

**PENERAPAN *DATA MINING* METODE DESKRIPSI UNTUK
MENGETAHUI PERBEDAAN PENGGUNA KASUAL DENGAN *ANNUAL*
MEMBER LAYANAN *BIKE-SHARING CYCLISTIC***

SKRIPSI

Diajukan untuk memenuhi salah satu syarat mencapai derajat Sarjana Program Studi
Informatika



disusun oleh
BINTANG BUANA RAJASA
19.11.3118

Kepada

FAKULTAS ILMU KOMPUTER
UNIVERSITAS AMIKOM YOGYAKARTA
2023

**PENERAPAN *DATA MINING* METODE DESKRIPSI UNTUK
MENGETAHUI PERBEDAAN PENGGUNA KASUAL DENGAN *ANNUAL*
MEMBER LAYANAN *BIKE-SHARING CYCLISTIC***

SKRIPSI

Diajukan untuk memenuhi salah satu syarat mencapai derajat Sarjana Program Studi
Informatika



disusun oleh

BINTANG BUANA RAJASA

19.11.3118

Kepada

**FAKULTAS ILMU KOMPUTER
UNIVERSITAS AMIKOM YOGYAKARTA
2023**

HALAMAN PERSETUJUAN

SKRIPSI

**PENERAPAN *DATA MINING* METODE DESKRIPSI UNTUK
MENGETAHUI PERBEDAAN PENGGUNA KASUAL DENGAN *ANNUAL*
*MEMBER LAYANAN BIKE-SHARING CYCLISTIC***

yang disusun dan diajukan oleh

Bintang Buana Rajasa
19.11.3118

telah disetujui oleh Dosen Pembimbing Skripsi
pada tanggal 25 mei 2023

Dosen Pembimbing,



Kusnawi, S.Kom, M. Eng.
NIK. 190302112

HALAMAN PENGESAHAN

SKRIPSI

**PENERAPAN *DATA MINING* METODE DESKRIPSI UNTUK
MENGETAHUI PERBEDAAN PENGGUNA KASUAL DENGAN *ANNUAL*
*MEMBER LAYANAN BIKE-SHARING CYCLISTIC***

yang disusun dan diajukan oleh

Bintang Buana Rajasa
19.11.3118

Telah dipertahankan di depan Dewan Penguji
pada tanggal 25 mei 2023

Susunan Dewan Penguji

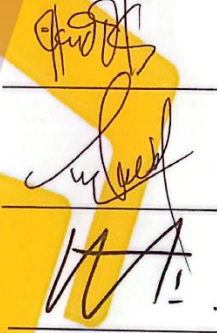
Nama Penguji

Anna Baita, M.Kom
NIK. 190302290

Subektiningsih, M.Kom
NIK. 190302413

Kusnawi, S.Kom, M. Eng.
NIK. 190302112

Tanda Tangan



Skripsi ini telah diterima sebagai salah satu persyaratan
untuk memperoleh gelar Sarjana Komputer
Tanggal 25 mei 2023

DEKAN FAKULTAS ILMU KOMPUTER



Hanif Al Fatta, S.Kom., M.Kom.
NIK. 190302096

HALAMAN PERNYATAAN KEASLIAN SKRIPSI

Yang bertandatangan di bawah ini,

Nama mahasiswa : Bintang Buana Rajasa
NIM : 19.11.3118

Menyatakan bahwa Skripsi dengan judul berikut:

Tuliskan Judul Skripsi

Dosen Pembimbing : Khusnawi, S.Kom, M. Eng.

1. Karya tulis ini adalah benar-benar ASLI dan BELUM PERNAH diajukan untuk mendapatkan gelar akademik, baik di Universitas AMIKOM Yogyakarta maupun di Perguruan Tinggi lainnya.
2. Karya tulis ini merupakan gagasan, rumusan dan penelitian SAYA sendiri, tanpa bantuan pihak lain kecuali arahan dari Dosen Pembimbing.
3. Dalam karya tulis ini tidak terdapat karya-atau pendapat orang lain, kecuali secara tertulis dengan jelas dicantumkan sebagai acuan dalam naskah dengan disebutkan nama pengarang dan disebutkan dalam Daftar Pustaka pada karya tulis ini.
4. Perangkat lunak yang digunakan dalam penelitian ini sepenuhnya menjadi tanggung jawab SAYA, bukan tanggung jawab Universitas AMIKOM Yogyakarta.
5. Pernyataan ini SAYA buat dengan sesungguhnya, apabila di kemudian hari terdapat penyimpangan dan ketidakbenaran dalam pernyataan ini, maka SAYA bersedia menerima SANKSI AKADEMIK dengan pencabutan gelar yang sudah diperoleh, serta sanksi lainnya sesuai dengan norma yang berlaku di Perguruan Tinggi.

Yogyakarta, <tanggal lulus ujian skripsi>

Yang Menyatakan,



Bintang Buana Rajasa

KATA PENGANTAR

Puji syukur penulis panjatkan pada Tuhan YME, atas berkat dan penyertaannya sehingga penelitian yang berjudul “Penerapan Data Mining Metode Deskripsi Untuk Mengetahui Perbedaan Pengguna Kasual Dengan *Annual Member* Layanan *Bike-Sharing* Cyclistic.” dapat diselesaikan dengan baik. Penelitian ini dapat terlaksana dengan baik dan lancar berkat dukungan yang diberikan oleh beberapa pihak tertentu diantaranya adalah:

1. Bapak Khusnawi, M.Kom, M. Eng. , selaku dosen pembimbing yang telah banyak memberikan wawasan, bimbingan, dan dorongan dalam menyelesaikan laporan penelitian ini.
2. Ibu dirumah selaku orang tua peneliti yang telah memberikan dukungan materil.
3. Teman-teman seperjuangan terutama Caroline Sentiaji.
4. Seluruh pihak yang telah membantu secara langsung dan tidak langsung dalam penyelesaian tugas akhir ini yang namanya tidak dapat disebutkan satu per satu oleh penulis.

Saya ucapkan terima kasih atas dukungan yang telah diberikan. Dengan segala kerendahan hati, penelitian ini masih jauh dari sempurna karena keterbatasan sumber daya dan pengetahuan yang penulis miliki. Semoga skripsi ini dapat bermanfaat bagi pembaca.

Yogyakarta, 18 Juni 2023

Penulis,

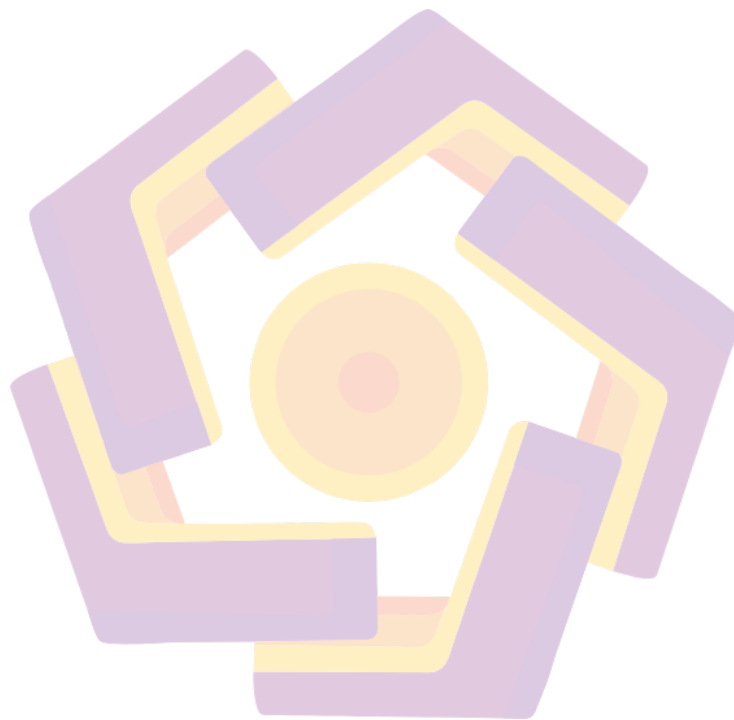
Bintang Buana Rajasa

DAFTAR ISI

HALAMAN JUDUL.....	i
HALAMAN PERSETUJUAN.....	ii
HALAMAN PENGESAHAN.....	iii
HALAMAN PERNYATAAN KEASLIAN SKRIPSI	iv
KATA PENGANTAR	v
DAFTAR ISI.....	vi
DAFTAR TABEL.....	ix
DAFTAR GAMBAR	x
DAFTAR LAMPIRAN.....	xi
DAFTAR LAMBANG DAN SINGKATAN	xii
INTISARI.....	xiii
ABSTRACT	xiv
BAB I	1
1.1 Latar Belakang	1
1.2 Rumusan Masalah.....	3
1.3 Batasan Masalah	3
1.4 Tujuan Penelitian	4
1.5 Manfaat Penelitian	4
1.6 Sistematika Penulisan	4
BAB II.....	6
2.1 Studi Literatur	6
2.2 Tabel Keaslian Penelitian.....	11
2.3 Dasar Teori.....	18
2.3.1 Bike-sharing.....	18
2.3.2 Cyclistic	19
2.3.3 Divvy.....	19
2.3.4 Data Mining	20
2.3.5 Metodologi CRISP-DM	21

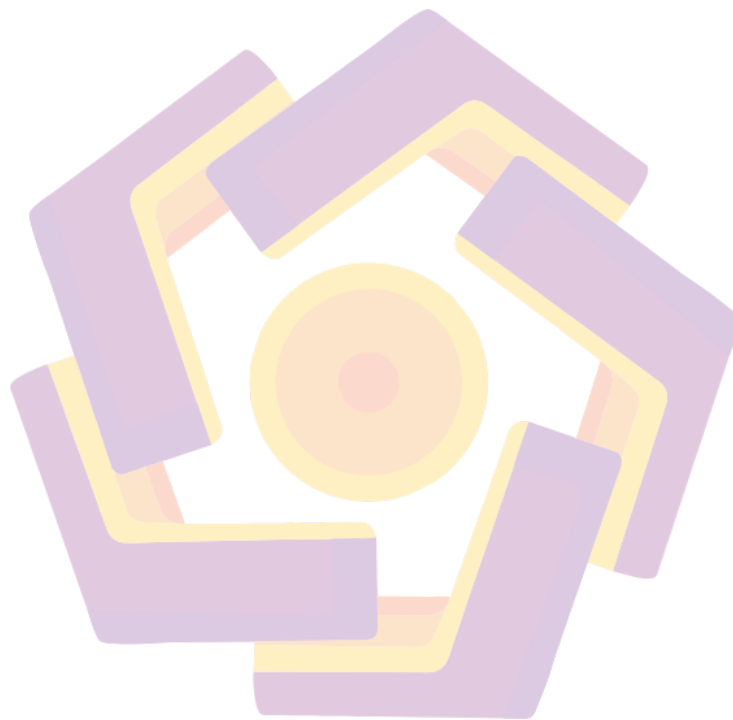
2.3.6	Metode Deskripsi <i>Data Mining</i>	23
2.3.7	K-Prototype <i>Algorithm</i>	24
2.3.8	<i>Silhouette Coefficient</i>	26
2.3.9	Davies-Bouldini <i>Index</i>	27
2.3.10	Rand <i>Index</i>	29
BAB III		30
3.1	Objek Penelitian.....	30
3.2	Alur Penelitian	30
3.2.1	Studi Literatur	31
3.2.2	Perencanaan Penelitian	31
3.2.3	<i>Data Mining</i> (CRISP-DM).....	31
3.3	Alat Dan Bahan.....	35
BAB IV		37
4.1	Data Mining (CRISP-DM).....	37
4.1.1	<i>Understanding of Business</i>	37
4.1.2	<i>Understanding of Data</i>	37
4.1.2.1	Deskripsi Data.....	37
4.1.2.2	Eksplorasi Data	38
4.1.3	<i>Data Preparation</i>	40
4.1.3.1	Pemilihan Data.....	40
4.1.3.2	Rekonstruksi Data.....	40
4.1.4.1	Pembersihan Data	41
4.1.4	<i>Modeling</i>	43
4.1.4.2	K-Prototype.....	43
4.1.5	<i>Evaluation</i>	47
4.1.5.1	<i>Silhouette Coefficient</i>	47
4.1.5.1	Davies-Bouldin <i>Index</i>	48
4.1.5.2	Rand <i>Index</i>	48
4.1.6	<i>Deployment</i>	49
BAB V.....		50

5.1	Kesimpulan	50
5.2	Saran	51
DAFTAR PUSTAKA		52
LAMPIRAN.....		55



DAFTAR TABEL

Tabel 3.1 Alat Penelitian.....	35
Tabel 4.1 Deskripsi Data.....	38
Tabel 4.2 Eksplorasi Data	38
Tabel 4.3 Verifikasi Data	39

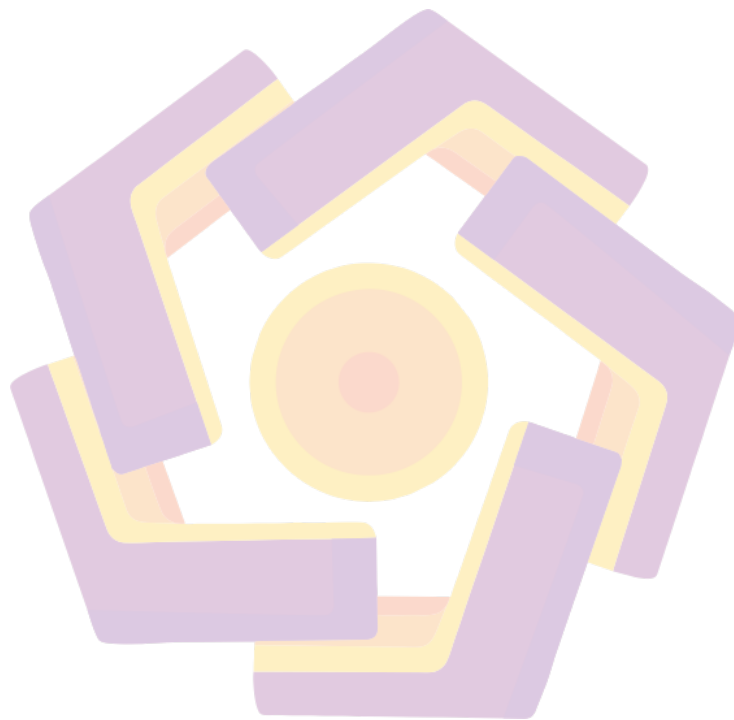


DAFTAR GAMBAR

Gambar 2.1 <i>Bike-sharing</i>	18
Gambar 2.2 <i>Bike-sharing</i> World Map.....	19
Gambar 2.3 Divvy Homepage	20
Gambar 3.1 Diagram Alur Penelitian.....	30
Gambar 3.2 Alur Proses Bisnis	32
Gambar 4.1 Eksplorasi Data	38
Gambar 4.2 Penggabungan Tabel	39
Gambar 4.3 Rekonstruksi Nama Kolom.....	40
Gambar 4.4 Kolom Menit	40
Gambar 4.5 Kolom Hari.....	41
Gambar 4.6 Hapus Kolom	41
Gambar 4.7 <i>Null</i>	41
Gambar 4.8 Data Duplikat	42
Gambar 4.9 <i>Outlier</i>	42
Gambar 4.10 Hapus <i>Outlier</i>	43
Gambar 4.11 Mengambil Sampel Data.....	43
Gambar 4.12 Implementasi K-Prototype	44
Gambar 4.13 Jumlah Penggunaan Berdasarkan <i>Cluster</i>	44
Gambar 4.14 Jumlah Penggunaan Berdasarkan Tipe Pengguna.....	45
Gambar 4.15 Distribusi Durasi Penggunaan <i>Cluster</i>	45
Gambar 4.16 Grafis Tipe Sepeda & Hari.....	46
Gambar 4.17 Mean & Median <i>Cluster</i>	47
Gambar 4.18 Evaluasi Silhouette Coefficient.....	47
Gambar 4.19 Evaluasi Davies-Boulden Index	48
Gambar 4.20 Evaluasi Rand Index.....	48

DAFTAR LAMPIRAN

Lampiran 1 <i>Understanding of Data</i>	55
Lampiran 2 <i>Data Preparation</i>	57



DAFTAR LAMBANG DAN SINGKATAN

DM	<i>Data Mining</i>
CRISP-DM	<i>Cross-Industry Standard Process for Data Mining</i>
SEMMA	<i>Sample, Explore, Modify, Model, Assess</i>
SC	<i>Silhouette Coefficient</i>
DBI	<i>Davies-Bouldini Index</i>
RI	<i>Rand-Index</i>
k	Jumlah <i>cluster</i>
$S(i)$	Nilai <i>Silhouette Coefficient</i>
SW_k	Mean dari jarak tiap data ke <i>centroid k</i> didalam <i>cluster k</i> .
DB	Nilai <i>Davies-Bouldini Index</i>
RI	Nilai <i>Rand-Index</i>



INTISARI

Banyaknya transportasi modern berupa kendaraan bermotor pada zaman modern menyebabkan polusi udara pada lingkungan. Banyaknya polusi udara tersebut disertai dengan kemacetan lalu lintas di kota-kota besar mempunyai pengaruh pada kesehatan fisik dan mental. Salah satu usaha yang berhasil mengurangi dampak negatif di atas adalah dengan penggunaan transportasi sepeda dengan layanan *bike-sharing*. Meskipun demikian, diantara pengguna *bike-sharing* masih ada yang bukan *annual member/anggota* anual dan hanya pengguna biasa/kasual layanan *bike-sharing*. Hal ini dapat diartikan bahwa penggunaan layanan *bike-sharing* masih dapat dimaksimalkan.

Karena itu, peneliti menganalisis *dataset* Cyclistic yang berisikan data penggunaan layanan *bike-sharing*, menggunakan *data mining* (DM) metodologi CRISP-DM untuk mengidentifikasi perbedaan *annual member* dan pengguna casual, serta hal-hal apa yang membuat seseorang menjadi *annual member* layanan *bike-sharing*. Dengan mengerti hal-hal tersebut, dapat diketahui tindakan apa yang perlu dilakukan untuk meningkatkan jumlah *annual member*, yang akan secara efektif meningkatkan jumlah penggunaan layanan *bike-sharing*. Peneliti menggunakan metode deskripsi DM dan Metode *clustering* algoritma K-Prototype. Untuk evaluasi digunakan metode *Silhouette Coefficient*, *Rand Index*, dan *Davies-Bouldini Index*.

Hasil penelitian yang dilakukan menunjukkan karakteristik perbedaan antara *annual member* dan pengguna casual layanan *bike-sharing* beserta orang-orang seperti apa yang menjadi *annual member* layanan *bike-sharing*. Dari hasil tersebut, juga ditemukan bagaimana caranya meningkatkan jumlah *annual member* pada layanan *bike-sharing* Cyclistic.

Kata Kunci: *bike-sharing*, metode deskripsi, *data mining*, K-Prototype.

ABSTRACT

The large number of modern transportations in the form of motorized vehicles in this day and ages causes a lot of air pollution in the environment. The amount of air pollution accompanied by traffic jams in big cities has an impact on our physical and mental health. One of the efforts that has succeeded in reducing the negative impacts is by using bicycles with a bike-sharing service. Even so, among bike-sharing users there are still those who are not annual members and are only ordinary/casual users of bike-sharing services. This means that the use of bike-sharing services can still be maximized.

Therefore, researcher analyzed the Cyclistic dataset which contains data on the use of bike-sharing services, using the CRISP-DM data mining (DM) method to identify differences between annual members and regular users, as well as what factors that made someone an annual member of bike-sharing services. By understanding these, it can be known what actions need to be taken to increase the number of annual members, which will effectively increase the use of bike-sharing services. Researcher used the DM description method and the K-Prototype algorithm clustering method. For evaluation, the Silhouette Coefficient method, Rand Index, and Davies-Bouldini Index was used.

The results show the characteristic differences between annual members and casual users of bike-sharing services and what kind of people are annual members of bike-sharing services. From these results, it was also discovered how to increase the number of annual members in the Cyclistic bike-sharing service.

Keyword: *bike-sharing, description method, data mining, K-Prototype*