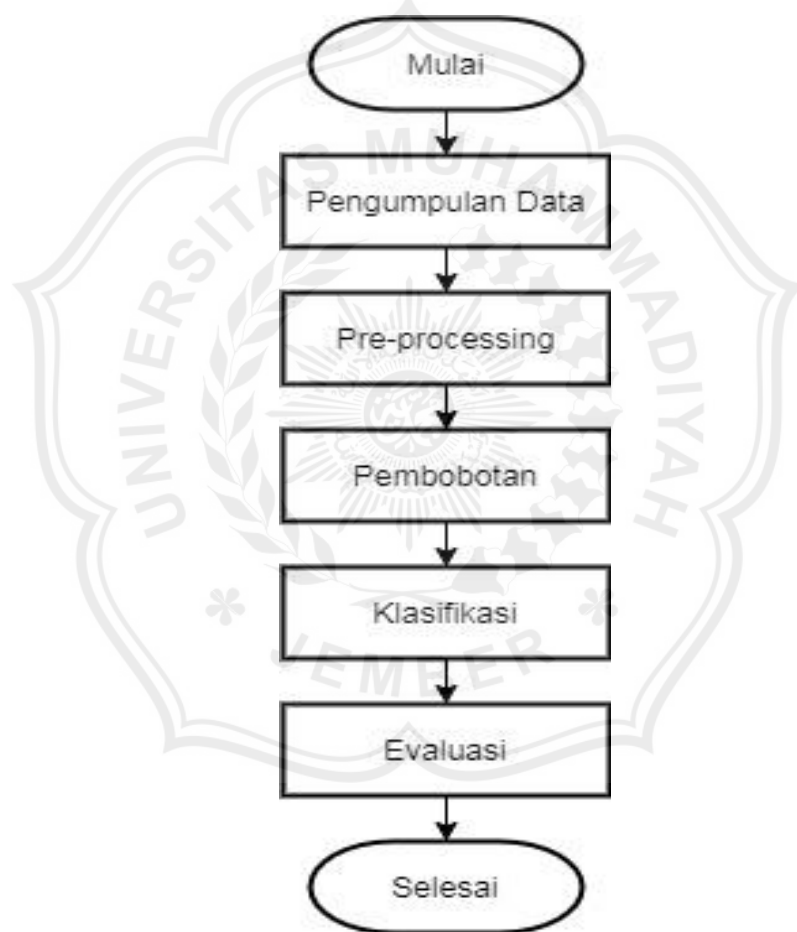


BAB III METODE PENELITIAN

3.1 Rancangan Penelitian

Rancangan penelitian dalam penelitian klasifikasi teks ini antara lain pengumpulan data, *pre-processing*, pembobotan, klasifikasi, dan evaluasi. Berikut gambaran rancangan penelitian seperti yang ditunjukkan pada Gambar 3.1.



Gambar 3.1 Tahapan Rancangan Penelitian

3.2 Pengumpulan Data

Pengumpulan data dilakukan untuk memperoleh data abstrak yang akan digunakan sebagai objek penelitian. Pengambilan data pada *Repository* Universitas Muhammadiyah Jember dilakukan dengan cara manual atau mengambil satu per satu file PDF abstrak tiap judul tugas akhir. Data abstrak yang telah dikumpulkan akan dijadikan satu file yang tersimpan dalam format Excel (.xlsx) agar dapat memudahkan dalam *proses pre-processing*.

3.3 Pelabelan Data

Pelabelan data dilakukan untuk mendapatkan label kelas setiap data abstrak tugas akhir. Proses ini dilakukan secara manual dengan mengacu pada pengelompokan awal tugas akhir berdasarkan program studi di *Repository* Universitas Muhammadiyah Jember.

3.4 *Pre-processing*

Pre-processing dilakukan untuk mempersiapkan data abstrak sehingga bersih dari *noise* dan sesuai dengan ketentuan klasifikasi teks. Dalam proses *pre-processing* terdapat 4 tahapan yaitu *Casefolding*, *Tokenizing*, *Filtering* dan *Stemming*. Beberapa proses pada persiapan data ini ditunjukkan pada Gambar 3.2.



Gambar 3.2 Alur *Text Pre-processing*

Berikut contoh data abstrak yang telah melewati proses *pre-processing*.

A. Teknik Elektro

Dokumen 1: Tuna Netra sudah kita kenal dengan baik, merupakan istilah yang digunakan untuk orang yang mengalami gangguan penglihatan. Berdasarkan tingkat gangguannya, Tuna Netra dibagi menjadi dua jenis yaitu : Total Blind di mana penyandang sama sekali tidak dapat melihat dan Low Vision di mana penyandang

masih memiliki sedikit kemampuan untuk melihat. Penyandang Tuna Netra, karena ketidak mampuannya melihat, memiliki banyak keterbatasan dalam melaksanakan aktivitas sehari-harinya. Khususnya yang beragama islam (muslim) adalah ketika menjalankan ibadah shalat maka mereka sulit bahkan tidak mampu menentukan arah kiblat. Untuk mengatasi permasalahan tersebut maka pada penelitian ini dibuatlah sebuah sistem alat bantu shalat untuk penyandang tuna netra.

B. Teknik Sipil

Dokumen 2: Jembatan secara umum adalah suatu konstruksi yang berfungsi untuk menghubungkan dua bagian jalan yang terputus oleh adanya rintangan. Dalam pembuatan jembatan perlu memperhatikan beberapa aspek seperti lokasi jembatan, kelas jembatan, jenis jembatan, tipe jembatan, dan kelas jalan. Dalam pemilihan tipe jembatan yang ada tidak membahas secara detail tentang jembatan yang lebih efisien dan optimal dalam penggunaannya maka tugas akhir ini membahas tentang optimalisasi tipe jembatan yang digunakan pada studi kasus jembatan Kencong dengan panjang bentang 60 meter.

C. Teknik Informatika

Dokumen 3: Perencanaan merupakan fungsi penting dalam pengembangan suatu organisasi atau lembaga. Salah satunya adalah perencanaan pembiayaan. Perencanaan pembiayaan yang sesuai sangat menunjang visi organisasi. Salah satu lembaga yang perlu melakukan fungsi perencanaan pembiayaan adalah perguruan tinggi. Menggunakan single exponential smoothing dan single moving average yang bertujuan untuk memprediksi jumlah data yang akan muncul pada masa yang akan datang dari perhitungan data masa lalu. Perbandingan dua algoritma tersebut yang akan dipersentasikan dalam bentuk MAPE, sehingga terpilih satu algoritma yang memiliki persentase kesalahan terkecil.

D. Teknik Mesin

Dokumen 4: Skuter 150 cc 2 tak adalah motor bakar 2 langkah yang sistem pemasukan bahan bakarnya menggunakan crankshaft valve, lubang masuk bahan bakar menuju crankshaft pada Skuter 150 cc 2 Tak ini sering mengalami kebocoran/keausan karena penggunaan oli mesin bekas sebagai pengganti oli samping dan seringnya pemakaian yang mengakibatkan borosnya konsumsi bahan bakar. Penelitian ini menggunakan metode ekperimental dengan membandingkan konsumsi bahan bakar pada motor Skuter 150 cc 2 Tak standart dan dengan penambahan Membran Untuk Motor 110 cc, dengan menghitung waktu yang di butuhkan untuk menghabiskan bahan bakar 30 ml (fuel consumption) pada putaran motor 1000, 2000, 3000, 4000, dan 5000 rpm.

E. Manajemen Informatika

Dokumen 5: Toko buku Mulya Kecamatan Besuki, Kabupaten Situbondo merupakan suatu perusahaan yang bergerak dibidang penjualan dan pendistribusian buku jasa di wilayah Situbondo, Namun dalam hal pemesanan buku perlu ada penambahan sistem informasi yang memadai agar suatu efektifitas waktu terjaga dengan baik dan tidak menimbulkan waktu yang lama dalam pemesanan. Oleh karena itu solusi yang memberikan dampak yang begitu luas pada masyarakat dengan memanfaatkan sistem informasi pembelian pada toko buku Mulya dibantu pembuatannya menggunakan PHP dan MY SQL.

F. Data Uji

Data uji diambil secara random untuk dilakukan uji coba dengan menggunakan algoritma.

Contoh: Teknologi dan informasi merupakan dua hal yang tidak dapat dipisahkan, hal ini terlihat dari proses mendapatkan informasi yang dapat diperoleh secara cepat, tepat dan akurat dengan

dukungan teknologi yang semakin canggih. Kemajuan teknologi ini membuat banyak organisasi dan lembaga khususnya pendidikan menggunakan teknologi berbasis komputer dan jaringan untuk membantu pekerjaannya karena bersifat efektif dan efisien. Sistem Informasi Pengolaha Data Presensi yang akan dibangun pada SMPN 1 Sukorambi sesuai dengan penelitian yang telah dilakukan. Sistem Informasi ini dikembangkan dengan Barcode dan SMS Gateway, bahasa pemrograman yang dipakai yaitu PHP dan Database MySQL.

3.5 Pembobotan

Teknik pembobotan dilakukan untuk menghitung bobot nilai suatu kata menggunakan algoritma TF-IDF. Proses pembobotan TF-IDF dimulai dari menghitung tiap *term*/kata yang ada pada setiap dokumen (TF). Kemudian proses dilanjutkan dengan menghitung jumlah dokumen yang mempunyai *term* tertentu (DF). Setelah itu proses menghitung *Inverse Document Frequency* (IDF) dan yang terakhir yaitu mengkalikan nilai TF dengan IDF.

Berikut adalah alur dan contoh dari proses pembobotan kata. Langkah awal pembobotan adalah menghitung *term frequency* pada setiap dokumen abstrak. Hasil perhitungan *term frequency* ditunjukkan pada Tabel 3.1.

Tabel 3.1 Perhitungan *Term Frequency*

KATA	TF						DF
	D1	D2	D3	D4	D5	DU	
tuna	4	0	0	0	0	0	1
netra	4	0	0	0	0	0	1
istilah	1	0	0	0	0	0	1
orang	1	0	0	0	0	0	1
ganggu	2	0	0	0	0	0	1
lihat	1	0	0	0	0	0	1
dasar	1	0	0	0	0	0	1
tingkat	1	0	0	0	0	0	1
bagi	1	0	0	0	0	0	1
total	1	0	0	0	0	0	1
blind	1	0	0	0	0	0	1

KATA	TF						DF
	D1	D2	D3	D4	D5	DU	
low	1	0	0	0	0	0	1
vision	1	0	0	0	0	0	1
sandang	2	0	0	0	0	0	1
memiliki	1	0	0	0	0	0	1
batas	2	0	0	0	0	0	1
alami	1	0	0	1	0	0	2
jalan	1	1	0	0	0	0	2
ibadah	1	0	0	0	0	0	1
shalat	2	0	0	0	0	0	1
tentu	1	0	0	0	0	0	1
arah	1	0	0	0	0	0	1
kilat	1	0	0	0	0	0	1
atas	1	0	0	0	0	0	1
masalah	1	0	0	0	0	0	1
buat	1	0	0	0	0	0	1
sistem	1	0	0	0	2	2	3
alat	1	0	1	0	0	0	2
bantu	1	0	1	0	1	0	3
jembatan	0	6	0	0	0	0	1
konstruksi	0	1	0	0	0	0	1
hubung	0	1	0	0	0	0	1
pilih	0	1	0	0	0	0	1
tipe	0	4	0	0	0	0	1
bahas	0	2	0	0	0	0	1
optimalisasi	0	1	0	0	0	0	1
strukture	0	1	0	0	0	0	1
steel	0	1	0	0	0	0	1
design	0	1	0	0	0	0	1
bentang	0	1	0	0	0	0	1
warren	0	1	0	0	0	0	1
pratt	0	1	0	0	0	0	1
profil	0	2	0	0	0	0	1
wf	0	2	0	0	0	0	1
gelagar	0	2	0	0	0	0	1
mamanjang	0	1	0	0	0	0	1
jarak	0	2	0	0	0	0	1
meter	0	2	0	0	0	0	1
lintang	0	1	0	0	0	0	1
single	0	0	3	0	0	0	1

KATA	TF						DF
	D1	D2	D3	D4	D5	DU	
exponential	0	0	2	0	0	0	1
smoothing	0	0	2	0	0	0	1
moving	0	0	1	0	0	0	1
average	0	0	1	0	0	0	1
prediksi	0	0	2	0	0	0	1
data	0	0	1	0	1	1	3
muncul	0	0	1	0	0	0	1
algoritma	0	0	3	0	0	0	1
presentasi	0	0	1	0	0	0	1
mape	0	0	2	0	0	0	1
hasil	0	0	1	0	0	0	1
ukur	0	0	1	0	0	0	1
kerja	0	0	1	0	0	0	1
metode	0	0	1	1	0	0	2
uji	0	0	1	0	0	0	1
time	0	0	1	0	0	0	1
series	0	0	1	0	0	0	1
analysis	0	0	1	0	0	0	1
aplikasi	0	0	1	0	0	0	1
pom	0	0	1	0	0	0	1
qm	0	0	1	0	0	0	1
kategori	0	0	1	0	0	0	1
bagus	0	0	1	0	0	0	1
presentase	0	0	1	0	0	0	1
alpha	0	0	1	0	0	0	1
bayar	0	0	1	0	0	0	1
spp	0	0	1	0	0	0	1
semester	0	0	1	0	0	0	1
genap	0	0	1	0	0	0	1
skuter	0	0	0	3	0	0	1
cc	0	0	0	4	0	0	1
motor	0	0	0	3	0	0	1
bakar	0	0	0	5	0	0	1
langkah	0	0	0	1	0	0	1
pasu	0	0	0	1	0	0	1
bahan	0	0	0	4	0	0	1
crankshaft	0	0	0	2	0	0	1
valve	0	0	0	1	0	0	1
lubang	0	0	0	1	0	0	1

KATA	TF						DF
	D1	D2	D3	D4	D5	DU	
masuk	0	0	0	1	0	0	1
guna	0	0	0	1	0	0	1
mesin	0	0	0	1	0	0	1
bekas	0	0	0	1	0	0	1
pakai	0	0	0	1	0	0	1
boros	0	0	0	1	0	0	1
konsumsi	0	0	0	2	0	0	1
eksperimental	0	0	0	1	0	0	1
banding	0	0	0	1	0	0	1
standart	0	0	0	1	0	0	1
tambah	0	0	0	1	1	0	2
membran	0	0	0	1	0	0	1
toko	0	0	0	0	1	0	1
buku	0	0	0	0	3	0	1
mulya	0	0	0	0	1	0	1
usaha	0	0	0	0	1	0	1
jasa	0	0	0	0	1	0	1
mesan	0	0	0	0	3	0	1
informasi	0	0	0	0	2	3	2
pada	0	0	0	0	1	0	1
efektifitas	0	0	0	0	1	0	1
jaga	0	0	0	0	1	0	1
timbul	0	0	0	0	1	0	1
solusi	0	0	0	0	1	0	1
manfaat	0	0	0	0	1	0	1
beli	0	0	0	0	1	0	1
php	0	0	0	0	1	1	2
my	0	0	0	0	1	0	1
sql	0	0	0	0	1	0	1
efektif	0	0	0	0	1	1	2
kelola	0	0	0	0	1	0	1
produk	0	0	0	0	1	0	1
pesan	0	0	0	0	1	0	1
lapor	0	0	0	0	1	0	1
proses	0	0	0	0	0	1	1
oleh	0	0	0	0	0	1	1
cepat	0	0	0	0	0	1	1
akurat	0	0	0	0	0	1	1
dukung	0	0	0	0	0	1	1

KATA	TF						DF
	D1	D2	D3	D4	D5	DU	
teknologi	0	0	0	0	0	2	1
canggih	0	0	0	0	0	1	1
maju	0	0	0	0	0	1	1
lembaga	0	0	0	0	0	1	1
didik	0	0	0	0	0	1	1
bas	0	0	0	0	0	1	1
komputer	0	0	0	0	0	1	1
jaring	0	0	0	0	0	1	1
bersifat	0	0	0	0	0	1	1
efisien	0	0	0	0	0	1	1
pengolaha	0	0	0	0	0	1	1
presensi	0	0	0	0	0	1	1
sms	0	0	0	0	0	1	1
gateway	0	0	0	0	0	1	1
bahasa	0	0	0	0	0	1	1
pemrograman	0	0	0	0	0	1	1
database	0	0	0	0	0	1	1
mysql	0	0	0	0	0	1	1

Langkah berikutnya adalah mencari nilai N/df_{ti}

Keterangan: N : banyaknya dokumen.

df_{ti} : jumlah dokumen dimana teks muncul.

$$\text{Contoh} = \text{tuna} : \frac{6}{1} = 6$$

$$\text{sistem} : \frac{6}{3} = 2$$

Selanjutnya adalah melakukan perhitungan *Inverse Document Frequency* (IDF). Berikut contoh perhitungan IDF pada kata 'tuna' dan 'sistem' yang terdapat pada data 1 dan hasil perhitungan yang lain selengkapnya ditunjukkan pada Tabel 3.2

$$IDF = \log \frac{N}{df_{ti}}$$

$$\text{Contoh} = \text{tuna} : IDF = \log \frac{N}{df_{ti}}$$

$$= \log \frac{6}{1}$$

$$= \log 6$$

$$\begin{aligned}
 &= (1,77815125) \\
 \text{sistem : } \quad IDF &= \log \frac{N}{df_{ti}} \\
 &= \log \frac{6}{3} \\
 &= \log 2 \\
 &= (1,301029996)
 \end{aligned}$$

Tabel 3.2 Perhitungan Pembobotan IDF

KATA	TF						DF	N/DF	IDF
	D1	D2	D3	D4	D5	DU			
tuna	4	0	0	0	0	0	1	6	1.77815125
netra	4	0	0	0	0	0	1	6	1.77815125
istilah	1	0	0	0	0	0	1	6	1.77815125
orang	1	0	0	0	0	0	1	6	1.77815125
ganggu	2	0	0	0	0	0	1	6	1.77815125
lihat	1	0	0	0	0	0	1	6	1.77815125
dasar	1	0	0	0	0	0	1	6	1.77815125
tingkat	1	0	0	0	0	0	1	6	1.77815125
bagi	1	0	0	0	0	0	1	6	1.77815125
total	1	0	0	0	0	0	1	6	1.77815125
blind	1	0	0	0	0	0	1	6	1.77815125
low	1	0	0	0	0	0	1	6	1.77815125
vision	1	0	0	0	0	0	1	6	1.77815125
sandang	2	0	0	0	0	0	1	6	1.77815125
memiliki	1	0	0	0	0	0	1	6	1.77815125
batas	2	0	0	0	0	0	1	6	1.77815125
alami	1	0	0	1	0	0	2	3	1.477121255
jalan	1	1	0	0	0	0	2	3	1.477121255
ibadah	1	0	0	0	0	0	1	6	1.77815125
shalat	2	0	0	0	0	0	1	6	1.77815125
tentu	1	0	0	0	0	0	1	6	1.77815125
arah	1	0	0	0	0	0	1	6	1.77815125
kilat	1	0	0	0	0	0	1	6	1.77815125
atas	1	0	0	0	0	0	1	6	1.77815125
masalah	1	0	0	0	0	0	1	6	1.77815125
buat	1	0	0	0	0	0	1	6	1.77815125
sistem	1	0	0	0	2	2	3	2	1.301029996
alat	1	0	1	0	0	0	2	3	1.477121255
bantu	1	0	1	0	1	0	3	2	1.301029996

KATA	TF						DF	N/DF	IDF
	D1	D2	D3	D4	D5	DU			
jembatan	0	6	0	0	0	0	1	6	1.77815125
konstruksi	0	1	0	0	0	0	1	6	1.77815125
hubung	0	1	0	0	0	0	1	6	1.77815125
pilih	0	1	0	0	0	0	1	6	1.77815125
tipe	0	4	0	0	0	0	1	6	1.77815125
bahas	0	2	0	0	0	0	1	6	1.77815125
optimalisasi	0	1	0	0	0	0	1	6	1.77815125
strukture	0	1	0	0	0	0	1	6	1.77815125
steel	0	1	0	0	0	0	1	6	1.77815125
design	0	1	0	0	0	0	1	6	1.77815125
bentang	0	1	0	0	0	0	1	6	1.77815125
warren	0	1	0	0	0	0	1	6	1.77815125
pratt	0	1	0	0	0	0	1	6	1.77815125
profil	0	2	0	0	0	0	1	6	1.77815125
wf	0	2	0	0	0	0	1	6	1.77815125
gelagar	0	2	0	0	0	0	1	6	1.77815125
mamanjang	0	1	0	0	0	0	1	6	1.77815125
jarak	0	2	0	0	0	0	1	6	1.77815125
meter	0	2	0	0	0	0	1	6	1.77815125
lintang	0	1	0	0	0	0	1	6	1.77815125
single	0	0	3	0	0	0	1	6	1.77815125
exponential	0	0	2	0	0	0	1	6	1.77815125
smoothing	0	0	2	0	0	0	1	6	1.77815125
moving	0	0	1	0	0	0	1	6	1.77815125
average	0	0	1	0	0	0	1	6	1.77815125
prediksi	0	0	2	0	0	0	1	6	1.77815125
data	0	0	1	0	1	1	3	2	1.301029996
muncul	0	0	1	0	0	0	1	6	1.77815125
algoritma	0	0	3	0	0	0	1	6	1.77815125
presentasi	0	0	1	0	0	0	1	6	1.77815125
mape	0	0	2	0	0	0	1	6	1.77815125
hasil	0	0	1	0	0	0	1	6	1.77815125
ukur	0	0	1	0	0	0	1	6	1.77815125
kerja	0	0	1	0	0	0	1	6	1.77815125
metode	0	0	1	1	0	0	2	3	1.477121255
uji	0	0	1	0	0	0	1	6	1.77815125
time	0	0	1	0	0	0	1	6	1.77815125
series	0	0	1	0	0	0	1	6	1.77815125
analysis	0	0	1	0	0	0	1	6	1.77815125

KATA	TF						DF	N/DF	IDF
	D1	D2	D3	D4	D5	DU			
aplikasi	0	0	1	0	0	0	1	6	1.77815125
pom	0	0	1	0	0	0	1	6	1.77815125
qm	0	0	1	0	0	0	1	6	1.77815125
kategori	0	0	1	0	0	0	1	6	1.77815125
bagus	0	0	1	0	0	0	1	6	1.77815125
presentase	0	0	1	0	0	0	1	6	1.77815125
alpha	0	0	1	0	0	0	1	6	1.77815125
bayar	0	0	1	0	0	0	1	6	1.77815125
spp	0	0	1	0	0	0	1	6	1.77815125
semester	0	0	1	0	0	0	1	6	1.77815125
genap	0	0	1	0	0	0	1	6	1.77815125
skuter	0	0	0	3	0	0	1	6	1.77815125
cc	0	0	0	4	0	0	1	6	1.77815125
motor	0	0	0	3	0	0	1	6	1.77815125
bakar	0	0	0	5	0	0	1	6	1.77815125
langkah	0	0	0	1	0	0	1	6	1.77815125
pasu	0	0	0	1	0	0	1	6	1.77815125
bahan	0	0	0	4	0	0	1	6	1.77815125
crankshaft	0	0	0	2	0	0	1	6	1.77815125
valve	0	0	0	1	0	0	1	6	1.77815125
lubang	0	0	0	1	0	0	1	6	1.77815125
masuk	0	0	0	1	0	0	1	6	1.77815125
guna	0	0	0	1	0	0	1	6	1.77815125
mesin	0	0	0	1	0	0	1	6	1.77815125
bekas	0	0	0	1	0	0	1	6	1.77815125
pakai	0	0	0	1	0	0	1	6	1.77815125
boros	0	0	0	1	0	0	1	6	1.77815125
konsumsi	0	0	0	2	0	0	1	6	1.77815125
eksperimental	0	0	0	1	0	0	1	6	1.77815125
banding	0	0	0	1	0	0	1	6	1.77815125
standart	0	0	0	1	0	0	1	6	1.77815125
tambah	0	0	0	1	1	0	2	3	1.477121255
membran	0	0	0	1	0	0	1	6	1.77815125
toko	0	0	0	0	1	0	1	6	1.77815125
buku	0	0	0	0	3	0	1	6	1.77815125
mulya	0	0	0	0	1	0	1	6	1.77815125
usaha	0	0	0	0	1	0	1	6	1.77815125
jasa	0	0	0	0	1	0	1	6	1.77815125
mesan	0	0	0	0	3	0	1	6	1.77815125

KATA	TF						DF	N/DF	IDF
	D1	D2	D3	D4	D5	DU			
informasi	0	0	0	0	2	3	2	3	1.477121255
pada	0	0	0	0	1	0	1	6	1.77815125
efektifitas	0	0	0	0	1	0	1	6	1.77815125
jaga	0	0	0	0	1	0	1	6	1.77815125
timbul	0	0	0	0	1	0	1	6	1.77815125
solusi	0	0	0	0	1	0	1	6	1.77815125
manfaat	0	0	0	0	1	0	1	6	1.77815125
beli	0	0	0	0	1	0	1	6	1.77815125
php	0	0	0	0	1	1	2	3	1.477121255
my	0	0	0	0	1	0	1	6	1.77815125
sql	0	0	0	0	1	0	1	6	1.77815125
efektif	0	0	0	0	1	1	2	3	1.477121255
kelola	0	0	0	0	1	0	1	6	1.77815125
produk	0	0	0	0	1	0	1	6	1.77815125
pesan	0	0	0	0	1	0	1	6	1.77815125
lapor	0	0	0	0	1	0	1	6	1.77815125
proses	0	0	0	0	0	1	1	6	1.77815125
oleh	0	0	0	0	0	1	1	6	1.77815125
cepat	0	0	0	0	0	1	1	6	1.77815125
akurat	0	0	0	0	0	1	1	6	1.77815125
dukung	0	0	0	0	0	1	1	6	1.77815125
teknologi	0	0	0	0	0	2	1	6	1.77815125
canggih	0	0	0	0	0	1	1	6	1.77815125
maju	0	0	0	0	0	1	1	6	1.77815125
lembaga	0	0	0	0	0	1	1	6	1.77815125
didik	0	0	0	0	0	1	1	6	1.77815125
bas	0	0	0	0	0	1	1	6	1.77815125
komputer	0	0	0	0	0	1	1	6	1.77815125
jaring	0	0	0	0	0	1	1	6	1.77815125
bersifat	0	0	0	0	0	1	1	6	1.77815125
efisien	0	0	0	0	0	1	1	6	1.77815125
pengolaha	0	0	0	0	0	1	1	6	1.77815125
presensi	0	0	0	0	0	1	1	6	1.77815125
sms	0	0	0	0	0	1	1	6	1.77815125
gateway	0	0	0	0	0	1	1	6	1.77815125
bahasa	0	0	0	0	0	1	1	6	1.77815125
pemrograman	0	0	0	0	0	1	1	6	1.77815125
database	0	0	0	0	0	1	1	6	1.77815125
mysql	0	0	0	0	0	1	1	6	1.77815125

Berdasarkan Tabel 3.1 dan Tabel 3.2, dilakukan pembobotan *TF-IDF* sehingga dihasilkan bobot dari masing-masing *term* sebagaimana Tabel 3.3.

Tabel 3.3 Hasil Perhitungan *TF-IDF*

KATA	W					
	D1	D2	D3	D4	D5	DU
tuna	7.112605	0	0	0	0	0
netra	7.112605	0	0	0	0	0
istilah	1.778151	0	0	0	0	0
orang	1.778151	0	0	0	0	0
ganggu	3.556303	0	0	0	0	0
lihat	1.778151	0	0	0	0	0
dasar	1.778151	0	0	0	0	0
tingkat	1.778151	0	0	0	0	0
bagi	1.778151	0	0	0	0	0
total	1.778151	0	0	0	0	0
blind	1.778151	0	0	0	0	0
low	1.778151	0	0	0	0	0
vision	1.778151	0	0	0	0	0
sandang	3.556303	0	0	0	0	0
memiliki	1.778151	0	0	0	0	0
batas	3.556303	0	0	0	0	0
alami	1.477121	0	0	1.477121	0	0
jalan	1.477121	1.477121	0	0	0	0
ibadah	1.778151	0	0	0	0	0
shalat	3.556303	0	0	0	0	0
tentu	1.778151	0	0	0	0	0
arah	1.778151	0	0	0	0	0
kilat	1.778151	0	0	0	0	0
atas	1.778151	0	0	0	0	0
masalah	1.778151	0	0	0	0	0
buat	1.778151	0	0	0	0	0
sistem	1.30103	0	0	0	2.60206	2.60205999
alat	1.477121	0	1.477121	0	0	0
bantu	1.30103	0	1.30103	0	1.30103	0
jembatan	0	10.66891	0	0	0	0
konstruksi	0	1.778151	0	0	0	0
hubung	0	1.778151	0	0	0	0

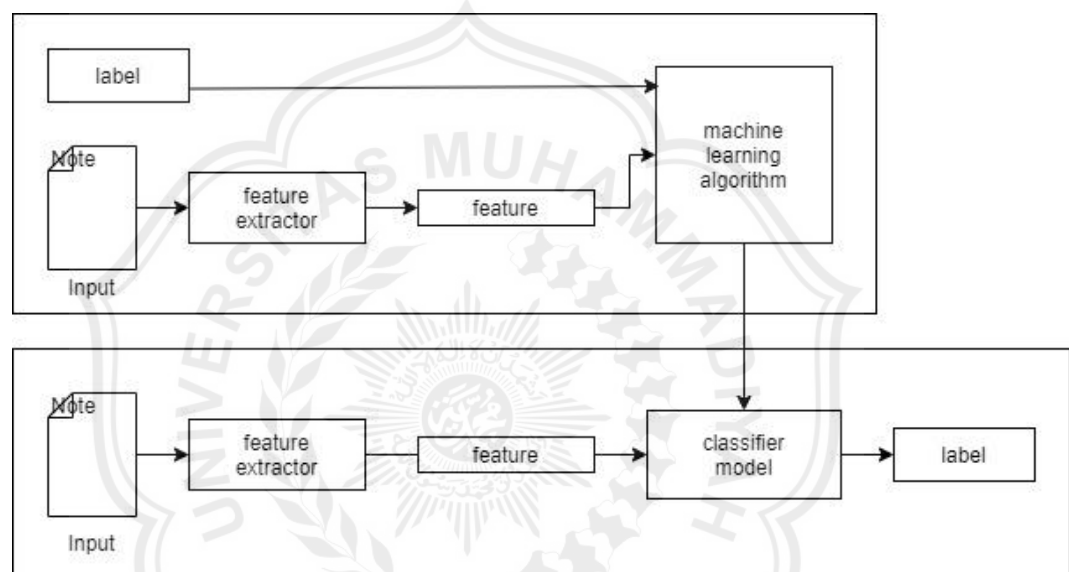
KATA	W					
	D1	D2	D3	D4	D5	DU
pilih	0	1.778151	0	0	0	0
tipe	0	7.112605	0	0	0	0
bahas	0	3.556303	0	0	0	0
optimalisasi	0	1.778151	0	0	0	0
strukture	0	1.778151	0	0	0	0
steel	0	1.778151	0	0	0	0
design	0	1.778151	0	0	0	0
bentang	0	1.778151	0	0	0	0
warren	0	1.778151	0	0	0	0
pratt	0	1.778151	0	0	0	0
profil	0	3.556303	0	0	0	0
wf	0	3.556303	0	0	0	0
gelagar	0	3.556303	0	0	0	0
mamanjang	0	1.778151	0	0	0	0
jarak	0	3.556303	0	0	0	0
meter	0	3.556303	0	0	0	0
lintang	0	1.778151	0	0	0	0
single	0	0	5.334454	0	0	0
exponential	0	0	3.556303	0	0	0
smoothing	0	0	3.556303	0	0	0
moving	0	0	1.778151	0	0	0
average	0	0	1.778151	0	0	0
prediksi	0	0	3.556303	0	0	0
data	0	0	1.30103	0	1.30103	1.30103
muncul	0	0	1.778151	0	0	0
algoritma	0	0	5.334454	0	0	0
presentasi	0	0	1.778151	0	0	0
mape	0	0	3.556303	0	0	0
hasil	0	0	1.778151	0	0	0
ukur	0	0	1.778151	0	0	0
kerja	0	0	1.778151	0	0	0
metode	0	0	1.477121	1.477121	0	0
uji	0	0	1.778151	0	0	0
time	0	0	1.778151	0	0	0
series	0	0	1.778151	0	0	0
analysis	0	0	1.778151	0	0	0
aplikasi	0	0	1.778151	0	0	0
pom	0	0	1.778151	0	0	0
qm	0	0	1.778151	0	0	0

KATA	W					
	D1	D2	D3	D4	D5	DU
kategori	0	0	1.778151	0	0	0
bagus	0	0	1.778151	0	0	0
presentase	0	0	1.778151	0	0	0
alpha	0	0	1.778151	0	0	0
bayar	0	0	1.778151	0	0	0
spp	0	0	1.778151	0	0	0
semester	0	0	1.778151	0	0	0
genap	0	0	1.778151	0	0	0
skuter	0	0	0	5.334454	0	0
cc	0	0	0	7.112605	0	0
motor	0	0	0	5.334454	0	0
bakar	0	0	0	8.890756	0	0
langkah	0	0	0	1.778151	0	0
pasu	0	0	0	1.778151	0	0
bahan	0	0	0	7.112605	0	0
crankshaft	0	0	0	3.556303	0	0
valve	0	0	0	1.778151	0	0
lubang	0	0	0	1.778151	0	0
masuk	0	0	0	1.778151	0	0
guna	0	0	0	1.778151	0	0
mesin	0	0	0	1.778151	0	0
bekas	0	0	0	1.778151	0	0
pakai	0	0	0	1.778151	0	0
boros	0	0	0	1.778151	0	0
konsumsi	0	0	0	3.556303	0	0
eksperimental	0	0	0	1.778151	0	0
banding	0	0	0	1.778151	0	0
standart	0	0	0	1.778151	0	0
tambah	0	0	0	1.477121	1.477121	0
membran	0	0	0	1.778151	0	0
toko	0	0	0	0	1.778151	0
buku	0	0	0	0	5.334454	0
mulya	0	0	0	0	1.778151	0
usaha	0	0	0	0	1.778151	0
jasa	0	0	0	0	1.778151	0
mesan	0	0	0	0	5.334454	0
informasi	0	0	0	0	2.954243	4.43136376
pada	0	0	0	0	1.778151	0
efektifitas	0	0	0	0	1.778151	0

KATA	W					
	D1	D2	D3	D4	D5	DU
jaga	0	0	0	0	1.778151	0
timbul	0	0	0	0	1.778151	0
solusi	0	0	0	0	1.778151	0
manfaat	0	0	0	0	1.778151	0
beli	0	0	0	0	1.778151	0
php	0	0	0	0	1.477121	1.47712125
my	0	0	0	0	1.778151	0
sql	0	0	0	0	1.778151	0
efektif	0	0	0	0	1.477121	1.47712125
kelola	0	0	0	0	1.778151	0
produk	0	0	0	0	1.778151	0
pesan	0	0	0	0	1.778151	0
lapor	0	0	0	0	1.778151	0
proses	0	0	0	0	0	1.77815125
oleh	0	0	0	0	0	1.77815125
cepat	0	0	0	0	0	1.77815125
akurat	0	0	0	0	0	1.77815125
dukung	0	0	0	0	0	1.77815125
teknologi	0	0	0	0	0	3.5563025
canggih	0	0	0	0	0	1.77815125
maju	0	0	0	0	0	1.77815125
lembaga	0	0	0	0	0	1.77815125
didik	0	0	0	0	0	1.77815125
bas	0	0	0	0	0	1.77815125
komputer	0	0	0	0	0	1.77815125
jaring	0	0	0	0	0	1.77815125
bersifat	0	0	0	0	0	1.77815125
efisien	0	0	0	0	0	1.77815125
pengolaha	0	0	0	0	0	1.77815125
presensi	0	0	0	0	0	1.77815125
sms	0	0	0	0	0	1.77815125
gateway	0	0	0	0	0	1.77815125
bahasa	0	0	0	0	0	1.77815125
pemrograman	0	0	0	0	0	1.77815125
database	0	0	0	0	0	1.77815125
mysql	0	0	0	0	0	1.77815125

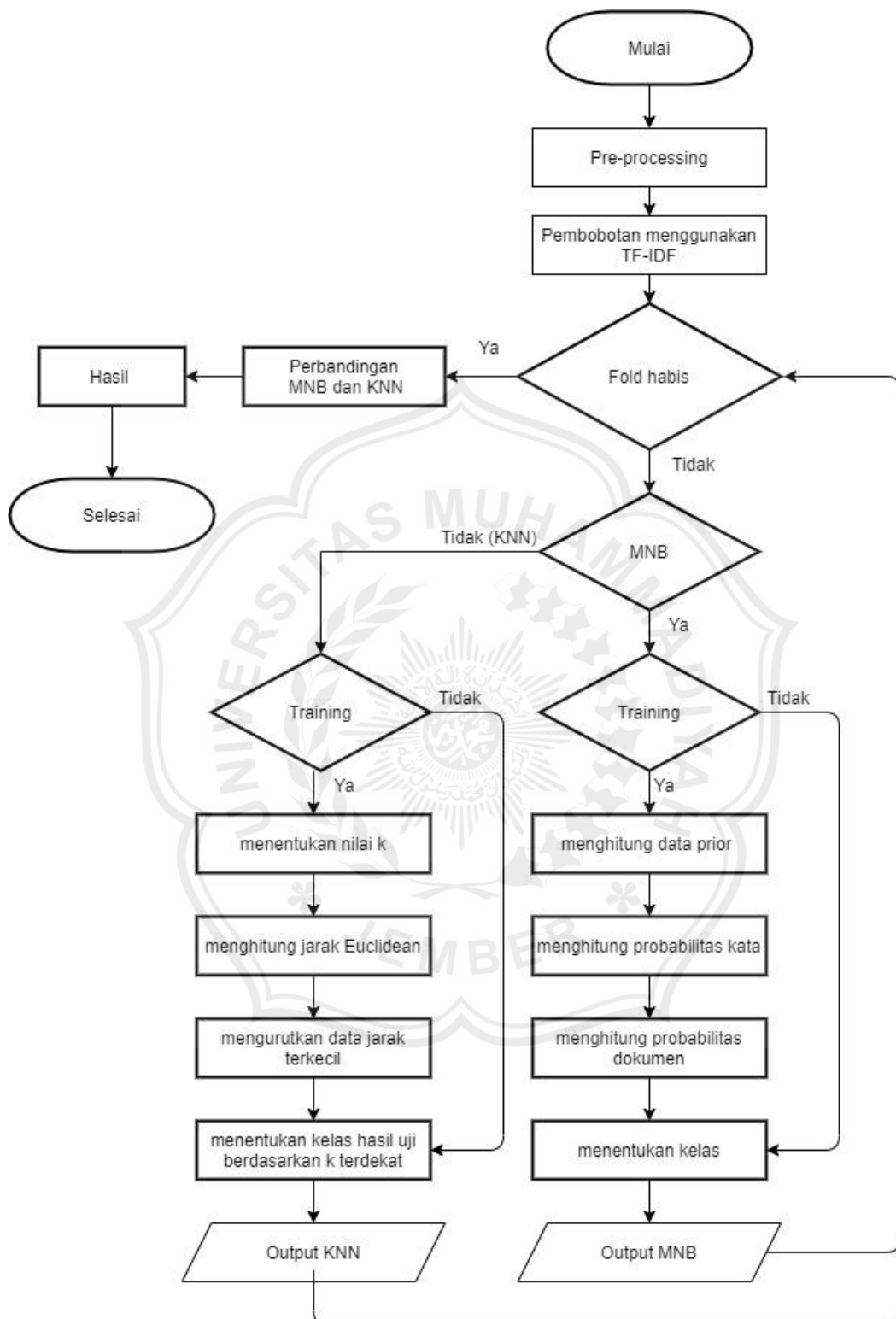
3.6 Klasifikasi

Setelah di dapatkan hasil dari *pre-processing* data dan pembobotan kata pada data, selanjutnya akan dilakukan proses pengklasifikasian pada data. Klasifikasi merupakan sebuah proses pengelompokan sesuatu ke dalam kelas-kelas atau kategori yang berbeda dalam kasus ini yang diklasifikasikan adalah abstrak atau dokumen. Berikut merupakan gambaran umum dari proses klasifikasi yang di tunjukan pada gambar 3.3.



Gambar 3.3. Alur proses klasifikasi

Proses klasifikasi yang digunakan pada penelitian ini menggunakan metode *Multinomial Naive Bayes* dan *K-Nearest Neighbour*. Tahap klasifikasi dibagi menjadi dua tahapan, yaitu tahap pelatihan (*training*) dan pengujian (*testing*). Berikut adalah alur klasifikasi dari algoritma *Multinomial Naive Bayes* dan *K-Nearest Neighbour* yang akan di tunjukan pada Gambar 3.4.



Gambar 3.4. Flowchart alur klasifikasi MNB dan KNN

Alur klasifikasi algoritma dimulai dari dengan *pre-processing* data. Terdapat 4 tahap yaitu *casefolding*, *tokenizing*, *filtering* dan *stemming*. Data yang telah di *pre-processing* selanjutnya akan di lakukan proses pembobotan dengan menggunakan *TF-IDF*. Setelah proses *TF-IDF* akan dilakukan proses pembagian data dengan menggunakan *K-Fold*. *K-Fold* membagi data antara data *training* dan data *testing*, *Fold K* yang di gunakan antara lain 2, 4, 5 dan 10. Setelah itu dilakukan proses klasifikasi dengan menggunakan algoritma *Multinomial Naïve Bayes* dan *K-Nearest Neighbour*.

Dalam proses klasifikasi pada algoritma *Multinomial Naïve Bayes* di awali dengan mengitung data prior dari masing-masing kelas pada data *training* dan data *testing* lalu dilakukan penghitungan probabilitas kata ke-n pada data dan menghitung probabilitas suatu dokumen masuk ke dalam suatu kelas, yang terakhir yaitu menentukan kelas dokumen uji dengan memilih probabilitas tertinggi. Sedangkan dalam algoritma *K-Nearest Neighbour* di awali dengan menentukan nilai *K* dari KNN lalu di lanjutkan dengan menghitung jarak *Euclidean* antara data *testing* dan data *training* lalu di urutkan ke data yang mempunyai jarak terkecil, dan yang terakhir yaitu menentukan kategori kelas dari data hasil uji berdasarkan label dari *K* tetangga terdekat.

Hasil dari masing-masing algoritma dengan menggunakan *Fold k = 2, 4, 5* dan 10 akan di bandingkan sehingga di dapatkan algoritma yang memiliki hasil klasifikasi tertinggi.

3.6.1 Pelatihan Menggunakan Metode *Multinomial Naive Bayes*

Data *training* yang telah melewati seleksi fitur kemudian digunakan sebagai bahan pembelajaran pada proses *testing*. Proses selanjutnya, yaitu melakukan pembelajaran *Multinomial Naive Bayes* untuk data latih dengan menghitung nilai $P(w_k | v_j)$ ditentukan dengan persamaan:

$$P(w_k | v_j) = \frac{W_{ct} + 1}{(\sum_{W' \in V} W'_{ct}) + B'}$$

Contoh:

Untuk melakukan pembelajaran *Multinomial Naïve Bayes* diambil satu contoh kata yang terdapat pada dokumen yaitu kata “tuna”.

$$\begin{aligned}
 \text{tuna} &= \frac{W_{ct}+1}{(\sum_{W' \in V} W'_{ct})+B'} \\
 &= \frac{7,11261+1}{(255,77047+67,4906)} \\
 &= 0,0251
 \end{aligned}$$

3.6.2 Pengujian Menggunakan Metode *Multinomial Naive Bayes*

Selanjutnya, setelah didapatkan data latih, maka tahapan dilanjutkan ke dalam proses klasifikasi untuk menguji model yang telah dibangun kepada data uji yang telah melewati *text pre-processing* dan kemudian dihitung nilai tiap fitur kata yang akan digunakan untuk proses klasifikasi menggunakan metode *Multinomial Naive Bayes*.

Tahapan pertama, menghitung data prior masing-masing kelas dengan menggunakan rumus. Probabilitas prior kelas c ditentukan dengan rumus:

$$P(c) = \frac{N_c}{N}$$

Keterangan:

N_c = Jumlah kelas c pada seluruh dokumen.

N = Jumlah seluruh dokumen.

Setelah nilai probabilitas ditentukan, selanjutnya menghitung probabilitas kata dan menghitung probabilitas suatu dokumen masuk ke dalam suatu kelas dengan rumus berikut:

$$P(c|\text{term dokumen } a) = P(c) \times P(t_1|c) \times P(t_2|c) \times P(t_3|c) \times \dots \times P(t_n|c)$$

Keterangan:

$P(c)$ = Probabilitas *prior* dari kelas c .

t_n = Kata dokumen a ke- n .

$P(c|term\ dokumen\ d)$ = Probabilitas suatu dokumen termasuk kelas c .

$P(t_n|c)$ = Probabilitas kata ke - n dengan diketahui kelas c .



Tabel 3.4 Hasil Perhitungan MNB

<i>MULTINOMIAL NAÏVE BAYES</i>									
P(0)	0.2	P(1)	0.2	P(2)	0.2	P(3)	0.2	P(4)	0.2
P(tn c)									
sistem	0.011142883	sistem	0.01133776	Sistem	0.011071671	sistem	0.011049486	sistem	0.009916965
data	0.007118179	data	0.007242668	Data	0.007072688	data	0.007058516	data	0.006335051
infromasi	0.016801789	infromasi	0.017095633	Infromasi	0.016694412	infromasi	0.01666096	infromasi	0.014953289
php	0.007662913	php	0.007796929	Php	0.007613941	php	0.007598684	php	0.006819854
efektif	0.007662913	efektif	0.007796929	Efektif	0.007613941	efektif	0.007598684	efektif	0.006819854
proses	0.008594142	proses	0.008744444	Proses	0.008539218	proses	0.008522108	proses	0.010861892
oleh	0.008594142	oleh	0.008744444	Oleh	0.008539218	oleh	0.008522108	oleh	0.007648631
cepat	0.008594142	cepat	0.008744444	Cepat	0.008539218	cepat	0.008522108	cepat	0.007648631
akurat	0.008594142	akurat	0.008744444	Akurat	0.008539218	akurat	0.008522108	akurat	0.007648631
dukung	0.008594142	dukung	0.008744444	dukung	0.008539218	dukung	0.008522108	dukung	0.007648631
teknologi	0.014094808	teknologi	0.014341311	teknologi	0.014004731	teknologi	0.013976669	teknologi	0.012544126
canggih	0.008594142	canggih	0.008744444	canggih	0.008539218	canggih	0.008522108	canggih	0.007648631
maju	0.008594142	maju	0.008744444	Maju	0.008539218	maju	0.008522108	maju	0.007648631
lembaga	0.008594142	lembaga	0.008744444	lembaga	0.008539218	lembaga	0.008522108	lembaga	0.007648631

didik	0.008594142	didik	0.008744444	Didik	0.008539218	didik	0.008522108	didik	0.007648631
bas	0.008594142	bas	0.008744444	Bas	0.008539218	bas	0.008522108	bas	0.007648631
komputer	0.008594142	komputer	0.008744444	komputer	0.008539218	komputer	0.008522108	komputer	0.007648631
jaring	0.008594142	jaring	0.008744444	Jaring	0.008539218	jaring	0.008522108	jaring	0.007648631
bersifaat	0.008594142	bersifaat	0.008744444	bersifaat	0.008539218	bersifaat	0.008522108	bersifaat	0.007648631
efisien	0.008594142	efisien	0.008744444	Efisien	0.008539218	efisien	0.008522108	efisien	0.007648631
pengolaha	0.008594142	pengolaha	0.008744444	pengolaha	0.008539218	pengolaha	0.008522108	pengolaha	0.007648631
presensi	0.008594142	presensi	0.008744444	presensi	0.008539218	presensi	0.008522108	presensi	0.007648631
sms	0.008594142	sms	0.008744444	Sms	0.008539218	sms	0.008522108	sms	0.007648631
gateway	0.008594142	gateway	0.008744444	gateway	0.008539218	gateway	0.008522108	gateway	0.007648631
bahasa	0.008594142	bahasa	0.008744444	Bahasa	0.008539218	bahasa	0.008522108	bahasa	0.007648631
pemrograman	0.008594142	Pemrograman	0.008744444	Pemrograman	0.008539218	pemrograman	0.008522108	pemrograman	0.007648631
database	0.008594142	database	0.008744444	database	0.008539218	database	0.008522108	database	0.007648631
mysql	0.008594142	mysql	0.008744444	Mysql	0.008539218	mysql	0.008522108	mysql	0.007648631
P(0 term dok d) = P(0) *P(tn 0)	7.87144E-59	P(1 term dok d) = P(1) *P(tn 1)	1.27904E-58	P(2 term dok d) = P(2) *P(tn 2)	6.57796E-59	P(3 term dok d) = P(3) *P(tn 3)	6.21871E-59	P(4 term dok d) = P(4) *P(tn 4)	4.2761 2E-60
Klasifikasi Data Uji	1.27904E-58								

Peluang kemunculan kata yang besar akan menghasilkan probabilitas yang tinggi, sehingga dokumen data uji akan diklasifikasikan ke dalam karakter dengan probabilitas yang paling tinggi. Pada kasus data uji diatas, dapat disimpulkan bahwa probabilitas yang paling tinggi adalah kelas 1, maka dokumen diklasifikasikan ke dalam kelas Teknik Elektro.

3.6.3 Pelatihan Menggunakan Metode *K-Nearest Neighbour*

Pada tahap pelatihan ini dilakukan untuk mendapatkan pemodelan dalam proses klasifikasi ke dalam lima kelas dengan mencari jarak *Euclidean* pada data latih. Data uji yang digunakan adalah data uji yang telah melalui tahap *preprocessing* dan pembobotan tf-idf.

Untuk mendapatkan jarak *Euclidean* tiap dokumen, digunakan persamaan Sebagai contoh dilakukan perhitungan mencari kedekatan antara (D1) dan (DU) sebagai berikut:

$$d(x_i, x_j) = \sqrt{(0 - 50,58915)^2 + (0 - 3,161822)^2 + (0 - 12,64729)^2 + \dots + (x_{ip_n} - x_{jp_n})^2}$$

$$d(x_1, x_j) = 18,06964499$$

Langkah-langkah untuk menemukan *Euclidean* pada dokumen kedua dan ketiga sama saja seperti menentukan *Euclidean* pertama. Berikut ini merupakan nilai *Euclidean* untuk (D2), (D3), (D4) dan (D5):

$$d(x_2, x_j) = 19,87696493$$

$$d(x_3, x_j) = 17,20747867$$

$$d(x_4, x_j) = 20,68608001$$

$$d(x_5, x_j) = 14,10416915$$

3.6.4 Pengujian Menggunakan Metode *K-Nearest Neighbour*

Setelah dilakukan pencarian jarak *Euclidean*, langkah selanjutnya adalah mengurutkan jarak *Euclidean* tersebut dari nilai yang paling kecil sampai nilai

yang terbesar dengan $K=1$. Sebagai data uji dilakukan pada Dokumen Uji (DU) dengan jarak *Euclidean* seperti yang ditunjukkan pada Tabel 3.5 berikut:

Tabel 3.5 Rank kedekatan jarak *euclidean*

DOKUMEN	KEDEKATAN
D1	18.06964499
D2	19.87696493
D3	17.20747867
D4	20.68608001
D5	14.10416915

Jarak *Euclidean* terkecil berada pada dokumen 5 (D5), maka data uji Dokumen Uji (DU) merupakan dokumen yang memiliki kelas yang sama dengan dokumen 5 (D5) yaitu kelas Manajemen Informatika.

3.7 Evaluasi

Proses evaluasi dilakukan terhadap kumpulan dataset abstrak tugas akhir mahasiswa Program Studi Teknik Informatika Universitas Muhammadiyah Jember untuk mendapatkan performansi akurasi yang optimal. Proses evaluasi menggunakan *Confusion Matrix*. *Confusion Matrix* sangat berguna untuk menganalisis kualitas *classifier* dalam mengenali tuple-tuple dari kelas yang ada. Untuk menilai atau mengevaluasi suatu model klasifikasi dapat menggunakan akurasi atau tingkat pengenalan dapat menggunakan rumus $\frac{TP+TN}{TP+FP+TN+FN}$. Akurasi atau tingkat pengenalan menyatakan presentase dari jumlah tuple dalam data uji yang diklasifikasi dengan benar oleh *classifier* (Suyanto, 2017).