

Rancang Bangun *Website* Klasifikasi Untuk Pencarian Produk Pasar *Online* Menggunakan Algoritma *K-Nearest Neighbor*

Danny Sebastian^{#1}

¹Teknik Informatika, Universitas Kristen Duta Wacana
Jln Dr Wahidin Sudirohusodo 5-25, Yogyakarta

¹danny.sebastian@staff.ukdw.ac.id

Abstract— In the past few years, online marketplace has improved so fast. There are many products listed at online marketplace. This growth makes a new problem for buyers, it's too much information. Because of many online marketplaces, the buyers might confuse to select the best one. Classification is a method for categorize product into related class. *K-Nearest Neighbor* is a powerful and efficient method for classification. From the problems, author gets the idea for making a website that could classified products and help user to search multiple online marketplace at once, and help the user to find the desired products. This study will produce the website design.

Keywords— classification, *K-Nearest Neighbor*, e-commerce, online marketplace

I. PENDAHULUAN

A. Latar belakang

Pasar yang memanfaatkan media *online* atau *online marketplace* di Indonesia sangat berkembang, dan terus bertambah jumlahnya dalam beberapa tahun terakhir [1]. Beberapa pasar *online* yang banyak digunakan orang Indonesia adalah Bukalapak, Tokopedia, Elevation, JD.id, dan mataharimall. Dengan semakin banyak nya pasar *online*, setiap orang dapat dengan mudah melakukan pemasaran secara *online*. Hal ini mengakibatkan jumlah penjual yang melakukan pemasaran secara *online* semakin banyak jumlahnya.

Pada sisi lain, muncul permasalahan dari sisi pembeli. Dimana calon pembeli mengalami kesulitan untuk melakukan pencarian produk yang hendak dibeli karena banyaknya produk yang dijual di pasar *online*.

Kesulitan berbelanja *online* dikarenakan banyaknya pasar *online* dan penjual yang menggunakan pasar *online* tersebut. Kesulitan mencari produk juga dapat muncul dari kesalahan penjual, misalnya pada penamaan sebuah produk yang tidak sesuai, pemberian kategori/pengelompokan produk yang tidak sesuai.

Ketika seorang pembeli mencari sebuah produk, muncul banyak sekali produk yang memiliki judul/nama tidak sesuai

dan belum tentu sesuai dengan keinginan pembeli. Sebagai contoh, ketika mencari sebuah handphone dengan kata kunci "Xiaomi Redmi Note 3 Pro", produk yang dapat muncul adalah semua produk dengan kata kunci "Xiaomi", "Redmi", "Note", dan "Pro". Sehingga pengguna perlu usaha lebih untuk mencari produk yang memang sesuai dengan yang diinginkan.

Klasifikasi merupakan metode untuk mengelompokkan dokumen kedalam kelas-kelas yang sesuai dengan kriteria. Salah satu metode melakukan klasifikasi adalah *K-Nearest Neighbor* atau *KNN*.

Dengan memanfaatkan klasifikasi, produk dari pasar *online* dapat dikelompokkan secara otomatis kedalam kelas-kelas yang sesuai. Pembeli hanya perlu mencari produk apa yang diinginkan melalui kelas yang sudah ditentukan. Produk yang sudah dikelompokkan dapat membantu pembeli dalam mencari produk yang hendak dibeli.

Berdasarkan permasalahan yang ada, perlu dikembangkan sebuah *website* yang dapat memudahkan pengguna dalam melakukan pencarian produk dengan cara mengelompokkan produk yang dijual pasar *online* berdasarkan kelas-kelas produk untuk memudahkan pencarian oleh pengguna.

B. Rumusan Masalah

Berdasarkan uraian latar belakang diatas, maka rumusan masalah yang mendasari penelitian ini adalah:

1. Sulitnya melakukan pencarian produk secara langsung ke beberapa *website* pasar *online*.
2. Diperlukan pengelompokan produk secara otomatis yang memudahkan pencarian di pasar *online*.
3. Diperlukan pemodelan sebelum membangun sistem *website*.

C. Tujuan Penelitian

Tujuan penelitian ini adalah:

1. Melakukan perancangan *website* klasifikasi yang mempermudah proses pencarian produk di pasar *online*.
2. Melakukan perancangan sistem klasifikasi produk

yang dijual di pasar *online* secara otomatis.

II. METODE PENELITIAN

Penelitian ini merupakan tahap awal dalam pengembangan keseluruhan *website* klasifikasi. Secara keseluruhan, penelitian ini akan menggunakan metode *System Development Lifecycle* dengan pendekatan *waterfall model*.

Pada penelitian ini metode penelitian yang digunakan terdiri dari 2 cara, yaitu:

1. Requirement Gathering

Requirement gathering atau pengumpulan daftar kebutuhan sistem *website* bertujuan untuk mendapatkan kebutuhan fungsional sistem *website* yang akan dibangun.

Metode yang digunakan pada tahap ini adalah studi pustaka dan observasi. Studi pustaka dilakukan dengan cara mempelajari bahan tertulis terkait penelitian ini. Sedangkan pengamatan atau observasi dilakukan dengan cara melihat data produk di pasar *online*, khususnya tokopedia dan bukalapak.

2. Design

Design atau perancangan bertujuan membuat perancangan sistem *website* yang akan dibangun. Proses perancangan dilakukan dengan memodelkan kebutuhan sistem *website*. Pemodelan yang akan dilakukan adalah *usecase*, *activity diagram*, dan *entity relationship diagram (ERD)*. Selain itu pada tahap desain, akan dilakukan perancangan *input* dan *output* sistem *website*.

III. TINJAUAN PUSTAKA

Klasifikasi merupakan salah satu permasalahan yang ada di dunia digital karena cepatnya pertumbuhan *data digital* [2]. Oleh karena itu dibutuhkan otomatisasi dalam melakukan klasifikasi. *K-Nearest Neighbor* atau *KNN* merupakan metode yang banyak digunakan dalam melakukan klasifikasi teks. Metode *KNN* dan *TF-IDF* memberikan hasil yang baik dalam proses klasifikasi dokumen secara otomatis.

K-Nearest Neighbor dapat digunakan untuk melakukan klasifikasi dokumen berita [3]. Disini metode *KNN* digunakan untuk melengkapi atau menyelesaikan proses klustering *Suffix Tree Clustering (STC)*.

K-Nearest Neighbor pernah diteliti penggunaannya untuk kasus identifikasi biometric telapak kaki manusia[4]. Hasil dari penelitian ini adalah metode *KNN* dengan pendekatan *euclidian distance* dan *cosine distance* memiliki akurasi terbesar bernilai 98%.

Penelitian yang dilakukan oleh Andro [5] pada tahun 2017 melakukan klasifikasi komentar pada *dataset* pemilu presiden Indonesia 2014. Metode yang digunakan adalah *improved K-Nearest Neighbor*. Hasil dari penelitian ini adalah *improved KNN* meningkatkan akurasi sebesar 0.49% dibandingkan *KNN*.

Penelitian aktual akan menggunakan metode klasifikasi *K-Nearest Neighbor* untuk melakukan klasifikasi produk yang dijual di pasar *online*. Perbedaan penelitian aktual dengan penelitian terdahulu dapat dilihat pada TABEL I.

TABEL I
PERBEDAAN PENELITIAN AKTUAL DENGAN PENELITIAN TERDAHULU

Judul	Penggunaan <i>KNN</i> untuk Klasifikasi Teks Berita yang tak-terkelompokkan pada saat Pengklasteran oleh <i>STC</i>	Analisis dan Implementasi Klasifikasi <i>K-Nearest Neighbor</i> pada system identifikasi Biometrik Telapak Kaki Manusia	Klasifikasi Komentar pada Dataset Pemilu Presiden Indonesia 2014 dengan Metode <i>Improved K-Nearest Neighbor</i>	Rancang Bangun <i>Website</i> Klasifikasi Untuk Pencarian Produk Pasar <i>Online</i> Menggunakan Metode <i>K-Nearest Neighbor</i>
Oleh	Edi Winarko Jumadi	Armanda Nur Fadillah, Ledy Novamizanti, dan Ratri Dwi Atmaja	Andro Ardiyanto	Danny Sebastian
Tahun	2015	2015	2017	2017
Metode	<i>K-Nearest Neighbor</i>	<i>K-Nearest Neighbor</i>	<i>Improved K-Nearest Neighbor</i> dan <i>K-Nearest Neighbor</i>	<i>K-Nearest Neighbor</i>
Objek	Teks berita	Citra biometric telapak kaki manusia	Data sentipol Pemilu presiden 2014	Produk yang dijual di pasar <i>online</i>
Akurasi	Tidak disebutkan	Akurasi terbaik 98%	<i>Improved KNN</i> :66.62% <i>KNN</i> : 66.13%	-

IV. LANDASAN TEORI

A. Klasifikasi

Klasifikasi teks merupakan kegiatan untuk menentukan sebuah dokumen tergabung kedalam sebuah kelas dokumen

[6,7]. Kegiatan mengklasifikasikan dokumen memiliki tantangan dari sisi besarnya jumlah dokumen, hal ini membuat *automatic classification* (klasifikasi yang dilakukan oleh komputer) perlu dikembangkan [6].

Berdasarkan jumlah kelas terdapat 2 tipe klasifikasi [7], yaitu:

- *Binary Classification*, yaitu melakukan klasifikasi objek ke salah satu dari 2 kelas yang ditentukan.
- *Multi-class classification*, yaitu melakukan klasifikasi objek ke satu atau lebih kelas.

Proses klasifikasi teks memerlukan pemahaman bagaimana cara melakukan klasifikasi dokumen, oleh karena itu input object dibagi menjadi *training* dan *testing data*[7]. *Training data* adalah dokumen yang sudah diberi label, sedangkan *testing data* adalah dokumen yang belum diberi label. Tujuannya adalah mendapatkan pengetahuan cara melakukan klasifikasi dari *training data* dan menerapkannya ke *testing data* secara akurat.

B. K-Nearest Neighbor

K-Nearest Neighbor atau *KNN* merupakan metode klasifikasi dengan berdasarkan jarak data baru ke beberapa data atau tetangga terdekat [8]. Pendekatan *Nearest Neighbor* adalah pendekatan untuk mencari kasus dengan menghitung kedekatan antara kasus baru dengan kasus lama, yaitu berdasarkan pada pencocokan bobot dari sejumlah fitur yang ada[8,9]. Tujuan dari algoritma *Nearest Neighbor*, jarak antara satu data dengan data yang lain dapat dihitung. Nilai jarak hasil perhitungan dapat digunakan sebagai nilai kedekatan atau kemiripan antara data uji dengan data latih [8].

Nilai *K* pada *Nearest Neighbor* berarti *K*-data terdekat dari data uji [8]. Jika *K* bernilai 2, akan diambil 2 tetangga terdekat dari data latih, begitu juga bila *K* bernilai *n* maka akan diambil sejumlah *n* tetangga terdekat dari data latih. Permasalahan dari metode *nearest neighbor* adalah pemilihan nilai *K* yang tepat.

Pengkategorian teks adalah proses pengelompokan document teks menjadi satu atau lebih kategori yang telah ditetapkan berdasarkan isi kontennya[8,10]. Sejumlah teknik klasifikasi statistik dan mesin pembelajaran telah ditetapkan untuk pengkategorian teks. Langkah pertama dalam pengkategorian teks adalah mengubah dokumen yang biasanya merupakan string karakter menjadi sebuah representasi yang cocok untuk algoritma pembelajaran dan tugas pengklasifikasian.

Langkah dari metode *KNN* [5] adalah:

1. Diasumsikan ada *j* training kategori (C_1, C_2, \dots, C_j) dan total seluruh sampel training adalah *N*. Kedua komponen tersebut menjadi *m-dimension feature vector*, yaitu sebuah bidang *vector* dimana hasil klasifikasinya akan ditandai di bidang tersebut.
2. Jadikan sampel *X* menjadi sama dengan *feature selection vector* pada bidang tersebut (X_1, X_2, \dots, X_m)
3. Hitung kesamaan antara seluruh training sampel dengan *X*. Untuk menghitung kesamaan dapat dilakukan dengan rumus *Cosine Similarity* dibawah ini[11,5].

$$\text{CosSim}(\vec{x}, \vec{y}) = \frac{\vec{x} \cdot \vec{y}}{|\vec{x}| |\vec{y}|} = \frac{\sum_{i=1}^n x_i y_i}{\sqrt{\sum_{i=1}^n x_i^2} \cdot \sqrt{\sum_{i=1}^n y_i^2}} \quad (1)$$

Keterangan:

x_i = Data uji

y_i = Data latih
 n = Jumlah data

4. Lakukan perhitungan kecenderungan dari *X* untuk setiap kategori menggunakan rumus dibawah ini [12,5].

$$P(X, C_j) = \sum_d \text{CosSim}(x, d_i) \cdot y(y_i, C_j) \quad (2)$$

Keterangan:

$y(y_i, C_j)$ = fungsi kategori atribut

5. Tentukan *X* masuk pada kategori apa berdasarkan bilangan terbesar pada perhitungan $P(X, C_j)$.

C. Unified Modelling Language (UML)

Unified Modelling Language atau *UML* adalah metode untuk melakukan pemodelan dalam *software engineering* [13]. *UML* merupakan cara untuk menstandarisasi model dengan tujuan memudahkan pemahaman orang yang membaca. *UML* diciptakan oleh tiga *modelling specialist*, James Rumbaugh, Grady Booch, dan Ivar Jacobson. *UML* berbasis objek dapat digunakan pada berbagai macam bahasa pemrograman dan teknologi [14].

UML memiliki 3 kategori diagram, yaitu[15]:

1. Structural Diagram

Menggambarkan elemen dari spesifikasi yang mengabaikan waktu. Terdiri atas: *class diagram*, *object diagram*, *deployment diagram*, *composite structure diagram* dan *package diagram*.

2. Behavior Diagram

Menggambarkan ciri-ciri *behavior* atau *method* atau fungsi dari sebuah sistem atau bisnis proses. Terdiri dari: *usecase diagram*, *activity diagram*, dan *state machine diagram*.

3. Interaction Diagram

Merupakan bagian dari *Behavior diagram* yang menggambarkan interaksi objek. Terdiri dari *communication diagram*, *interaction overview diagram*, *sequence diagram*, dan *timing diagram*.

D. Usecase

Usecase merupakan metode pemodelan yang menunjukkan interaksi antara sistem dengan pengguna. *Usecase diagram* menguraikan siapa saja yang akan menggunakan sistem dan dengan cara apa pemakai dapat saling berhubungan dengan sistem [16]. Dalam pemodelan *usecase* terdapat 2 hal penting, yaitu *usecase diagram* dan deskripsi *usecase diagram*.

Komponen dari *usecase diagram* adalah:

1. Aktor

Aktor atau *actor* merupakan *user* yang menggunakan sistem. Aktor dapat berupa manusia atau sistem yang lain [13].

2. Relationship

Relationship atau relasi menggambarkan hubungan antara aktor dengan sistem, hubungan aktor dengan aktor, dan hubungan antara *usecase* dengan *usecase* lain.

Deskripsi *usecase* berisi semua informasi yang dibutuhkan untuk menghasilkan diagram use case. Deskripsi

usecase berisi langkah-langkah yang dilakukan untuk mencapai sebuah hasil akhir. Pada deskripsi usecase terdapat beberapa komponen, yaitu nama usecase, aktor, hasil akhir, prasyarat, tahapan, dan ekstensi.

E. Activity diagram

Activity diagram digunakan untuk memodelkan sebuah aliran proses [13]. Activity diagram berisikan kumpulan action dan control flows. Action merupakan sebuah langkah kegiatan, dan digambarkan sebagai rounded rectangle. Sedangkan control flow adalah aliran proses dari awal hingga akhir dan digambarkan sebagai tanda panah. Simbol-simbol yang digunakan untuk menggambarkan activity diagram adalah:

1. *Start point*, digunakan untuk menggambarkan titik awal proses
2. *End point*, digunakan untuk menggambarkan titik akhir proses
3. *Activity*, menggambarkan sebuah aktivitas atau kegiatan bisnis.
4. *Fork*, digunakan untuk menggambarkan proses yang dikerjakan secara paralel.
5. *Decision point*, digunakan untuk menggambarkan proses percabangan
6. *Swimlane*, digunakan untuk menunjukkan siapa melakukan apa pada activity diagram.

F. Entity Relationship Diagram

Entity Relationship Diagram atau ERD merupakan notasi grafis dalam memodelkan data konseptual yang berhubungan dengan penyimpanan data [14,17]. Komponen dari ERD adalah [17]:

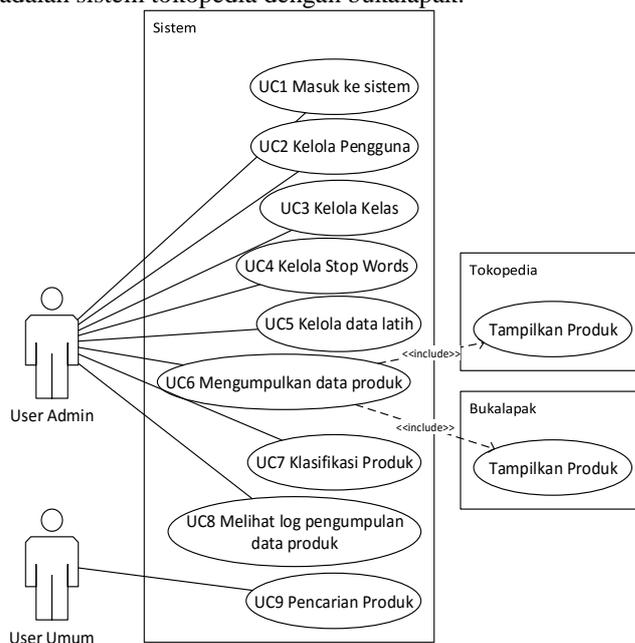
1. *Entities*
Entity merupakan objek yang ada dalam suatu sistem nyata maupun abstrak dimana data tersimpan dan diberi nama dengan kata benda. Didalam perancangan database, merupakan tabel. Sedangkan dalam pemodelan menggunakan ERD, digambarkan dengan persegi panjang.
2. *Attributes*
Attributes merupakan karakteristik dari entitas. Dalam perancangan database attribute merupakan kolom tabel.
3. *Relationship*
Relationship merupakan hubungan antar entitas, digambarkan menggunakan garis. Terdapat 2 cara memodelkan relationship, yaitu menggunakan notasi *chen* atau menggunakan *crow's foot*.

V. HASIL PENELITIAN

A. Usecase

Usecase diagram menggambarkan interaksi antara aktor dengan sistem yang akan dibangun. Dalam perancangan ini, aktor dari website yang akan dibangun adalah *user admin*

dan *user umum*. Sedangkan sistem eksternal yang ada adalah sistem tokopedia dengan bukalapak.



Gambar 1. Usecase diagram website

Dari Gambar 1 diatas adalah diagram usecase website yang akan dibangun. Deskripsi usecase diperlukan dalam proses perancangan karena memberikan gambaran detail mengenai interaksi antara aktor dengan sistem yang akan dibangun.

Tahapan utama dalam menggunakan sistem website ini adalah proses klasifikasi produk. Ada beberapa proses/tahapan yang harus dilakukan sebelum dapat melakukan klasifikasi, yaitu:

1. Menentukan kelas klasifikasi apa saja yang ada. Langkah mengelola kelas klasifikasi dapat dilihat pada TABEL II.
2. Menentukan stop words apa saja yang digunakan. Langkah mengelola stop words dapat dilihat pada TABEL III.
3. Menentukan data latih yang digunakan pada masing-masing kelas. Langkah mengelola data latih dapat dilihat pada TABEL IV.
4. Mengumpulkan data produk dari pasar online. Langkah mengumpulkan data produk dari pasar online dapat dilihat pada TABEL V.

Setelah keempat proses/tahapan sudah dilakukan, maka proses klasifikasi dapat dilakukan. Langkah dalam melakukan klasifikasi dapat dilihat pada TABEL VI.

TABEL II
DESKRIPSI USECASE KELOLA KELAS KLASIFIKASI

Nama use case	:	UC 3 Kelola Kelas Klasifikasi
Aktor	:	User Admin
Hasil akhir	:	• Admin berhasil melakukan pengelolaan

		kelas
Prasyarat	:	• Sudah masuk kedalam sistem
Tahapan	:	1. <i>User</i> mengakses menu kelola kelas 2. Sistem menampilkan halaman pencarian kelas 3. Sistem menampilkan daftar kelas 4. <i>User</i> melakukan pengelolaan kelas
Ekstensi	:	4.A. <i>User</i> memilih menu tambah kelas 4.A.1. Sistem menampilkan halaman form kelas 4.A.2. <i>User</i> mengisikan data kelas 4.A.3 Sistem menampilkan pesan sukses menambahkan kelas Use case selesai. 4.B. <i>User</i> memilih menghapus salah satu kelas 4.B.1. Sistem menampilkan pesan konfirmasi penghapusan 4.B.2. <i>User</i> mengkonfirmasi penghapusan 4.B.3. Sistem menampilkan pesan sukses menghapus kelas Use case selesai.

TABEL III
DESKRIPSI USECASE KELOLA STOP WORDS

Nama use case	:	UC 4 Kelola Stop Words
Aktor	:	<i>User Admin</i>
Hasil akhir	:	• <i>Admin</i> berhasil melakukan pengelolaan stop words
Prasyarat	:	• Sudah masuk kedalam sistem
Tahapan	:	1. <i>User</i> mengakses menu kelola stop words 2. Sistem menampilkan halaman pencarian stop words 3. Sistem menampilkan daftar stop words 4. <i>User</i> melakukan pengelolaan stop words
Ekstensi	:	4.A. <i>User</i> memilih menu tambah stop words 4.A.1. Sistem menampilkan halaman form stop words 4.A.2. <i>User</i> mengisikan data stop words 4.A.3 Sistem menampilkan pesan sukses menambahkan stop words Use case selesai. 4.B. <i>User</i> memilih menghapus salah satu stop words 4.B.1. Sistem menampilkan pesan konfirmasi penghapusan 4.B.2. <i>User</i> mengkonfirmasi penghapusan 4.B.3. Sistem menampilkan pesan sukses menghapus stop words Use case selesai.

TABEL IV
DESKRIPSI USECASE KELOLA DATA LATIH

Nama use case	:	UC 5 Kelola data latihan
Aktor	:	<i>User Admin</i>
Hasil akhir	:	• <i>Admin</i> berhasil melakukan pengelolaan data latihan
Prasyarat	:	• Sudah masuk kedalam sistem
Tahapan	:	1. <i>User</i> mengakses menu kelola data latihan 2. Sistem menampilkan halaman pencarian

		data latihan 3. Sistem menampilkan daftar data latihan 4. <i>User</i> melakukan pengelolaan data latihan
Ekstensi	:	4.A. <i>User</i> memilih menu tambah data latihan 4.A.1. Sistem menampilkan halaman form data latihan 4.A.2. <i>User</i> mengisikan data data latihan 4.A.3 Sistem menampilkan pesan sukses menambahkan data latihan Use case selesai. 4.B. <i>User</i> memilih menghapus salah satu data latihan 4.B.1. Sistem menampilkan pesan konfirmasi penghapusan 4.B.2. <i>User</i> mengkonfirmasi penghapusan 4.B.3. Sistem menampilkan pesan sukses menghapus data latihan Use case selesai.

TABEL V
DESKRIPSI USECASE MENGUMPULKAN DATA DARI PASAR ONLINE

Nama use case	:	UC 6 Mengumpulkan data dari pasar online
Aktor	:	<i>User Admin</i>
Hasil akhir	:	• Data produk berhasil dikumpulkan dari pasar online
Prasyarat	:	• Sudah masuk kedalam sistem
Tahapan	:	1. <i>User</i> memilih menu pengumpulan data 2. Sistem menampilkan form pengumpulan data 3. <i>User</i> memilih pasar online, kategori dan mengisikan halaman yang akan diambil datanya. 4. Sistem mengumpulkan data dari pasar online, kategori, dan halaman yang sesuai 5. Sistem menampilkan hasil pengumpulan data.
Ekstensi	:	3.A. Data halaman tidak sesuai 3.A.1. Sistem menampilkan form pengumpulan data dan menampilkan pesan error. 3.A.2. <i>User</i> Mengisikan halaman yang benar Langkah 3.A.1 sampai 3.A.2 diulang sampai pengisian data sesuai. Melanjutkan ke langkah 4 use case. 4.A Mengalami gangguan jaringan pada saat mengakses pasar online 4.A.1. Sistem membatalkan proses pengumpulan data 4.A.2. Sistem menampilkan pesan error. Usecase selesai.

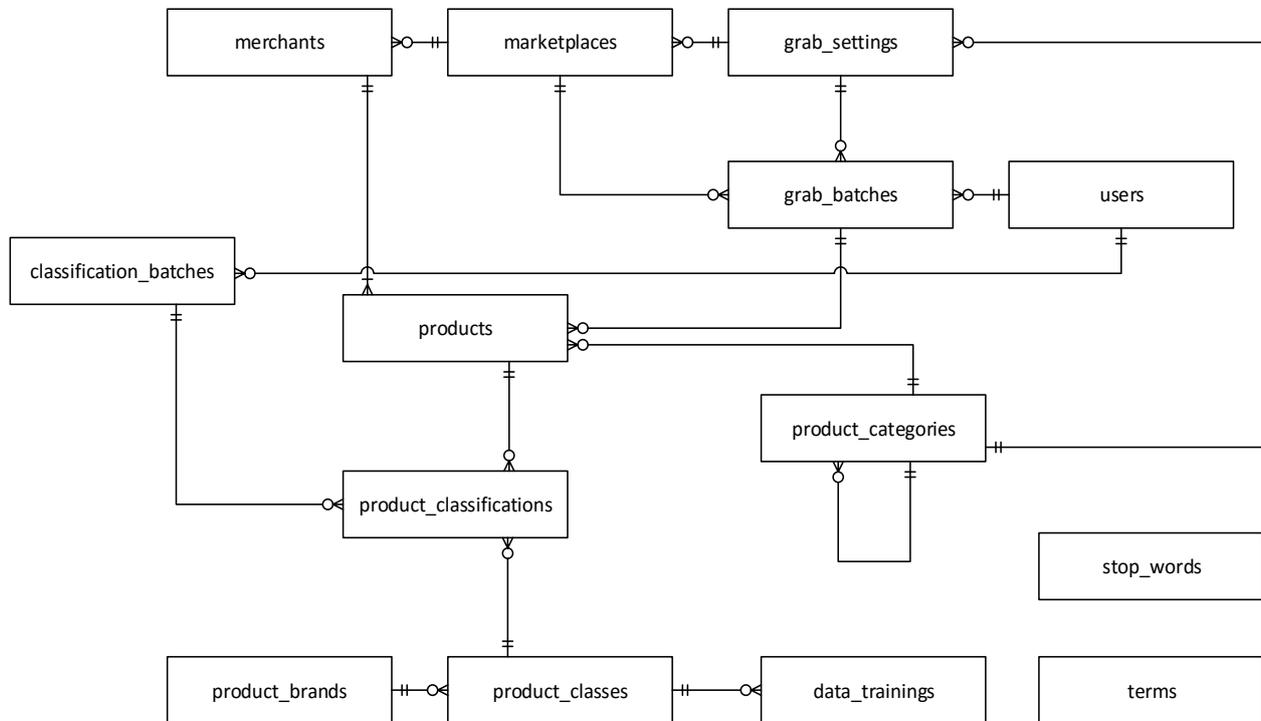
TABEL VI
DESKRIPSI USECASE PROSES KLASIFIKASI PRODUK

Nama use case	:	UC 8 Proses klasifikasi produk
Aktor	:	<i>User Admin</i>
Hasil akhir	:	• Data produk berhasil diklasifikasi sesuai kelas yang ditentukan
Prasyarat	:	• Sudah masuk kedalam sistem • Sudah melakukan pengumpulan data produk
Tahapan	:	1. <i>User</i> memilih menu klasifikasi produk 2. Sistem menampilkan halaman index

	3. klasifikasi produk 4. <i>User</i> memilih tombol tambah klasifikasi 5. Sistem menampilkan halaman form klasifikasi produk 6. <i>User</i> memilih produk yang akan diklasifikasi, kemudian memilih tombol proses klasifikasi 7. Sistem melakukan klasifikasi dan menampilkan hasil klasifikasi.
--	---

B. Perancangan Basis Data

Perancangan basis data akan berisikan *Entity Relationship Diagram* (ERD) dan struktur tabel yang digunakan pada aplikasi. *Entity Relationship Diagram* (ERD) sistem yang akan dibangun, dapat dilihat pada Gambar 2. ERD pada penelitian ini digambarkan secara konseptual, dan dilengkapi dengan struktur tabel.



Gambar 2. Entity Relationship Diagram (ERD) website yang akan dibangun

Berdasarkan ERD diatas, terdapat 13 tabel yang ada didalam sistem *website* yang akan dibangun. Tabel VII sampai Tabel XIX menunjukkan struktur tabel *website* yang akan dibangun.

TABEL VII merupakan struktur tabel yang digunakan untuk menyimpan data pengguna yang dapat masuk kedalam sistem *website*.

TABEL VII
STRUKTUR TABEL *USERS*

Nama kolom	Tipe data	Keterangan
id	int (10)	PRIMARY KEY, AUTO INCREMENT
name	varchar (191)	
email	varchar (191)	
password	varchar (191)	
role	enum	Enum value 'ADMIN', 'EDITOR'
remember_token	varchar(100)	
created_at	timestamp	
updated_at	timestamp	

Tabel VIII merupakan struktur tabel yang digunakan untuk menyimpan data pasar *online* yang dapat diakses oleh sistem *website*.

TABEL VIII
STRUKTUR TABEL *MARKETPLACES*

Nama kolom	Tipe data	Keterangan
id	int (10)	PRIMARY KEY, AUTO INCREMENT
name	varchar (191)	
code	varchar (191)	
url	varchar (191)	
base_grab_url	text	
deleted_at	timestamp	
created_at	timestamp	
updated_at	timestamp	

TABEL IX merupakan struktur tabel yang digunakan untuk menyimpan data *stop words* atau kata-kata yang akan dihilangkan dari proses klasifikasi.

TABEL IX
STRUKTUR TABEL *STOP WORDS*

Nama kolom	Tipe data	Keterangan
id	int (10)	PRIMARY KEY, AUTO INCREMENT
term	varchar (191)	
deleted_at	timestamp	
created_at	timestamp	
updated_at	timestamp	

TABEL X merupakan struktur tabel yang digunakan untuk menyimpan data *merchants* atau penjual yang terdaftar di *sistem website*.

TABEL X
STRUKTUR TABEL *MERCHANTS*

Nama kolom	Tipe data	Keterangan
id	int (10)	PRIMARY KEY; AUTO INCREMENT
marketplace_id	int (10)	INDEX
marketplace_merchant_id	int (10)	
name	varchar (191)	
url	varchar (191)	
status	enum	INDEX; enum value 'GOLD', 'PREMIUM', 'NONE'
deleted_at	timestamp	
created_at	timestamp	
updated_at	timestamp	

TABEL XI merupakan struktur tabel yang menyimpan informasi merk dari produk yang diambil dari pasar *online*. Merk merupakan bagian dari alamat yang digunakan untuk mengumpulkan data produk dari pasar *online*.

TABEL XI
STRUKTUR TABEL *PRODUCT_BRANDS*

Nama kolom	Tipe data	Keterangan
id	int (10)	PRIMARY KEY; AUTO INCREMENT
name_id	varchar (191)	
deleted_at	timestamp	
created_at	timestamp	
updated_at	timestamp	

TABEL XII merupakan struktur tabel yang menyimpan daftar kelas yang digunakan untuk klasifikasi. Daftar kelas klasifikasi pada kasus ini adalah nama produk.

TABEL XII
STRUKTUR TABEL *PRODUCT_CLASSES*

Nama kolom	Tipe data	Keterangan
id	int (10)	PRIMARY KEY; AUTO INCREMENT
product_brand_id	int (10)	
name	varchar (191)	
deleted_at	timestamp	
created_at	timestamp	
updated_at	timestamp	

TABEL XIII merupakan struktur tabel yang menyimpan informasi kategori produk yang ada di pasar *online*. Kategori produk digunakan untuk mendapatkan alamat pasar *online* dalam proses pengumpulan *data* produk.

TABEL XIII
STRUKTUR TABEL *PRODUCT_CATEGORIES*

Nama kolom	Tipe data	Keterangan
id	int (10)	PRIMARY KEY; AUTO INCREMENT
name	varchar (191)	
parent_id	int (10)	
deleted_at	timestamp	
created_at	timestamp	
updated_at	timestamp	

TABEL XIV merupakan struktur tabel yang menyimpan pengaturan pada proses pengumpulan daftar produk dari pasar *online*.

TABEL XIV
STRUKTUR TABEL *GRAB_SETTINGS*

Nama kolom	Tipe data	Keterangan
id	int (10)	PRIMARY KEY; AUTO INCREMENT
marketplace_id	int (10)	
product_category_id	int (10)	
category_name	varchar (191)	
uri	varchar (191)	
deleted_at	timestamp	
created_at	timestamp	
updated_at	timestamp	

TABEL XV merupakan struktur tabel yang mencatat proses pengumpulan *data* yang sudah pernah dilakukan. Tabel ini merupakan tabel *parent* proses pengumpulan *data*.

TABEL XV
STRUKTUR TABEL *GRAB_BATCHES*

Nama kolom	Tipe data	Keterangan
id	int (10)	PRIMARY KEY; AUTO INCREMENT
marketplace_id	int (10)	
grab_setting_id	int (10)	
user_id	int (10)	
page	int (10)	
per_page	int (10)	
status	enum	Enum value 'SUCCESS', 'ERROR', 'PAGE_NOT_FOUND', 'ON_GRABBING'
is_classified	tinyint(1)	
deleted_at	timestamp	
created_at	timestamp	
updated_at	timestamp	

Tabel XVI merupakan struktur tabel yang menyimpan daftar produk yang berhasil diambil dari pasar *online*.

TABEL XVI
STRUKTUR TABEL *PRODUCTS*

Nama kolom	Tipe data	Keterangan
id	int (10)	PRIMARY KEY; AUTO INCREMENT
merchant_id	int (10)	
product_category_id	int (10)	
grab_batch_id	int (10)	
name	varchar(191)	
price	int (11)	
image_source_url	varchar(191)	
image_url	varchar(191)	
url	text	
condition	enum	Enum value 'BARU','BEKAS'
description	text	
rating	double(2,1)	
count_review	int(10)	
deleted_at	timestamp	
created_at	timestamp	
updated_at	timestamp	

TABEL XVII merupakan struktur tabel yang menyimpan *terms* atau kata-kata hasil *text preprocessing* pada proses klasifikasi.

TABEL XVII
STRUKTUR TABEL *TERMS*

Nama kolom	Tipe data	Keterangan
id	int (10)	PRIMARY KEY; AUTO INCREMENT
tf_doc	text	serialize array: docs 1 : 1 docs 2 : 1 docs 3 : 2
idf	double (10,8)	
deleted_at	timestamp	
created_at	timestamp	
updated_at	Timestamp	

TABEL XVIII merupakan struktur tabel yang menyimpan daftar klasifikasi yang pernah dilakukan. Tabel ini merupakan tabel *parent* yang mencatat proses klasifikasi yang pernah dilakukan.

TABEL XVIII
STRUKTUR TABEL *CLASSIFICATION_BATCHES*

Nama kolom	Tipe data	Keterangan
id	int (10)	PRIMARY KEY; AUTO INCREMENT
user_id	int (10)	
data_start	timestamp	
data_end	timestamp	
time	int(11)	
deleted_at	timestamp	
created_at	timestamp	
updated_at	timestamp	

TABEL XIX merupakan struktur tabel yang menyimpan hasil detil klasifikasi yang sudah dilakukan. Tabel ini merupakan tabel detil yang mencatat relasi antara produk dengan hasil klasifikasi untuk produk tersebut.

TABEL XIX
STRUKTUR TABEL *PRODUCT_CLASSIFICATIONS*

Nama kolom	Tipe data	Keterangan
id	int (10)	PRIMARY KEY; AUTO INCREMENT
classification_batch_id	int(10)	
product_class_id	int (10)	
product_id	int (10)	
deleted_at	timestamp	
created_at	timestamp	
updated_at	timestamp	

C. Activity diagram

Activity diagram dalam proses rancang bangun aplikasi akan digunakan untuk menggambarkan proses yang ada didalam sistem. Proses umum dalam sistem ini dapat dilihat pada Gambar 3. Langkah pertama yang dilakukan adalah pengumpulan data produk dari pasar *online*. Kemudian *text preprocessing* dapat dilakukan untuk mendapatkan data yang siap diproses ke langkah pembobotan *TF-IDF* dan Klasifikasi *KNN*. Sedangkan langkah terakhir adalah evaluasi confusion matrix.

Gambar 4 merupakan gambar diagram aktifitas pengambilan data dari pasar *online*. Pengambilan data dimulai dengan pemilihan pasar *online* dan kategori, serta mengisikan nomor halaman. Kemudian sistem akan mengakses URL dari pasar *online*. Langkah berikutnya adalah sistem akan melakukan parsing daftar produk. Untuk setiap data produk, lakukan proses pengecekan apakah sudah ada merchant yang sama terdaftar di dalam *database*, apabila belum ada maka merchant akan diisikan kedalam *database*. Kemudian pengecekan selanjutnya adalah apabila data produk yang sudah diparsing belum ada didalam *database*, data produk tersebut akan diisikan kedalam *database*. Proses dilakukan sampai daftar produk habis.

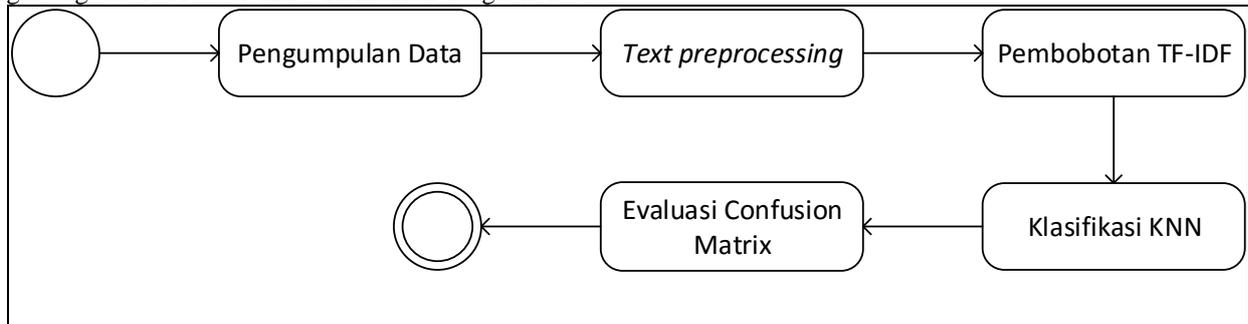
Setelah pengumpulan data, langkah selanjutnya adalah *text preprocessing* dapat dilihat pada Gambar 5. Pada tahap ini, produk yang telah dikumpulkan, akan dilakukan proses *cleansing*. Proses *cleansing* bertujuan untuk mengurangi

noise pada data produk. Langkah selanjutnya adalah proses *case folding*, dimana semua teks dalam data produk akan dirubah menjadi huruf kecil. Kemudian dilanjutkan dengan *filtering* untuk menghapus semua *stopword* yang terdaftar di *stoplist*.

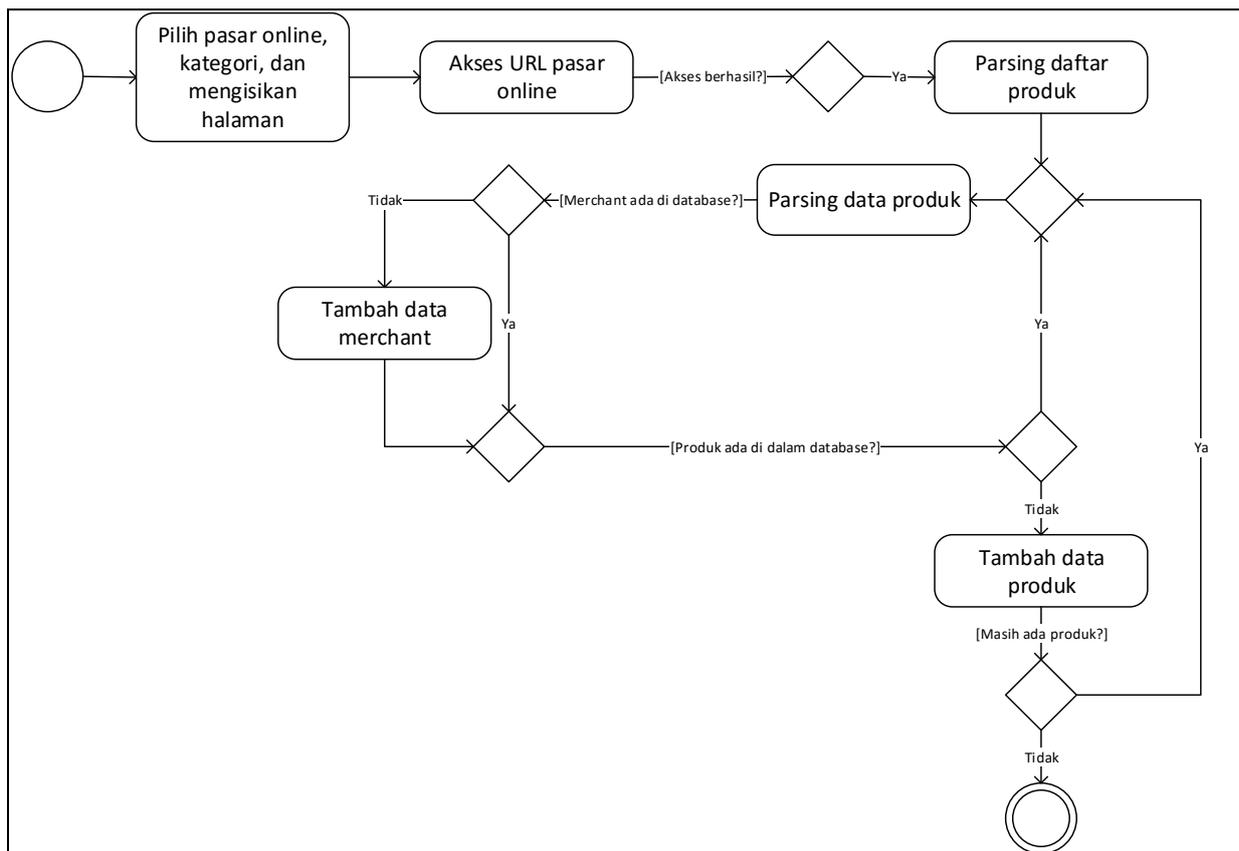
Gambar 6 merupakan gambar diagram aktifitas pembobotan dokumen menggunakan metode pembobotan *TF-IDF*. Hasil dari tahap ini adalah sebuah angka yang dapat digunakan untuk membandingkan satu dokumen dengan dokumen lain. Langkah pertama dilakukan dengan menghitung nilai TF dan IDF. Kemudian langkah kedua

adalah menghitung bobot *W* untuk setiap term pada dokumen.

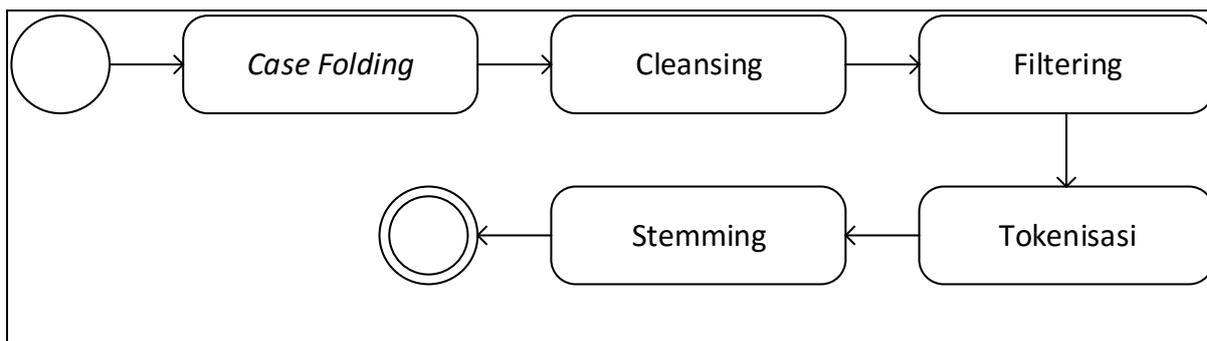
Gambar 7 merupakan diagram aktifitas proses klasifikasi produk. Nilai *W* yang didapatkan dari proses pembobotan *TF-IDF* akan diproses lebih lanjut pada tahap klasifikasi produk. Langkah pertama adalah dengan melakukan perhitungan *cos-similarity*. Kemudian menghitung probabilitas masing-masing dokumen dengan kategori. Probabilitas paling tinggi adalah kelas klasifikasi untuk dokumen.



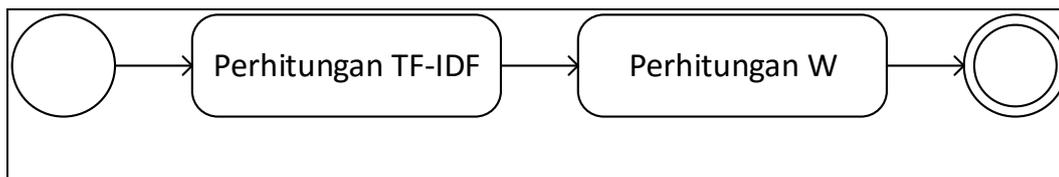
Gambar 3. Diagram aktifitas proses umum system



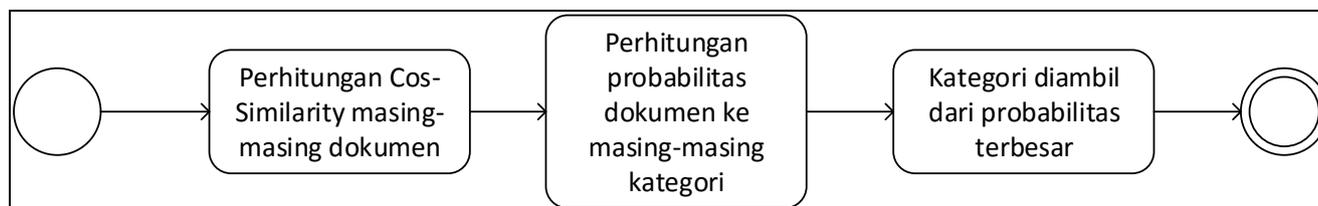
Gambar 4. Diagram aktifitas proses pengambilan data pasar online



Gambar 5. Diagram aktifitas proses text-preprocessing



Gambar 6. Diagram aktifitas proses pembobotan dokumen

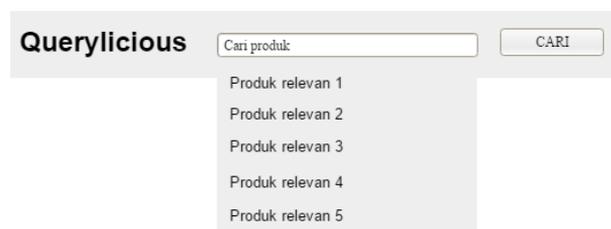


Gambar 7. Diagram aktifitas proses klasifikasi produk

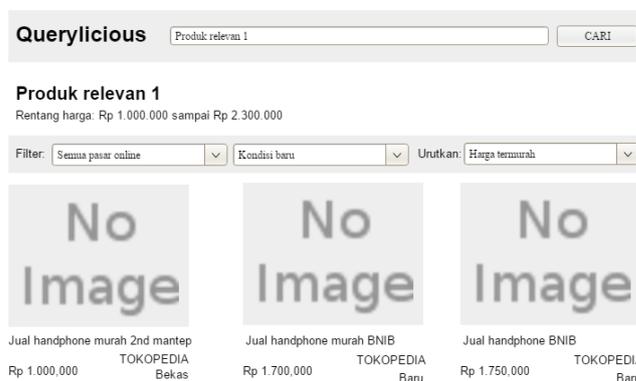
D. Perancangan antarmuka

Sistem yang dibangun akan memiliki 2 muka *website*, yaitu muka *website* umum dan muka *website admin*. Muka *website* umum terdiri dari 2 halaman, yaitu halaman pencarian produk dan halaman detail produk. Sedangkan muka *website admin* terdiri dari 14 halaman.

Gambar 8 adalah rancangan antar muka yang digunakan untuk *user* umum melakukan pencarian produk. Pada halaman ini *user* dapat melakukan pencarian produk dengan cara mengisikan nama produk yang diinginkan, kemudian sistem akan menampilkan daftar produk yang relevan dengan pencarian *user*. Setelah *user* memilih salah satu produk yang diinginkan, maka akan muncul halaman detail produk (Gambar 9). Pada halaman detail produk, akan ditampilkan daftar produk dari pasar *online* yang sudah diklasifikasikan ke kelas produk yang dipilih.



Gambar 8. Rancangan antarmuka halaman pencarian produk



Gambar 9. Rancangan antarmuka halaman detail produk

Gambar 10 adalah rancangan antarmuka halaman login. Halaman login digunakan oleh *admin* untuk melakukan akses ke *website admin*. Di dalam *website admin*, *user admin* dapat melakukan pengaturan.

Gambar 10.Rancangan antarmuka halaman login

Setelah login, menu yang dapat diakses oleh *user admin* dapat dilihat pada Gambar 11.

- Pengguna
- Grab
- Produk
- Kelas
- Data Latih
- Stop Word
- Pengaturan

Gambar 11.Rancangan antarmuka menu *website admin*

Gambar 12 dan Gambar 13 adalah rancangan antarmuka halaman *admin* yang digunakan untuk mengelola pengguna. Halaman cari pengguna (Gambar 12) adalah halaman awal ketika *user admin* mengakses menu “Pengguna”. Pada halaman cari pengguna, *user admin* dapat melihat daftar pengguna yang dapat login ke sistem. Selain itu *user admin* dapat menambahkan pengguna baru, mengubah dan menghapus *user* yang terdaftar. Penambahan dan perubahan pengguna dilakukan melalui form pengguna (Gambar 13).

Pengguna

No	Nama	Email	Action
1	Pengguna 1	email@domain.com	Edit, Hapus
2	Pengguna 2	email@domain.com	Edit, Hapus
3	Pengguna 3	email@domain.com	Edit, Hapus

Gambar 12.Rancangan antarmuka halaman *admin*, cari pengguna

Pengguna

Gambar 13.Rancangan antarmuka halaman *admin*, form pengguna

Gambar 14 dan Gambar 15 adalah rancangan antarmuka halaman *admin* yang digunakan untuk mengelola kelas. Halaman cari kelas (Gambar 14) merupakan halaman awal ketika *user admin* melakukan akses ke menu “Kelas”. Pada halaman cari kelas, *user admin* dapat melihat daftar kelas yang ada di sistem dan digunakan untuk proses klasifikasi *KNN*. Selain itu *user admin* dapat menambahkan kelas baru, mengubah dan menghapus kelas yang terdaftar. Penambahan dan perubahan pengguna dilakukan melalui form kelas (Gambar 15).

Kelas Klasifikasi

No	Nama	Action
1	Kelas Klasifikasi 1	Hapus
2	Kelas Klasifikasi 2	Hapus
3	Kelas Klasifikasi 3	Hapus

Gambar 14.Rancangan antarmuka halaman *admin*, cari kelas klasifikasi

Kelas Klasifikasi

Form Kelas Klasifikasi

Nama Kelas Klasifikasi:

Gambar 15. Rancangan antarmuka halaman *admin*, form kelas klasifikasi

Gambar 16 dan Gambar 17 adalah rancangan antarmuka halaman *admin* yang digunakan untuk mengelola data latih. Halaman cari data latih (Gambar 16) merupakan halaman awal ketika *user admin* melakukan akses ke menu “Data Latih”. Pada halaman cari data latih, *user admin* dapat melihat daftar data latih yang ada di sistem dan digunakan untuk proses klasifikasi *KNN*. Selain itu *user admin* dapat menambahkan data latih baru, mengubah dan menghapus data latih yang terdaftar. Penambahan dan perubahan pengguna dilakukan melalui form data latih (Gambar 17).

Data Latih

Tambah Data Latih

Tampilkan data Cari:

No	Kelas	Teks	Action
1	Kelas 1	Lorem ipsum dolor sit amet, consectetur adipiscing elit.	Hapus
2	Kelas 2	Nam pulvinar sapien luctus, convallis massa in, finibus eros.	Hapus
3	Kelas 3	Integer omare ipsum non elit placerat, eu iaculis metus faucibus.	Hapus

Gambar 16. Rancangan antarmuka halaman *admin*, cari data latih

Data Latih

Form Data Latih

Teks:

Kelas :

Gambar 17. Rancangan antarmuka halaman *admin*, form data latih

Gambar 19 dan Gambar 20 adalah rancangan antarmuka halaman *admin* yang digunakan untuk mengelola data *stop word*. Halaman cari data *stop word* (Gambar 18) merupakan halaman awal ketika *user admin* melakukan akses ke menu “Stop Word”. Pada halaman cari data latih, *user admin* dapat melihat daftar data *stop word* yang ada di sistem dan digunakan untuk proses *text preprocessing*. Selain itu *user admin* dapat menambahkan data *stop word* baru, mengubah dan menghapus data *stop word* yang terdaftar. Penambahan dan perubahan pengguna dilakukan melalui form data *stop word* (Gambar 19).

Stop Word

Tambah Stop Word

Tampilkan data Cari:

No	Kata	Dibuat pada	Action
1	BNIB	28 Jun 2017	Hapus
2	Brand New	28 Jun 2017	Hapus
3	In Box	28 Jun 2017	Hapus

Gambar 18. Rancangan antarmuka halaman *admin*, cari *stop word*

Stop Word

Form Stop Word

Term/Kata :

Gambar 19. Rancangan antarmuka halaman *admin*, form *stop word*

Gambar 20 adalah halaman *admin* yang digunakan untuk mengumpulkan data produk. Pengumpulan data produk membutuhkan masukan berupa pilihan pasar *online*, pilihan kategori, dan halaman yang akan diambil datanya. Data produk yang sudah diambil, dapat dilihat pada halaman daftar produk (Gambar 21).

PRODUK

Pengumpulan Data Produk

Pasar Online:

Kategori:

Halaman:
 sampai

Gambar 20. Rancangan antarmuka halaman admin, pengumpulan data produk

PRODUK

Tampilkan data
 Cari:

Nr	Produk	Harga	Kategori	Penjual	Pasar Online	Batch
1	Handphone BNIB	1.000.000	Handphone	HP Store	Tokopedia	1
2	Handphone BNOB	1.500.000	Handphone	HP holic	Bukalapak	1
3	Handphone 2nd	750.000	Handphone	HP holic	Bukalapak	1

Gambar 21. Rancangan antarmuka halaman admin, daftar produk

Gambar 22 adalah halaman pencarian batch klasifikasi yang digunakan untuk melakukan pencarian batch klasifikasi yang sudah pernah dilakukan. Dari masing-masing baris data, dapat dilihat detail produk hasil klasifikasi untuk batch tersebut pada halaman detail klasifikasi (Gambar 23). Pada halaman pencarian batch klasifikasi, terdapat tombol “Tambah Klasifikasi” yang digunakan untuk masuk ke halaman form klasifikasi (Gambar 24). Untuk melakukan klasifikasi, diperlukan pemilihan produk yang mau diklasifikasi.

Klasifikasi

Klasifikasi Batch

Tampilkan data
 Cari:

Batch	Jumlah data	Lama Waktu	Action
1	30 data	2 menit 10 detik	Detail
2	30 data	2 menit 10 detik	Detail
3	30 data	2 menit 10 detik	Detail
4	60 data	3 menit 15 detik	Detail
5	70 data	4 menit 11 detik	Detail

Gambar 22. Rancangan antarmuka halaman admin, pencarian batch klasifikasi

Klasifikasi

Klasifikasi Detil

Batch Klasifikasi : 1
 Jumlah data: 30 data
 Waktu mulai : 10 Januari 2017 20:00:00
 Waktu selesai : 10 Januari 2017 20:02:10
 Lama waktu: 2 menit 10 detik

Tampilkan data
 Cari:

Id	Produk	Harga	Kategori	Penjual	Pasar Online	Kelas
1	Handphone BNIB	1.000.000	Handphone	HP Store	Tokopedia	Kelas handphone 1
2	Handphone BNOB	1.500.000	Handphone	HP holic	Bukalapak	Kelas handphone 1
3	Handphone 2nd	750.000	Handphone	HP holic	Bukalapak	Kelas handphone 2
4	Handphone Bagus	775.000	Handphone	HP Jogja	Bukalapak	Kelas handphone 3
5	Handphone Langka	2.000.000	Handphone	HP Jogja	Tokopedia	Kelas handphone 3

Gambar 23. Rancangan antarmuka halaman admin, detail klasifikasi

Klasifikasi

Form Klasifikasi

Kata Kunci :

Pasar Online :

Tampilkan data

<input type="checkbox"/>	Id	Produk	Harga	Kategori	Penjual	Pasar Online
<input checked="" type="checkbox"/>	1	Handphone BNIB	1.000.000	Handphone	HP Store	Tokopedia
<input checked="" type="checkbox"/>	2	Handphone BNOB	1.500.000	Handphone	HP holic	Bukalapak
<input type="checkbox"/>	3	Handphone 2nd	750.000	Handphone	HP holic	Bukalapak
<input type="checkbox"/>	4	Handphone Bagus	775.000	Handphone	HP Jogja	Bukalapak
<input type="checkbox"/>	5	Handphone Langka	2.000.000	Handphone	HP Jogja	Tokopedia
<input type="checkbox"/>	6	Handphone Emas	1.775.000	Handphone	HP holic	Bukalapak

Gambar 24. Rancangan antarmuka halaman admin, form klasifikasi

E. Contoh Kasus Klasifikasi KNN

Daftar kelas yang dipilih adalah:

1. Oppo Find 5 Mini
2. Oppo N1 Mini
3. Xiaomi Mi5

Untuk masing-masing daftar kelas klasifikasi yang ada, ditentukan daftar data latih yang ada. Data latih yang ada dapat dilihat pada TABEL XX.

TABEL XX
 DAFTAR KELAS DAN DATA LATIH CONTOH KASUS KNN

	Data latih	Kelas Klasifikasi
D1	Oppo Find 5 Mini	Oppo Find 5 Mini
D2	Oppo Find5 Mini	Oppo Find 5 Mini
D3	Oppo Fine 5Mini	Oppo Find 5 Mini
D4	Oppo N1 Mini	Oppo N1 Mini
D5	Oppo N 1 Mini	Oppo N1 Mini
D6	Oppo N 1Mini	Oppo N1 Mini

D7	Xiaomi Mi 5	Xiaomi Mi5
D8	Xiaomi Mi5	Xiaomi Mi5
D9	Mi5	Xiaomi Mi5

Data uji (x): WTS Oppo Find 5 Mini lengkap.

Langkah pertama adalah data latih akan dipecah menjadi kata-kata/terms yang akan dihitung *df* dan *idf* nya. Daftar terms dan nilai *df* dan *idf* untuk *terms* dapat dilihat pada TABEL XXI.

TABEL XXI
TERMS & PERHITUNGAN *DF* DAN *IDF* PADA DATA LATIH

Terms	tf									df	idf	
	D1	D2	D3	D4	D5	D6	D7	D8	D9			
Oppo	1	1	1	1	1	1				6	0.1761	
Find	1		1			1				3	0.4771	
5	1						1			2	0.6532	
Mini	1	1		1	1					4	0.3522	
Find5		1								1	0.9542	
5Mini			1							1	0.9542	
N1				1						1	0.9542	
N					1					1	0.9542	
l					1					1	0.9542	
lMini						1				1	0.9542	
Xiaomi							1	1		2	0.6532	
Mi								1		1	0.9542	
Mi5									1	1	2	0.6532

Data uji (x) tersebut selanjutnya akan dilakukan *text preprocessing* dan terbentuk *token-token* seperti pada TABEL XXII. Masing-masing *token* dari data uji akan dibandingkan dengan *token-token* pada data latih.

TABEL XXII
JUMLAH TERMS DAN PERHITUNGAN *TF* PADA DATA UJI DAN DATA LATIH

Terms	tf									x
	D1	D2	D3	D4	D5	D6	D7	D8	D9	
Oppo	1	1	1	1	1	1				1
Find	1		1			1				1
5	1						1			1
Mini	1	1		1	1					1
Find5		1								
5Mini			1							
N1				1						
N					1					
l					1					
lMini						1				
Xiaomi							1	1		
Mi								1		
Mi5									1	1
WTS										1
Lengkap										1

Kemudian, akan dihitung bobot (*W*) dari masing-masing *token*. Bobot dihitung dengan mengalikan *tf* dengan *idf*. Hasil perhitungan dapat dilihat pada TABEL XXIII.

TABEL XXIII
PERHITUNGAN BOBOT SETIAP DOKUMEN

Terms	W									x
	D1	D2	D3	D4	D5	D6	D7	D8	D9	
Oppo	0.176	0.176	0.176	0.176	0.176	0.176				0.176
Find	0.477		0.477			0.477				0.477
5	0.653						0.653			0.653
Mini	0.352	0.352		0.352	0.352					0.352
Find5		0.954								
5Mini			0.954							
N1				0.954						
N					0.954					
l					0.954					
lMini						0.954				
Xiaomi							0.653	0.653		
Mi							0.954			
Mi5								0.653	0.653	
WTS										0
Lengkap										0

Setelah mendapatkan bobot dari setiap dokumen, kalikan bobot dari data yang diuji dengan bobot data latih yang ada. Bobot pada x yang memiliki nilai, dikalikan dengan bobot lainnya yang sebaris dengannya. Kemudian jumlahkan bobot pada setiap dokumen, sehingga didapatkan total bobot dari setiap dokumen. Hasil perhitungan dapat dilihat pada TABEL XXIV.

TABEL XXIV
HASIL PENJUMLAHAN DARI PERKALIAN BOBOT DATA UJI DENGAN SELURUH DATA LATIH

Terms	W uji * W latih								
	D1	D2	D3	D4	D5	D6	D7	D8	D9
Oppo	0.031	0.031	0.031	0.031	0.031	0.031			
Find	0.228		0.228			0.228			
5	0.427						0.427		
Mini	0.124	0.124		0.124	0.124				
Find5		0.000							
5Mini			0.000						
N1				0.000					
N					0.000				
l					0.000				
lMini						0.000			
Xiaomi							0.000	0.000	
Mi							0.000		
Mi5								0.000	0.000
WTS									
Lengkap									
SUM	0.809	0.155	0.259	0.155	0.155	0.259	0.427	0.000	0.000

Setelah mendapatkan total penjumlahan dari setiap kolom, maka langkah selanjutnya adalah melakukan perhitungan kuadrat bobot dari setiap data uji dan data latih. Dari jumlah total kuadrat bobot dari setiap data akan di akar dan menghasilkan sebuah nilai. Hasil perhitungan dapat dilihat pada TABEL XXV.

TABEL XXV
PENJUMLAHAN TOTAL KUADRAT BOBOT DATA LATIH DAN UJI, HASIL AKAR DARI MASING-MASING TOTAL BOBOT

Terms	Vector									
	D1	D2	D3	D4	D5	D6	D7	D8	D9	x
Oppo	0.031	0.031	0.031	0.031	0.031	0.031				0.031
Find	0.228		0.228			0.228				0.228
5	0.427						0.427			0.427
Mini	0.124	0.124		0.124	0.124					0.124
Find5		0.911								
5Mini			0.911							
N1				0.911						
N					0.911					
l					0.911					
lMini						0.911				
Xiao mi							0.427	0.427		
Mi							0.911			
Mi5								0.427	0.427	
WTS										
Lengkap										
SUM	0.809	1.066	1.169	1.066	1.976	1.169	1.764	0.853	0.427	0.809
Root	0.89	1.03	1.08	1.03	1.40	1.08	1.32	0.92	0.65	0.89
	97	23	13	23	58	13	81	38	32	97

Langkah selanjutnya adalah menghitung *cosine similarity*. Perhitungan *cosine similarity* dilakukan menggunakan rumus nomor (1). Lakukan perhitungan *cosine similarity* untuk masing-masing dokumen latih yang ada. Kemudian diurutkan secara *descending*. Maka dihasilkan nilai *cosine similarity* dan urutan seperti pada TABEL XXVI.

TABEL XXVI
PERHITUNGAN NILAI *COSINE SIMILARITY* DAN PENGURUTAN SECARA *DESCENDING*

Peringkat	Data Latih	Kelas Klasifikasi	Cosine Similarity
1	D1	Oppo Find 5 Mini	0.8715
2	D7	Oppo N1 Mini	0.3478
3	D3	Oppo Find 5 Mini	0.2317
4	D6	Xiaomi Mi5	0.1801
5	D2	Oppo Find 5 Mini	0.1389
Tidak diperhitungkan	D4	Oppo N1 Mini	0.1068
Tidak diperhitungkan	D5	Oppo N1 Mini	0.1020
Tidak diperhitungkan	D8	Xiaomi Mi5	0

Tidak diperhitungkan	D9	Xiaomi Mi5	0
----------------------	----	------------	---

Setelah mendapatkan hasil tersebut, selanjutnya menghitung nilai kecenderungan untuk data uji ke masing-masing kelas berdasarkan nilai K. Dalam menghitung nilai kecenderungan, digunakan rumus nomor (2).

Dalam contoh ini digunakan K=5, sehingga yang diambil adalah 5 nilai *cosine similarity* tertinggi. Dari 5 data tersebut, dihitung nilai kecenderungan-nya dengan cara menjumlahkan data yang memiliki kelas yang sama. Hasil perhitungan dapat dilihat pada TABEL XXVII

TABEL XXVII
PERHITUNGAN KECENDERUNGAN DOKUMEN UJI KE KELAS KLASIFIKASI

Kelas Klasifikasi	Nilai Kecenderungan
Oppo Find 5 Mini	1.2421
Oppo N1 Mini	0.3478
Oppo Find 5 Mini	0.1801

Berdasarkan perhitungan nilai kecenderungan, nilai tertinggi adalah 1.2421 yang dimiliki oleh kelas Oppo Find 5 Mini. Sehingga Dokumen uji diklasifikasikan ke kelas Oppo Find 5 Mini. Hasil klasifikasi menggunakan perhitungan ini sesuai dengan yang seharusnya. Dimana dokumen uji memang masuk ke kelas Oppo Find 5 Mini.

VI. KESIMPULAN

A. Kesimpulan

Pada penelitian kali ini, sudah didapatkan rancangan antarmuka, alur proses, serta perancangan fitur sistem yang akan digunakan sebagai acuan pembuatan *website*. Semua aktor yang terlibat sudah memiliki proses sesuai dengan kebutuhan masing-masing.

Berdasarkan hasil studi literatur dan contoh kasus, metode klasifikasi *K-Nearest Neighbor* dapat digunakan untuk kasus pengelompokan *handphone*.

UCAPAN TERIMA KASIH

Saya mengucapkan terima kasih kepada Fakultas Teknologi Informasi Universitas Kristen Duta Wacana dan Lembaga Penelitian dan Pengabdian Masyarakat yang telah memberikan dukungan kepada saya dalam menyelesaikan penelitian ini.

DAFTAR PUSTAKA

- [1] Hugh Harsono. (2016, July) TechCrunch. [Online]. <https://techcrunch.com/2016/07/29/indonesia-will-be-asias-next-biggest-e-commerce-market/>
- [2] Bruno Trstenjak, Sasa Mikac, and Dzenana Donko, "KNN with TF-IDF Based Framework for Text Categorization," in *Procedia Engineering*, 2014, pp. 1356-1364.
- [3] Edi Winarko Jumadi, "Penggunaan KNN (K-Nearest Neighbor) untuk Klasifikasi Teks Berita yang Tak-terkelompokkan pada saat Pengklasteran oleh STC (Suffix Tree Clustering)," *ISTEK Jurnal*

- Kajian Islam, Sains, dan Teknologi*, vol. IX, no. 1, 2015.
- [4] Armanda Nur Fadhillah, Ledy Novamizanti, and Ratri Dwi Atmaja, "Analisis dan Implementasi Klasifikasi *K-Nearest Neighbor* (K-NN) pada sistem identifikasi Biometrik Telapak Kaki Manusia," Universitas Telkom, Skripsi 2015.
- [5] Andro Ardiyanto, "Klasifikasi Komentar pada Dataset Pemilu Presiden Indonesia 2014 dengan Metode Improved *K-Nearest Neighbor*," Universitas Kristen Duta Wacana, Yogyakarta, Skripsi 2017.
- [6] Euihong Han, George Karypis, and Vipin Kumar, "Text Categorization Using Weight Adjusted *K-Nearest Neighbor* Classification," Department of Computer Science, University of Minnesota, Minneapolis, 1999.
- [7] Bharath Sriram, Dave Fuhry, Engin Demir, Hakan Ferhastosmanoglu, and Murat Demirbas, "Short Text Classification in Twitter to Improve Information Filtering," in *SIGIR '10 Proceedings of the 33rd international ACM SIGIR conference on Research and development in information retrieval*, Geneva, 2010, pp. 841-842.
- [8] Edi Winarko Jumadi, "Penggunaan *KNN* (*K-Nearest Neighbor*) Untuk Klasifikasi Teks Berita yang Tak-Terkelompokkan pada Saat Pengklasteran Oleh STC (Suffix Tree Clustering)," *ISTEK*, vol. IX, pp. 50-81, Juni 2015.
- [9] Kusri and Luthfi, *Algoritma Data Mining*. Yogyakarta: Andi Offset, 2009.
- [10] Y Liao. (2002) Review of *K-Nearest Neighbor* Text Categorization Method. [Online].
https://www.usenix.org/legacy/event/sec02/full_papers/liao/liao_html/node4.html
- [11] C.D. Manning and H Schutze, *Foundation of Statistical Natural Language Processing*. Cambridge: MIT Press, 2000.
- [12] N Suguna and K Tanushkodi, "An Improved *K-Nearest Neighbor* Classification Using Genetic Algorithm," *International Journal of Computer Science Issue*, vol. 7, no. 2, pp. 18-21, 2010.
- [13] Bimlesh Wadhwa, Soo Yuen Jien, and Damith C Rajapakse, *A First Course in Object-Oriented Software Engineering*, 2nd ed. Asia: McGraw-Hill Education (Asia), 2010.
- [14] Marini, "Rancang Bangun Sistem Informasi Kenaikan Pangkat Pegawai pada Kantor Kecamatan Bukit Intan," *Jurnal Teknik Informatika dan Sistem Informasi*, vol. 3, no. 1, pp. 1-16, April 2017.
- [15] Prabowo Pudjo Widodo and Herlawati, *Menggunakan UML, Unified Modelling Language*. Bandung, Indonesia: Penerbit Informatika, 2011.
- [16] Adi Nugroho, *Perancangan dan Implementasi Basis Data*. Yogyakarta, Indonesia: Penerbit Andi, 2011.
- [17] Carlos Coronel, Steven Morris, and Peter Rob, *Database Principles Fundamentals of Design, Implementation, and Management*, 10th ed., 2013.