



SKRIPSI

**ANALISIS POLA *PAIN POINTS* DARI DATA
KUALITATIF DALAM *UX RESEARCH*
MENGUNAKAN K-MEANS DAN NLTK**

IKHYA ULUMMUDDIN

NPM 20081010120

DOSEN PEMBIMBING

Dr. Eng. Ir. Anggraini Puspita Sari, ST., MT.

Made Hanindia Prami Swari, S.Kom, M.Cs

KEMENTERIAN PENDIDIKAN TINGGI, SAINS, DAN TEKNOLOGI

UNIVERSITAS PEMBANGUNAN NASIONAL VETERAN JAWA TIMUR

FAKULTAS ILMU KOMPUTER

PROGRAM STUDI INFORMATIKA

SURABAYA

2025



SKRIPSI

ANALISIS POLA *PAIN POINTS* DARI DATA KUALITATIF DALAM *UX RESEARCH* MENGUNAKAN K-MEANS DAN NLTK

IKHYA ULUMMUDDIN

NPM 20081010120

DOSEN PEMBIMBING

Dr. Eng. Ir. Anggraini Puspita Sari, ST., MT.

Made Hanindia Prami Swari, S.Kom, M.Cs

KEMENTERIAN PENDIDIKAN TINGGI, SAINS, DAN TEKNOLOGI

UNIVERSITAS PEMBANGUNAN NASIONAL VETERAN JAWA TIMUR

FAKULTAS ILMU KOMPUTER

PROGRAM STUDI INFORMATIKA

SURABAYA

2025

LEMBAR PENGESAHAN

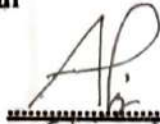
ANALISIS POLA PAIN POINTS DARI DATA KUALITATIF DALAM UX RESEARCH MENGGUNAKAN K-MEANS DAN NLTK

Oleh:
IKHYA ULUMMUDDIN
NPM. 20081010120

Telah dipertahankan di hadapan dan diterima oleh Tim Penguji Skripsi Prodi Informatika, Fakultas Ilmu Komputer, Universitas Pembangunan Nasional Veteran Jawa Timur pada tanggal

Menyetujui

Dr. Eng. Ir. Anggraini Puspita Sari, ST., MT.
NPT. 222198 60 816400


..... (Pembimbing I)

Made Hanindia Prami Swari, S.Kom, M.Cs
NIP. 19890205 2018032 001


..... (Pembimbing II)

Yisti Vita Via, S. ST., M.Kom.
NIP. 19860425 202121 2 001


..... (Ketua Penguji)

Afina Lina Nurlaili, S.Kom., M.Kom.
NIP. 1993121 3202203 2 010


..... (Anggota Penguji II)

Mengetahui,
Dekan Fakultas Ilmu Komputer,


Prof. Dr. Ir. Novirina Hendrasarie, MT.
NIP. 19681126 199403 2 001

LEMBAR PERSETUJUAN

ANALISIS POLA *PAIN POINTS* DARI DATA KUALITATIF DALAM *UX* *RESEARCH* MENGGUNAKAN K-MEANS DAN NLTK

Oleh:

IKHYA ULUMMUDDIN

NPM. 20081010120

Menyetujui

Koordinator Program Studi Informatika,
Fakultas Ilmu Komputer



Fetty Tri Anggraeny, S.Kom., M.Kom.

NIP. 19820211 202121 2 005

SURAT PERNYATAAN BEBAS PLAGIASI

Saya yang bertanda tangan di bawah ini:

Nama : Ikhya Ulummuddin
NPM : 20081010120
Program : Sarjana (S1)
Program Studi : Informatika
Fakultas : Ilmu Komputer

Menyatakan bahwa dalam dokumen ilmiah Skripsi ini tidak terdapat bagian dari karya ilmiah lain yang telah diajukan untuk memperoleh gelar akademik di suatu lembaga pendidikan tinggi dan juga tidak terdapat karya atau pendapat yang pernah ditulis atau diterbitkan oleh orang atau lembaga lain kecuali yang secara tertulis disitasi dalam dokumen ini dan disebutkan secara lengkap dalam daftar pustaka.

Dan saya menyatakan bahwa dokumen ilmiah ini bebas dari unsur-unsur plagiasi. Jika dikemudian hari ditemukan indikasi plagiat pada Skripsi ini, maka saya bersedia menerima sanksi sesuai dengan peraturan perundang-undangan yang berlaku.

Demikian surat pernyataan ini saya buat dengan sesungguhnya tanpa ada paksaan dari siapapun juga dan untuk dipergunakan sebagaimana mestinya.

Surabaya, 16 Mei 2025

Yang membuat pernyataan,



Ikhya Ulummuddin

NPM. 20081010120

ABSTRAK

Nama Mahasiswa / NPM : Ikhya Ulummuddin / 20081010120
Judul Skripsi : Analisis Pola Pain Points Dari Data Kualitatif
Dalam *UX Research* Menggunakan *K-Means* dan
NLTK
Dosen Pembimbing : 1. Dr. Eng. Ir. Anggraini Puspita Sari, ST., MT.
2. Made Hanindia Prami S, S.Kom, M.Cs

Dalam *framework design thinking*, tahap *empathize* dan *define* merupakan bagian dari proses *User Experience (UX) research* yang bertujuan untuk menganalisis *pain points* pengguna melalui data kualitatif. Namun, proses tersebut selalu dilakukan secara manual, sehingga membutuhkan sumber daya dan waktu yang lebih. Penelitian ini mengembangkan sistem klasterisasi berbasis *unsupervised machine learning* untuk menganalisis data kualitatif berdasarkan topik *pain points*-nya menggunakan algoritma *K-Means*. Sebagai objek penelitian, digunakan data ulasan pengguna aplikasi Gojek versi 4.9.0, 4.9.1, dan 4.9.3 (November 2021 hingga Januari 2024). Proses *text preprocessing* dilakukan menggunakan pendekatan *Natural Language Processing (NLP)* dengan pustaka NLTK. Sistem diuji dengan variasi parameter jumlah klaster dan reduksi komponen *Principal Component Analysis (PCA)*. Evaluasi hasil klasterisasi dilakukan menggunakan *silhouette score*, *Davies-Bouldin Index*, dan *Calinski-Harabasz Index*. Hasil terbaik diperoleh pada jumlah klaster sebanyak 8 dengan *silhouette score* sebesar 0.6512, *Davies-Bouldin Index* sebesar 1.0818, *Calinski-Harabasz Index* sebesar 4.9126. Rata-rata waktu komputasi yang dibutuhkan sistem tercatat selama 7.85 detik. Hasil ini menunjukkan bahwa sistem cukup efektif dalam mengelompokkan data kualitatif berdasarkan pola *pain points*. Sistem yang dikembangkan dapat dimanfaatkan oleh *UX researcher* untuk analisis data secara otomatis dan efisien.

Kata kunci: *User experience research, pain points, K-Means, Natural Language Processing (NLP)*

ABSTRACT

Nama Mahasiswa / NPM : Ikhya Ulummuddin / 20081010120
Judul Skripsi : Analisis Pola Pain Points Dari Data Kualitatif
Dalam *UX Research* Menggunakan *K-Means* dan
NLTK
Dosen Pembimbing : 1. Dr. Eng. Ir. Anggraini Puspita Sari, ST., MT.
2. Made Hanindia Prami S, S.Kom, M.Cs

Within the design thinking framework, the empathize and define stages are integral parts of the User Experience (UX) research process, aimed at analyzing user pain points through qualitative data. However, this process is often conducted manually, requiring considerable time and resources. This study develops a clustering system based on unsupervised machine learning to analyze qualitative data by identifying topics of user pain points using the K-Means algorithm. The dataset used consists of user reviews from the Gojek application versions 4.9.0, 4.9.1, and 4.9.3, collected between November 2021 and January 2024. Text preprocessing was carried out using Natural Language Processing (NLP) techniques with the NLTK library. The system was tested with variations in the number of clusters and the number of components in Principal Component Analysis (PCA) for dimensionality reduction. The clustering results were evaluated using the silhouette score, Davies-Bouldin Index, and Calinski-Harabasz Index. The best results were obtained with 8 clusters, yielding a silhouette score of 0.6512, a Davies-Bouldin Index of 1.0818, and a Calinski-Harabasz Index of 4.9126. The average system computation time was recorded at 7.85 seconds. These results indicate that the system is reasonably effective in clustering qualitative data based on patterns of user pain points. The developed system can assist UX researchers in conducting data analysis in an automated and efficient manner.

Keywords: *User experience research, pain points, K-Means, Natural Language Processing (NLP)*

KATA PENGANTAR

Segala puji dan syukur penulis panjatkan ke hadirat Allah SWT atas limpahan petunjuk, pertolongan, serta nikmat-Nya yang tiada henti. Berkat izin dan ridho-Nya, penulis akhirnya dapat menyelesaikan penyusunan laporan skripsi yang berjudul **“ANALISIS POLA PAIN POINTS DARI DATA KUALITATIF DALAM UX RESEARCH MENGGUNAKAN K-MEANS DAN NLTK”**.

Ada cukup bantuan dan dukungan yang didapatkan dari berbagai pihak selama melakukan penelitian serta menyelesaikan laporan skripsi ini. Penulis memahami bahwa laporan skripsi ini belum sepenuhnya sempurna, baik dari aspek isi maupun cara penyajiannya. Oleh karena itu, penulis sangat mengharapkan kritik dan saran yang membangun dari semua pembaca agar dapat mengembangkan dan membangun penelitian lebih baik ke depannya.

Semoga laporan skripsi ini bermanfaat untuk menambah wawasan dan pengetahuan para pembaca agar bisa mengembangkan dan meningkatkan ilmu pengetahuan.

Surabaya, 16 Mei 2025

Ikhya Ulummuddin

UCAPAN TERIMA KASIH

Penyusunan dan penulisan laporan skripsi ini tidak lepas dari arahan, bantuan, serta dukungan dari berbagai pihak. Untuk itu, penulis ingin menyampaikan terima kasih yang sebesar-besarnya kepada Allah SWT yang senantiasa memberikan kemudahan, kelancaran, kekuatan, serta sumber segala petunjuk. Penulis juga menyampaikan apresiasi setinggi-tingginya kepada pihak-pihak terhormat yang telah memberikan kontribusi dalam proses ini.

1. Kedua orang tua, saudara, dan seluruh keluarga besar, atas segala doa, kasih sayang, serta dukungan yang tak pernah putus dalam mendampingi penulis menempuh pendidikan di jenjang perguruan tinggi sejak awal hingga saat ini.
2. Bapak Prof. Dr. Ir. Akhmad Fauzi, M.MT., IPU., selaku Rektor Universitas Pembangunan Nasional “Veteran” Jawa Timur.
3. Ibu Prof. Dr. Ir. Novirina Hendrasarie, MT., selaku Dekan Fakultas Ilmu Komputer, Universitas Pembangunan Nasional “Veteran” Jawa Timur.
4. Ibu Fetty Tri Anggraeny, S.Kom, M.Kom., selaku Koordinator Program Studi Informatika, Fakultas Ilmu Komputer, Universitas Pembangunan Nasional “Veteran” Jawa Timur.
5. Ibu Dr. Eng. Ir. Anggraini Puspita Sari, ST., MT., sebagai dosen pembimbing I dan Ibu Made Hanindia Prami S, S.Kom, M.Cs, sebagai dosen pembimbing II yang telah dengan sabar meluangkan waktu serta tenaga untuk memberi arahan dan bimbingan yang berharga dalam penyusunan skripsi ini.
6. Seluruh dosen dan staf Program Studi Informatika Universitas Pembangunan Nasional “Veteran” Jawa Timur, atas ilmu pengetahuan, arahan, dan pengalaman yang diberikan selama perkuliahan.
7. Rekan-rekan di dalam maupun di luar lingkungan perkuliahan yang telah membantu, memotivasi, dan memberi semangat kepada penulis dalam menyelesaikan skripsi.
8. Jacques Bermon Webster II, Brando Franco Windah, Diwantara Anugrah Putra, dan Gema Cita Andika yang telah memotivasi penulis agar tetap semangat untuk menyelesaikan laporan skripsi ini.

9. Semua pihak yang tidak dapat penulis tuliskan baik secara sengaja maupun tidak sengaja dalam membantu, menyemangati, dan mendoakan penulis dalam menyelesaikan laporan skripsi ini.
10. Om Hari dan Tante Ninik yang selalu memberi semangat dan motivasi untuk menyelesaikan laporan skripsi.

Atas segala bantuan, dukungan, doa, dan dorongan semangat yang telah diberikan dari seluruh pihak tersebut, penulis mengucapkan terima kasih yang tulus. Semoga segala kebaikan yang telah diberikan bisa mendapatkan balasan yang sepadan dari Allah SWT dan semoga segala upaya dalam menyelesaikan penelitian serta penulisan laporan skripsi ini menjadi amal yang bernilai.

Surabaya, 16 Mei 2025

Ikhya Ulummuddin

DAFTAR ISI

LEMBAR PENGESAHAN.....	ii
LEMBAR PERSETUJUAN.....	iii
SURAT PERNYATAAN BEBAS PLAGIAT.....	iv
ABSTRAK.....	v
KATA PENGANTAR.....	vii
UCAPAN TERIMA KASIH.....	viii
DAFTAR ISI.....	x
DAFTAR TABEL.....	xiii
DAFTAR GAMBAR.....	xiv
DAFTAR NOTASI.....	xvi
BAB I PENDAHULUAN.....	1
1.1. Latar Belakang.....	1
1.2. Rumusan Masalah.....	3
1.3. Tujuan Penelitian.....	3
1.4. Manfaat Penelitian.....	4
1.5. Batasan Masalah.....	4
BAB II KAJIAN PUSTAKA.....	5
2.1. Penelitian Terdahulu.....	5
2.2. <i>UX Research</i>	7
2.3. <i>Design Thinking</i>	8
2.3.1. <i>Empathize</i>	9
2.3.2. <i>Define</i>	10
2.3.3. <i>Ideate</i>	11
2.3.4. <i>Prototype</i>	11
2.3.5. <i>Test</i>	12
2.4. Python.....	13
2.5. Google Colaboratory.....	13
2.6. NLP.....	14
2.7. <i>Preprocessing Data Teks</i>	15
2.7.1. <i>Case Folding</i>	15

2.7.2. <i>Stopword Removal</i>	16
2.7.3. <i>Stemming</i>	16
2.7.4. <i>Tokenizing</i>	17
2.7.5. <i>Text Cleaning</i>	17
2.7.6. <i>Vektorisasi Teks</i>	17
2.8. <i>WordCloud</i>	18
2.9. <i>TF-IDF Vectorizer</i>	18
2.10. <i>Silhouette Score</i>	19
2.11. <i>Davies-Bouldin Index</i>	20
2.12. <i>Calinski-Harabasz Index</i>	21
2.13. <i>K-Means Clustering</i>	22
2.14. <i>Natural Language Toolkit</i>	23
2.15. <i>Principal Component Analysis</i>	24
BAB III METODOLOGI PENELITIAN	25
3.1. <i>Analisis Kebutuhan</i>	25
3.2. <i>Pengumpulan Data</i>	26
3.2. <i>Pengolahan Data</i>	27
3.2.1. <i>Preprocessing Data</i>	27
3.2.2. <i>Vektorisasi TF-IDF</i>	31
3.3. <i>Pengelompokkan Data</i>	34
3.3.1. <i>Menentukan Jumlah Kluster Optimal</i>	35
3.3.2. <i>Menentukan Nilai Titik Pusat Secara Acak</i>	36
3.3.3. <i>Menghitung Jarak Data Dengan Titik Pusat</i>	36
3.3.4. <i>Membuat Titik Pusat Baru</i>	37
3.3.5. <i>Iterasi Perhitungan Jarak dan Titik Pusat Baru</i>	37
3.4. <i>Pengujian Hasil Pengelompokkan</i>	39
BAB IV HASIL DAN PEMBAHASAN	45
4.1. <i>Hasil Penelitian</i>	45
4.1.1. <i>Persiapan Dataset</i>	45
4.1.2. <i>Text preprocessing</i>	46
4.1.3. <i>Vektorisasi TF-IDF</i>	49
4.1.4. <i>Klasterisasi Algoritma K-Means</i>	50

4.2. Pembahasan.....	54
4.2.1. Kriteria Pengujian.....	54
4.2.2. Hasil Pengujian.....	57
BAB V PENUTUP.....	87
5.1. Kesimpulan.....	87
5.2. Saran.....	88
DAFTAR PUSTAKA.....	89

DAFTAR TABEL

Tabel 3.1. Persebaran jumlah data pada setiap rating ulasan.....	26
Tabel 3.2. Persebaran jumlah data pada setiap rating ulasan yang terfilter.....	27
Tabel 3.3. Data sampel untuk perhitungan.....	28
Tabel 3.4. Hasil proses tokenisasi.....	29
Tabel 3.5. Hasil proses <i>stopwords removal</i>	30
Tabel 3.6. Hasil proses <i>stemming</i>	30
Tabel 3.7. Hasil proses <i>text cleaning</i>	31
Tabel 3.8. Hasil perhitungan TF.....	32
Tabel 3.9. Hasil perhitungan IDF.....	33
Tabel 3.10. Hasil vektorisasi numerik dengan TF-IDF.....	33
Tabel 3.11. Hasil nilai jarak setiap titik data dengan titik pusat.....	36
Tabel 3.12. Hasil nilai jarak setiap titik data dengan titik pusat baru.....	37
Tabel 3.13. Hasil klasterisasi sampel data.....	38
Tabel 3.14. Hasil rata-rata setiap klaster.....	38
Tabel 3.15. Hasil nilai $a(i)$ pada setiap klaster.....	40
Tabel 3.16. Hasil nilai $b(i)$ pada setiap data <i>point</i>	40
Tabel 3.17. Hasil nilai <i>silhouette score</i> pada setiap data point.....	41
Tabel 3.18. Hasil nilai <i>compactness</i> pada setiap klaster.....	42
Tabel 3.19. Hasil nilai <i>separation</i> pada setiap klaster.....	42
Tabel 3.20. Hasil nilai rasio <i>compactness</i> dan <i>separation</i> pada setiap klaster.....	42
Tabel 3.21. Hasil nilai SSW pada setiap klaster.....	43
Tabel 3.22. Hasil nilai SSB pada setiap klaster.....	44
Tabel 3.23. Hasil metrik pengujian dari pengelompokan sampel data.....	44
Tabel 4.1. Nilai <i>silhouette score</i> dari setiap kombinasi pengujian.....	78
Tabel 4.2. Nilai <i>Davies-Bouldin Index</i> dari setiap kombinasi pengujian.....	79
Tabel 4.3. Nilai <i>Calinski-Harabasz Index</i> dari setiap kombinasi pengujian.....	80
Tabel 4.4. Waktu yang dibutuhkan sistem dalam mengelompokkan data.....	81

DAFTAR GAMBAR

Gambar 2.1. Proses <i>design thinking</i>	8
Gambar 2.2. Ilustrasi tahap <i>empathize</i>	9
Gambar 2.3. Ilustrasi tahap <i>define</i>	10
Gambar 2.4. Ilustrasi tahap <i>ideate</i>	11
Gambar 2.5. Ilustrasi tahap <i>prototype</i>	12
Gambar 2.6. Ilustrasi tahap <i>test</i>	12
Gambar 2.7. Logo Python.....	13
Gambar 2.8. Tampilan utama Google Colaboratory.....	14
Gambar 2.9. Tampilan <i>wordcloud</i>	18
Gambar 2.10. <i>Natural Language Toolkit</i>	23
Gambar 3.1. Alur penelitian.....	25
Gambar 3.2. Alur preproceasing data.....	28
Gambar 3.3. Alur vektorisasi TF-IDF.....	32
Gambar 3.4. Alur <i>K-Means clustering</i>	35
Gambar 4.1. Beberapa baris data ulasan pengguna.....	46
Gambar 4.2. Hasil <i>filtering</i> data ulasan pengguna.....	48
Gambar 4.3. Beberapa data dari hasil proses <i>text preprocessing</i>	49
Gambar 4.4. Hasil vektorisasi TF-IDF kolom topik.....	50
Gambar 4.5. Hasil dari proses <i>K-Means clustering</i>	52
Gambar 4.6. Total data pada setiap klaster.....	53
Gambar 4.7. Metode <i>elbow</i> untuk pengelompokkan data ulasan pengguna.....	58
Gambar 4.8. Hasil metrik evaluasi dengan jumlah klaster 2.....	59
Gambar 4.9. Jumlah data setiap klaster dengan jumlah klaster 2.....	59
Gambar 4.10. <i>Scatter plot</i> dengan jumlah klaster 2.....	59
Gambar 4.11. Waktu komputasi dengan jumlah klaster 2.....	60
Gambar 4.12. Hasil metrik dengan jumlah klaster 3.....	60
Gambar 4.13. Jumlah data setiap klaster dengan jumlah klaster 3.....	61
Gambar 4.14. <i>Scatter plot</i> dengan jumlah klaster 3.....	61
Gambar 4.15. Waktu komputasi dengan jumlah klaster 3.....	61
Gambar 4.16. Hasil metrik dengan jumlah klaster 4.....	62

Gambar 4.17. Jumlah data setiap klaster dengan jumlah klaster 4.....	63
Gambar 4.18. <i>Scatter plot</i> dengan jumlah klaster 4.....	63
Gambar 4.19. Waktu komputasi dengan jumlah klaster 4.....	63
Gambar 4.20. Hasil metrik dengan jumlah klaster 5.....	64
Gambar 4.21. Jumlah data setiap klaster dengan jumlah klaster 5.....	65
Gambar 4.22. <i>Scatter plot</i> dengan jumlah klaster 5.....	65
Gambar 4.23. Waktu komputasi dengan jumlah klaster 5.....	65
Gambar 4.24. Hasil metrik dengan jumlah klaster 6.....	66
Gambar 4.25. Jumlah data setiap klaster dengan jumlah klaster 6.....	67
Gambar 4.26. <i>Scatter plot</i> dengan jumlah klaster 6.....	67
Gambar 4.27. Waktu komputasi dengan jumlah klaster 6.....	67
Gambar 4.28. Hasil metrik dengan jumlah klaster 7.....	68
Gambar 4.29. Jumlah data setiap klaster dengan jumlah klaster 7.....	69
Gambar 4.30. <i>Scatter plot</i> dengan jumlah klaster 7.....	69
Gambar 4.31. Waktu komputasi dengan jumlah klaster 7.....	69
Gambar 4.32. Hasil metrik dengan jumlah klaster 8.....	70
Gambar 4.33. Jumlah data setiap klaster dengan jumlah klaster 8.....	71
Gambar 4.34. <i>Scatter plot</i> dengan jumlah klaster 8.....	71
Gambar 4.35. Waktu komputasi dengan jumlah klaster 8.....	71

DAFTAR NOTASI

t	:	Kata yang ingin dihitung TF-IDF-nya
d	:	Dokumen yang ingin dihitung TF-IDF-nya
n_i	:	Jumlah indeks data <i>point</i> di dalam klaster C_i
n	:	Jumlah total data
$n(t, d)$	:	Jumlah kemunculan kata t dalam dokumen d
N	:	Jumlah dokumen dalam <i>corpus</i>
$N(d)$	:	Jumlah kata dalam dokumen d
$df(t)$	:	Jumlah dokumen yang mengandung kata t
i	:	Indeks data <i>point</i>
$s(i)$	:	<i>Silhouette score</i> untuk data <i>point</i> i
$a(i)$	:	Rata-rata jarak antara data <i>point</i> i dengan semua objek lain dalam cluster yang sama
$b(i)$	:	Rata-rata jarak antara objek i dengan semua data <i>point</i> dalam cluster lain dan dipilih yang terendah
C_i	:	Klaster yang memiliki indeks data <i>point</i> i
j	:	Indeks data <i>point</i> lain
C_k	:	Klaster lain terdekat dengan C_i yang bukan klaster dari i
$distance(i, j)$	:	Jarak antara indeks data <i>point</i> i dengan j
k	:	Indeks data <i>point</i> dari klaster lain yang berbeda dengan klaster dari i
S_i	:	Rata-rata jarak intraklaster
$ C_i $	:	Jumlah indeks data <i>point</i> i di dalam klaster C_i
x	:	Index data <i>point</i> di dalam klaster
c_i	:	titik pusat dari klaster C_i

$\ x - c_i\ $	:	Jarak antara indeks data <i>point</i> x dengan titik pusat c_i
M	:	Titik pusat keseluruhan data
$d(x, y)$	:	Jarak antara data <i>point</i> x dan y
x_i & y_i	:	Nilai atribut ke-i dari data <i>point</i> x dan y
x_j	:	Data <i>point</i> ke-j dalam cluster ke-i