

BAB III

METODOLOGI PENELITIAN

3.1 Metode Penelitian

Metodologi sendiri merujuk kepada alur pemikiran umum atau menyeluruh (general logic) dan gagasan teoritis (theoretic perspectives) suatu penelitian. Sedangkan kata 'metode' menunjuk pada teknik yang digunakan dalam penelitian seperti survey, wawancara dan observasi. Suatu penelitian yang baik senantiasa memperhatikan kesesuaian antara teknik yang digunakan dengan alur pemikiran umum serta gagasan teoritis. Tahapan-tahapannya dapat dilihat pada gambar berikut.



Gambar 3.1 Alur Penelitian

Tahapan pada gambar diatas akan di uraikan sebagai berikut:

3.1.1 Pengumpulan Data

Pada tahapan pengumpulan data adalah tahapan awal pada penelitian ini yang mana dataset yang akan digunakan di ambil dari komentar pengguna aplikasi ID Express yang di dapatkan melalui *website playstore* dimana komentar yang di ambil hanya berupa komentar yang menggunakan Bahasa Indonesia saja. Dibawah ini di tunjukan code untuk pengambilan dataset dengan menggunakan *Google Colab* dengan Bahasa pemograman *Python*.

```
[23] app_packages = [
      'com.idexpress.customerapp',
    ]

    app_infos = []

    for application in tqdm(app_packages):
        info = app(application, lang='id', country='id')
        del info['comments']
        app_infos.append(info)
```

100% | 1/1 [00:00<00:00, 4.04

Gambar 3.2 Pengambilan dataset menggunakan Python

Gambar diatas menunjukkan code untuk pengambilan dataset yang di ambil dari website playstore ID Express. Data yang digunakan pada penelitian ini, sebanyak 676 Komentar data dalam kurung waktu tanggal 02-10-2020 sampai 12-05-2023 melalui *scraping google Playstore* menggunakan *google colab*.

3.1.2 Preprocessing

Pada tahap ini dilakukan klasifikasi teks atau sentimen dengan tahapan preprocessing agar teks yang memiliki isi yang tidak sempurna seperti data yang hilang, data yang tidak valid atau juga hanya sekedar salah ketik. Selain itu, ada juga atribut-atribut data yang tidak relevan. Data tersebut lebih baik dibuang karena keberadaannya bisa mengurangi mutu atau akurasi. Teks yang belum diolah biasanya memiliki karakteristik dimensi yang tinggi, terdapat noise pada data dan terdapat struktur teks yang tidak baik. Untuk itu, dalam pengolahan data awal, text mining harus melalui beberapa tahapan yang disebut dengan pre-processing. Tahapan pre-processing yang dapat dilakukan yaitu.

Pada tahap ini, data yang digunakan sebanyak 676 komentar berupa data text. Dan Nantinya akan melalui tahap pre-processing yang akan dibagi menjadi data testing dan data training.

a) *Cleansing Data*

Tahap Cleansing data dilakukan sebagai tahap awal yang sangat penting dalam penelitian ini. Hal ini dikarenakan data yang didapatkan dari komentar *Playstore* masih dalam bentuk teks yang tidak sesuai kaidah dan tidak lengkap. Normalisasi adalah proses penskalaan nilai atribut dari data sehingga terletak pada rentang tertentu.

Selain itu pada tahap normalisasi juga dilakukan secara Cleansing data menggunakan software *Python*, hal ini bertujuan untuk memastikan isi dari data tersebut. Cleaning data yang akan menghapus atribut-atribut yang tidak diperlukan di dalam data seperti, tanda baca, emot ikon, angka dan lain-lain. Atribut-atribut yang akan dihapus contohnya “@[A-Za-z0-9]+, [0-9]+, #, [^\w], ‘_’, [\n]+, :, ‘RT[\s]+, ^https?:\V\.[\r\n], ^http?:\V\.[\r\n]”. Gambar 3.4 Cleansing Data dengan *Python*.

```
[ ] def clean_tweet(tweet):
    if type(tweet) == np.float:
        return ""
    temp = tweet.lower()
    temp = re.sub("","", temp)
    temp = re.sub("@[A-Za-z0-9_]+","", temp)
    temp = re.sub("#[A-Za-z0-9_]+","", temp)
    temp = re.sub(r'http\S+', '', temp)
    temp = re.sub('(!)?', ' ', temp)
    temp = re.sub('\.[*?\\]', ' ', temp)
    temp = re.sub("[^a-z0-9]", " ", temp)
    temp = temp.split()
    temp = [w for w in temp if not w in stopwords]
    temp = " ".join(word for word in temp)
    return temp

tw['cn_english'] = tw['in_english'].apply(clean_tweet)
tw.head(20)
```

Gambar 3.3 Proses *cleansing* data menggunakan *python*

Pada gambar diatas menunjukkan code proses *Cleansing* data Menggunakan *python* dapat di lihat pada gambar 3.3 dan hasil dari proses *cleansing* data dapat dilihat pada gambar 3.4 dibawah ini.

	content	in_english	cn_english
0	Bad service	Bad service	bad service
2	Biasa aja	Just the same	
4	Kecewa Ada masalah Saat Proses pengiriman kurir ke customer... padahal udh sampe di andong dan enggak tertau jauh dari rumah saya tapi pengiriman kurir lambat banget	Disappointed there was a problem during the courier delivery process to the customer... even though I arrived at Andong and not too far from my house but the shipping Kurir was very slow	disappointed problem courier delivery process customer even though arrived andong far house shipping kurir slow
6	Aku mesan barang di shopee lewat pengiriman id express. Udah 4 hari gak dikrim kerumah. Padahal statusnya udah sampe. Dia kota aku!	I ordered goods at Shopee via Express ID Delivery It's been 4 days not sent home.Even though the status has arrived.He's my city!	ordered goods shopee via express id delivery 4 days sent home even though status arrived hes city
8	aplikasi nya bagus pelayanannya terbaik dan memuaskan, jasa pengirimannya juga super cepet dan kurir nya juga rapi" dan sopan" semangat ID Express The Best lah pake nya patut di coba pake jasa ekspedit ini	The application is good for the best and satisfying service, the shipping service is also super fast and the courier is also neat "and polite". The spirit of ID Express The Best is that the pk is worth trying to use this expedition service.	application good best satisfying service shipping service also super fast courier also neat polite spirit id express best pk worth trying use expedition service
9	Aplikasinya sangat memudahkan dan pengirimannya cepet. Selain ongkirnya murah, wilayah" pelosok lg bisa dijangkau dgn cepet	The application is very easy and the delivery is fast Besides the cheap postage, the area "remote areas can also be reached quickly	application easy delivery fast besides cheap postage area remote areas also reached quickly
10	Ongkirnya murah kurirnya pun ramah 🍷🍷🍷	The postage is cheap, the courier is also ramh.	postage cheap courier also ramh
12	Semoga sukses trus id mempermudah	Good luck then the ID makes it easier	good luck id makes easier
14	Pengiriman cepet, Harga lebih murah dibandingkan yg lain. Yang lain boleh sama murah nya tapi pengiriman lama sekali. Kurirnya juga ramah", Thanks ID EXPRESS	Fast shipping, prices are cheaper than others. Others may be as cheap but a long delivery.The courier is also friendly", thanks ID EXPRESS	fast shipping prices cheaper others others may cheap long delivery courier also friendly thanks id express
15	Baik	Good	good
17	Harga nya paling murah se Indonesia raya dan pelayan kurir pickup nya ramah2 sekali fast respon pokok nya. Pengiriman nya cepet bgt. Recommended buat kalian2 yg pengen paket nya cepet sampai rumah...buat id express semoga tambah jaya.	The price is the cheapest in Indonesia, and the pickup courier service is very friendly, the main response... the delivery is really fast... Recommended for those of you who want the package to get fast to home... for ID Express, hopefully it will be more victorious.	price cheapest Indonesia pickup courier service friendly main response delivery really fast recommended want package get fast home id express hopefully victorious
18	Aplikasi ini sangat membantu sekali untuk mencari paket, pelayanannya kurirnya juga ramah" murah sejujurnya, paketnya datangnya cepet harga nya juga murah sekali, terima kasih ID EXPRESS	This application is very helpful to find packages, the courier service is also friendly "*, smiling, the package comes quickly the price is also very cheap, thank you ID Express	application helpful find packages courier service also friendly smiling package comes quickly price also cheap thank id express

Gambar 3.4 Hasil proses *cleansing* data

b) Labeling Data

Labeling data adalah tahap memberikan label pada setiap data yang sudah dibersihkan atau telah melalui Cleansing data. Dalam penelitian ini data akan dilabeli melalui Natural Language Processing yang ada di Library Python yaitu Sastrawi. Proses Labeling melalui Sastrawi dilakukan dalam Bahasa Indonesia dimana data yang akan diberi label harus melalui tahap pemenggalan kata (stemming). Stemming sendiri merupakan proses menghilangkan imbuhan atau awalan kata sehingga hanya tersisa kata dasar. Penggunaan stemming ini dapat membantu dalam mengenali pola sentimen dalam kalimat tanpa harus

memperhitungkan perbedaan kata berimbuhan. Proses Labeling juga dilakukan dengan berkonsultasi kepada ahli Bahasa yaitu Bapak Dr. Muhammad Sukirlan, M.A. selaku kepala UPT Bahasa di Unila, dan didapatkan hasil bahwa:

Tabel 3.1 Penjelasan Masing-Masing Label

No.	Sentimen	Penjelasan
1.	Negatif	Pernyataan yang sangat tidak mendukung terhadap konten yang dimaksudkan.
2.	Netral	Pernyataan yang tidak termasuk dalam mendukung dan setuju terhadap konten serta tidak termasuk kedalam pernyataan yang menentang atau tidak mendukung konten.
3.	Positif	Pernyataan yang sangat mendukung atau mengajak terhadap konten yang dimaksudkan.

Menurut Bapak Dr. Muhammad Sukirlan, M.A. labeling dilakukan dengan melihat kata kerja, kata bantu, kata sifat dan juga kata sambung di kalimatnya, contoh beberapa kata yang bisa diindikasikan sebagai kata yang negatif atau positif:

Tabel 3.2 Contoh Kata Positif dan Kata Negatif

Positif	Negatif
Senang	Benci
Ayo	Tua
Dukung	Tidak
Ajak	Tertindas
Percaya	Khawatir
Perlu	Bukan
Menaati	Melanggar

Beliau juga mengatakan bahwa jika ada kata Positif namun diawali dengan kata “Tidak” maka kata tersebut akan berlabel Negatif, begitu juga sebaliknya jika ada kata Negatif yang diawali dengan kata “Tidak” maka kata tersebut akan berlabel Positif, sedangkan untuk kata Netral beliau mengatakan bahwa kalimatnya akan mengandung

keduanya atau tidak mengandung keduanya dan juga bisa berupa pertanyaan, berikut contoh kalimatnya:

Tabel 3.3 Cara Melabeli Kalimat

Kalimat	Sentimen
Senang sekali membaca tulisan dari tentang RKHUP dan perjalanannya dari rumah ke rumah seperti lagunya	Positif
Harkristuti Harkrisnowo berpendapat bahwa KUHP yang diterapkan sekarang sudah terlalu tua dan bukan merupakan buatan Indonesia	Negatif
Mana suaranya yang sepakat mempertemukan dan untuk debat RKUHP	Netral

Dalam penelitian ini data akan dilabeli secara manual dengan mengambil data 20% dari data yang akan di labeli.

c) Pembobotan Kata

Pada tahap pembobotan kata atau Terms Weight ini akan dilakukan perhitungan secara otomatis menggunakan python. Namun pada kali ini akan memberikan contoh dan Contoh akan dibagi menjadi 4 dimana 3 adalah data training dan 1 adalah data testing. Berikut adalah perhitungan manual dari pembobotan kata, yaitu:

Tabel 3.4 Contoh Data Pembobotan Kata

Ket	Content	Sentimen
D1	kecewa proses pengiriman kurir customer udh sampe andong rumah pengiriman kurim lambat banget	Negatif
D2	mesan barang shopee pengiriman id express udah 4 gak dikirim kerumah statusnya udah sampe kota	Positif
D3	mengecewakan	Negatif

D4	aplikasi nya bagus pelayanannya terbaik memuaskan jasa pengiriman nya super cepet kurir nya rapi sopan semangat id express the best pkonya patut coba pakai jasa expedisi	?
----	---	---

Selanjutnya melakukan tokenisasi dan lakukan *stopwords*, tokenisasi dilakukan untuk memisahkan kalimat menjadi kata-kata dan *stopwords* digunakan untuk membuang kata-kata yang dirasa memiliki makna yang kurang berarti seperti: “yang”, “dan”, “atau” dan lain-lain.

3.1.3 Analisis Data

Tahap analisis merupakan suatu prosedur atau proses sistematis pengombinasian metode atau teknik untuk menentukan data yang sesuai dan cara terbaik untuk memanfaatkannya. Tahap analisis data dilakukan dari awal sampai akhir adalah membandingkan hasil eksperimen. Data yang telah dikumpulkan akan dieksperimen dengan melakukan cleaning dan normalisasi. Selanjutnya tahapan yang dilakukan dengan transformasi data pengklusteran kemudian dibagi beberapa kelompok data atau group. Setelah data sudah dikelompokkan maka selanjutnya akan dilakukan pemodelan data yang diubah menjadi nilai-nilai yang telah dipisahkan dengan bentuk pelebelan data, atau Comma Sparated Value (CSV) sebagai format data input ke dalam *Machine Learning*.

Setelah proses memasukkan data atau pemodelan data ke dalam suatu database selesai, selanjutnya proses yang akan dilakukan yaitu menerapkan model algoritma metode *Naïve Bayes* dan *K-Nearest Neighbor* yang selanjutnya akan di optimasi menggunakan *feature Particle Swarm Optimization (PSO)*.

3.1.4 Eksperimen dan pengujian Metode

Tahap ini akan menjelaskan eksperimen dan teknik pengujian yang akan digunakan untuk menguji dan mengevaluasi suatu metode atau teknik dengan menggunakan data yang telah dikumpulkan dari observasi atau percobaan. Eksperimen dan pengujian metode merupakan bagian penting dari proses penelitian, karena dengan melakukannya maka akan tercipta suatu metode yang dapat digunakan secara efektif dan efisien untuk menyelesaikan suatu masalah.

Eksperimen dan pengujian metode juga bertujuan untuk mengetahui kelebihan dan kekurangan suatu metode atau teknik, serta untuk menemukan cara-cara baru yang lebih baik dalam menyelesaikan masalah. Oleh karena itu, eksperimen dan pengujian metode sangat penting dilakukan agar dapat memperoleh hasil yang bermutu dan dapat

dipertanggung jawabkan. Berikut penjelasan dan perhitungan masing-masing metode yang akan digunakan.

3.1.4.1 Naïve Bayes

Berikut ini adalah perhitungan manual penyelesaian menggunakan metode *Naïve Bayes* dalam klasifikasi ulasan.

Tabel 3.5 Komentar Data *Training*

No	Komentator	Komentar	Kelas
1	Nk Rara	Cepat tanggap,aman dan Ramah pelayanannya 😊	Positif
2	Juwita Dwi Kartika	OH INI YANG LELET JASA PENGIRIMAN NYA!	Negatif
3	R G N Family	Cek tracking di website ko ga bisa* ya ... dimohon penjelasannya	Netral

Preprocessing

1. *Cleaning* : Membersihkan karakter atau symbol (@#\$%^&*[]?)

Tabel 3.6 *Clening*

No	Komentar	Hasil <i>Cleaning</i>
1	Cepat tanggap,aman dan Ramah pelayanannya 😊	Cepat tanggap,aman dan Ramah pelayanannya
2	OH INI YANG LELET JASA PENGIRIMAN NYA!	OH INI YANG LELET JASA PENGIRIMAN NYA
3	Cek tracking di website ko ga bisa* ya ... dimohon penjelasannya	Cek tracking di website ko ga bisa ya dimohon penjelasannya

2. *Case Folding* : mengubah semua huruf menjadi huruf kecil

Tabel 3.7 *Case Folding*

No	Komentar	Hasil <i>Case Folding</i>
1	Cepat tanggap,aman dan Ramah pelayanannya	cepat tanggap,aman dan ramah pelayanannya
2	OH INI YANG LELET JASA PENGIRIMAN NYA	oh ini yang lelet jasa pengiriman nya
3	Cek tracking di website ko ga bisa ya dimohon penjelasannya	cek tracking di website ko ga bisa ya dimohon penjelasannya

3. *Filtering/Stopword* : Merupakan tahapan untuk mengilangkan kata-kata yang kurang penting yang di khawatirkan akan mengganggu proses klasifikasi

Tabel 3.8 *Filtering/Stopword*

No	Komentar	Hasil Case Folding
1	cepat tanggap,aman dan ramah pelayanannya	tanggap,aman dan ramah pelayanannya
2	oh ini yang lelet jasa pengiriman nya	ini yang lelet jasa pengiriman nya
3	cek tracking di website ko ga bisa ya dimohon penjelasannya	cek tracking di website ga bisa dimohon penjelasannya

4. *Stemming* : Mengembalikan kata ke bentuk dasar, akhiran secara langsung ditambahkan pada kata dasar, serta awalan yang dapat langsung diberikan pada kata dasar murni

Tabel 3.9 *Filtering/Stopword*

No	Komentar	Hasil Case Folding
1	tanggap,aman dan ramah pelayanannya	Tanggap aman ramah pelayanannya
2	ini yang lelet jasa pengiriman nya	ini yang lelet jasa pengiriman nya
3	cek tracking di website ga bisa dimohon penjelasannya	cek tracking di website ga bisa mohon penjelasannya

5. *Tokenizing* : Pemecahan Teks/Kalimat menjadi Kumpulan kata agar mudah dalam melakukan pembobotan tiap kata.

Tabel 3.10 *Tokenizing*

Data Komentar 1	Data Komentar 2	Data Komentar 3
Tanggap	Ini	Cek
Aman	Yang	Tracking
Ramah	Lelet	Diwebsite
Pelayanannya	Jasa	Ga
	Pengirimannya	Bisa
		Mohon
		Penjelasannya

6. Pembobotan : menggunakan tf (*term frekuensi*) / jumlah kemunculan kata

Tabel 3.11 Pembobotan

No	Kosa Kata	tf(Positif)	tf(Neutral)	tf(Negatif)
1	Tanggap	1	0	0
2	Aman	1	0	0
3	Ramah	1	0	0
4	Pelayanannya	1	0	0
5	Cek	0	1	0
6	Tracking	0	1	0
7	Diwebsite	0	1	0
8	Ga	0	1	0
9	Bisa	0	1	0
10	Mohon	0	1	0
11	Penjelasannya	0	1	0
12	Ini	0	0	1
13	Yang	0	0	1
14	Lelet	0	0	1
15	Jasa	0	0	1
16	Pengirimannya	0	0	1
Jumlah term		4	7	5

Diperoleh :

Count Potif = 4, *Count* netral = 7, *Count* negative = 5

A) Proses Klasifikasi Data Latih

1. Hitung probabilitas *prior* setiap kategori, yang menjadi kategori ada 3 kategori yaitu kelas positif, netral, negatif

$$P\left(\frac{\text{positif}}{\text{negatif}}\right) = \frac{x(\text{positif/netral/negatif})}{|c|} \dots \dots \dots (6)$$

$$P(\text{positif}) = \frac{f x(\text{positif})}{|c|} = \frac{1}{3} = 0,333333333$$

$$P(\text{netral}) = \frac{f x(\text{netral})}{|c|} = \frac{1}{3} = 0,333333333$$

$$P(\text{negatif}) = \frac{f x(\text{negatif})}{|c|} = \frac{1}{3} = 0,333333333$$

2. Menghitung probabilitas *likelihood* setiap *term* dari semua dokumen. Jumlah seluruh kata 16, 4 *term* dari kelas positif, 7 *term* dari kelas netral, 5 *term* dari kelas negative. Banyaknya *term* tergantung pada hasil praposes data.

$$P(w|\text{positif/netral/negatif}) = \frac{nk(\text{positif/netral/negatif}) + 1}{n.\text{positif/netral/negatif} + |r|} \dots \dots \dots (7)$$

- 1) probabilitas kata “tanggap”

$$P(\text{tanggap}|\text{positif}) = \frac{1 + 1}{4 + 16} = 0,1$$

$$P(\text{tanggap}|\text{netral}) = \frac{0 + 1}{7 + 16} = 0,0303030$$

$$P(\text{tanggap}|\text{negatif}) = \frac{0 + 1}{5 + 16} = 0,047619$$

- 2) probabilitas kata “aman”

$$P(\text{aman}|\text{positif}) = \frac{1 + 1}{4 + 16} = 0,1$$

$$P(\text{aman}|\text{netral}) = \frac{0 + 1}{7 + 16} = 0,0303030$$

$$P(\text{aman}|\text{negatif}) = \frac{0 + 1}{5 + 16} = 0,047619$$

- 3) probabilitas kata “ramah”

$$P(\text{ramah}|\text{positif}) = \frac{1 + 1}{4 + 16} = 0,1$$

$$P(\text{ramah}|\text{netral}) = \frac{0 + 1}{7 + 16} = 0,0303030$$

$$P(\text{ramah}|\text{negatif}) = \frac{0 + 1}{5 + 16} = 0,047619$$

- 4) probabilitas kata “pelayanannya”

$$P(\text{pelayanannya}|\text{positif}) = \frac{1 + 1}{4 + 16} = 0,1$$

$$P(\text{pelayanannya}|\text{netral}) = \frac{0 + 1}{7 + 16} = 0,0303030$$

$$P(\text{pelayanannya}|\text{negatif}) = \frac{0+1}{5+16} = 0,047619$$

- 5) probabilitas kata “cek”

$$P(\text{cek}|\text{positif}) = \frac{0 + 1}{4 + 16} = 0,05$$

$$P(\text{cek}|\text{netral}) = \frac{1 + 1}{7 + 16} = 0,08695652$$

$$P(\text{cek}|\text{negatif}) = \frac{0 + 1}{5 + 16} = 0,047619$$

6) probabilitas kata “tracking”

$$P(\text{tracking}|\text{positif}) = \frac{0 + 1}{4 + 16} = 0,05$$

$$P(\text{tracking}|\text{netral}) = \frac{1 + 1}{7 + 16} = 0,08695652$$

$$P(\text{tracking}|\text{negatif}) = \frac{0 + 1}{5 + 16} = 0,047619$$

7) probabilitas kata “diwebsite”

$$P(\text{diwebsite}|\text{positif}) = \frac{0 + 1}{4 + 16} = 0,05$$

$$P(\text{diwebsite}|\text{netral}) = \frac{1 + 1}{7 + 16} = 0,08695652$$

$$P(\text{diwebsite}|\text{negatif}) = \frac{0 + 1}{5 + 16} = 0,047619$$

8) probabilitas kata “ga”

$$P(\text{ga}|\text{positif}) = \frac{0 + 1}{4 + 16} = 0,05$$

$$P(\text{ga}|\text{netral}) = \frac{1 + 1}{7 + 16} = 0,08695652$$

$$P(\text{ga}|\text{negatif}) = \frac{0 + 1}{5 + 16} = 0,047619$$

9) probabilitas kata “bisa”

$$P(\text{bisa}|\text{positif}) = \frac{0 + 1}{4 + 16} = 0,05$$

$$P(\text{bisa}|\text{netral}) = \frac{1 + 1}{7 + 16} = 0,08695652$$

$$P(\text{bisa}|\text{negatif}) = \frac{0 + 1}{5 + 16} = 0,047619$$

10) probabilitas kata “mohon”

$$P(\text{mohon}|\text{positif}) = \frac{0 + 1}{4 + 16} = 0,05$$

$$P(\text{mohon}|\text{netral}) = \frac{1 + 1}{7 + 16} = 0,08695652$$

$$P(\text{mohon}|\text{negatif}) = \frac{0 + 1}{5 + 16} = 0,047619$$

11) probabilitas kata “penjelasannya”

$$P(\text{penjelasannya}|\text{positif}) = \frac{0 + 1}{4 + 16} = 0,05$$

$$P(\text{penjelasannya}|\text{netral}) = \frac{1 + 1}{7 + 16} = 0,08695652$$

$$P(\text{penjelasannya}|\text{negatif}) = \frac{0 + 1}{5 + 16} = 0,047619$$

12) probabilitas kata “ini”

$$P(\text{ini}|\text{positif}) = \frac{0 + 1}{4 + 16} = 0,05$$

$$P(\text{ini}|\text{netral}) = \frac{0 + 1}{7 + 16} = 0,0303030$$

$$P(\text{ini}|\text{negatif}) = \frac{1 + 1}{5 + 16} = 0,095238095$$

13) probabilitas kata “yang”

$$P(\text{yang}|\text{positif}) = \frac{0 + 1}{4 + 16} = 0,05$$

$$P(\text{yang}|\text{netral}) = \frac{0 + 1}{7 + 16} = 0,0303030$$

$$P(\text{yang}|\text{negatif}) = \frac{1 + 1}{5 + 16} = 0,095238095$$

14) probabilitas kata “lelet”

$$P(\text{lelet}|\text{positif}) = \frac{0 + 1}{4 + 16} = 0,05$$

$$P(\text{lelet}|\text{netral}) = \frac{0 + 1}{7 + 16} = 0,0303030$$

$$P(\text{lelet}|\text{negatif}) = \frac{1 + 1}{5 + 16} = 0,095238095$$

15) probabilitas kata “jasa”

$$P(\text{jasa}|\text{positif}) = \frac{0 + 1}{4 + 16} = 0,05$$

$$P(\text{jasa}|\text{netral}) = \frac{0 + 1}{7 + 16} = 0,0303030$$

$$P(\text{jasa}|\text{negatif}) = \frac{1 + 1}{5 + 16} = 0,095238095$$

16) probabilitas kata “pengirimannya”

$$P(\text{pengirimannya}|\text{positif}) = \frac{0 + 1}{4 + 16} = 0,05$$

$$P(\text{pengirimannya}|\text{netral}) = \frac{0 + 1}{7 + 16} = 0,0303030$$

$$P(\text{pengirimannya}|\text{negatif}) = \frac{1 + 1}{5 + 16} = 0,095238095$$

Tabel 3.12 Data Uji

Komentator	Komentar	Kelas Asli
Dasit Widodo	Bagus dan cepat	Positif
agustina ina_	Expedisi terburuk di Indonesia	Negatif

Tabel 3.13 Data Uji setelah *preprocessing*

Komentator	Komentar	Kelas Asli
Dasit Widodo	bagus dan cepat	Positif
agustina ina	expedisi terburuk di indonesia	Negatif

Tabel 3.14 Data Uji dan Data Latih

No	Komentator	Komentar	Kelas Asli
1	Nk Rara	cepat tanggap,aman dan ramah pelayanannya	Positif
2	Juwita Dwi Kartika	oh ini yang lelet jasa pengirimannya	Negatif
3	R G N Family	cek tracking di website ko ga bisa ya dimohon penjelasannya	Netral
Data Uji			
4	Dasit Widodo	bagus dan cepat	?
5	agustina ina	expedisi terburuk di indonesia	?

B) Proses Klasifikasi Data Uji

Proses klasifikasi data uji dengan mengalikan semua nilai peluang. Nilai yang lebih tinggi merupakan kelas baru dari data tersebut.

$$P\left(\frac{\text{positif}}{\text{netral}}\right) | d = P\left(\frac{\text{positif}}{\text{netral}}\right) \times \pi p(w | \text{positif/netral/negatif}) \dots\dots\dots (8)$$

1. Pada data uji “bagus dan cepat” yang termasuk ke dalam data *training* adalah kata “bagus” dan “cepat”

$$P(\text{uji}|\text{positif})$$

$$= P(\text{positif}) \times P(\text{bagus}|\text{positif}) \times (\text{cepat}|\text{positif})$$

$$= 0,3333333 \times 0,22 \times 0,22$$

$$= 0,015972$$

$$P(\text{uji}|\text{netral})$$

$$= P(\text{netral}) \times P(\text{bagus}|\text{netral}) \times (\text{cepat}|\text{netral})$$

$$= 0,3333333 \times 0,1111 \times 0,1111$$

$$= 0,0041152$$

$$P(\text{uji}|\text{negatif})$$

$$= P(\text{negatif}) \times P(\text{bagus}|\text{negatif}) \times (\text{cepat}|\text{negatif})$$

$$= 0,3333333 \times 0,1111 \times 0,1111$$

$$= 0,0041152$$

Kesimpulan : Nilai probabilitas tertinggi yaitu sebesar 0,015972 pada P(uji | positif) sehingga komentar tersebut di klasifikasikan kedalam kelas "Positif"

2. Pada data uji “expedisi terburuk di indonesia” yang termasuk ke dalam data *training* adalah kata “expedisi” dan “terburuk”.

$$P(\text{uji}|\text{positif})$$

$$= P(\text{positif}) \times P(\text{expedisi}|\text{positif}) \times (\text{terburuk}|\text{positif})$$

$$= 0,3333333 \times 0,1111 \times 0,1111$$

$$= 0,0041152$$

$$P(\text{uji}|\text{netral})$$

$$= P(\text{netral}) \times P(\text{expedisi}|\text{netral}) \times (\text{terburuk}|\text{netral})$$

$$= 0,3333333 \times 0,1111 \times 0,1111$$

$$= 0,0041152$$

$$P(\text{uji}|\text{negatif})$$

$$= P(\text{negatif}) \times P(\text{expedisi}|\text{negatif}) \times (\text{netral}|\text{negatif})$$

$$= 0,3333333 \times 0,22 \times 0,22$$

$$= 0,015972$$

Kesimpulan : Nilai probabilitas tertinggi yaitu sebesar 0,015972 pada P(uji | negatif) sehingga komentar tersebut di klasifikasikan kedalam kelas "Negatif"

3.1.4.2 K-Nearest Neighbor

Berikut ini adalah perhitungan manual penyelesaian menggunakan metode *K-Nearest Neighbor* dalam klasifikasi ulasan.

A) Preprocessing Data

Tahapan yang dilakukan dalam preprocessing yaitu:

1) Case Folding

Tabel 3.15 Proses case folding

No	Komentar asli	Komentar setelah case Folding
1	Bagus dan cepat	bagus dan cepat
2	Ongkirnya murah kurirnya pun ramah	ongkirnya murah kurirnya pun ramah
3	Lelet,, pengiriman lama	lelet,, pengiriman lama
4	Buruk sekali tinggal selangkah doang tapi mandek gak karuannn	buruk sekali tinggal selangkah doang tapi mandek gak karuannn

2) Tokenizing

Pada tahap ini akan dilakukan penghapusan delimiter yaitu karakter angka dan karakter simbol kecuali karakter huruf dengan cara membuat daftar kode karakter yang diperlukan. Contoh karakter yang ada pada daftar kode Tokenizing adalah:

Tabel 3.16 Hasil tokenizing

No	Komentar	Kelas
1	bagus//dan//cepat	Positif
2	ongkirnya//murah//kurirnya//pun//ramah	Positif
3	lelet//pengiriman//lama	Negatif
4	buruk//sekali//tinggal//selangkah//doang//tapi mandek//gak//karuannn	Negatif

3) Stopword Removal (Filtering)

Langkah-langkah untuk proses Filtering adalah sebagai berikut:

- a. Mempersiapkan kata kunci yang diperlukan untuk penghapusan proses Filtering.
- b. Proses pencarian pada kata-kata yang ada didalam data komentar.
- c. Jika ditemukan kata yang sesuai dengan kata kunci maka akan dihapus, namun jika kata kunci stopwords tidak ada pada kata yang ada di dalam komentar maka kata tersebut tidak dihapus.

4) Stemming

Tahap ini melakukan Stemming dengan menggunakan algoritma porter, langkah pertama mempersiapkan kamus kata dasar untuk proses Stemming yang diambil dari Kamus Besar Bahasa Indonesia, setelah itu mengecek kata yang ada disistem dalam kamus kata dasar, jika ditemukan maka diasumsikan kata tersebut adalah kata dasar, maka algoritma berhenti, jika tidak ditemukan maka lakukan penghapusan sisa huruf yang tidak ada pada kamus kata dasar tersebut.

B) Pembobotan Kata TF-IDF

Pembobotan kata dimulai dari mencari TF dengan cara menghitung jumlah kata yang muncul pada dokumen, proses selanjutnya setelah didapatkan nilai TF adalah mencari nilai IDF, dimana IDF merupakan hasil invers dari

DF, cara untuk menentukan IDF adalah jumlah seluruh dokumen yaitu D1, D2, D3, D4, D5, Dx dibagi jumlah dokumen yang mengandung sebuah kata (DF) kemudian hasilnya di log-kan.

$$IDF(w) = \log\left(\frac{N}{DF(w)}\right) \dots \dots \dots (9)$$

Dari nilai idf tersebut akan dicari lagi nilai TF-IDF yaitu perkalian antara hasil dari frekuensi kemunculan kata pada setiap dokumen (TF) dengan bobot kata dalam seluruh dokumen (IDF).

Berikut merupakan contoh perhitungan TF-IDF secara manual. Perhitungan dilakukan terhadap empat dokumen yang sudah melalui tahap text preprocessing dan siap untuk diproses lebih lanjut.

Tabel 3.17 Dokumen yang akan digunakan perthitungan

Dokumen 1	bagus dan cepat
Dokumen 2	ongkirnya murah kurirnya pun ramah
Dokumen 3	lelet pengiriman lama
Dokumen 4	buruk sekali tinggal selangkah doang tapi mandek gak karuan

Tabel 3.18 Perhitungan Manual TF-IDF

Token	tf				df	D/df	IDF(LOG D/df)	W			
	D1	D2	D3	D4				D1	D2	D3	D4
bagus	1	0	0	0	1	4	0,60206	0,60206	0	0	0
cepat	1	0	0	0	1	4	0,60206	0,60206	0	0	0
ongkir	0	1	0	0	1	4	0,60206	0	0,60206	0	0
murah	0	1	0	0	1	4	0,60206	0	0,60206	0	0
kurir	0	1	0	0	1	4	0,60206	0	0,60206	0	0
ramah	0	1	0	0	1	4	0,60206	0	0,60206	0	0
lelet	0	0	1	0	1	4	0,60206	0	0	0,60206	0
pengiriman	0	0	1	0	1	4	0,60206	0	0	0,60206	0
lama	0	0	1	0	1	4	0,60206	0	0	0,60206	0
buruk	0	0	0	1	1	4	0,60206	0	0	0	0,60206
sekali	0	0	0	1	1	4	0,60206	0	0	0	0,60206

tinggal	0	0	0	1	1	4	0,60206	0	0	0	0,60206
selangkah	0	0	0	1	1	4	0,60206	0	0	0	0,60206
doang	0	0	0	1	1	4	0,60206	0	0	0	0,60206
mendek	0	0	0	1	1	4	0,60206	0	0	0	0,60206
gak	0	0	0	1	1	4	0,60206	0	0	0	0,60206
karuan	0	0		1	1	4	0,60206	0	0	0	0,60206

Setelah melakukan pembobotan term yang terdapat dalam setiap dokumen, langkah selanjutnya yaitu menghitung kemiripan dokumen yang akan diklasifikasi terhadap dokumen lainnya. Perhitungan ini terlebih dahulu akan dilakukan dengan menggunakan rumus cosine similarity untuk mendapatkan nilai panjang vektor.

Tabel 3.19 Bobot Term

NO	Dokumen	Bobot Term																	Sentimen
		bagus	cepat	ongkir	murah	kurir	ramah	lelet	pengiriman	lama	buruk	sekali	tinggal	selangkah	doang	mandek	gak	karuan	
1	Dokumen 1	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	Positif
2	Dokumen 2	0	0	0	1	1	1	0	0	0	0	0	0	0	0	0	0	0	Positif
3	Dokumen 3	0	0	0	0	0	0	1	1	1	1	1	0	0	0	0	0	0	Negatif
4	Dokumen 4	0	0	0	0	0	0	0	0	0	0	0	1	1	1	1	1	1	Negatif
5	Dokumen 5	1	1	0	1	0	1	0	0	0	0	0	0	0	0	0	0	0	?

Langkah-langkah :

1. Tentukan $K = 3$ (K sebaiknya bernilai ganjil)
2. Hitung *Similarity/Distance* (Metode *Euclidian Distance*)

Tabel 3.20 Perhitungan Distance

No	Dokumen	Distance
1	d1,d5	1,414213562
2	d2,d5	1,732050808
3	d3,d5	3
4	d4,d5	3,16227766

3. Urutkan Dokumen Berdasarkan Similaritas

Selanjutnya kita mengurutkan dokumen berdasarkan nilai *Distance* yang di dapat, semakin kecil nilai *Distance* yang dihasilkan maka semakin mirip kedekatannya.

Tabel 3.21 Perangkingan

No	Dokumen	Distance	ranking
1	d1,d5	1,414213562	1
2	d2,d5	1,732050808	2
3	d3,d5	3	3
4	d4,d5	3,16227766	4

4. Tentukan K klasifikasi (K terbaik berdasarkan Ranking)

Karena K yang di gunakan ada $K=3$ maka kita mengambil data ranking 3 teratas, dapat di lihat dari tabel di bawah ini.

Tabel 3.22 Rangking Terbaik

No	Dokumen	Distance	ranking
1	d1,d5	1,414213562	1
2	d2,d5	1,732050808	2
3	d3,d5	3	3

5. Tentukan klasifikasi berdasarkan Mayoritas

Tabel 3.236 Pelabelan Rangking Terbaik

No	Dokumen	Distance	Ranking	Sentimen
1	d1,d5	1,414213562	1	positif
2	d2,d5	1,732050808	2	positif
3	d3,d5	3	3	negatif

Kesimpulan : karena mayoritas dari ke-5 langkah adalah "Positif" maka dokumen 5 dilabelkan ke dalam sentimen "Positif"

3.1.5 Evaluasi dan validasi Hasil

Tahap ini akan menyimpulkan dari hasil evaluasi dari eksperimen yang dilakukan dari dataset yang telah dikumpulkan. Evaluasi dan validasi hasil merupakan proses yang digunakan untuk mengevaluasi dan memvalidasi hasil suatu proyek atau kegiatan. Evaluasi adalah proses menilai keberhasilan suatu proyek atau kegiatan dengan

mengukur sejauh mana tujuan yang telah ditetapkan tercapai. Validasi adalah proses menguji kebenaran atau keabsahan suatu hasil.

Evaluasi dan validasi hasil sangat penting dilakukan untuk memastikan bahwa proyek atau kegiatan yang dilakukan telah sesuai dengan tujuan yang telah ditetapkan dan hasil yang diperoleh dapat dipertanggungjawabkan. Dengan melakukan evaluasi dan validasi hasil, maka akan tercipta suatu proyek atau kegiatan yang berkualitas dan dapat memberikan manfaat yang optimal bagi yang bersangkutan.

Pada tahapan ini diusulkan model untuk data review pengguna aplikasi ID Express Customer Particle Swarm Optimize (PSO). dengan menerapkan algoritma Naïve Bayes dan K-Nearest Neighbor (KNN) dengan menambahkan feature selection Particle Swarm Optimize (PSO).

3.2 Alat dan Bahan

Penelitian ini, menggunakan perangkat keras Laptop ASUS VivoBook dengan Intel® Core™ i3 8250U Processor (6M Cache, up to 3.40 GHz), RAM 4,00 GB, sedangkan perangkat lunak yang digunakan Microsoft Excel, Python versi 3 dan Google Colab untuk pengolahan data. Sedangkan bahan yang digunakan diambil dari hasil proses *scrapping* data di aplikasi ID Express Customer di Gloogle Playstore.