

**CLUSTERING PENGGUNA UNTUK REKOMENDASI  
PENGEMBANGAN FITUR PADA APLIKASI CODR  
MENGUNAKAN METODE K-MEANS**

**SKRIPSI**

**Oleh :**

**Agung Wahyu Prayogo**

**201910225300**



-

**PROGRAM STUDI INFORMATIKA  
FAKULTAS ILMU KOMPUTER  
UNIVERSITAS BHAYANGKARA JAKARTA RAYA  
2023**

## LEMBAR PERSETUJUAN PEMBIMBING

Judul Tugas Akhir : *Clustering* Pengguna Untuk  
Rekomendasi Pengembangan Fitur Pada  
Aplikasi Codr Menggunakan Metode K-  
Means

Nama Mahasiswa : Agung Wahyu Prayogo

Nomor Pokok Mahasiswa : 201910225300

Program Studi / Fakultas : Informatika / Ilmu Komputer

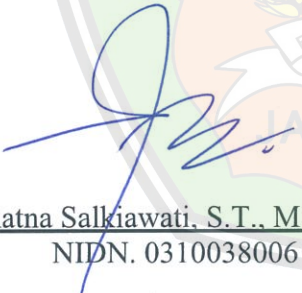
Tanggal Lulus Ujian Tugas Akhir : 15 Juli 2023

Bekasi, 31 Juli 2023

Menyetujui,

Pembimbing I

Pembimbing II

  
(Ratna Salkiawati, S.T., M.Kom)  
NIDN. 0310038006

  
(Khairunnisa Fadhilla Ramdhan, S.Si., M.Si)  
NIDN. 0328039201

## LEMBAR PENGESAHAN

Judul Tugas Akhir : *Clustering Pengguna Untuk Rekomendasi Pengembangan Fitur Pada Aplikasi Codr Menggunakan Metode K-Means*

Nama Mahasiswa : Agung Wahyu Prayogo

Nomor Pokok Mahasiswa : 201910225300

Program Studi/ Fakultas : Informatika/ Ilmu Komputer

Tanggal Lulus Ujian Tugas Akhir : 15 Juli 2023

Bekasi, 31 Juli 2023

Mengesahkan,

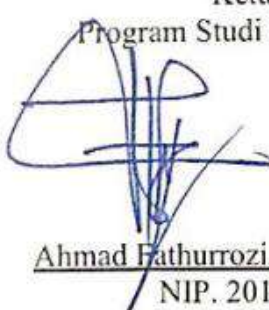
Ketua Tim Penguji : Wowon Priatna, S.T., M.Ti  
NIDN : 0429118007

Penguji I : Rafika Sari, S.Si., M.Si  
NIDN : 0329098902

Penguji II : Ratna Salkiawati, S.T., M.Kom  
NIDN : 0310038006

Mengetahui,

Ketua  
Program Studi Informatika



Ahmad Fathurrozi, S.E., M.M.S.I  
NIP. 2012486

Dekan  
Fakultas Ilmu Komputer



Dr. Dra. Tyastuti Sri Lestari, M.M.  
NIP. 1408206



UNIVERSITAS BHAYANGKARA JAKARTA RAYA  
FAKULTAS ILMU KOMPUTER  
PROGRAM STUDI INFORMATIKA

## LEMBAR PERNYATAAN BUKAN PLAGIASI

Yang bertanda tangan dibawah ini :

Nama : Agung Wahyu Prayogo  
NPM : 201910225300  
Program Studi : Informatika  
Fakultas : Ilmu Komputer  
Judul Tugas Akhir : *Clustering* Pengguna Untuk Rekomendasi  
Pengembangan Fitur Pada Aplikasi Codr  
Menggunakan Metode K-Means

Dengan ini menyatakan bahwa hasil penulisan skripsi yang telah saya buat ini merupakan **hasil karya saya sendiri dan benar keasliannya**. Apabila dikemudian hari penulisan skripsi ini merupakan plagiat atau penjiplakan terhadap karya orang lain, maka saya bersedia mempertanggungjawabkan sekaligus bersedia menerima sanksi berdasarkan tata tertib di Universitas Bhayangkara Jakarta Raya.

Demikian pernyataan ini saya buat dalam keadaan sadar dan tidak dipaksakan dari pihak manapun.

Bekasi, 31 Juli 2023

Penulis



Agung Wahyu Prayogo

## ABSTRAK

**Agung Wahyu Prayogo, 201910225300.** *Clustering Pengguna Untuk Rekomendasi Pengembangan Fitur Pada Aplikasi Codr Menggunakan Metode K-Means.* Universitas Bhayangkara Jakarta Raya. 2023

Penelitian ini bertujuan untuk merekomendasikan fitur yang perlu dikembangkan pada aplikasi Codr, sebuah aplikasi untuk pembelajaran algoritma dan pemrograman berbasis iOS dengan menggunakan metode k-means berdasarkan pola perilaku pengguna dalam menggunakan aplikasi tersebut. Data yang digunakan dalam penelitian ini adalah *UserFlow.csv*, yang merekam sejauh mana pengguna dalam menggunakan aplikasi.

Metode yang digunakan dalam pengolahan data adalah CRISP-DM (*cross-industry standard process for data mining*), yang terdiri dari beberapa tahap yaitu *Business Understanding*, *Data Understanding*, *Data Preparation*, *Modelling*, *Evaluation*, dan *Result*. Metode k-means digunakan untuk mengelompokkan pola perilaku pengguna berdasarkan interaksi mereka dengan aplikasi Codr. Penggunaan k-means memerlukan penentuan jumlah *cluster* (*k*) *elbow method* yang digunakan untuk menentukan 'k' yang optimal. Selain itu, evaluasi kualitas *cluster* dilakukan menggunakan *Davies-Bouldin Index*.

Hasil dari penelitian ini diharapkan dapat memberikan rekomendasi kepada tim pengembangan aplikasi Codr berdasarkan pola perilaku pengguna dan aplikasi dapat memberikan pembelajaran pemrograman yang lebih menarik sesuai dengan kebutuhan pengguna.

**Kata Kunci:** Rekomendasi Aplikasi, K-Means, Pola Perilaku Pengguna, CRISP-DM, *Elbow Method*, *Davies-Bouldin Index*.



## LEMBAR PERNYATAAN PERSETUJUAN PUBLIKASI KARYA ILMIAH UNTUK KEPENTINGAN AKADEMIK

Sebagai sivitas akademik Universitas Bhayangkara Jakarta Raya, saya yang bertanda tangan di bawah ini :

Nama : Agung Wahyu Prayogo  
NPM : 201910225300  
Program Studi : Informatika  
Fakultas : Ilmu Komputer  
Jenis Karya : Skripsi

Demi pengembangan ilmu pengetahuan, menyetujui untuk memberikan kepada Universitas Bhayangkara Jakarta Raya **Hak Bebas Royalti Non-Eksklusif (*Non-Exclusive Royalty-Free Right*)**, atas karya ilmiah saya yang berjudul :

**“Clustering Pengguna Untuk Rekomendasi Pengembangan Fitur Pada Aplikasi Codr Menggunakan Metode K-Means”** beserta perangkat yang ada (bila diperlukan). Dengan hak bebas royalti non-eksklusif ini, Universitas Bhayangkara Jakarta Raya berhak menyimpan, mengalihmediakan, mengelolanya dalam bentuk pangkalan data (*database*), mendistribusikannya dan mempublikasikannya di Internet atau media lain untuk kepentingan akademis tanpa perlu meminta ijin dari saya selama tetap mencantumkan nama saya sebagai penulis dan sebagai pemilik hak cipta.

Segala bentuk tuntutan hukum yang timbul atas pelanggaran hak cipta dalam karya ilmiah ini menjadi tanggung jawab saya pribadi

Demikian pernyataan ini saya buat dengan sebenarnya.

Dibuat di : Bekasi  
Pada tanggal : 30 Juli 2023  
Yang Menyatakan



Agung Wahyu Prayogo

## KATA PENGANTAR

Puji syukur kepada Allah SWT berkat Rahmat, Hidayah, dan Karunia-Nya kepada kita semua sehingga peneliti dapat menyelesaikan Skripsi yang disusun sebagai salah satu syarat untuk mencapai gelar sarjana Strata Satu (S1) yang berjudul “*Clustering* Pengguna Untuk Rekomendasi Pengembangan Fitur Pada Aplikasi Codr Menggunakan Metode K-Means”. Dalam penyusunan Skripsi ini peneliti mendapat banyak sekali bantuan, dukungan, dan bimbingan. Oleh karena itu, dalam kesempatan ini ingin menyampaikan terima kasih kepada :

1. Bapak Irjen Polisi (Purn) Prof. Dr. Drs. Bambang Karsono, S.H., M.M. Selaku Rektor Universitas Bhayangkara Jakarta Raya.
2. Ibu Dr. Dra. Tyastuti Sri Lestari, M.M. Selaku Dekan Fakultas Ilmu Komputer Universitas Bhayangkara Jakarta Raya.
3. Bapak Ahmad Fathurrozi, S.E., M.M.S.I. Selaku kepala Program Studi Informatika Universitas Bhayangkara Jakarta Raya.
4. Ibu Ratna Salkiawati, S.T., M.Kom. Selaku pembimbing I yang telah memberikan arahan serta motivasi yang mudah dipahami saat sesi bimbingan.
5. Mrs Khairunnisa Fadhillah Ramdhanisa, S.Si., M.Si. Selaku pembimbing II yang telah pengertian terhadap waktu bimbingan.
6. Kedua orang tua saya. Yang sudah memberikan saya semangat dan untuk tidak mudah menyerah selama proses kegiatan berlangsung.

Bekasi, 31 Juli 2023

Peneliti,



Agung Wahyu Prayogo

201910225300

## DAFTAR ISI

|                                                      |             |
|------------------------------------------------------|-------------|
| <b>LEMBAR PERSETUJUAN PEMBIMBING .....</b>           | <b>i</b>    |
| <b>LEMBAR PENGESAHAN .....</b>                       | <b>ii</b>   |
| <b>LEMBAR PERNYATAAN BUKAN PLAGIASI .....</b>        | <b>iii</b>  |
| <b>ABSTRAK .....</b>                                 | <b>iv</b>   |
| <b>ABSTRACT .....</b>                                | <b>v</b>    |
| <b>LEMBAR PERNYATAAN PERSETUJUAN PUBLIKASI .....</b> | <b>vi</b>   |
| <b>KATA PENGANTAR.....</b>                           | <b>vii</b>  |
| <b>DAFTAR ISI.....</b>                               | <b>viii</b> |
| <b>DAFTAR GAMBAR.....</b>                            | <b>xi</b>   |
| <b>DAFTAR TABEL .....</b>                            | <b>xiv</b>  |
| <b>BAB I PENDAHULUAN.....</b>                        | <b>1</b>    |
| 1.1 Latar Belakang .....                             | 1           |
| 1.2 Identifikasi Masalah .....                       | 4           |
| 1.3 Rumusan Masalah .....                            | 5           |
| 1.4 Batasan Masalah.....                             | 5           |
| 1.5 Tujuan Penelitian.....                           | 5           |
| 1.6 Manfaat Penelitian.....                          | 6           |
| 1.7 Sistematika Tugas Akhir .....                    | 6           |
| <b>BAB II TINJAUAN PUSTAKA.....</b>                  | <b>8</b>    |
| 2.1 Kajian Literatur .....                           | 8           |
| 2.2 Algoritma dan Pemrograman .....                  | 9           |
| 2.3 Google Colab.....                                | 10          |
| 2.4 Python.....                                      | 10          |
| 2.5 K-Means .....                                    | 11          |



|                |                                                                                |           |
|----------------|--------------------------------------------------------------------------------|-----------|
| 2.6            | <i>Clustering</i> .....                                                        | 14        |
| 2.7            | <i>Elbow Method</i> .....                                                      | 15        |
| 2.8            | <i>Davies-Bouldin Index</i> .....                                              | 15        |
| 2.9            | <i>Flowchart</i> .....                                                         | 16        |
| 2.10           | <i>CRISP-DM</i> .....                                                          | 18        |
| <b>BAB III</b> | .....                                                                          | <b>22</b> |
| 3.1            | Objek Penelitian .....                                                         | 22        |
| 3.2            | Kerangka Penelitian .....                                                      | 23        |
| 3.3            | Metode Pengumpulan Data .....                                                  | 25        |
| 3.3.1          | <i>App Store Connect</i> .....                                                 | 25        |
| 3.3.2          | <i>Mixpanel</i> .....                                                          | 25        |
| 3.4            | Perbedaan <i>Elbow Method</i> , DBI, dan K-Means dalam <i>Clustering</i> ..... | 25        |
| <b>BAB IV</b>  | .....                                                                          | <b>27</b> |
| 4.1            | Hasil Penelitian.....                                                          | 27        |
| 4.2            | Pengumpulan Data .....                                                         | 27        |
| 4.3            | <i>Business Understanding</i> .....                                            | 28        |
| 4.4            | <i>Data Understanding</i> .....                                                | 29        |
| 4.5            | <i>Exploratory Data Analysis (EDA)</i> .....                                   | 32        |
| 4.6            | <i>Data Preparation</i> .....                                                  | 34        |
| 4.7            | <i>Feature Engineering</i> .....                                               | 35        |
| 4.8            | <i>Preprocessing Modelling</i> .....                                           | 36        |
| 4.8.1          | <i>Elbow Method</i> .....                                                      | 37        |
| 4.8.2          | <i>Davies-Bouldin Score</i> .....                                              | 43        |
| 4.8.3          | Menghitung Pusat <i>Centroid</i> .....                                         | 47        |
| 4.9            | <i>Modelling</i> .....                                                         | 48        |
| 4.9.1          | <i>Cluster 0</i> .....                                                         | 53        |

|                       |                             |           |
|-----------------------|-----------------------------|-----------|
| 4.9.2                 | <i>Cluster 1</i> .....      | 54        |
| 4.9.3                 | <i>Cluster 2</i> .....      | 57        |
| 4.9.4                 | <i>Cluster 3</i> .....      | 59        |
| 4.9.5                 | <i>Cluster 4</i> .....      | 61        |
| 4.10                  | Pembahasan Penelitian ..... | 63        |
| <b>BAB V</b>          | .....                       | <b>72</b> |
| 5.1                   | Kesimpulan.....             | 72        |
| 5.2                   | Saran.....                  | 72        |
| <b>DAFTAR PUSTAKA</b> | .....                       | <b>73</b> |
| <b>LAMPIRAN</b>       | .....                       | <b>74</b> |



## DAFTAR GAMBAR

|                                                                              |    |
|------------------------------------------------------------------------------|----|
| Gambar 1.1 Visualisasi Data <i>First Time Downloads</i> .....                | 2  |
| Gambar 1.2 Visualisasi Data <i>Installations</i> .....                       | 2  |
| Gambar 1. 3 Visualisasi Data <i>User Flow</i> .....                          | 3  |
| Gambar 2. 1 Cara Kerja Algoritma K-Means .....                               | 12 |
| Gambar 2. 2 Metode CRISP-DM.....                                             | 19 |
| Gambar 3. 1 Logo Aplikasi Codr.....                                          | 22 |
| Gambar 3. 2 Kerangka Penelitian .....                                        | 23 |
| Gambar 4. 1 Dashboard Mixpanel .....                                         | 27 |
| Gambar 4. 2 Alur Penggunaan Aplikasi Codr .....                              | 28 |
| Gambar 4. 3 Dataset <i>UserFlow.csv</i> .....                                | 30 |
| Gambar 4. 4 Grafik Pengguna Dalam Menggunakan Fitur di Codr .....            | 32 |
| Gambar 4. 6 Persentase Pengguna Dalam menggunakan Aplikasi (2) .....         | 34 |
| Gambar 4. 7 <i>Script</i> Menghapus Kolom <i>User ID</i> .....               | 35 |
| Gambar 4. 8 Menghapus kolom <i>User ID</i> .....                             | 35 |
| Gambar 4. 9 Mengubah <i>Data Frame</i> Menjadi <i>Array</i> .....            | 36 |
| Gambar 4. 10 Mencari Tahu Nilai 'K' terbaik dengan <i>Elbow Method</i> ..... | 37 |
| Gambar 4. 11 Visualisasi <i>Elbow Method</i> .....                           | 39 |
| Gambar 4. 12 Contoh 5 Data teratas untuk Mencari <i>Elbow Method</i> .....   | 40 |
| Gambar 4. 13 Contoh Pusat <i>Centroid</i> pada Python .....                  | 40 |
| Gambar 4. 14 <i>Script Davies-Bouldin Score</i> dan Skor yang Muncul .....   | 46 |
| Gambar 4. 15 Data Acak Sebagai Pusat <i>Centroid</i> .....                   | 47 |
| Gambar 4. 16 <i>Script K-Modes</i> .....                                     | 50 |
| Gambar 4. 17 Menambah Kolom <i>Cluster</i> .....                             | 51 |
| Gambar 4. 18 <i>Script</i> Melihat Total Pengguna Per <i>Cluster</i> .....   | 52 |
| Gambar 4. 19 Total Pengguna Per <i>Cluster</i> .....                         | 52 |

|                                                                                         |    |
|-----------------------------------------------------------------------------------------|----|
| Gambar 4. 20 Menampilkan Data Pengguna Yang Masuk <i>Cluster</i> 0 .....                | 53 |
| Gambar 4. 21 <i>Script</i> Menampilkan Karakter Pengguna di <i>Cluster</i> 0 .....      | 53 |
| Gambar 4. 22 Visualisasi Data <i>Cluster</i> 0 .....                                    | 54 |
| Gambar 4. 23 <i>Script</i> Menampilkan Pengguna di <i>Cluster</i> 1 .....               | 55 |
| Gambar 4. 24 Data Pengguna Yang Masuk Pada <i>Cluster</i> 1 .....                       | 55 |
| Gambar 4. 25 <i>Script</i> Melihat Berapa Banyak Pengguna di <i>Cluster</i> 1.....      | 55 |
| Gambar 4. 26 <i>Script</i> Menampilkan Karakter Pengguna di <i>Cluster</i> 1 .....      | 56 |
| Gambar 4. 27 Visualisasi Karakteristik Pengguna <i>Cluster</i> 1.....                   | 56 |
| Gambar 4. 28 <i>Script</i> Menampilkan Data Pengguna di <i>Cluster</i> 2.....           | 57 |
| Gambar 4. 29 Data Pengguna yang Masuk Pada <i>Cluster</i> 2 .....                       | 57 |
| Gambar 4. 30 <i>Script</i> Melihat Total Pengguna di <i>Cluster</i> 2 .....             | 57 |
| Gambar 4. 31 <i>Script</i> Melihat Karakteristik Pengguna di <i>Cluster</i> 2 .....     | 58 |
| Gambar 4. 32 Visualisasi Data Karakteristik Pengguna <i>Cluster</i> 2 .....             | 58 |
| Gambar 4. 33 <i>Script</i> Menampilkan Data <i>Cluster</i> 3.....                       | 59 |
| Gambar 4. 34 Data Pengguna Yang Masuk Pada <i>Cluster</i> 3 .....                       | 59 |
| Gambar 4. 35 <i>Script</i> Menampilkan Total Pengguna <i>Cluster</i> 3.....             | 59 |
| Gambar 4. 36 <i>Script</i> Menampilkan Karakter Pengguna <i>Cluster</i> 3 .....         | 60 |
| Gambar 4. 37 Visualisasi Data Pengguna <i>Cluster</i> 3 .....                           | 60 |
| Gambar 4. 38 <i>Script</i> Menampilkan Pengguna di <i>Cluster</i> 4 .....               | 61 |
| Gambar 4. 39 Pengguna Yang Masuk Pada <i>Cluster</i> 4 .....                            | 61 |
| Gambar 4. 40 <i>Script</i> Menampilkan Total Pengguna <i>Cluster</i> 4 .....            | 62 |
| Gambar 4. 41 <i>Script</i> Menampilkan Karakteristik Pengguna di <i>Cluster</i> 4 ..... | 62 |
| Gambar 4. 42 Karakteristik Pengguna di <i>Cluster</i> 4 .....                           | 62 |
| Gambar 4. 43 Pusat <i>Centroid</i> di Python .....                                      | 66 |
| Gambar 4. 44 Pusat <i>Centroid</i> Tiap <i>Cluster</i> Pada Perhitungan Excel .....     | 67 |
| Gambar 4. 45 Karakter Pengguna pada <i>Cluster</i> 0 .....                              | 68 |



|                                                                 |    |
|-----------------------------------------------------------------|----|
| Gambar 4. 46 Karakteristik Pengguna pada <i>Cluster</i> 1 ..... | 68 |
| Gambar 4. 47 Karakteristik Pengguna pada <i>Cluster</i> 2 ..... | 69 |
| Gambar 4. 48 Karakteristik Pengguna pada <i>Cluster</i> 3 ..... | 70 |
| Gambar 4. 48 Karakteristik Pengguna pada <i>Cluster</i> 4 ..... | 70 |



## DAFTAR TABEL

|                                                                                          |    |
|------------------------------------------------------------------------------------------|----|
| Tabel 2. 1 Tinjauan Pustaka .....                                                        | 8  |
| Tabel 2. 2 Simbol <i>Flowchart</i> dan Fungsinya .....                                   | 17 |
| Tabel 4. 1 Penjelasan Mengenai <i>Script Elbow Method</i> Dengan Python.....             | 38 |
| Tabel 4. 2 Mencari jarak terdekat antara data dengan pusat <i>centroid</i> .....         | 41 |
| Tabel 4. 3 Mengkuadrat nilai minimum data ke-1 hingga ke-n .....                         | 42 |
| Tabel 4. 4 Menjumlahkan semua hasil kuadrat.....                                         | 42 |
| Tabel 4. 5 Mencari SSW dari setiap data .....                                            | 44 |
| Tabel 4. 6 Menghitung SSB ketika $K=2$ .....                                             | 45 |
| Tabel 4. 7 Matrix SSB $K = 2$ .....                                                      | 45 |
| Tabel 4. 8 Matrix Rasio <i>Cluster 1</i> dan <i>Cluster 2</i> .....                      | 46 |
| Tabel 4. 9 <i>Centroid cluster 0</i> hingga <i>cluster 4</i> setelah iterasi kedua ..... | 47 |
| Tabel 4. 10 Data teratas dari dataset <i>UserFlow.csv</i> .....                          | 48 |
| Tabel 4. 11 <i>Centroid Cluster 0</i> hingga <i>Cluster 4</i> .....                      | 49 |
| Tabel 4. 12 Hasil dari data ke-0 dikali <i>cluster 0</i> hingga 4 .....                  | 50 |
| Tabel 4. 13 <i>Clustering</i> iterasi ke-2 .....                                         | 50 |
| Tabel 4. 14 <i>Clustering</i> Iterasi ke-2.....                                          | 51 |
| Tabel 4. 15 Total Jumlah Data yang Masuk Tiap <i>Cluster</i> .....                       | 52 |
| Tabel 4. 16 Tabel Penggunaan Fitur dan Penurunan .....                                   | 63 |
| Tabel 4. 17 Skor <i>Davies Boulding Index</i> .....                                      | 64 |
| Tabel 4. 18 Hasil <i>Clustering</i> .....                                                | 65 |
| Tabel 4. 19 Perbandingan perhitungan Python dan Excel .....                              | 65 |
| Tabel 4. 20 Perbandingan perhitungan DBI Python dan Excel .....                          | 66 |
| Tabel 4. 21 Perbandingan Perhitungan Python dan Excel .....                              | 67 |