

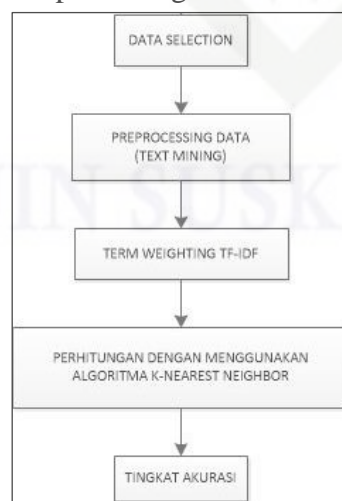
## BAB IV

### ANALISA

Analisa merupakan sebuah tahapan dimana dilakukan pembahasan untuk memahami permasalahan yang dihadapi dalam penelitian yang dilakukan. Dengan adanya tahapan analisa ini maka inti penelitian yang ada dapat dijelaskan secara detail dan dapat dipelajari secara mendalam sehingga akan dapat dipahami dengan baik.

#### 4.1. Analisa Penerapan KNN untuk Klasifikasi Kepribadian

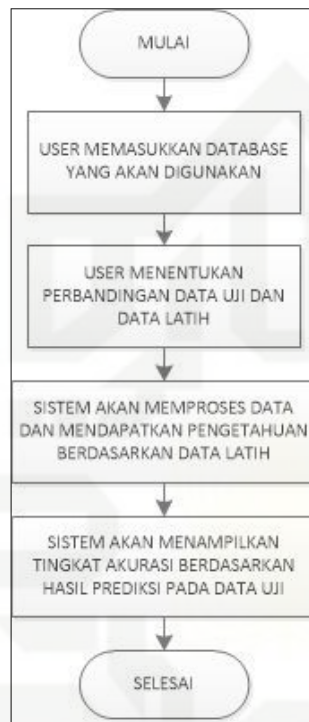
Dalam menerapkan algoritma *K-Nearest Neighbor* untuk mengklasifikasikan kepribadian, dilakukan beberapa proses penting yaitu *selection data*, *preprocessing data*, pengujian dengan menggunakan algoritma *K-nearest Neighbor*. Tahap *selection data* dilakukan untuk memilih atribut yang dibutuhkan untuk diproses, pada tahapan *preprocessing* dilakukan pemrosesan data sebelum dilakukan pengujian, adapun proses yang dilakukan adalah proses *text mining* pada umumnya akan tetapi ditahapan ini peneliti hanya akan melakukan tahap *case folding*, *tokenizing* dan *filtering*. Pada tahap pengujian dilakukan pembobotan kata terlebih dahulu dengan metode pembobotan *TF-IDF*. Adapun rincian dalam tahap analisa dan perancangan adalah sebagai berikut:



**Gambar 4.1 Alur Penerapan KNN untuk Klasifikasi Kepribadian**

Gambar 4.1 merupakan alur utama atau proses yang melibatkan perhitungan pada penerapan algoritma *KNN* untuk klasifikasi kepribadian.

*Flowchart* pengujian *KNN* untuk mengklasifikasikan kepribadian dapat dilihat pada gambar 4.2 berikut:



**Gambar 4.2** *Flowchart* Pengujian Algoritma *KNN*

Pada gambar 4.2 diatas dapat dilihat bahwa *flowchart* dimulai dengan melakukan proses memasukkan database yang kan digunakan kedalam alat pengujian yang akan digunakan yaitu *spyder*, lalu user menentukan jumlah presentasi data latih dan data uji yang akan digunakan, program akan menganalisa data latih untuk di gunakan melakukan pengujian terhadap data uji yang telah ditentukan dan akan menampilkan tingkat akurasinya.

## 4.2. Pengumpulan Data

Data yang digunakan pada penelitian ini adalah data sekunder yang di dapatkan dari peneliti sebelumnya (Syafrianto, 2018). Data terdiri dari 15 akun Twitter dari orang yang telah mengisi angket tes kepribadian dengan teori kepribadian *Big Five Personality*, dengan perincian setiap akun memiliki 150

**Hak Cipta Dilindungi Undang-Undang**

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
  - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
  - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

*tweet*. Data penelitian terdiri dari 7 atribut, data tersebut dapat dilihat pada tabel 4.1 berikut.

**Tabel 4.1 Atribut Data Awal Penelitian**

No	Variabel	Deskripsi
1	<i>user_id</i>	User ID akun twitter
2	<i>screen_name</i>	Nama tampilan yang terdapat di akun twitter
3	teks	<i>Tweet</i> yang diambil sebelum proses <i>text mining</i>
4	preprocess	<i>Tweet</i> yang telah dilakukan proses <i>text mining</i>
5	tanggal	Tanggal <i>tweet</i> di posting oleh pengguna
6	keterangan	Merengkan bahwa data merupakan data uji atau data latih
7	kelas	Merupakan hasil tes psikologi yang di isi melalui angket yang disebar oleh peneliti sebelumnya, yang terdiri dari 5 class yaitu <i>Openness, Conscientiouness, Extraversion, Agreeableness, Neuroticism</i> yang disingkat menjadi O, C, E, A, N

Berdasarkan atribut data yang terlihat pada tabel 4.1 diatas, maka contoh data klasifikasi kepribadian berdasarkan *tweet* yang digunakan dalam penelitian ini akan dijabarkan pada tabel 4.2 berikut ini :



Hak Cipta dilindungi Undang-Undang  
 1. Dilarang mengutip, sebagian atau seluruhnya, atau membuat karya tulis tanpa mencantumkan dan menyebutkan sumber:  
 a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, pen-  
 b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.  
 2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruhnya karya tulis ini dalam bentuk apapun tanpa izin

**Tabel 4.2 Data Awal Penelitian**

No	User ID	Screen Name	Tweet Sebelum Proses	Tweet Setelah Proses	Tanggal Tweet	Keterangan	Kelas
1	297190207	Ameliahilyasa	Jangan balik ke dia ya, eh kok jahat..	jangan balik ke dia ya kok jahat	2015-02-20	Uji	O
2	297190207	Ameliahilyasa	Tak selamanya penampilan luar mencerminkan isi hati seseorang..	tak lama tampil luar cermin isi hati	2015-02-22	Uji	O
3	297190207	Ameliahilyasa	Luarnya aja yg kekar tapi di dalamnya rapuh..	luar aja kekar tapi di dalam rapuh	2015-02-28	Uji	O
4	1672477808	pratiwihenny12	Hari ini hari yang kau tunggu, bertambah satu tahun usiamu,â€ <a href="https://t.co/BUJkQjITCW">https://t.co/BUJkQjITCW</a>	hari ini hari yang kau tunggu tambah satu tahun usia	2016-04-21	Latih	C
5	1672477808	pratiwihenny12	Apalah daya jika waktu dan jarak memisahkan kita, hanya doa dan kerinduan yang bisaâ€ (w/ Oki Apri Nanda) [pic] â€ <a href="https://t.co/TO6YocQPUF">https://t.co/TO6YocQPUF</a>	apa daya jika waktu dan jarak pisah kita hanya doa dan rindu yang bisa	2016-04-24	Latih	C
6	1672477808	pratiwihenny12	Ada kalanya kita jenuh dengan aktivitas yang kita lakukan saat ini,â€ <a href="https://t.co/KHkS1FCA3T">https://t.co/KHkS1FCA3T</a>	ada kala kita jenuh dengan aktivitas yang kita laku saat ini	2016-05-14	Latih	C
7	561027897	ResdifaEza	Walau badai menerjang kami tetap satu, Laf you my girls ðŸ™ðŸ– <a href="https://t.co/dBGcYDeCdC">https://t.co/dBGcYDeCdC</a>	walau badai terjang kami tetap satu	2016-12-25	Uji	E
8	561027897	ResdifaEza	Ingatlah rencana allah lebih indah daripada yang kita inginkan, ðŸ–ðŸ– <a href="https://t.co/Jz8Sgu9nF1">https://t.co/Jz8Sgu9nF1</a>	ingat rencana lebih indah daripada yang kita ingin	2017-01-25	Uji	E
9	561027897	ResdifaEza	Aku akan tetap menjadi aku, Berusaha sekuatnya untuk melakukan ygâ€ <a href="https://t.co/5CoaxvrSMj">https://t.co/5CoaxvrSMj</a>	aku akan tetap jadi aku usaha kuat untuk laku	2017-02-02	Uji	E





Hak Cipta Dilindungi Undang-Undang  
 1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:  
 a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, pen-  
 b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.  
 2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin

No	User ID	Screen Name	Tweet Sebelum Proses	Tweet Setelah Proses	Tanggal Tweet	Keterangan	Kelas
10	895199329	Mellysyahfitrii	Bisa saja dengan mudah memaafkan, tapi dengan kata maaf itu tidak akan pernah mengembalikan keadaan seperti sedia kala:)	bisa saja dengan mudah maaf tapi dengan kata maaf itu tidak akan pernah kembali seperti sedia kala	2014-07-18	Latih	A
11	895199329	Mellysyahfitrii	Kalau jenuh dengan sifatku katakanlah, tak perlu risau memikirkankuu. Aku akan jauh lebih tegar dari apa yang kamu bayangkan:))	kalau jenuh dengan sifat katak tak perlu risau aku akan jauh lebih tegar dari apa yang kamu bayang	2014-07-18	Latih	A
12	895199329	Mellysyahfitrii	Hohooo enggakkokk;p"@safirashiela: Syhm? Hahacie"@mellysyahfitrii: Dibilang sayang sih enggak terlalu, tapi adalah rasa takut kehilangan	bilang sayang sih enggak lalu tapi adalah rasa takut hilang	2014-07-18	Latih	A
13	229008946	Jaakkaa	bberapa hari nyo, sabar :D "@alysadriana Tapi sepi jd nya jek wkwkwk @Jaakkaa: hahaa, skaliÂ² nyo :D @alysadriana Gak jek, ditinggal sndiri	hari sabar tapi sepi tinggal	2014-03-28	Uji	N
14	229008946	Jaakkaa	halahh, di Batam banyak nyo mang :D "@petakers17 nde dn ma arokan HP lo ny mang @Jaakkaa: siiip, dn tunggu samsung S4 untuak dn lah :D"	di banyak mang lo mang tunggu lah	2014-03-30	Uji	N
15	229008946	Jaakkaa	yoo sukses lah kawan lu, lai ka den tungguan nyo :D "@petakers17 y bnyak e,ptih e mang @Jaakkaa: halahh, di Batam banyak nyo mang :D"	sukses lah kawan lai ka den tunggu mang di banyak mang	2014-03-30	Uji	N



Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
  - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, pen-  
gutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
  - b. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izi-  
n UIN Suska Riau.

No	User ID	Screen Name	Tweet Sebelum Proses	Tweet Setelah Proses	Tanggal Tweet	Keterangan	Kelas
1500	390585245	zheptia_nj	Pagi yang indah bersama senyum manis	pagi yang indah sama senyum manis	2017-11-01	Uji	A

Untuk data selengkapnya dapat dilihat **dilampiran A**

### 4.3. Analisa Preprocessing

Berdasarkan analisa kebutuhan data yang telah dijelaskan diatas maka pada bagian ini dijelaskan teknik-teknik yang akan dilakukan terhadap data penelitian, penjelasannya sebagai berikut.

#### 4.3.1 Selection Data

*Selection Data* merupakan tahap pemilihan data/atribut yang akan dipakai untuk melakukan pengujian. Dari data awal yang memiliki 7 atribut yaitu *user\_id*, *screen\_name*, *teks*, *preprocess*, *tanggal*, *keterangan* dan *kelas* dipilih menjadi hanya 2 atribut saja yang digunakan yaitu atribut *teks* dan *kelas*.

#### 4.3.2 Text Mining

*Text Mining* adalah proses pengolahan data sebelum dilakukan pengujian. Pada proses ini dilakukan tahapan seperti *case folding*, *tokenizing*, *filtering*.

Contoh :

Pada tabel 4.3 dibawah ini terdapat dokumen yang belum melalui tahapan *text mining*.

**Tabel 4.3 Data sebelum proses *text mining***

No.	username	tweet
1	ameliahilyasa	Serasa mau pecah kepala kami ha pak karna tugas bapak-_-

Pada tahap *case folding* huruf S besar pada kata *serasa* akan dirubah menjadi huruf kecil, berikutnya pada tahap *tokenizing* karakter selain dari huruf a-z akan dihapus, dan pada tahap *filtering* susunan kalimat pada dokumen akan di pisahkan menjadi kata per kata dan dilakukan proses *stoplist removal* atau membuang kata-kata yang tidak penting berdasarkan kamus bahasa indonesia, sehingga hasil dari proses ini dapat dilihat pada tabel 4.4 berikut :

**Tabel 4.4 Data setelah melalui tahap *text mining***

Kata
pecah
kepala

<b>Kata</b>
tugas

### 4.3.3 Term frequention Dan Term Weighting Dengan TF-IDF

*Term frequention* adalah proses penghitungan kemunculan kata pada keseluruhan dokumen. *Term weighting* adalah proses dimana tiap kata yang muncul pada tiap dokemen akan diberi nilai bobot berdasarkan presentase jumlah kemunculannya.

Tabel 4.5 berikut merupakan 10 contoh document/data yang digunakan dalam penelitian ini dan telah melalui tahapan *text mining*, data terdiri atas 2 kelas *Openess*, 2 kelas *Neuroticism*, 2 kelas *Extraversion*, 2 kelas *Agreeablenes*, dan 2 kelas *Conscientiouness*.

**Tabel 4.5 Data yang akan beri pembobotan**

No.	username	tweet	Kelas
1	ameliahilyasa	rasa mau pecah kepala kami ha pak tugas bapak	O
2	ameliahilyasa	tau kakak mau kasih puji apa ha buat makan puasa puasa puji kak makan makan kapan	O
3	CNanadya_	kasi tinggal hitung hari	N
4	CNanadya_	lewat tanggung aja deh ya hajar pagi	N
5	mazaya_rizda	kakak moga makin sukses lulus masuk ingin traktir ya	E
6	mazaya_rizda	maaf ya bisa ini aja baru siap beres rumah maaf ya bilang juga iya	E
7	mellysyahfitrii	bisa saja dengan mudah maaf tapi dengan kata maaf itu tidak akan pernah kembali seperti sedia kala	A
8	mellysyahfitrii	kalau jenuh dengan sifat katak tak perlu risau aku akan jauh lebih tegar dari apa yang kamu bayang	A
9	pratiwihenny12	jadi ingat tegang hadap uli bapak	C
10	pratiwihenny12	satu kita teguh cerai kita runtuh tunjuk bakat dan tekad	C

Dari 10 data diatas, penelti akan melakukan pengujian untuk mengklasifikasikan kelas dari data pada tabel 4.6 berikut :



**Tabel 4.6 Tabel Data Uji**

No.	username	tweet	Kelas
1	windisisilia	moga besok lancar ya ada salah sedikit pun moga suara besok nyaring luar ya amin	E

Pada tahap ini semua kata yang ada didalam *document* telah diproses dengan prosedur *text mining* akan dipisahkan kata per kata dan di buat frekuensi kemunculan nya seperti pada tabel 4.7 dibawah ini.

**Tabel 4.7 Tabel Hasil Term Frequency**

Term	TF												IDF
	dx	d1	d2	d3	d4	d5	d6	d7	d8	d9	d10	df	log(n/df)
Moga	2	0	0	0	0	1	0	0	0	0	0	3	0,564271
Besok	2	0	0	0	0	0	0	0	0	0	0	2	0,740363
Lancar	1	0	0	0	0	0	0	0	0	0	0	1	1,041393
Salah	1	0	0	0	0	0	0	0	0	0	0	1	1,041393
Suara	1	0	0	0	0	0	0	0	0	0	0	1	1,041393
Nyaring	1	0	0	0	0	0	0	0	0	0	0	1	1,041393
Luar	1	0	0	0	0	0	0	0	0	0	0	1	1,041393
Pecah	0	1	0	0	0	0	0	0	0	0	0	1	1,041393
Kepala	0	1	0	0	0	0	0	0	0	0	0	1	1,041393
Tugas	0	1	0	0	0	0	0	0	0	0	0	1	1,041393
Kasih	0	0	1	0	0	0	0	0	0	0	0	1	1,041393
Puji	0	0	2	0	0	0	0	0	0	0	0	2	0,740363
Makan	0	0	2	0	0	0	0	0	0	0	0	2	0,740363
Puasa	0	0	2	0	0	0	0	0	0	0	0	2	0,740363
Kasi	0	0	0	1	0	0	0	0	0	0	0	1	1,041393
Tinggal	0	0	0	1	0	0	0	0	0	0	0	1	1,041393
Hitung	0	0	0	1	0	0	0	0	0	0	0	1	1,041393
tanggung	0	0	0	0	1	0	0	0	0	0	0	1	1,041393
hajar	0	0	0	0	1	0	0	0	0	0	0	1	1,041393
pagi	0	0	0	0	1	0	0	0	0	0	0	1	1,041393
sukses	0	0	0	0	0	1	0	0	0	0	0	1	1,041393

- Hak Cipta Dilindungi Undang-Undang
1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
    - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
    - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
  2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

Term	TF											IDF	
	dx	d1	d2	d3	d4	d5	d6	d7	d8	d9	d10	df	log(n/df)
lulus	0	0	0	0	0	1	0	0	0	0	0	1	1,041393
masuk	0	0	0	0	0	1	0	0	0	0	0	1	1,041393
traktir	0	0	0	0	0	1	0	0	0	0	0	1	1,041393
maaf	0	0	0	0	0	0	2	2	0	0	0	4	0,439333
siap	0	0	0	0	0	0	1	0	0	0	0	1	1,041393
beres	0	0	0	0	0	0	1	0	0	0	0	1	1,041393
rumah	0	0	0	0	0	0	1	0	0	0	0	1	1,041393
bilang	0	0	0	0	0	0	1	0	0	0	0	1	1,041393
mudah	0	0	0	0	0	0	0	1	0	0	0	1	1,041393
pernah	0	0	0	0	0	0	0	1	0	0	0	1	1,041393
kembali	0	0	0	0	0	0	0	1	0	0	0	1	1,041393
sedia	0	0	0	0	0	0	0	1	0	0	0	1	1,041393
jenuh	0	0	0	0	0	0	0	0	1	0	0	1	1,041393
sifat	0	0	0	0	0	0	0	0	1	0	0	1	1,041393
katak	0	0	0	0	0	0	0	0	1	0	0	1	1,041393
perlu	0	0	0	0	0	0	0	0	1	0	0	1	1,041393
risau	0	0	0	0	0	0	0	0	1	0	0	1	1,041393
tegar	0	0	0	0	0	0	0	0	1	0	0	1	1,041393
bayang	0	0	0	0	0	0	0	0	1	0	0	1	1,041393
tegang	0	0	0	0	0	0	0	0	0	1	0	1	1,041393
hadap	0	0	0	0	0	0	0	0	0	1	0	1	1,041393
teguh	0	0	0	0	0	0	0	0	0	0	1	1	1,041393
cerai	0	0	0	0	0	0	0	0	0	0	1	1	1,041393
runtuh	0	0	0	0	0	0	0	0	0	0	1	1	1,041393
bakat	0	0	0	0	0	0	0	0	0	0	1	1	1,041393
tekad	0	0	0	0	0	0	0	0	0	0	1	1	1,041393

Berdasarkan tabel diatas, asumsikan bahwa kolom dx merupakan data uji/dokumen uji sedangkan kolom d1-d10 merupakan data latih, angka pada tabel dx-d10 merupakan kemunculan kata pada tiap dokumen, kolom df merupakan

jumlah munculnya kata tersebut, sedangkan kolom IDF merupakan hasil penghitungan dari log jumlah dokumen dibagi dengan jumlah munculnya kata.

Setelah di hitung frekuensi kemunculannya, selanjutnya setiap kata akan diberi pembobotan dengan metode *term weighting TF-IDF* seperti pada tabel 4.8 berikut.

**Tabel 4.8 Tabel Perhitungan Term Weighting**

WDT=TF.IDF										
dx	d1	d2	d3	d4	d5	d6	d7	d8	d9	d10
1,12854	0	0	0	0	0,56427	0	0	0	0	0
1,48073	0	0	0	0	0	0	0	0	0	0
1,04139	0	0	0	0	0	0	0	0	0	0
1,04139	0	0	0	0	0	0	0	0	0	0
1,04139	0	0	0	0	0	0	0	0	0	0
1,04139	0	0	0	0	0	0	0	0	0	0
1,04139	0	0	0	0	0	0	0	0	0	0
0	1,04139	0	0	0	0	0	0	0	0	0
0	1,04139	0	0	0	0	0	0	0	0	0
0	1,04139	0	0	0	0	0	0	0	0	0
0	0	1,04139	0	0	0	0	0	0	0	0
0	0	1,48073	0	0	0	0	0	0	0	0
0	0	1,48073	0	0	0	0	0	0	0	0
0	0	1,48073	0	0	0	0	0	0	0	0
0	0	0	1,04139	0	0	0	0	0	0	0
0	0	0	1,04139	0	0	0	0	0	0	0
0	0	0	1,04139	0	0	0	0	0	0	0
0	0	0	0	1,04139	0	0	0	0	0	0
0	0	0	0	1,04139	0	0	0	0	0	0
0	0	0	0	1,04139	0	0	0	0	0	0
0	0	0	0	0	1,04139	0	0	0	0	0
0	0	0	0	0	1,04139	0	0	0	0	0
0	0	0	0	0	1,04139	0	0	0	0	0
0	0	0	0	0	1,04139	0	0	0	0	0

WDT=TF.IDF												
dx	d1	d2	d3	d4	d5	d6	d7	d8	d9	d10		
0	0	0	0	0	0	0,8787	0,8786654	0	0	0		
0	0	0	0	0	0	1,0414	0	0	0	0		
0	0	0	0	0	0	1,0414	0	0	0	0		
0	0	0	0	0	0	1,0414	0	0	0	0		
0	0	0	0	0	0	1,0414	0	0	0	0		
0	0	0	0	0	0	0	1,0413927	0	0	0		
0	0	0	0	0	0	0	1,0413927	0	0	0		
0	0	0	0	0	0	0	1,0413927	0	0	0		
0	0	0	0	0	0	0	1,0413927	0	0	0		
0	0	0	0	0	0	0	0	1,04139	0	0		
0	0	0	0	0	0	0	0	1,04139	0	0		
0	0	0	0	0	0	0	0	1,04139	0	0		
0	0	0	0	0	0	0	0	1,04139	0	0		
0	0	0	0	0	0	0	0	1,04139	0	0		
0	0	0	0	0	0	0	0	1,04139	0	0		
0	0	0	0	0	0	0	0	1,04139	0	0		
0	0	0	0	0	0	0	0	0	1,0414	0		
0	0	0	0	0	0	0	0	0	1,0414	0		
0	0	0	0	0	0	0	0	0	0	1,0414		
0	0	0	0	0	0	0	0	0	0	0	1,0414	
0	0	0	0	0	0	0	0	0	0	0	0	1,0414
0	0	0	0	0	0	0	0	0	0	0	0	1,0414
0	0	0	0	0	0	0	0	0	0	0	0	1,0414
0	0	0	0	0	0	0	0	0	0	0	0	1,0414

Nilai-nilai pada tabel diatas didapatkan dari hasil perkalian kemunculan kata pada tiap dokumen (TF) dan kolom IDF pada Tabel 4.7, hasil dari tabel ini nanti akan digunakan untuk mencari panjang vektor.

Berikutnya, hitung perkalian skalar antara data uji dengan semua data latih yaitu dx pada tabel 4.8 dikalikan dengan semua dokumen (d1-d10) dan didapatkan hasil seperti pada tabel 4.9 berikut.







#### 4.3.4 Proses Perhitungan Panjang Vektor Setiap Dokumen

Pada tahapan ini dilakukan perhitungan panjang vektor dari setiap dokumen yaitu dengan mengkuadratkan nilai bobot setiap kata pada dokumen di tabel 4.7 lalu menjumlahkan keseluruhan nilai bobot dalam dokumen tersebut, jumlah inilah yang kemudian di akarkan dan akan digunakan untuk menghitung kemiripan antar dokumen dengan rumus *cosine similarity*. Untuk lebih jelasnya dapat dilihat pada tabel 4.10 berikut :

##### Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
  - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
  - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.





**Tabel 4.8 Perhitungan Panjang Vektor**

panjang vektor										
dx	d1	d2	d3	d4	d5	d6	d7	d8	d9	d10
1,273609	0	0	0	0	0,318402	0	0	0	0	0
2,192548	0	0	0	0	0	0	0	0	0	0
1,084499	0	0	0	0	0	0	0	0	0	0
1,084499	0	0	0	0	0	0	0	0	0	0
1,084499	0	0	0	0	0	0	0	0	0	0
1,084499	0	0	0	0	0	0	0	0	0	0
0	1,084499	0	0	0	0	0	0	0	0	0
0	1,084499	0	0	0	0	0	0	0	0	0
0	1,084499	0	0	0	0	0	0	0	0	0
0	0	1,084499	0	0	0	0	0	0	0	0
0	0	2,192548	0	0	0	0	0	0	0	0
0	0	2,192548	0	0	0	0	0	0	0	0
0	0	2,192548	0	0	0	0	0	0	0	0
0	0	0	1,084499	0	0	0	0	0	0	0

Hak Cipta Dilindungi Undang-Undang  
 1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:  
 a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, pen-  
 b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.  
 2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izn





Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
  - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, pen-
  - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izi-

Panjang vektor											
dx	d1	d2	d3	d4	d5	d6	d7	d8	d9	d10	
0	0	0	1,084499	0	0	0	0	0	0	0	0
0	0	0	1,084499	0	0	0	0	0	0	0	0
0	0	0	0	1,084499	0	0	0	0	0	0	0
0	0	0	0	1,084499	0	0	0	0	0	0	0
0	0	0	0	1,084499	0	0	0	0	0	0	0
0	0	0	0	0	1,084499	0	0	0	0	0	0
0	0	0	0	0	0	1,084499	0	0	0	0	0
0	0	0	0	0	0	1,084499	0	0	0	0	0
0	0	0	0	0	0	1,084499	0	0	0	0	0
0	0	0	0	0	0	0	0,772053	0,772053	0	0	0
0	0	0	0	0	0	0	1,084499	0	0	0	0
0	0	0	0	0	0	0	1,084499	0	0	0	0
0	0	0	0	0	0	0	1,084499	0	0	0	0
0	0	0	0	0	0	0	1,084499	0	0	0	0
0	0	0	0	0	0	0	0	1,084499	0	0	0
0	0	0	0	0	0	0	0	1,084499	0	0	0
0	0	0	0	0	0	0	0	1,084499	0	0	0
0	0	0	0	0	0	0	0	1,084499	0	0	0
0	0	0	0	0	0	0	0	1,084499	0	0	0



- Hak Cipta Dilindungi Undang-Undang
1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
    - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, pen-  
gutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
    - b. Dilarang mengumpukan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izi-  
n UIN Suska Riau.

panjang vektor										
dx	d1	d2	d3	d4	d5	d6	d7	d8	d9	d10
0	0	0	0	0	0	0	0	1,084499	0	0
0	0	0	0	0	0	0	0	1,084499	0	0
0	0	0	0	0	0	0	0	1,084499	0	0
0	0	0	0	0	0	0	0	1,084499	0	0
0	0	0	0	0	0	0	0	1,084499	0	0
0	0	0	0	0	0	0	0	1,084499	0	0
0	0	0	0	0	0	0	0	0	1,084499	0
0	0	0	0	0	0	0	0	0	1,084499	0
0	0	0	0	0	0	0	0	0	0	1,084499
0	0	0	0	0	0	0	0	0	0	1,084499
0	0	0	0	0	0	0	0	0	0	1,084499
0	0	0	0	0	0	0	0	0	0	1,084499
0	0	0	0	0	0	0	0	0	0	1,084499
8,88865	3,253496	7,662142	3,253496	3,253496	4,656397	5,110048	5,110048	7,591491	2,168997	5,422494
2,981384	1,803745	2,768057	1,803745	1,803745	2,157869	2,260541	2,260541	2,755266	1,472752	2,328625

- Hak Cipta Dilindungi Undang-Undang
1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
    - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
    - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
  2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

Baris dengan warna abu-abu merupakan penjumlahan dari nilai bobot kata dari setiap dokumen, sedangkan baris berwarna hijau merupakan hasil pengakaran dari baris di atasnya, nilai ini akan digunakan untuk mencari proses penghitungan kemiripan data uji dengan data latih..

### 4.3.5 Proses Penghitungan Kemiripan Data Uji Dengan Semua Data Latih

Selanjutnya Selanjutnya cari kemiripan antara data uji dengan semua data latih dengan rumus *cosine similarity* yaitu dengan rumus persamaan (2.4) Pertama kita cari kemiripan data uji (dx) dengan dokumen 1/data latih 1(d1)

$$(dx,d1) = \frac{0}{2,981384 \times 1,803745} = 0$$

Kemiripan dx dengan d2

$$(dx,d1) = \frac{0}{2,981384 \times 2,768057} = 0$$

Kemiripan dx dengan d3

$$(dx,d3) = \frac{0}{2,981384 \times 1,803745} = 0$$

Kemiripan dx dengan d4

$$(dx,d4) = \frac{0}{2,981384 \times 1,803745} = 0$$

Kemiripan dx dengan d5

$$(dx,d5) = \frac{0}{2,981384 \times 2,157869} = 0,098984$$

Kemiripan dx dengan d6

$$(dx,d6) = \frac{0}{2,981384 \times 2,260541} = 0$$

Kemiripan dx dengan d7

$$(dx,d7) = \frac{0}{2,981384 \times 2,260541} = 0$$

Kemiripan dx dengan d8

$$(dx,d8) = \frac{0}{2,981384 \times 2,755266} = 0$$

Kemiripan dx dengan d9

$$(dx,d9) = \frac{0}{2,981384 \times 1,472752} = 0$$

Kemiripan dx dengan d10

$$(dx,d10) = \frac{0}{2,981384 \times 2,328625} = 0$$

Untuk hasil penghitungannya dapat dilihat pada tabel 4.11

**Tabel 4.11 Hasil Penghitungan Kemiripan Data Uji Dengan Data Latih**

	d1	d2	d3	d4	d5	d6	d7	d8	d9	d10
<i>cosine similarity</i>	0	0	0	0	<b>0,098984</b>	0	0	0	0	0

Dari tabel 4.11 diatas dapat dilihat bahwa hanya d5 yang memiliki nilai kemiripan dengan data uji, jadi d5 merupakan tetangga terdekat dari data uji. Jadi hasil klasifikasi akan jatuh kepada kelas yang sama dengan d5 yaitu kelas E (*Extraversion*), disini hasil klasifikasi dengan menggunakan algoritma *K-Nearest Neighbor* secara manual terbukti benar karena kelas dari data uji adalah kelas E (*Extraversion*).

**Hak Cipta Dilindungi Undang-Undang**

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
  - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
  - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.