

**PENERAPAN ALGORITME *K-NEAREST NEIGHBOR*  
UNTUK MELAKUKAN KLASIFIKASI PRODUK  
PADA *E-MARKETPLACE SHOPEE***

**SKRIPSI**



disusun oleh

**Adilla Laela Tusifaiyah**

**17.12.0212**

**PROGRAM SARJANA  
PROGRAM STUDI SISTEM INFORMASI  
FAKULTAS ILMU KOMPUTER  
UNIVERSITAS AMIKOM YOGYAKARTA  
YOGYAKARTA  
2021**

**PENERAPAN ALGORITME *K-NEAREST NEIGHBOR*  
UNTUK MELAKUKAN KLASIFIKASI PRODUK  
PADA *E-MARKETPLACE SHOPEE***

**SKRIPSI**



disusun oleh

**Adilla Laela Tusifaiyah**

**17.12.0212**

**PROGRAM SARJANA  
PROGRAM STUDI SISTEM INFORMASI  
FAKULTAS ILMU KOMPUTER  
UNIVERSITAS AMIKOM YOGYAKARTA  
YOGYAKARTA  
2021**

# **PERSETUJUAN**

## **SKRIPSI**

### **PENERAPAN ALGORITME *K-NEAREST NEIGHBOR* UNTUK MELAKUKAN KLASIFIKASI PRODUK PADA *E-MARKETPLACE SHOPEE***

yang dipersiapkan dan disusun oleh

**Adilla Laela Tusifaiyah**

**17.12.0212**

telah disetujui oleh Dosen Pembimbing Skripsi  
pada tanggal 29 Maret 2021

**Dosen Pembimbing,**

**Sumarni Adi, S.Kom., M.Cs**  
**NIK. 190302256**

**PENGESAHAN**

**SKRIPSI**

**PENERAPAN ALGORITME *K-NEAREST NEIGHBOR*  
UNTUK MELAKUKAN KLASIFIKASI PRODUK  
PADA *E-MARKETPLACE SHOPEE***

yang dipersiapkan dan disusun oleh

**Adilla Laela Tusifaiyah**

**17.12.0212**

telah dipertahankan di depan Dewan Penguji  
pada tanggal 18 Maret 2021

**Susunan Dewan Penguji**

**Nama Penguji**

**Tanda Tangan**

**Anggit Dwi Hartanto, M.Kom**  
**NIK. 190302163**

\_\_\_\_\_

**Acihmah Sidauruk, M.Kom**  
**NIK. 190302238**

\_\_\_\_\_

**Sumarni Adi, S.Kom., M.Cs**  
**NIK. 190302256**

\_\_\_\_\_

Skripsi ini telah diterima sebagai salah satu persyaratan  
untuk memperoleh gelar Sarjana Komputer  
Tanggal 29 Maret 2021

**DEKAN FAKULTAS ILMU KOMPUTER**

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

## PERNYATAAN

Saya yang bertandatangan dibawah ini menyatakan bahwa, skripsi ini merupakan karya saya sendiri (ASLI), dan isi dalam skripsi ini tidak terdapat karya yang pernah diajukan oleh orang lain untuk memperoleh gelar akademis di suatu institusi pendidikan tinggi manapun, dan sepanjang pengetahuan saya juga tidak terdapat karya atau pendapat yang pernah ditulis dan/atau diterbitkan oleh orang lain, kecuali yang secara tertulis diacu dalam naskah ini dan disebutkan dalam daftar pustaka.

Segala sesuatu yang terkait dengan naskah dan karya yang telah dibuat adalah menjadi tanggungjawab saya pribadi.

Magelang, 22 Maret 2021



Adilla Laela Tusifaiyah

NIM. 17.12.0212

## MOTTO

Karena sesungguhnya sesudah kesulitan itu ada kemudahan (Al-Insyirah:5).  
Hasil tidak akan mengkhianati usaha.



## PERSEMBAHAN

Alhamdulillah rabbi 'alamin

Sujud syukur kupersembahkan kepada Allah SWT, Tuhan Yang Maha Esa, atas takdirMu telah kau jadikan saya manusia yang senantiasa berpikir, berilmu, beriman, dan bersyukur dalam menjalani kehidupan ini. Semoga keberhasilan ini menjadi satu langkah awal dalam meraih cita-cita untuk masa depan.

Dengan ini saya persembahkan karya ini untuk:

1. Kedua orang tua saya yang selalu mendoakan, menyayangi, memberi dukungan semangat serta nasihat selama ini.
2. Kakak dan adik tersayang, yang telah memotivasi saya agar semangat belajar.
3. Nenek yang selalu mendoakan, memberi dukungan dan semangat.
4. Nares yang sudah mau menjadi sahabat, partner terbaikku, dan pendengar yang baik.
5. Teman-teman terbaikku: Laila, Fina, Ahmad, Sefri, Linda, Ma'ruf, Huda, Nana, Herni, Latifa yang sudah sering direpotin.
6. Teman-teman kelas 17-SI-04 yang telah menjadi bagian dalam proses menempuh gelar sarjana ini. Semoga ilmu yang kita dapatkan menjadi berkah dan menjadi bekal untuk kesuksesan. Aamiin
7. Pengurus Taekwondo Amikom periode 2018/2019 yang telah memberikan banyak pengalaman baru di dalam organisasi.
8. Teman-teman Asisten Dosen yang telah berbagi semangat untuk selalu belajar.

## KATA PENGANTAR

Assalamu'alaikum Wr.Wb

Segala puji bagi Allah, Tuhan semesta alam, atas segala rahmat dan karunia-Nya penulis mampu menyelesaikan skripsi dengan judul “Penerapan Algoritme *K-Nearest Neighbor* Untuk Melakukan Klasifikasi Produk Pada *E-marketplace* Shopee” dengan baik dan lancar. Tak lupa shalawat serta salam semoga senantiasa tercurah kepada junjungan kita Nabi Muhammad SAW, keluarganya, sahabatnya, dan juga pengikutnya hingga akhir zaman.

Dalam penyelesaian skripsi ini, banyak hambatan dan kekurangan yang penulis hadapi. Namun pada akhirnya dapat terlampaui dengan adanya bimbingan, dukungan, serta bantuan dari berbagai pihak. Untuk itu pada kesempatan ini penulis menyampaikan terimakasih kepada:

1. Prof, Dr. M. Suyanto, MM selaku Rektor Universitas AMIKOM Yogyakarta.
2. Bapak Hanif Al Fatta, S.Kom., M.Kom selaku Dekan Fakultas Ilmu Komputer Universitas AMIKOM Yogyakarta.
3. Ibu Sumarni Adi, S.Kom. M.Cs selaku dosen pembimbing.
4. Bapak dan Ibu Dosen Universitas AMIKOM Yogyakarta yang telah memberikan ilmunya selama perkuliahan.
5. Kedua orang tua yang tidak lelah mendoakan putrinya agar dapat menyelesaikan studi.
6. Kakak, adik, nenek tersayang.
7. Teman-teman terbaikku yang telah mendoakan, berbagi ilmu dan semangat.
8. Teman-teman kelas 17-SI-04 yang telah berjuang bersama selama ini.
9. Semua pihak yang tidak dapat disebutkan satu persatu yang telah membantu, dan mendoakan penyelesaian skripsi ini.

Pembaca mungkin menemukan berbagai kekurangan dan kesalahan dalam penulisan skripsi ini, untuk itu penulis menerima segala kritik dan saran yang membangun dari pembaca. Sehingga dapat disempurnakan oleh peneliti



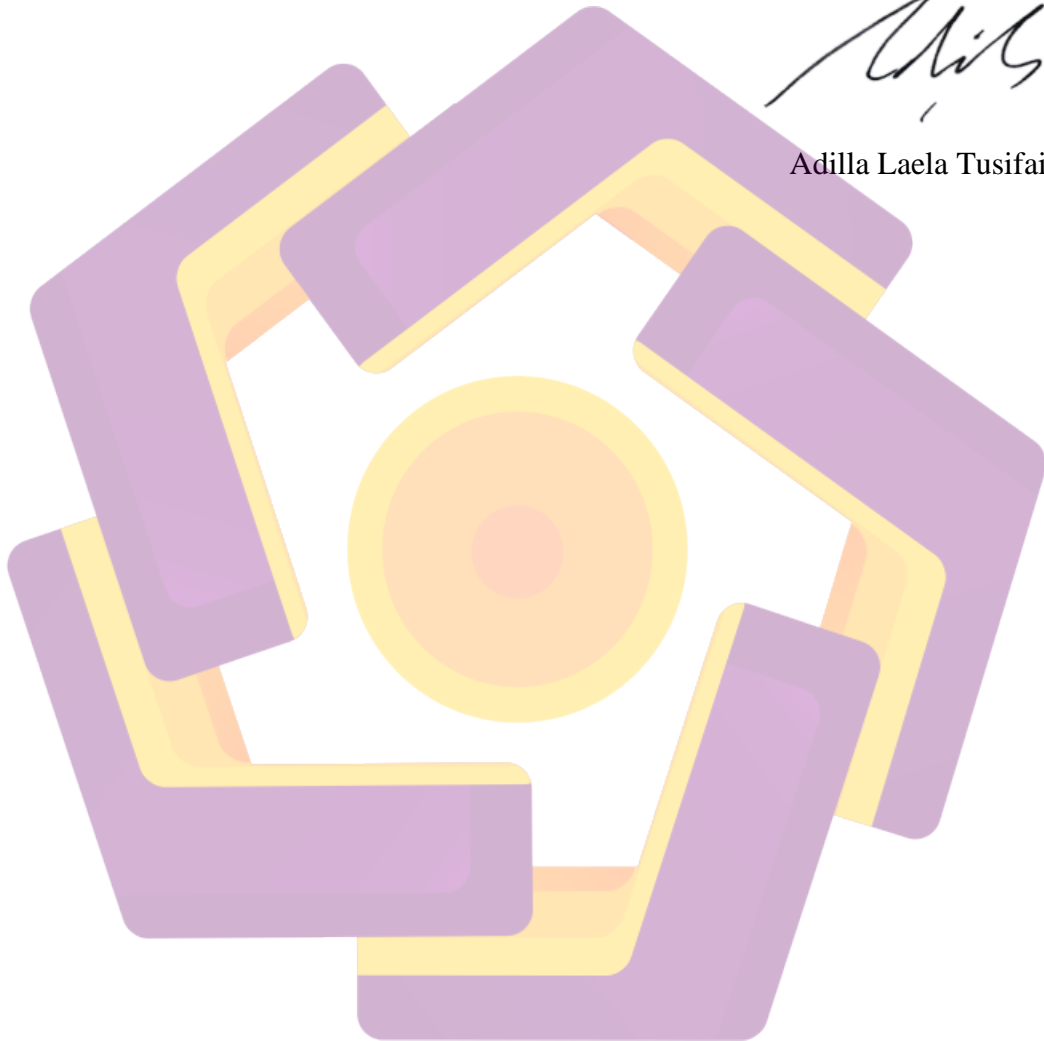
selanjutnya, dan semoga karya tulis ini bisa bermanfaat bagi semua pihak yang membacanya. Aamiin

Wassalamu'alaikum Wr.Wb

Magelang, 20 Maret 2021



Adilla Laela Tusifaiyah



## DAFTAR ISI

JUDUL .....	i
PERSETUJUAN .....	ii
PENGESAHAN .....	iii
PERNYATAAN .....	iiiv
MOTTO .....	v
PERSEMBAHAN .....	vi
KATA PENGANTAR .....	vii
DAFTAR ISI .....	ix
DAFTAR TABEL .....	xiii
DAFTAR GAMBAR .....	xiv
DAFTAR ISTILAH .....	xviii
INTISARI .....	xx
ABSTRACT .....	xxi
BAB I PENDAHULUAN .....	1
1.1 Latar Belakang .....	1
1.2 Rumusan Masalah .....	4
1.3 Batasan Masalah .....	4
1.4 Maksud dan Tujuan Penelitian .....	5
1.5 Manfaat Penelitian .....	5
1.6 Metode Penelitian .....	6

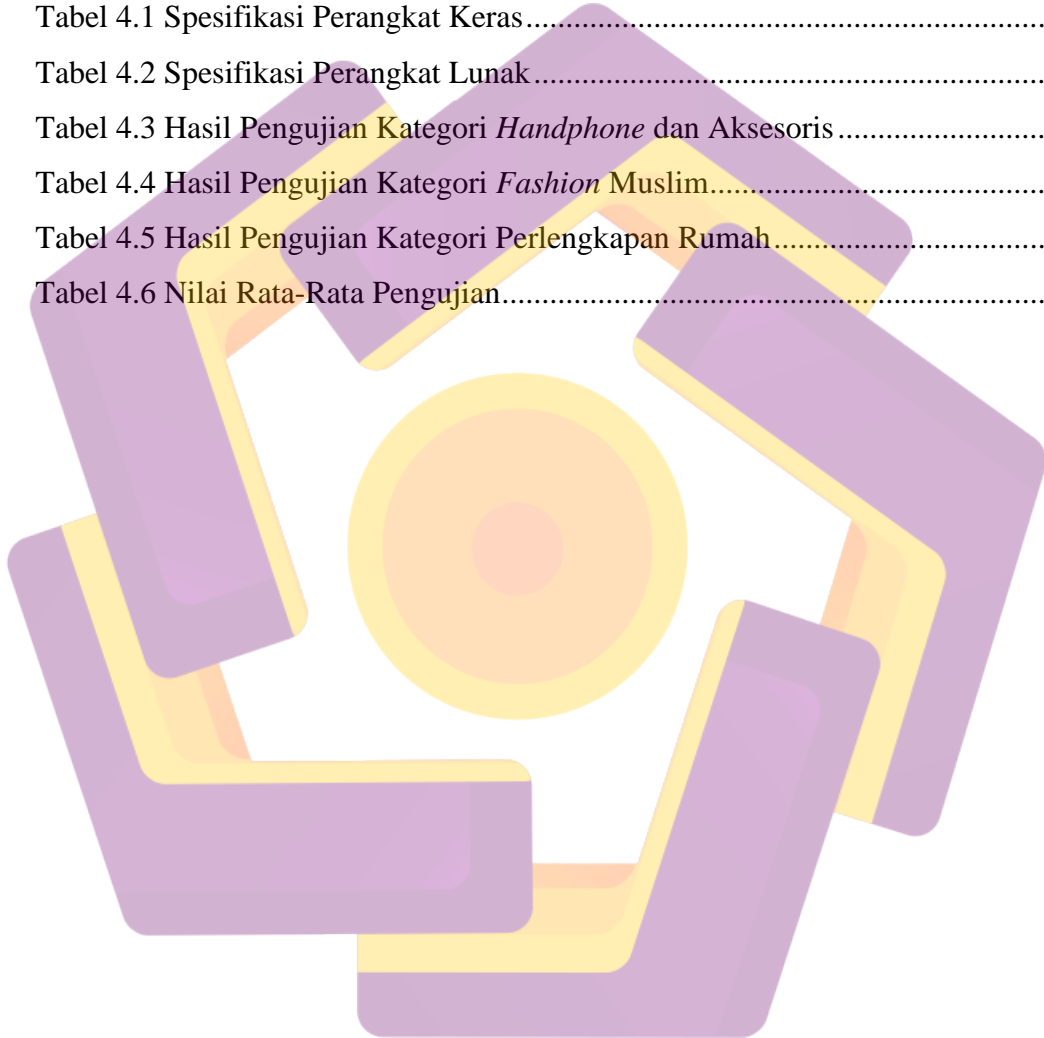
1.6.1	Metode Pengumpulan Data.....	6
1.6.2	Metode Studi Pustaka.....	6
1.6.3	Metode Analisis .....	6
1.6.4	Metode Perancangan .....	6
1.6.5	Metode Implementasi.....	6
1.6.6	Metode Pengujian.....	7
1.7	Sistematika Penulisan .....	8
<b>BAB II LANDASAN TEORI.....</b>		<b>10</b>
2.1	Kajian Pustaka.....	10
2.2	Text Mining.....	16
2.3	Text Pre-Processing .....	17
2.4	Term Frequency-Inverse Document Frequency (TF-IDF) .....	18
2.5	Cosine Similarity.....	19
2.6	Confusion Matrix .....	20
2.7	Classification.....	22
2.7.1	K-Nearest Neighbor .....	22
2.8	E-marketplace .....	23
2.8.1	Keuntungan E-marketplace.....	23
2.9	Shopee.com .....	24
2.10	Web Scraping .....	25
<b>BAB III ANALISA DAN PERANCANGAN SISTEM.....</b>		<b>26</b>
3.1	Pengumpulan Data .....	26
3.2	Perancangan Sistem .....	29
3.2.1	Arsitektur Umum .....	29
3.2.2	Perhitungan Manual Algoritme K-Nearest Neighbor .....	32
3.2.3	Perhitungan Manual Evaluasi Algoritme.....	38

3.3	Analisa Kebutuhan Sistem .....	40
3.3.1	Use Case Diagram.....	40
3.3.2	Activity Diagram.....	40
3.3.3	Class Diagram .....	47
3.3.4	Sequence Diagram .....	47
3.4	Perancangan Flowchart .....	54
3.4.1	Flowchart Kerja Algoritme K-Nearest Neighbor.....	54
3.4.2	Flowchart Sistem.....	55
3.5	Perancangan Antar Muka (interface) .....	55
3.5.1	Rancangan Halaman User.....	55
3.5.2	Rancangan Halaman Admin .....	56
<b>BAB IV IMPLEMENTASI DAN PEMBAHASAN .....</b>		<b>66</b>
4.1	Implementasi Sistem .....	66
4.1.1	Spesifikasi Perangkat Keras (Hardware) .....	66
4.1.2	Spesifikasi Perangkat Lunak (Software).....	66
4.2	Implementasi Database .....	67
4.2.1	Tabel Admin.....	67
4.2.2	Tabel Basic Words .....	67
4.2.3	Tabel Case Folding .....	68
4.2.4	Tabel Data Latih.....	68
4.2.5	Tabel Filtering.....	68
4.2.6	Tabel Index.....	69
4.2.7	Tabel Kelas.....	69
4.2.8	Tabel Kategori.....	69
4.2.9	Tabel Query.....	70
4.2.10	Tabel q_casefolding .....	70
4.2.11	Tabel q_filtering .....	70
4.2.12	Tabel q_index.....	71
4.2.13	Tabel q_stemming.....	71

4.2.14	Tabel q_token.....	71
4.2.15	Tabel Stemming .....	72
4.2.16	Tabel Token .....	72
4.3	Implementasi Antar Muka (interface).....	72
4.3.1	Implementasi Antar Muka User (frontend).....	72
4.3.2	Implementasi Antar Muka Admin (backend) .....	76
4.4	Implementasi Web Scraping .....	85
4.5	Implementasi Klasifikasi Teks Menggunakan Algoritme K-Nearest Neighbor.....	86
4.5.1	Script PHP Preprocessing .....	87
4.5.2	Script PHP Pembobotan TF-IDF .....	89
4.5.3	Script PHP Cosine Similarity.....	90
4.5.4	Script PHP Algoritme K-Nearest Neighbor.....	91
4.6	Hasil Pengujian dan Pembahasan.....	92
4.6.1	Pengujian Algoritme .....	92
BAB V PENUTUP.....		95
5.1	Kesimpulan .....	95
5.2	Saran.....	95
DAFTAR PUSTAKA .....		97

## DAFTAR TABEL

Tabel 2.1 Perbandingan penelitian terkait dengan sistem yang dibangun .....	12
Tabel 2.2 Confusion Matrix .....	20
Tabel 3.1 Confusion Matrix .....	38
Tabel 4.1 Spesifikasi Perangkat Keras .....	66
Tabel 4.2 Spesifikasi Perangkat Lunak .....	66
Tabel 4.3 Hasil Pengujian Kategori <i>Handphone</i> dan Aksesoris .....	92
Tabel 4.4 Hasil Pengujian Kategori <i>Fashion Muslim</i> .....	93
Tabel 4.5 Hasil Pengujian Kategori Perlengkapan Rumah .....	93
Tabel 4.6 Nilai Rata-Rata Pengujian .....	94



## DAFTAR GAMBAR

Gambar 1.1 <i>E-commerce Activities In Indonesia</i> .....	1
Gambar 1.2 <i>E-commerce Spend By Category</i> .....	2
Gambar 3.1 Data Kategori <i>Handphone</i> dan Aksesoris .....	27
Gambar 3.2 Data Kategori <i>Fashion Muslim</i> .....	27
Gambar 3.3 Data Kategori <i>Perlengkapan Rumah</i> .....	28
Gambar 3.4 Kelas Kategori <i>Handphone</i> dan Aksesoris .....	28
Gambar 3.5 Kelas Kategori <i>Fashion Muslim</i> .....	28
Gambar 3.6 Kelas Kategori <i>Perlengkapan Rumah</i> .....	29
Gambar 3.7 Alur Kerja <i>Web Scraping</i> .....	29
Gambar 3.8 <i>Arsitektur Umum</i> .....	30
Gambar 3.9 Data <i>Handphone</i> .....	33
Gambar 3.10 Data Latih.....	33
Gambar 3.11 Data Uji .....	33
Gambar 3.12 Hasil <i>Case Folding</i> .....	34
Gambar 3.13 Hasil <i>Tokenizing</i> .....	34
Gambar 3.14 Hasil <i>Filtering</i> .....	34
Gambar 3.15 Hasil <i>Stemming</i> .....	34
Gambar 3.16 Pembobotan Term .....	35
Gambar 3.17 Perhitungan $WX*WDi$ .....	35
Gambar 3.18 Perhitungan Panjang Vektor.....	36
Gambar 3.19 Perhitungan <i>Cosine Similarity</i> .....	36
Gambar 3.20 Pengurutan Hasil <i>Cosine Similarity</i> .....	37
Gambar 3.21 Data Sebanyak Nilai $k$ .....	37
Gambar 3.22 Klasifikasi <i>K-Nearest Neighbor</i> .....	38
Gambar 3.23 Hasil Klasifikasi Algoritme <i>K-Nearest Neighbor</i> .....	38
Gambar 3.24 <i>Use Case Diagram</i> .....	40
Gambar 3.25 <i>Activity Diagram Login</i> .....	41
Gambar 3.26 <i>Activity Diagram Logout</i> .....	41
Gambar 3.27 <i>Activity Diagram Registrasi</i> .....	42

Gambar 3.28 <i>Activity Diagram</i> Kelola Data Latih .....	43
Gambar 3.29 <i>Activity Diagram</i> Kelola Kelas .....	44
Gambar 3.30 <i>Activity Diagram</i> Kelola Kategori .....	45
Gambar 3.31 <i>Activity Diagram</i> Kelola <i>Basic Words</i> .....	46
Gambar 3.32 <i>Activity Diagram</i> Klasifikasi Produk.....	46
Gambar 3.33 <i>Class Diagram</i> .....	47
Gambar 3.34 <i>Sequence Diagram</i> <i>Login</i> .....	48
Gambar 3.35 <i>Sequence Diagram</i> Registrasi.....	48
Gambar 3.36 <i>Sequence Diagram</i> Logout .....	49
Gambar 3.37 <i>Sequence Diagram</i> Kelola Data Latih .....	50
Gambar 3.38 <i>Sequence Diagram</i> Kelola Kelas .....	51
Gambar 3.39 <i>Sequence Diagram</i> Kelola Kategori .....	52
Gambar 3.40 <i>Sequence Diagram</i> Kelola <i>Basic Words</i> .....	53
Gambar 3.41 <i>Sequence Diagram</i> Klasifikasi Produk.....	53
Gambar 3.42 Flowchart Kerja Algoritme K-Nearest Neighbor.....	54
Gambar 3.43 <i>Flowchart</i> Sistem .....	55
Gambar 3.44 Halaman Beranda .....	56
Gambar 3.45 Halaman Klasifikasi .....	56
Gambar 3.46 Halaman <i>Login</i> .....	57
Gambar 3.47 Halaman Registrasi .....	57
Gambar 3.48 Halaman Beranda .....	58
Gambar 3.49 Halaman Data Latih .....	59
Gambar 3.50 Halaman Kelas .....	59
Gambar 3.51 Halaman Kategori .....	60
Gambar 3.52 Halaman <i>Basic Words</i> .....	60
Gambar 3.53 Halaman <i>Case Folding</i> .....	61
Gambar 3.54 Halaman <i>Tokenizing</i> .....	62
Gambar 3.55 Halaman <i>Filtering</i> .....	62
Gambar 3.56 Halaman <i>Stemming</i> .....	63
Gambar 3.57 Halaman Pembobotan TF - IDF .....	63
Gambar 3.58 Halaman <i>Import Data</i> .....	64



Gambar 3.59 Halaman Pembobotan dan <i>Cosine Similarity</i> .....	65
Gambar 3.60 Halaman Bantuan .....	65
Gambar 4.1 Tabel Admin.....	67
Gambar 4.2 Tabel <i>Basic Words</i> .....	68
Gambar 4.3 Tabel <i>Case Folding</i> .....	68
Gambar 4.4 Tabel Data Latih.....	68
Gambar 4.5 Tabel <i>Filtering</i> .....	69
Gambar 4.6 Tabel <i>Index</i> .....	69
Gambar 4.7 Tabel Kelas.....	69
Gambar 4.8 Tabel Kategori.....	69
Gambar 4.9 Tabel <i>Query</i> .....	70
Gambar 4.10 Tabel <i>q_casefolding</i> .....	70
Gambar 4.11 Tabel <i>q_filtering</i> .....	70
Gambar 4.12 Tabel <i>q_index</i> .....	71
Gambar 4.13 Tabel <i>q_stemming</i> .....	71
Gambar 4.14 Tabel <i>q_token</i> .....	72
Gambar 4.15 Tabel <i>Stemming</i> .....	72
Gambar 4.16 Tabel <i>Token</i> .....	72
Gambar 4.17 Halaman Beranda .....	73
Gambar 4.18 Halaman Klasifikasi Sebelum Proses.....	73
Gambar 4.19 Perhitungan Bobot Halaman <i>User</i> .....	74
Gambar 4.20 Perhitungan <i>Cosine Similarity</i> Halaman <i>User</i> .....	75
Gambar 4.21 Klasifikasi KNN Halaman <i>User</i> .....	76
Gambar 4.22 Halaman <i>Login</i> .....	77
Gambar 4.23 Halaman Registrasi .....	77
Gambar 4.24 Halaman Beranda .....	77
Gambar 4.25 Halaman Data Latih .....	78
Gambar 4.26 Halaman Kelas .....	79
Gambar 4.27 Halaman Kategori .....	79
Gambar 4.28 Halaman <i>Basic Words</i> .....	80
Gambar 4.29 Halaman <i>Case Folding</i> .....	80

Gambar 4.30 Halaman <i>Tokenizing</i> .....	81
Gambar 4.31 Halaman <i>Filtering</i> .....	81
Gambar 4.32 Halaman <i>Stemming</i> .....	82
Gambar 4.33 Pembobotan TF-IDF .....	82
Gambar 4.34 Halaman <i>Import Data</i> .....	83
Gambar 4.35 Halaman Pembobotan dan <i>Cosine Similarity</i> Sebelum Proses .....	83
Gambar 4.36 Halaman Pembobotan dan <i>Cosine Similarity</i> Setelah Proses.....	84
Gambar 4.37 Perhitungan Akurasi .....	85
Gambar 4.38 Halaman Bantuan .....	85
Gambar 4.39 Potongan Script PHP Scraping Website .....	86
Gambar 4.40 User Interface Scraping Website.....	86
Gambar 4.41 <i>Script PHP Case Folding</i> .....	87
Gambar 4.42 <i>Script PHP Tokenizing</i> .....	88
Gambar 4.43 <i>Script PHP Filtering</i> .....	88
Gambar 4.44 <i>Script PHP Stemming</i> .....	89
Gambar 4.45 <i>Script PHP</i> Pembobotan TF-IDF .....	89
Gambar 4.46 <i>Script PHP</i> Perhitungan $WX*WDi$ .....	90
Gambar 4.47 <i>Script PHP</i> Perhitungan Panjang Vektor .....	91
Gambar 4.48 <i>Script PHP</i> Klasifikasi Algoritme <i>K-Nearest Neighbor</i> .....	92

## DAFTAR ISTILAH

### ***E-marketplace***

Pasar online berbasis web yang digunakan sebagai tempat terjadinya jual beli.

### ***Customer engagement***

Proses interaksi yang terjalin antara pelanggan dan pihak perusahaan melalui berbagai media.

### ***Customer loyalty***

Kesetiaan pelanggan dalam menikmati produk dan layanan yang digunakan.

### ***Website***

Kumpulan informasi dalam bentuk halaman web yang saling tergabung di sebuah domain atau URL.

### ***Bag of words***

Sebuah gambaran sederhana yang digunakan dalam pengolahan bahasa alami dan pencarian informasi.

### **HTML (Hypertext Markup Language)**

Bahasa markah standar untuk dokumen yang dirancang untuk ditampilkan di peramban internet.

### **XHTML (eXtensible HyperText Markup Language)**

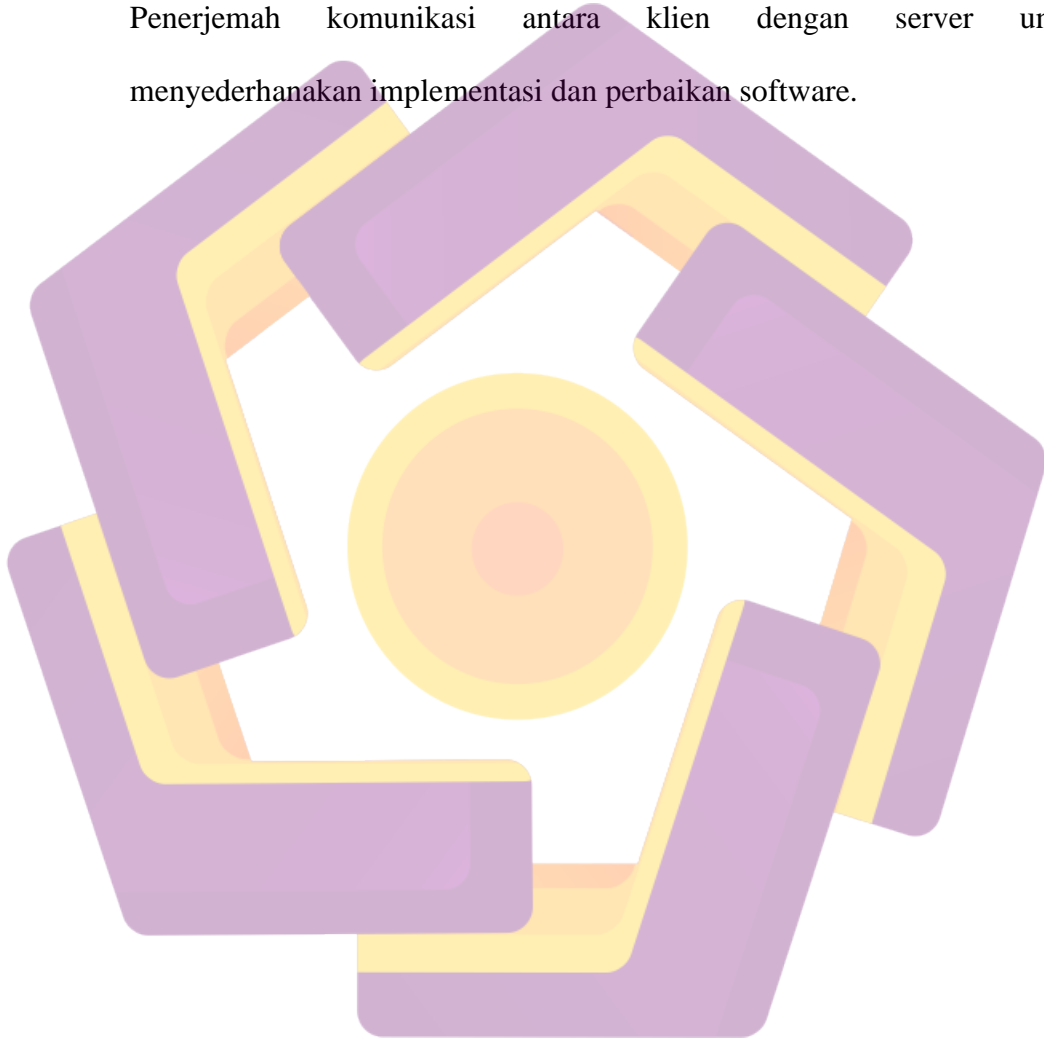
Bahasa markup penerus dan pengembangan dari HTML yang memiliki kemampuan yang kurang lebih mirip HTML, tetapi dengan aturan sintaks yang lebih ketat.

## **UML (Unified Modeling Language)**

Bahasa spesifikasi standar untuk mendokumentasikan, menspesifikasikan, dan membangun sistem perangkat lunak.

## **API (Antarmuka Pemrograman Aplikasi)**

Penerjemah komunikasi antara klien dengan server untuk menyederhanakan implementasi dan perbaikan software.



## INTISARI

Pesatnya pertumbuhan *e-commerce* di Indonesia banyak difasilitasi oleh kehadiran *e-marketplace*. Tren *e-marketplace* di Indonesia terus berkembang seiring perkembangan teknologi dan internet. Dalam perkembangannya, *e-marketplace* semakin banyak menawarkan produk. Akibatnya, pembeli membutuhkan lebih banyak usaha untuk mencari produk yang diinginkan. Agar dapat memudahkan pencarian produk tersebut, maka dilakukan klasifikasi produk.

Penelitian ini mengklasifikasikan produk pada *e-marketplace* Shopee menggunakan algoritme *K-Nearest Neighbor*. Data produk yang digunakan berasal dari *web scraping* pada kategori *handphone* dan aksesoris, *fashion muslim*, dan perlengkapan rumah. Tahapan dari sistem klasifikasi dimulai dengan tahap *preprocessing*, selanjutnya tahap pembobotan term menggunakan metode TF-IDF, berikutnya *cosine similarity* untuk menghitung jarak kemiripan antar dokumen, dan kemudian melakukan pengurutan hasil *cosine similarity* untuk diambil data sejumlah nilai  $k$ .

Berdasarkan pengujian pada 9 data produk dengan tiga nilai  $k$  berbeda. Diperoleh rata-rata yang menunjukkan hasil *accuracy*, *precision*, dan *recall* terendah ketika nilai  $k=3$ . Hasil *accuracy* sebesar 88.89%, *precision* sebesar 83.33%, dan *recall* sebesar 100% didapatkan ketika menggunakan nilai  $k=5$  atau  $k=7$ .

**Kata Kunci :** Klasifikasi, *E-marketplace*, *K-Nearest Neighbor*, *Text Mining*

## **ABSTRACT**

*The rapid growth of e-commerce in Indonesia has been largely facilitated by the presence of e-marketplaces. The e-marketplace trend in Indonesia continues to develop along with the development of technology and the internet. During its development, e-marketplaces offer more and more products. As a result, buyers need more effort to find the product they want. In order to facilitate the search for these products, a product classification is carried out.*

*This study classifies products in the Shopee e-marketplace using the K-Nearest Neighbor algorithm. The product data used comes from web scraping in the categories of cellphones and accessories, Muslim fashion, and home appliances. The stages of the classification system begin with the preprocessing stage, then the term weighting stage uses the TF-IDF method, then cosine similarity to calculate the similarity distance between documents, and then sorting the results of the cosine similarity to retrieve data for the number of k values.*

*Based on testing on 9 product data with three different k values. Obtained an average that shows the lowest accuracy, precision, and recall results when the value of  $k = 3$ . The accuracy result is 88.89%, precision is 83.33%, and a recall of 100% is obtained when using the value of  $k = 5$  or  $k = 7$ .*

**Keyword:** *Classification, E-marketplace, K-Nearest Neighbor, Text Mining*

