi Menggunakan R. Diakses pada 17 Desember 2022, dari <https://dqlab.id/analisis-pca-sederhanakan-data-dengan-reduksi-dimensi-menggunakan-r>.
- Siswanto. (2018). Analisis Cluster dengan Menggunakan metode K-Means dan K-Medoids. Diakses pada 17 Desember 2022, dari <https://swanstatistics.com/analisis-cluster-dengan-menggunakan-metode-k-means-dan-k-medoids/>.
- Tunguz, B. (2018). Big Five Personality Test. Diakses pada 17 Desember 2022, dari <https://www.kaggle.com/datasets/tunguz/big-five-personality-test>.
- Yuliasari, N., & Hermawan, H. (2017). Rancang Bangun Aplikasi Tes Kepribadian Berbasis Web Menggunakan Metode MBTI. *Jurnal Informatika dan Multimedia*, 9(2), 41-48. Diakses pada 17 Desember 2022, dari <https://docplayer.info/116850655-Rancang-bangun-aplikasi-tes-kepribadian-berbasis-web-menggunakan-metode-mbti.html>.