

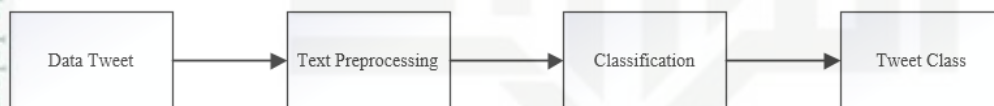
BAB IV

ANALISA DAN PERANCANGAN

Tahap Analisa dan perancangan akan menentukan hasil dari penelitian ini. Pada proses analisa berisi tentang tahapan-tahapan yang akan dilakukan pada sistem yang akan dibuat. Selanjutnya proses perancangan adalah proses merancang rincian sistem yang akan dibuat berdasarkan analisa yang telah dilakukan sebelumnya.

1.1 Analisa Data

Sistem ini bertujuan untuk mengklasifikasikan *tweet* kedalam lima kelas dengan menggunakan metode *Naïve Bayes*, dimana inputnya berupa teks dengan panjang maksimal 280 karakter atau yang disebut dengan *tweet* yang didapatkan melalui Twitter dengan menggunakan layanan API Twitter. Sebelum sistem melakukan proses klasifikasi dengan metode *Naïve Bayes*, *tweet* terlebih dahulu dilakukan *text preprocessing* terlebih dahulu seperti normalisasi teks, dan *stemming* tujuannya untuk meningkatkan akurasi dalam proses klasifikasi. Secara garis besar alur proses sistem ini dapat dilihat pada Gambar 4.1 berikut ini.



Gambar 4. 1 Gambaran Umum Sistem

1.2 Data Tweet Pelecehan *Online*

Data pelecehan *online* yang akan digunakan pada penelitian ini berjumlah 1500 *tweet*. Seluruh data *tweet* yang ada diberikan label kelas yang telah ditentukan yaitu *racist*, *embarrass*, *physically threatened*, *sexually harassed* dan *neutral*.

Data *tweet* yang diunduh adalah data *tweet* yang berasal dari Twitter berbahasa Indonesia cara yang digunakan dalam pengumpulan data adalah mengunduh data *tweet* menggunakan *script* PHP dengan layanan API (*Application Programming Interface*) Twitter. Data yang digunakan adalah data yang mengacu dari beberapa kasus pelecehan *online* yang populer dibicarakan di Twitter.

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.

Untuk mengumpulkan kata-kata pelecehan *online* peneliti membaca referensi dari berbagai media seperti berita terkait kategori pelecehan dan mengambil kata kunci yang biasa digunakan. Adapun klasifikasi kata-kata pelecehan *online* dapat dilihat pada Tabel 4.1 sebagai berikut:

Tabel 4. 1 Kata Kunci Pelecehan Online

| Kelas | Kata kunci Pelecehan Online |
|--|--|
| Kata-kata pelecehan terkait <i>Racist</i> | -aseng -pribumi -cina |
| Kata-kata pelecehan terkait <i>Physically treated</i> | -labrak -jambak -bunuh -jambak -gampar -bacok -tonjok -sajam -ganyang -tendang -jitak -cekek -santet -sleding -hajar |
| Kata-kata pelecehan terkait <i>Embarras</i> | -goblok -bego -babi -dasar ndeso -otak -anjing -taik -penjilat |
| Kata-kata pelecehan terkait <i>Sexually Harrasment</i> | -pelacur -lgbt -pantat -bokep -mesum -homo -lesbi -bugil -pelakor |

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.

1.3 Seleksi dan Pelabelan

Tahap selanjutnya setelah data berhasil dikumpulkan adalah seleksi data pada tahap ini data yang terkumpulkan kemudian diseleksi. Seleksi dilakukan karena tidak semua data yang berhasil diunduh dapat digunakan, seperti data yang menggunakan bahasa asing selain bahasa Indonesia harus dihilangkan. Dapat dilihat pada Gambar 4.2 data yang berwarna adalah contoh data tweet yang dihilangkan.

| | | |
|-----|--|--------|
| 112 | Dah la pakai tudung indon ngan seluar tracksuit ngan slipar ngan kaki terhincut hincut pi kfc pi dataran smh | Racist |
| 113 | lepas mamak tadi, waiter fav indon tu mai amik order wt : cam biasa ka ? me : yup..eh kelmarin sy ke medan wt : iâ€¦ https://t.co/QDSyJ9HVU4 | Racist |
| 114 | @shahruzimhdzr kalau awak nak tahu kasih sayang kita kat awak meruap ruap macam gunung berapi kat indon awak tu | Racist |
| 115 | Marathon filem indon dah 2 malam. Habis lah mimpi malam ni speaking indon takde subtitle | Racist |
| 116 | Nak faham tulisan kidal dia tu mcm baca bahasa indon :/ bertuah punya budak https://t.co/nDOycvflcN | Racist |
| 117 | Aku memang lag sikit bab tengok movie melayu or indon ni. First time tengok SYTD tadi, memang muka mei tu combo lofa€¦ https://t.co/1J4KsSdUTJ | Racist |
| 118 | @mhdshfq Hahahaha. Baik tanya kassim. Kassim indon gak kot @Hafizghanafii | Racist |
| 119 | Sementara itu, keluarga teroris hendak diberi fasilitas sama pemerintah supaya bisa mandiri, indon emang negara heba€¦ https://t.co/nkHn2y5cyk | Racist |
| 120 | Jarang sia bercakap ngan awek indon hahaha | Racist |
| 121 | @DivHumasPribumi Jangan ngomong aseng begini...aseng begitu...aseng bangsat.kalau masih mau kerja,jd babunya..cobaâ€¦ https://t.co/Agwt92yqez | Racist |

Gambar 4. 2 Tweet Yang Dihilangkan

Dan menghilangkan *tweet* yang serupa atau *retweet* merupakan proses untuk menyalin *tweet* dari orang lain untuk di *posting*-kan kembali, *retweet* ini ditandai dengan “RT” sebagai contoh “RT @ janesyar80 negeri ini rusak oleh aseng mata sipit”. Ini biasanya terjadi jika pengguna melihat *tweet* orang yang menarik untuk dibagikan. Berikut Gambar 4.3 proses penghapusan *retweet* dengan perintah query SQL.

```
Run SQL query/queries on table onlineharrasment.tweets: ⓘ
1 DELETE FROM `tweets` WHERE tweet LIKE '%RT%' |
```

Gambar 4. 3 Hapus Retweet

Proses pelabelan data dilakukan dengan beberapa filter sebagai berikut:

1. Pelabelan dilakukan berdasarkan kata kunci yang sudah mencirikan untuk setiap masing-masing kelas.
2. Banyaknya *tweet* yang mengandung kata kunci yang mencirikan beberapa kelas dalam sebuah *tweet* maka *tweet* ini akan dihapus contoh : “*pengen di jambak si cina dasar tolol hobi mesum*” ini merupakan *tweet* yang akan dihapus karena *tweet* tidak bisa di label ke kelas tertentu jika terdapat bagian dari kata kunci kelas lain didalamnya.
3. Selanjutnya memberi label pada seluruh data *tweet* sesuai dengan kelas yang telah ditentukan yaitu *racist*, *embarrass*, *physically threatened*, *sexually harassed* dan *neutral*.

Pada Tabel 4.2 adalah contoh *tweet* yang dikumpulkan dan sudah diberi label secara manual.

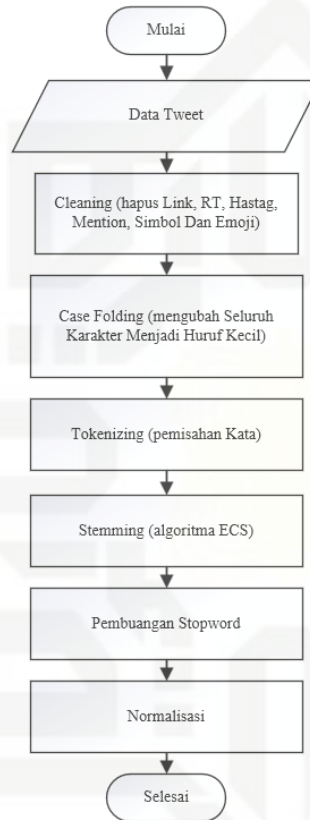
Tabel 4. 2 Contoh Tweet Berlabel

| Dok | <i>Tweet</i> | Kelas |
|-------|---|------------------------------|
| Dok 1 | Yang pasti Zumi zola yang digoreng Sudah Muslim, Pribumi pulak. Makanya kalao mau aman korupsi ya harus non-pribum... https://t.co/NIZVDu0mpi | <i>Racist</i> |
| Dok 2 | SAKA AGARI GA BUTUH ORANG KAYA LU. RAME NYA KAMPUNGAN NORAK DAN NGERUSAK HAJATAN ROBA. GA GUNA! Kzl bgt gue | <i>Embarrass</i> |
| Dok 3 | Ngadu lo ngadu. Awas lo ye, gue labrak lo di prosotan | <i>Physically threatened</i> |
| Dok 4 | Ooh @B12udy ngatain aku lonte pelacur bajingan, tp ternyata doyan jg bisa terima cewe itu sbg pacar yg dr awal sdh... https://t.co/czBEJg5jMZ | <i>Sexually harassed</i> |
| Dok 5 | Skandal seks dan sewa pelacur di Haiti, Oxfam jadi bulan-bulanan. Simak selengkapnya di sini: https://t.co/31LugUZMA3 | <i>Neutral</i> |

Proses seleksi dilakukan dengan cara manual. Data *tweet* diunduh lebih kurang 3000 *tweet* kemudian data diseleksi menyisakan data berjumlah 1500 *tweet* yang siap diolah untuk proses selanjutnya. Setelah data diberi label secara manual, barulah dapat dilakukan proses *text preprocessing*.

1.4 Text Preprocessing

Text Preprocessing merupakan langkah untuk membersihkan data dengan tujuan untuk menyeragamkan bentuk kata dan mengurangi volume kata dengan menghapus unsur-unsur yang tidak dibutuhkan. Adapun proses dapat dilakukan dalam *text preprocessing* data dapat dilihat pada Gambar 4.4 berikut :



Gambar 4. 4 *Flowchart Text Preprocessing*

1.4.1 Cleaning

Cleaning adalah proses pembersihan *tweet* dari kata-kata yang tidak digunakan untuk mengurangi *noise* pada proses klasifikasi. Kata-kata yang akan dibersihkan pada proses ini seperti *hashtag* (#), karakter HTML, RT (*retweet*), *username* (@), *emoticon* dan URL. Proses *cleaning* pada *tweet* dapat dilihat pada Tabel 4.3 berikut ini.

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. 3 Hasil *Cleaning*

| Dok | <i>Tweet</i> | Kelas |
|------------|--|------------------------------|
| Dok 1 | Yang pasti Zumi zola yang digoreng Sudah Muslim Pribumi pulak Makanya kalao mau aman korupsi ya harus non pribum | <i>Racist</i> |
| Dok 2 | SAKA AGARI GA BUTUH ORANG KAYA LU RAME NYA KAMPUNGAN NORAK DAN NGERUSAK HAJATAN ROBA GA GUNA Kzl bgt gue | <i>Embarrass</i> |
| Dok 3 | Tidak semua teman itu baik dan tidak semua musuh itu jahat Jagain pinutnya gua ya Awas klo disakitin Klo lu nyakitin dia bakal gua bunuh idup wkwk | <i>Physically threatened</i> |
| Dok 4 | Ooh ngatain aku lonte pelacur bajingan tp ternyata doyan jg 6ag nerima cewe itu sbg pacar yg dr awal sdh | <i>Sexually harassed</i> |
| Dok 5 | Skandal seks dan sewa pelacur di Haiti Oxfam jadi bulan bulanan Simak selengkapnya di sini | <i>Neutral</i> |

1.4.2 *Case folding*

Case folding adalah proses mengubah semua huruf kapital menjadi huruf kecil. Proses ini dilakukan karena tidak semua isi *tweet* konsisten dalam penggunaan huruf. Oleh karena itu proses *case folding* diperlukan untuk mengatasi kata atau *term* ganda hanya karena penulisan katanya yang tidak sama. Makanya untuk meratakan semua kata, dikonversikan semua huruf kapital ke bentuk standar (huruf kecil). Proses *case folding* dapat dilihat pada Tabel 4.4 berikut:

Tabel 4. 4 Hasil *Case Folding*

| Dok | <i>Tweet</i> | Kelas |
|------------|--|------------------------------|
| Dok 1 | yang pasti zumi zola yang digoreng sudah muslim pribumi pulak makanya kalao mau aman korupsi ya harus non pribum | <i>Racist</i> |
| Dok 2 | saka agari ga butuh orang kaya lu rame nya kampungan norak dan ngerusak hajatan roba 6ag una kzl bgt gue | <i>Embarrass</i> |
| Dok 3 | tidak semua teman itu baik dan tidak semua musuh itu jahat jagain pinutnya gua ya awas klo disakitin klo lu nyakitin dia bakal gua bunuh idup wkwk | <i>Physically threatened</i> |
| Dok 4 | Ooh ngatain aku lonte pelacur bajingan tp ternyata doyan jg bisa nerima cewe itu sbg pacar yg dr awal sdh | <i>Sexually harassed</i> |
| Dok 5 | skandal seks dan sewa pelacur di haiti oxfam jadi bulan bulanan simak selengkapnya di sini | <i>Neutral</i> |

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.

1.4.3 Tokenizing

Tokenizing adalah proses pemisahan kalimat menjadi kata perkata atau potongan kata tunggal. Contoh proses *tokenizing* seperti pada Tabel 4.5 berikut:

Tabel 4. 5 Hasil Tokenizing

| Dok 1 | Dok 2 | Dok 3 | Dok 4 | Dok 5 |
|----------|-----------|-----------|----------|--------------|
| yang | saka | tidak | ooh | skandal |
| pasti | agari | semua | ngatain | seks |
| zumi | ga | teman | aku | dan |
| zola | butuh | itu | lonte | sewa |
| yang | orang | baik | pelacur | pelacur |
| digoreng | kaya | dan | bajingan | di |
| sudah | lu | tidak | tp | haiti |
| muslim | rame | semua | ternyata | oxfam |
| pribumi | nya | musuh | doyan | jadi |
| pulak | kampungan | itu | jg | simak |
| makanya | norak | jahat | bisa | selengkapnya |
| kalao | dan | jagain | nerima | di |
| mau | ngerusak | pinutnya | cewe | sini |
| aman | hajatan | gua | itu | |
| korupsi | roba | ya | sbg | |
| ya | ga | awas | pacar | |
| harus | guna | klo | yg | |
| non | kzl | disakitin | dr | |
| pribum | bgt | klo | awal | |
| | gue | lu | sdh | |
| | | nyakitin | | |
| | | dia | | |
| | | bakal | | |
| | | gua | | |
| | | bunuh | | |
| | | idup | | |
| | | wkwk | | |

1.4.4 Normalisasi

Normalisasi adalah proses mengubah kata-kata yang tidak baku menjadi kata baku menurut kamus normalisasi yang telah dibuat. Sebagian besar mengacu pada kata baku menurut Kamus Besar Bahasa Indonesia dan ada juga penambahan kata yang sesuai dengan penelitian ini, seperti kata *ndeso* (sifat orang kampung, tapi

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.

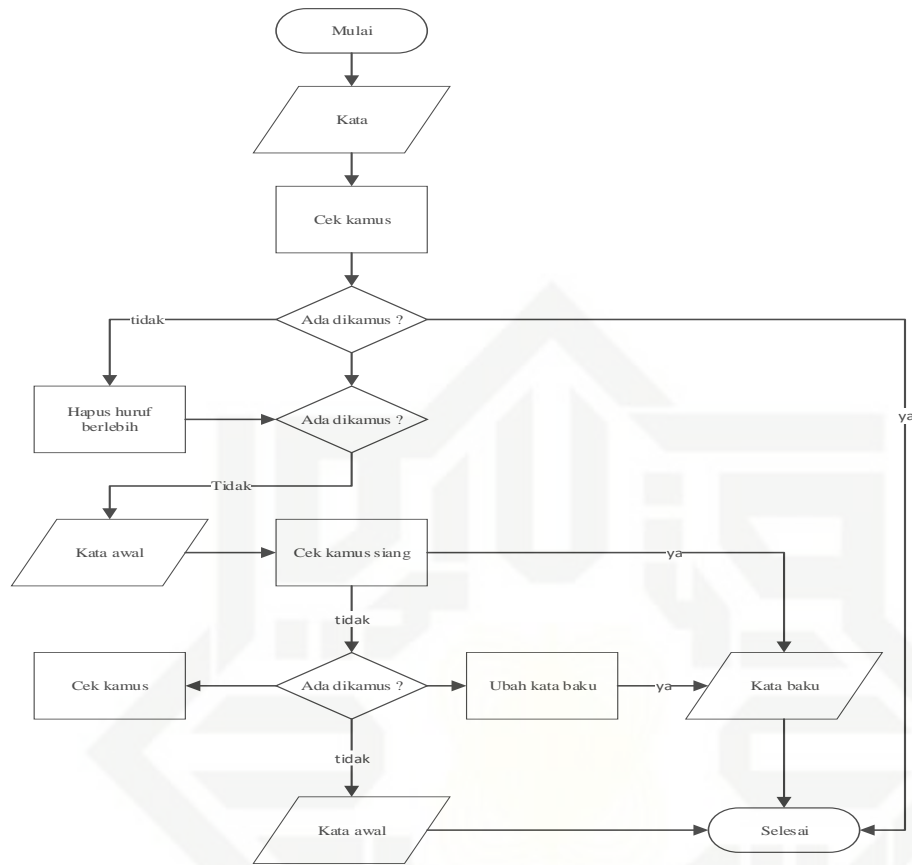
bukan orang kampung), *aseng* (simbol anti Cina) dan lain-lain. Adapun algoritma normalisasi yang dilakukan pada penelitian ini sebagai berikut:

1. Cari kata yang akan dinormalisasikan dalam kamus. Jika ditemukan maka diasumsikan bahwa kata tersebut adalah *root word*, maka algoritma terhenti.
2. Jika tidak ditemukan hapus huruf berlebihan dimulai untuk setiap huruf pada kata, periksa huruf pertama kata tersebut, kemudian *recoding*. Periksa huruf selanjutnya jika huruf sama dengan huruf sebelumnya maka hapus huruf tersebut, jika tidak simpan huruf lakukan hal yang sama pada huruf selanjutnya.
3. Melakukan *recoding*
4. Jika telah diperiksa untuk setiap huruf periksa kata harus proses sebelumnya pada kamus.
5. Jika ditemukan maka algoritma terhenti. Jika tidak ditemukan algoritma ini mengembalikan kata yang asli sebelum dilakukan penghapusan huruf berlebihan.
6. Selanjutnya periksa kata pada kamus normalisasi.
7. Jika ditemukan lakukan perubahan kata menjadi kata baku. Jika tidak ditemukan maka kembalikan kata pada *root word*.

Flowchart normalisasi dapat dilihat pada Gambar 4.5 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.



Gambar 4.5 Flowchart Normalisasi

1.4.5 Stopword Removal

Stopword Removal adalah proses penghilangan *stopword* atau kata yang dianggap tidak berguna dan tahapan ini juga dapat menekan penggunaan waktu dalam menentukan hasil. Contoh *stopword* bahasa Indonesia yaitu kata hubung seperti “akan”, “kemudian”, “yang”, “sebagai”, dan lain-lain, petunjuk waktu, serta kata tanya. Prosesnya dapat dilihat pada Tabel 4.6 berikut:

Tabel 4.6 Hasil Stopword Removal

| Dok 1 | Dok 2 | Dok 3 | Dok 4 | Dok 5 |
|----------|-----------|-------|----------|--------------|
| digoreng | butuh | teman | lonte | skandal |
| muslim | orang | musuh | pelacur | seks |
| pribumi | kaya | jahat | bajingan | sewa |
| aman | kampungan | awas | ternyata | pelacur |
| korupsi | norak | bakal | doyan | simak |
| | hajatan | bunuh | pacar | selengkapnya |
| | gue | hidup | | |

1.4.6 Stemming

Stemming adalah proses menghilangkan imbuhan pada kata berimbuhan menjadi kata dasar menggunakan algoritma *Enhanced Confix Stemming*. Penentuan kata dasar atau proses *stemming* ini mengacu pada kamus bahasa Indonesia sebagai kumpulan kata dasar. Analisa *stemming* ini mengikuti format penulisan bahasa Indonesia yang telah mengandung imbuhan (*affix*) awalan (*prefix*, akhiran (*suffix*) sisipan (*infix*), dan awalan akhiran (*konfix*) berikut contoh hasil proses *stemming*:

Tabel 4. 7 Hasil *Stemming*

| Dok 1 | Dok 2 | Dok 3 | Dok 4 | Dok 5 |
|---------|---------|-------|--------|---------|
| goreng | butuh | teman | lonte | skandal |
| muslim | orang | musuh | lacur | seks |
| pribumi | kaya | jahat | bajing | sewa |
| aman | kampung | awas | nyata | lacur |
| korupsi | norak | bakal | doyan | simak |
| | hajat | bunuh | pacar | lengkap |
| | | hidup | | |

1.5 Pembobotan

Pembobotan adalah proses merubah kata menjadi bentuk vektor. Ada beberapa tahapan yang harus dilalui agar tujuan itu bisa dicapai seperti menghitung *term frequency* (TF) jumlah kemuculan kata dalam dokumen.

Dalam penelitian tugas akhir ini fitur yang digunakan adalah *unigram* dengan pembobotan menggunakan TF (*Term Frequency*) Kata direpresentasi ke dalam bentuk vektor, dimana tiap kata dihitung sebagai suatu fitur. Adapun perhitungan bobot yang digunakan adalah *Term Frequency* (TF). Pada Tabel 4.7 adalah hasil pembobotan selengkapnya berdasarkan contoh *tweet* diatas. Proses pembobotan dalam penelitian ini akan dilakukan dengan *script* PHP.

Tabel 4. 8 Hasil Pembobotan TF

| No | Kata | Dok 1 | Dok 2 | Dok 3 | Dok 4 | Dok 5 |
|----|---------|-------|-------|-------|-------|-------|
| 1 | goreng | 1 | 0 | 0 | 0 | 0 |
| 2 | muslim | 1 | 0 | 0 | 0 | 0 |
| 3 | pribumi | 1 | 0 | 0 | 0 | 0 |
| 4 | aman | 1 | 0 | 0 | 0 | 0 |
| 5 | korupsi | 1 | 0 | 0 | 0 | 0 |
| 6 | butuh | 0 | 1 | 0 | 0 | 0 |

Hak Cipta Dilindungi Undang-Undang

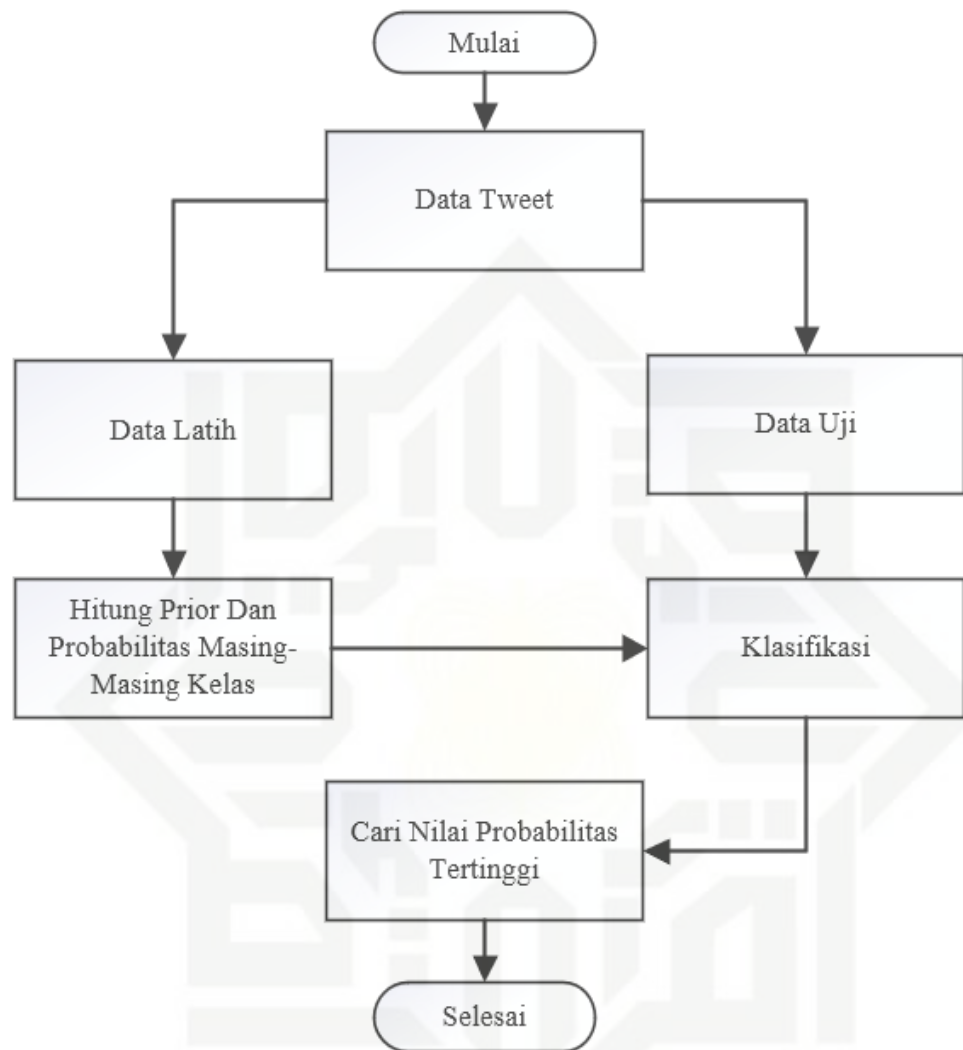
1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan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

| No | Kata | Dok 1 | Dok 2 | Dok 3 | Dok 4 | Dok 5 |
|----|---------|-------|-------|-------|-------|-------|
| 7 | orang | 0 | 1 | 0 | 0 | 0 |
| 8 | kaya | 0 | 1 | 0 | 0 | 0 |
| 9 | kampung | 0 | 1 | 0 | 0 | 0 |
| 10 | norak | 0 | 1 | 0 | 0 | 0 |
| 11 | hajat | 0 | 1 | 0 | 0 | 0 |
| 12 | teman | 0 | 0 | 1 | 0 | 0 |
| 13 | musuh | 0 | 0 | 1 | 0 | 0 |
| 14 | jahat | 0 | 0 | 1 | 0 | 0 |
| 15 | awas | 0 | 0 | 1 | 0 | 0 |
| 16 | bakal | 0 | 0 | 1 | 0 | 0 |
| 17 | bunuh | 0 | 0 | 1 | 0 | 0 |
| 18 | hidup | 0 | 0 | 1 | 0 | 0 |
| 19 | lonte | 0 | 0 | 0 | 1 | 0 |
| 20 | lacur | 0 | 0 | 0 | 1 | 1 |
| 21 | bajing | 0 | 0 | 0 | 1 | 0 |
| 22 | nyata | 0 | 0 | 0 | 1 | 0 |
| 23 | doyan | 0 | 0 | 0 | 1 | 0 |
| 24 | pacar | 0 | 0 | 0 | 1 | 0 |
| 25 | skandal | 0 | 0 | 0 | 0 | 1 |
| 26 | seks | 0 | 0 | 0 | 0 | 1 |
| 27 | sewa | 0 | 0 | 0 | 0 | 1 |
| 28 | simak | 0 | 0 | 0 | 0 | 1 |
| 29 | lengkap | 0 | 0 | 0 | 0 | 1 |

1.6 Analisa Algoritma *Naïve Bayes Classifier*

Berdasarkan data *tweet* pelecehan *online* yang telah dilakukan proses sebelumnya, maka pada tahapan ini akan dijelaskan teknik-teknik yang digunakan dalam klasifikasi *Naïve Bayes Classifier*.

Naïve Bayes merupakan algoritma *supervised learning* yang memanfaatkan teori probabilitas. Sehingga dibutuhkan pengetahuan awal untuk dapat mengambil keputusan. Berikut langkah-langkah proses algoritma *Naïve Bayes* dalam klasifikasi *tweet* pelecehan *online* dalam *flowchart* 4.6 berikut.



Gambar 4. 6 Flowchart Klasifikasi Pelecehan Online Metode Naïve Bayes

Pada proses dengan *Naïve Bayes Classifier* dibagi menjadi dua tahapan, pertama tahap *training* (latih) dan tahap *testing* (pengujian). Penjelasan tahapan ini dapat dilihat sebagai berikut:

1. Tahap *training*

Pada tahap *training* ini data yang telah didapatkan TF-nya seperti pada sampel Tabel 4.7 dijadikan data latih untuk menjadi acuan dalam membentuk model klasifikasi. Pada tahapan ini akan dicari nilai probabilitas kelas dari data latih.

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 mengemukakan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

Menghitung nilai probabilitas (persamaan 2.8 dan persamaan 2.9) setiap *term* untuk setiap kelas dari data latih. Perhitungannya adalah sebagai berikut:

- a. Pertama hitung probabilitas setiap kategori menggunakan persamaan 2.8, dalam penelitian ini yang menjadi kategori ada lima yaitu, kategori kelas *Racist*, *Embarrass*, *Physically threatened*, *Sexually harassed* dan kelas *Neutral*.

$$P(r/e/p/s/n) = \frac{d(r/e/p/s/n)}{|c|} = \frac{1}{5}$$

$$P(racist) = \frac{d(racist)}{|c|} = \frac{1}{5}$$

$$P(embarrass) = \frac{d(embarrass)}{|c|} = \frac{1}{5}$$

$$P(physically\ threatened) = \frac{d(physically\ threatened)}{|c|} = \frac{1}{5}$$

$$P(sexually\ harassed) = \frac{d(sexually\ harassed)}{|c|} = \frac{1}{5}$$

$$P(netral) = \frac{d(netral)}{|c|} = \frac{1}{5}$$

- b. Selanjutnya hitung probabilitas setiap *term* dari semua dokumen dengan menggunakan persamaan 2.12. sebagai contoh perhitungan dari 5 dokumen yang masing-masing terdiri 1 mewakili setiap kelas yang digunakan dalam perhitungan ini, memiliki banyak jumlah *term* sebanyak 29, 5 *term* dari kelas *racist*, 6 *term* dari kelas *embarrass*, 7 *term* dari kelas *physically threatened*, 6 *term* dari kelas *sexually harassed* dan 6 *term* dari kelas *neutral*. Banyaknya *term* tergantung dari proses *text preprocessing* dan *stemming*. Perhitungan probabilitas dari setiap *term* adalah sebagai berikut, berdasarkan sampelnya pada Tabel 4.8.

$$P(w_k|r/e/p/s/n) = \frac{(n_k, R/E/P/S/N) + 1}{(n, R/E/P/S/N) + |\text{kosakata}|}$$

Keterangan :

$P(w_k|r/e/p/s/n)$ = Peluang kemunculan kata-kata pada sebuah kategori/kelas, w_k adalah kata yang muncul pada sebuah kategori.

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.

$(n_k, R/E/P/S/N) + 1$ = Jumlah frekuensi kemunculan setiap kata pada sebuah kategori/kelas. Kemudian ditambah 1, 1 merupakan nilai mutlak dari rumusnya.

$(n, R/E/P/S/N)$ = Jumlah frekuensi kemunculan seluruh kata dari setiap kategori/kelas, n adalah kemunculan seluruh kata dari setiap kategori/kelas.

$|\text{kosakata}|$ = Jumlah semua kata dari semua kategori.

Diketahui $|\text{kosakata}|=30$

count Racist=5, count embarrass=6, count physically threatened =7, count sexually harassed=6, count neutral=6

Probabilitas kata “goreng”

$$P(\text{"goreng"}|R) = \frac{(\text{"goreng"}|R) + 1}{(R) + |\text{kosakata}|} = \frac{1 + 1}{5 + 29} = 0,0588235294117647$$

$$P(\text{"goreng"}|E) = \frac{(\text{"goreng"}|E) + 1}{(E) + |\text{kosakata}|} = \frac{0 + 1}{6 + 29} = 0,0285714285714286$$

$$P(\text{"goreng"}|P) = \frac{(\text{"goreng"}|P) + 1}{(P) + |\text{kosakata}|} = \frac{0 + 1}{7 + 29} = 0,0277777777777778$$

$$P(\text{"goreng"}|S) = \frac{(\text{"goreng"}|S) + 1}{(S) + |\text{kosakata}|} = \frac{0 + 1}{6 + 29} = 0,0285714285714286$$

$$P(\text{"goreng"}|N) = \frac{(\text{"goreng"}|N) + 1}{(N) + |\text{kosakata}|} = \frac{0 + 1}{6 + 29} = 0,0285714285714286$$

Probabilitas kata “muslim”

$$P(\text{"muslim"}|R) = \frac{(\text{"muslim"}|R) + 1}{(R) + |\text{kosakata}|} = \frac{1 + 1}{5 + 29} = 0,0588235294117647$$

$$P(\text{"muslim"}|E) = \frac{(\text{"muslim"}|E) + 1}{(E) + |\text{kosakata}|} = \frac{0 + 1}{6 + 29} = 0,0285714285714286$$

$$P(\text{"muslim"}|P) = \frac{(\text{"muslim"}|P) + 1}{(P) + |\text{kosakata}|} = \frac{0 + 1}{7 + 29} = 0,0277777777777778$$

$$P(\text{"muslim"}|S) = \frac{(\text{"muslim"}|S) + 1}{(S) + |\text{kosakata}|} = \frac{0 + 1}{6 + 29} = 0,0285714285714286$$

$$P(\text{"muslim"}|N) = \frac{(\text{"muslim"}|N) + 1}{(N) + |\text{kosakata}|} = \frac{0 + 1}{6 + 29} = 0,0285714285714286$$

Probabilitas kata “pribumi”

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.

$$P(\text{"pribumi"}|\text{R}) = \frac{(\text{"pribumi"}|\text{R}) + 1}{(\text{R}) + |\text{kosakata}|} = \frac{1 + 1}{5 + 29} = 0,0588235294117647$$

$$P(\text{"pribumi"}|\text{E}) = \frac{(\text{"pribumi"}|\text{E}) + 1}{(\text{E}) + |\text{kosakata}|} = \frac{0 + 1}{6 + 29} = 0,0285714285714286$$

$$P(\text{"pribumi"}|\text{P}) = \frac{(\text{"pribumi"}|\text{P}) + 1}{(\text{P}) + |\text{kosakata}|} = \frac{0 + 1}{7 + 29} = 0,0277777777777778$$

$$P(\text{"pribumi"}|\text{S}) = \frac{(\text{"pribumi"}|\text{S}) + 1}{(\text{S}) + |\text{kosakata}|} = \frac{0 + 1}{6 + 29} = 0,0285714285714286$$

$$P(\text{"pribumi"}|\text{N}) = \frac{(\text{"pribumi"}|\text{N}) + 1}{(\text{N}) + |\text{kosakata}|} = \frac{0 + 1}{6 + 29} = 0,0285714285714286$$

Probabilitas kata "aman"

$$P(\text{"aman"}|\text{R}) = \frac{(\text{"aman"}|\text{R}) + 1}{(\text{R}) + |\text{kosakata}|} = \frac{1 + 1}{5 + 29} = 0,0588235294117647$$

$$P(\text{"aman"}|\text{E}) = \frac{(\text{"aman"}|\text{E}) + 1}{(\text{E}) + |\text{kosakata}|} = \frac{0 + 1}{6 + 29} = 0,0285714285714286$$

$$P(\text{"aman"}|\text{P}) = \frac{(\text{"aman"}|\text{P}) + 1}{(\text{P}) + |\text{kosakata}|} = \frac{0 + 1}{7 + 29} = 0,0277777777777778$$

$$P(\text{"aman"}|\text{S}) = \frac{(\text{"aman"}|\text{S}) + 1}{(\text{S}) + |\text{kosakata}|} = \frac{0 + 1}{6 + 29} = 0,0285714285714286$$

$$P(\text{"aman"}|\text{N}) = \frac{(\text{"aman"}|\text{N}) + 1}{(\text{N}) + |\text{kosakata}|} = \frac{0 + 1}{6 + 29} = 0,0285714285714286$$

Probabilitas kata "korupsi"

$$P(\text{"korupsi"}|\text{R}) = \frac{(\text{"korupsi"}|\text{R}) + 1}{(\text{R}) + |\text{kosakata}|} = \frac{0 + 1}{5 + 29} = 0,0588235294117647$$

$$P(\text{"korupsi"}|\text{E}) = \frac{(\text{"korupsi"}|\text{E}) + 1}{(\text{E}) + |\text{kosakata}|} = \frac{0 + 1}{6 + 29} = 0,0285714285714286$$

$$P(\text{"korupsi"}|\text{P}) = \frac{(\text{"korupsi"}|\text{P}) + 1}{(\text{P}) + |\text{kosakata}|} = \frac{0 + 1}{7 + 29} = 0,0277777777777778$$

$$P(\text{"korupsi"}|\text{S}) = \frac{(\text{"korupsi"}|\text{S}) + 1}{(\text{S}) + |\text{kosakata}|} = \frac{0 + 1}{6 + 29} = 0,0285714285714286$$

$$P(\text{"korupsi"}|\text{N}) = \frac{(\text{"korupsi"}|\text{N}) + 1}{(\text{N}) + |\text{kosakata}|} = \frac{0 + 1}{6 + 29} = 0,0285714285714286$$

Probabilitas kata "butuh"

$$P(\text{"butuh"}|\text{R}) = \frac{(\text{"butuh"}|\text{R}) + 1}{(\text{R}) + |\text{kosakata}|} = \frac{0 + 1}{5 + 29} = 0,0294117647058824$$

$$P(\text{"butuh"}|\text{E}) = \frac{(\text{"butuh"}|\text{E}) + 1}{(\text{E}) + |\text{kosakata}|} = \frac{1 + 1}{6 + 29} = 0,0571428571428571$$

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.

$$P(\text{"butuh"}|E) = \frac{(\text{"butuh"}|P) + 1}{(E) + |\text{kosakata}|} = \frac{0 + 1}{7 + 29} = 0,02777777777777778$$

$$P(\text{"butuh"}|E) = \frac{(\text{"butuh"}|S) + 1}{(E) + |\text{kosakata}|} = \frac{0 + 1}{6 + 29} = 0,0285714285714286$$

$$P(\text{"butuh"}|E) = \frac{(\text{"butuh"}|N) + 1}{(E) + |\text{kosakata}|} = \frac{0 + 1}{6 + 29} = 0,0285714285714286$$

Probabilitas kata “orang”

$$P(\text{"orang"}|E) = \frac{(\text{"orang"}|R) + 1}{(E) + |\text{kosakata}|} = \frac{0 + 1}{5 + 29} = 0,0294117647058824$$

$$P(\text{"orang"}|E) = \frac{(\text{"orang"}|E) + 1}{(E) + |\text{kosakata}|} = \frac{1 + 1}{6 + 29} = 0,0571428571428571$$

$$P(\text{"orang"}|E) = \frac{(\text{"orang"}|P) + 1}{(E) + |\text{kosakata}|} = \frac{0 + 1}{7 + 29} = 0,02777777777777778$$

$$P(\text{"orang"}|E) = \frac{(\text{"orang"}|S) + 1}{(E) + |\text{kosakata}|} = \frac{0 + 1}{6 + 29} = 0,0285714285714286$$

$$P(\text{"orang"}|E) = \frac{(\text{"orang"}|N) + 1}{(E) + |\text{kosakata}|} = \frac{0 + 1}{6 + 29} = 0,0285714285714286$$

Probabilitas kata “kaya”

$$P(\text{"kaya"}|E) = \frac{(\text{"kaya"}|R) + 1}{(E) + |\text{kosakata}|} = \frac{0 + 1}{5 + 29} = 0,0294117647058824$$

$$P(\text{"kaya"}|E) = \frac{(\text{"kaya"}|E) + 1}{(E) + |\text{kosakata}|} = \frac{1 + 1}{6 + 29} = 0,0571428571428571$$

$$P(\text{"kaya"}|E) = \frac{(\text{"kaya"}|P) + 1}{(E) + |\text{kosakata}|} = \frac{0 + 1}{7 + 29} = 0,02777777777777778$$

$$P(\text{"kaya"}|E) = \frac{(\text{"kaya"}|S) + 1}{(E) + |\text{kosakata}|} = \frac{0 + 1}{6 + 29} = 0,0285714285714286$$

$$P(\text{"kaya"}|E) = \frac{(\text{"kaya"}|N) + 1}{(E) + |\text{kosakata}|} = \frac{0 + 1}{6 + 29} = 0,0285714285714286$$

Probabilitas kata “kampung”

$$P(\text{"kampung"}|E) = \frac{(\text{"kampung"}|R) + 1}{(E) + |\text{kosakata}|} = \frac{0 + 1}{5 + 29} = 0,0294117647058824$$

$$P(\text{"kampung"}|E) = \frac{(\text{"kampung"}|E) + 1}{(E) + |\text{kosakata}|} = \frac{1 + 1}{6 + 29} = 0,0571428571428571$$

$$P(\text{"kampung"}|E) = \frac{(\text{"kampung"}|P) + 1}{(E) + |\text{kosakata}|} = \frac{0 + 1}{7 + 29} = 0,02777777777777778$$

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.

$$P(\text{"kampung"}|E) = \frac{(\text{"kampung"}|S) + 1}{(E) + |\text{kosakata}|} = \frac{0 + 1}{6 + 29} = 0,0285714285714286$$

$$P(\text{"kampung"}|E) = \frac{(\text{"kampung"}|N) + 1}{(E) + |\text{kosakata}|} = \frac{0 + 1}{6 + 29} = 0,0285714285714286$$

Probabilitas kata “norak”

$$P(\text{"norak"}|E) = \frac{(\text{"norak"}|R) + 1}{(E) + |\text{kosakata}|} = \frac{0 + 1}{5 + 29} = 0,0294117647058824$$

$$P(\text{"norak"}|E) = \frac{(\text{"norak"}|E) + 1}{(E) + |\text{kosakata}|} = \frac{1 + 1}{6 + 29} = 0,0571428571428571$$

$$P(\text{"norak"}|E) = \frac{(\text{"norak"}|P) + 1}{(E) + |\text{kosakata}|} = \frac{0 + 1}{7 + 29} = 0,0277777777777778$$

$$P(\text{"norak"}|E) = \frac{(\text{"norak"}|S) + 1}{(E) + |\text{kosakata}|} = \frac{0 + 1}{6 + 29} = 0,0285714285714286$$

$$P(\text{"norak"}|E) = \frac{(\text{"norak"}|N) + 1}{(E) + |\text{kosakata}|} = \frac{0 + 1}{6 + 29} = 0,0285714285714286$$

Probabilitas kata “hajat”

$$P(\text{"hajat"}|E) = \frac{(\text{"hajat"}|R) + 1}{(E) + |\text{kosakata}|} = \frac{0 + 1}{5 + 29} = 0,0294117647058824$$

$$P(\text{"hajat"}|E) = \frac{(\text{"hajat"}|E) + 1}{(E) + |\text{kosakata}|} = \frac{1 + 1}{6 + 29} = 0,0571428571428571$$

$$P(\text{"hajat"}|E) = \frac{(\text{"hajat"}|P) + 1}{(E) + |\text{kosakata}|} = \frac{0 + 1}{7 + 29} = 0,0277777777777778$$

$$P(\text{"hajat"}|E) = \frac{(\text{"hajat"}|S) + 1}{(E) + |\text{kosakata}|} = \frac{0 + 1}{6 + 29} = 0,0285714285714286$$

$$P(\text{"hajat"}|E) = \frac{(\text{"hajat"}|N) + 1}{(E) + |\text{kosakata}|} = \frac{0 + 1}{6 + 29} = 0,0285714285714286$$

Probabilitas kata “teman”

$$P(\text{"teman"}|P) = \frac{(\text{"teman"}|R) + 1}{(P) + |\text{kosakata}|} = \frac{0 + 1}{5 + 29} = 0,0294117647058824$$

$$P(\text{"teman"}|P) = \frac{(\text{"teman"}|E) + 1}{(P) + |\text{kosakata}|} = \frac{0 + 1}{6 + 29} = 0,0285714285714286$$

$$P(\text{"teman"}|P) = \frac{(\text{"teman"}|P) + 1}{(P) + |\text{kosakata}|} = \frac{1 + 1}{7 + 29} = 0,0555555555555556$$

$$P(\text{"teman"}|P) = \frac{(\text{"teman"}|S) + 1}{(P) + |\text{kosakata}|} = \frac{0 + 1}{6 + 29} = 0,0285714285714286$$

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 mengemukakan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

$$P(\text{"teman"}|P) = \frac{(\text{"teman"}|"N) + 1}{(P) + |\text{kosakata}|} = \frac{0 + 1}{6 + 29} = 0,0285714285714286$$

Probabilitas kata “musuh”

$$P(\text{"musuh"}|P) = \frac{(\text{"musuh"}|"R) + 1}{(P) + |\text{kosakata}|} = \frac{0 + 1}{5 + 29} = 0,0294117647058824$$

$$P(\text{"musuh"}|P) = \frac{(\text{"musuh"}|"E) + 1}{(P) + |\text{kosakata}|} = \frac{0 + 1}{6 + 29} = 0,0285714285714286$$

$$P(\text{"musuh"}|P) = \frac{(\text{"musuh"}|"P) + 1}{(P) + |\text{kosakata}|} = \frac{1 + 1}{7 + 29} = 0,0555555555555556$$

$$P(\text{"musuh"}|P) = \frac{(\text{"musuh"}|"S) + 1}{(P) + |\text{kosakata}|} = \frac{0 + 1}{6 + 29} = 0,0285714285714286$$

$$P(\text{"musuh"}|P) = \frac{(\text{"musuh"}|"N) + 1}{(P) + |\text{kosakata}|} = \frac{0 + 1}{6 + 29} = 0,0285714285714286$$

Probabilitas kata “jahat”

$$P(\text{"jahat"}|P) = \frac{(\text{"jahat"}|"R) + 1}{(P) + |\text{kosakata}|} = \frac{0 + 1}{5 + 29} = 0,0294117647058824$$

$$P(\text{"jahat"}|P) = \frac{(\text{"jahat"}|"E) + 1}{(P) + |\text{kosakata}|} = \frac{0 + 1}{6 + 29} = 0,0285714285714286$$

$$P(\text{"jahat"}|P) = \frac{(\text{"jahat"}|"P) + 1}{(P) + |\text{kosakata}|} = \frac{1 + 1}{7 + 29} = 0,0555555555555556$$

$$P(\text{"jahat"}|P) = \frac{(\text{"jahat"}|"S) + 1}{(P) + |\text{kosakata}|} = \frac{0 + 1}{6 + 29} = 0,0285714285714286$$

$$P(\text{"jahat"}|P) = \frac{(\text{"jahat"}|"N) + 1}{(P) + |\text{kosakata}|} = \frac{0 + 1}{6 + 29} = 0,0285714285714286$$

Probabilitas kata “awas”

$$P(\text{"awas"}|P) = \frac{(\text{"awas"}|"R) + 1}{(P) + |\text{kosakata}|} = \frac{0 + 1}{5 + 29} = 0,0294117647058824$$

$$P(\text{"awas"}|P) = \frac{(\text{"awas"}|"E) + 1}{(P) + |\text{kosakata}|} = \frac{0 + 1}{6 + 29} = 0,0285714285714286$$

$$P(\text{"awas"}|P) = \frac{(\text{"awas"}|"P) + 1}{(P) + |\text{kosakata}|} = \frac{1 + 1}{7 + 29} = 0,0555555555555556$$

$$P(\text{"awas"}|P) = \frac{(\text{"awas"}|"S) + 1}{(P) + |\text{kosakata}|} = \frac{0 + 1}{6 + 29} = 0,0285714285714286$$

$$P(\text{"awas"}|P) = \frac{(\text{"awas"}|"N) + 1}{(P) + |\text{kosakata}|} = \frac{0 + 1}{6 + 29} = 0,0285714285714286$$

Probabilitas kata “bakal”

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.

$$P("bakal"|P) = \frac{("bakal"|"R) + 1}{(P) + |kosakata|} = \frac{0 + 1}{5 + 29} = 0,0294117647058824$$

$$P("bakal"|P) = \frac{("bakal"|"E) + 1}{(P) + |kosakata|} = \frac{0 + 1}{6 + 29} = 0,0285714285714286$$

$$P("bakal"|P) = \frac{("bakal"|"P) + 1}{(P) + |kosakata|} = \frac{1 + 1}{7 + 29} = 0,0555555555555556$$

$$P("bakal"|P) = \frac{("bakal"|"S) + 1}{(P) + |kosakata|} = \frac{0 + 1}{6 + 29} = 0,0285714285714286$$

$$P("bakal"|P) = \frac{("bakal"|"N) + 1}{(P) + |kosakata|} = \frac{0 + 1}{6 + 29} = 0,0285714285714286$$

Probabilitas kata “bunuh”

$$P("bunuh"|P) = \frac{("bunuh"|"R) + 1}{(P) + |kosakata|} = \frac{0 + 1}{5 + 29} = 0,0294117647058824$$

$$P("bunuh"|P) = \frac{("bunuh"|"E) + 1}{(P) + |kosakata|} = \frac{0 + 1}{6 + 29} = 0,0285714285714286$$

$$P("bunuh"|P) = \frac{("bunuh"|"P) + 1}{(P) + |kosakata|} = \frac{1 + 1}{7 + 29} = 0,0555555555555556$$

$$P("bunuh"|P) = \frac{("bunuh"|"S) + 1}{(P) + |kosakata|} = \frac{0 + 1}{6 + 29} = 0,0285714285714286$$

$$P("bunuh"|P) = \frac{("bunuh"|"N) + 1}{(P) + |kosakata|} = \frac{0 + 1}{6 + 29} = 0,0285714285714286$$

Probabilitas kata “hidup”

$$P("hidup"|P) = \frac{("hidup"|"R) + 1}{(P) + |kosakata|} = \frac{0 + 1}{5 + 29} = 0,0294117647058824$$

$$P("hidup"|P) = \frac{("hidup"|"E) + 1}{(P) + |kosakata|} = \frac{0 + 1}{6 + 29} = 0,0285714285714286$$

$$P("hidup"|P) = \frac{("hidup"|"P) + 1}{(P) + |kosakata|} = \frac{1 + 1}{7 + 29} = 0,0555555555555556$$

$$P("hidup"|P) = \frac{("hidup"|"S) + 1}{(P) + |kosakata|} = \frac{0 + 1}{6 + 29} = 0,0285714285714286$$

$$P("hidup"|P) = \frac{("hidup"|"N) + 1}{(P) + |kosakata|} = \frac{0 + 1}{6 + 29} = 0,0285714285714286$$

Probabilitas kata “lonte”

$$P("lonte"|S) = \frac{("lonte"|"R) + 1}{(S) + |kosakata|} = \frac{0 + 1}{5 + 29} = 0,0294117647058824$$

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.

$$P(\text{"lonte"}|S) = \frac{(\text{"lonte"}|"E) + 1}{(S) + |\text{kosakata}|} = \frac{0 + 1}{6 + 29} = 0,0285714285714286$$

$$P(\text{"lonte"}|S) = \frac{(\text{"lonte"}|"P) + 1}{(S) + |\text{kosakata}|} = \frac{0 + 1}{7 + 29} = 0,0277777777777778$$

$$P(\text{"lonte"}|S) = \frac{(\text{"lonte"}|"S) + 1}{(S) + |\text{kosakata}|} = \frac{1 + 1}{6 + 29} = 0,0571428571428571$$

$$P(\text{"lonte"}|S) = \frac{(\text{"lonte"}|"N) + 1}{(S) + |\text{kosakata}|} = \frac{0 + 1}{6 + 29} = 0,0285714285714286$$

Probabilitas kata “lacur”

$$P(\text{"lacur"}|S) = \frac{(\text{"lacur"}|"R) + 1}{(S) + |\text{kosakata}|} = \frac{0 + 1}{5 + 29} = 0,0294117647058824$$

$$P(\text{"lacur"}|S) = \frac{(\text{"lacur"}|"E) + 1}{(S) + |\text{kosakata}|} = \frac{0 + 1}{6 + 29} = 0,0285714285714286$$

$$P(\text{"lacur"}|S) = \frac{(\text{"lacur"}|"P) + 1}{(S) + |\text{kosakata}|} = \frac{2 + 1}{7 + 29} = 0,0833333333333333$$

$$P(\text{"lacur"}|S) = \frac{(\text{"lacur"}|"S) + 1}{(S) + |\text{kosakata}|} = \frac{1 + 1}{6 + 29} = 0,0571428571428571$$

$$P(\text{"lacur"}|S) = \frac{(\text{"lacur"}|"N) + 1}{(S) + |\text{kosakata}|} = \frac{0 + 1}{6 + 29} = 0,0285714285714286$$

Probabilitas kata “bajing”

$$P(\text{"bajing"}|S) = \frac{(\text{"bajing"}|"R) + 1}{(S) + |\text{kosakata}|} = \frac{0 + 1}{5 + 29} = 0,0294117647058824$$

$$P(\text{"bajing"}|S) = \frac{(\text{"bajing"}|"E) + 1}{(S) + |\text{kosakata}|} = \frac{0 + 1}{6 + 29} = 0,0285714285714286$$

$$P(\text{"bajing"}|S) = \frac{(\text{"bajing"}|"P) + 1}{(S) + |\text{kosakata}|} = \frac{0 + 1}{7 + 29} = 0,0277777777777778$$

$$P(\text{"bajing"}|S) = \frac{(\text{"bajing"}|"S) + 1}{(S) + |\text{kosakata}|} = \frac{1 + 1}{6 + 29} = 0,0571428571428571$$

$$P(\text{"bajing"}|S) = \frac{(\text{"bajing"}|"N) + 1}{(S) + |\text{kosakata}|} = \frac{0 + 1}{6 + 29} = 0,0285714285714286$$

Probabilitas kata “nyata”

$$P(\text{"nyata"}|S) = \frac{(\text{"nyata"}|"R) + 1}{(S) + |\text{kosakata}|} = \frac{0 + 1}{5 + 29} = 0,0294117647058824$$

$$P(\text{"nyata"}|S) = \frac{(\text{"nyata"}|"E) + 1}{(S) + |\text{kosakata}|} = \frac{0 + 1}{6 + 29} = 0,0285714285714286$$

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.

$$P(\text{"nyata"}|S) = \frac{(\text{"nyata"}|"P) + 1}{(S) + |\text{kosakata}|} = \frac{0 + 1}{7 + 29} = 0,02777777777777778$$

$$P(\text{"nyata"}|S) = \frac{(\text{"nyata"}|"S) + 1}{(S) + |\text{kosakata}|} = \frac{1 + 1}{6 + 29} = 0,0571428571428571$$

$$P(\text{"nyata"}|S) = \frac{(\text{"nyata"}|"N) + 1}{(S) + |\text{kosakata}|} = \frac{0 + 1}{6 + 29} = 0,0285714285714286$$

Probabilitas kata “doyan”

$$P(\text{"doyan"}|S) = \frac{(\text{"doyan"}|"R) + 1}{(S) + |\text{kosakata}|} = \frac{0 + 1}{5 + 29} = 0,0294117647058824$$

$$P(\text{"doyan"}|S) = \frac{(\text{"doyan"}|"E) + 1}{(S) + |\text{kosakata}|} = \frac{0 + 1}{6 + 29} = 0,0285714285714286$$

$$P(\text{"doyan"}|S) = \frac{(\text{"doyan"}|"P) + 1}{(S) + |\text{kosakata}|} = \frac{0 + 1}{7 + 29} = 0,02777777777777778$$

$$P(\text{"doyan"}|S) = \frac{(\text{"doyan"}|"S) + 1}{(S) + |\text{kosakata}|} = \frac{1 + 1}{6 + 29} = 0,0571428571428571$$

$$P(\text{"doyan"}|S) = \frac{(\text{"doyan"}|"N) + 1}{(S) + |\text{kosakata}|} = \frac{0 + 1}{6 + 29} = 0,0285714285714286$$

Probabilitas kata “pacar”

$$P(\text{"pacar"}|S) = \frac{(\text{"pacar"}|"R) + 1}{(S) + |\text{kosakata}|} = \frac{0 + 1}{5 + 29} = 0,0294117647058824$$

$$P(\text{"pacar"}|S) = \frac{(\text{"pacar"}|"E) + 1}{(S) + |\text{kosakata}|} = \frac{0 + 1}{6 + 29} = 0,0285714285714286$$

$$P(\text{"pacar"}|S) = \frac{(\text{"pacar"}|"P) + 1}{(S) + |\text{kosakata}|} = \frac{0 + 1}{7 + 29} = 0,02777777777777778$$

$$P(\text{"pacar"}|S) = \frac{(\text{"pacar"}|"S) + 1}{(S) + |\text{kosakata}|} = \frac{1 + 1}{6 + 29} = 0,0571428571428571$$

$$P(\text{"pacar"}|S) = \frac{(\text{"pacar"}|"N) + 1}{(S) + |\text{kosakata}|} = \frac{0 + 1}{6 + 29} = 0,0285714285714286$$

Probabilitas kata “skandal”

$$P(\text{"skandal"}|N) = \frac{(\text{"skandal"}|"R) + 1}{(N) + |\text{kosakata}|} = \frac{0 + 1}{5 + 29} = 0,0294117647058824$$

$$P(\text{"skandal"}|N) = \frac{(\text{"skandal"}|"E) + 1}{(N) + |\text{kosakata}|} = \frac{0 + 1}{6 + 29} = 0,0285714285714286$$

$$P(\text{"skandal"}|N) = \frac{(\text{"skandal"}|"P) + 1}{(N) + |\text{kosakata}|} = \frac{0 + 1}{7 + 29} = 0,02777777777777778$$

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.

$$P(\text{"skandal"}|N = \frac{(\text{"skandal"}|"S) + 1}{(N) + |\text{kosakata}|} = \frac{0 + 1}{6 + 29} = 0,0285714285714286$$

$$P(\text{"skandal"}|N = \frac{(\text{"skandal"}|"N) + 1}{(N) + |\text{kosakata}|} = \frac{1 + 1}{6 + 29} = 0,0571428571428571$$

Probabilitas kata “seks”

$$P(\text{"seks"}|N = \frac{(\text{"seks"}|"R) + 1}{(N) + |\text{kosakata}|} = \frac{0 + 1}{5 + 29} = 0,0294117647058824$$

$$P(\text{"seks"}|N = \frac{(\text{"seks"}|"E) + 1}{(N) + |\text{kosakata}|} = \frac{0 + 1}{6 + 29} = 0,0285714285714286$$

$$P(\text{"seks"}|N = \frac{(\text{"seks"}|"P) + 1}{(N) + |\text{kosakata}|} = \frac{0 + 1}{7 + 29} = 0,0277777777777778$$

$$P(\text{"seks"}|N = \frac{(\text{"seks"}|"S) + 1}{(N) + |\text{kosakata}|} = \frac{0 + 1}{6 + 29} = 0,0285714285714286$$

$$P(\text{"seks"}|N = \frac{(\text{"seks"}|"N) + 1}{(N) + |\text{kosakata}|} = \frac{1 + 1}{6 + 29} = 0,0571428571428571$$

Probabilitas kata “sewa”

$$P(\text{"sewa"}|N = \frac{(\text{"sewa"}|"R) + 1}{(N) + |\text{kosakata}|} = \frac{0 + 1}{5 + 29} = 0,0294117647058824$$

$$P(\text{"sewa"}|N = \frac{(\text{"sewa"}|"E) + 1}{(N) + |\text{kosakata}|} = \frac{0 + 1}{6 + 29} = 0,0285714285714286$$

$$P(\text{"sewa"}|N = \frac{(\text{"sewa"}|"P) + 1}{(N) + |\text{kosakata}|} = \frac{0 + 1}{7 + 29} = 0,0277777777777778$$

$$P(\text{"sewa"}|N = \frac{(\text{"sewa"}|"S) + 1}{(N) + |\text{kosakata}|} = \frac{0 + 1}{6 + 29} = 0,0285714285714286$$

$$P(\text{"sewa"}|N = \frac{(\text{"sewa"}|"N) + 1}{(N) + |\text{kosakata}|} = \frac{1 + 1}{6 + 29} = 0,0571428571428571$$

Probabilitas kata “simak”

$$P(\text{"simak"}|N = \frac{(\text{"simak"}|"R) + 1}{(N) + |\text{kosakata}|} = \frac{0 + 1}{5 + 29} = 0,0294117647058824$$

$$P(\text{"simak"}|N = \frac{(\text{"simak"}|"E) + 1}{(N) + |\text{kosakata}|} = \frac{0 + 1}{6 + 29} = 0,0285714285714286$$

$$P(\text{"simak"}|N = \frac{(\text{"simak"}|"P) + 1}{(N) + |\text{kosakata}|} = \frac{0 + 1}{7 + 29} = 0,0277777777777778$$

$$P(\text{"simak"}|N = \frac{(\text{"simak"}|"S) + 1}{(N) + |\text{kosakata}|} = \frac{0 + 1}{6 + 29} = 0,0285714285714286$$

$$P(\text{"simak"}|N) = \frac{(\text{"simak"}|N) + 1}{(N) + |\text{kosakata}|} = \frac{1 + 1}{6 + 29} = 0,0571428571428571$$

Probabilitas kata “lengkap”

$$P(\text{"lengkap"}|N) = \frac{(\text{"lengkap"}|R) + 1}{(N) + |\text{kosakata}|} = \frac{0 + 1}{5 + 29} = 0,0294117647058824$$

$$P(\text{"lengkap"}|N) = \frac{(\text{"lengkap"}|E) + 1}{(N) + |\text{kosakata}|} = \frac{0 + 1}{6 + 29} = 0,0285714285714286$$

$$P(\text{"lengkap"}|N) = \frac{(\text{"lengkap"}|P) + 1}{(N) + |\text{kosakata}|} = \frac{0 + 1}{7 + 29} = 0,0277777777777778$$

$$P(\text{"lengkap"}|N) = \frac{(\text{"lengkap"}|S) + 1}{(N) + |\text{kosakata}|} = \frac{0 + 1}{6 + 29} = 0,0285714285714286$$

$$P(\text{"lengkap"}|N) = \frac{(\text{"lengkap"}|N) + 1}{(N) + |\text{kosakata}|} = \frac{1 + 1}{6 + 29} = 0,0571428571428571$$

2. Tahap *testing*

Pada tahap ini dilakukan pengujian dengan memasukkan data uji ke dalam model yang telah dibentuk pada tahap *training* tadi. Pada Tabel 4.9 ada sebuah data yang akan dijadikan data uji.

Tabel 4. 9 Contoh Data *Testing*

| <i>Tweet</i> | Kelas |
|--|--------------|
| Kalo ngomongin seputaran orang 'pribumi', keknya orang yg berkoar-koar dia 'pribumi' harus ditanya "emang lo masih... https://t.co/BPVHmjTaLL | Racist |

Berdasarkan data uji dari Tabel 4.9 diatas, maka akan dihitung probabilitasnya berdasarkan nilai probabilitas setiap *term*. Perhitungan probabilitas dengan metode *Naïve Bayes Classifier* pada dokumen *testing* ini dilakukan dengan melakukan perkalian probabilitas semua kategori dengan probabilitas tiap *term* yang diambil dari semua data.

Sebelum masuk ketahapan *testing*, terlebih dahulu data uji akan melewati tahapan *text preprocessing*. Hasil data uji setelah *text preprocessing* dapat dilihat pada Tabel 4.10.

Tabel 4. 10 Hasil *Preprocessing Data Testing*

| <i>Tweet</i> | Kelas |
|---|---------------|
| putar orang pribumi kayak orang pribumi emang | <i>Racist</i> |

Selanjutnya menghitung nilai probabilitas yang akan dilakukan pengujian pada tiap-tiap kelas dalam data uji dengan menggunakan hasil probabilitas *term* yang diperoleh pada data latih dimana data latih dan data uji sudah melakukan tahap *text preprocessing* seperti Tabel 4.11 berikut:

Tabel 4. 11 Contoh Tabel Data Latih dan Data Uji

| Data Latih (n) | <i>Tweet</i> | Kelas |
|----------------|--|------------------------------|
| 1 | yang pasti zumi zola yang digoreng sudah muslim pribumi pulak makanya kalao mau aman korupsi ya harus non pribum | <i>Racist</i> |
| 2 | saka agari ga butuh orang kaya lu rame nya kampungan norak dan ngerusak hajatan roba ga guna kzl bgt gue | <i>Embarrass</i> |
| 3 | tidak semua teman itu baik dan tidak semua musuh itu jahat jagain pinutnya gua ya awas klo disakitin klo lu nyakitin dia bakal gua bunuh idup wkwk | <i>Physically threatened</i> |
| 4 | ooh ngatain aku lonte pelacur bajingan tp ternyata doyan jg bisa nerima cewe itu sbg pacar yg dr awal sdh | <i>Sexually harassed</i> |
| 5 | skandal seks dan sewa pelacur di haiti oxfam jadi bulan bulanan simak selengkapnya di sini | <i>Neutral</i> |
| Data Uji | | |
| 6 | korupsi super kakap etnis cina biar asing kondensat biar kabur | ? |

Probabilitas tiap-tiap kata pada kelas dilakukan menggunakan persamaan 2.2

$$P(w_k|r/e/p/s/n) = \frac{(n_k, R/E/P/S/N) + 1}{(n, R/E/P/S/N) + |\text{kosakata}|}$$

Diketahui |kosakata|= 29

count *Racist*=5, count *embarrass*=6, count *physically threatened* =7, count *sexually harassed*=6, count *neutral*=6

Kata data uji “korupsi super kakap etnis cina biar asing kondensat biar kabur”

$$P(\text{korupsi}|Racist) = (1+1)/(5+29) = 0,0588235294117647$$

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 mengemukakan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

| | |
|-------------------------------------|---------------------------------------|
| P(super Racist) | $= (0+1)/(5+29) = 0,0294117647058824$ |
| P(kakap Racist) | $= (0+1)/(5+29) = 0,0294117647058824$ |
| P(etnis Racist) | $= (0+1)/(5+29) = 0,0294117647058824$ |
| P(cina Racist) | $= (0+1)/(5+29) = 0,0294117647058824$ |
| P(asing Racist) | $= (0+1)/(5+29) = 0,0294117647058824$ |
| P(kondensat Racist) | $= (0+1)/(5+29) = 0,0294117647058824$ |
| P(biar Racist) | $= (0+1)/(5+29) = 0,0294117647058824$ |
| P(kabur Racist) | $= (0+1)/(5+29) = 0,0294117647058824$ |
| P(korupsi Embarrass) | $= (0+1)/(6+29) = 0,0285714285714286$ |
| P(super Embarrass) | $= (0+1)/(6+29) = 0,0285714285714286$ |
| P(kakap Embarrass) | $= (0+1)/(6+29) = 0,0285714285714286$ |
| P(etnis Embarrass) | $= (0+1)/(6+29) = 0,0285714285714286$ |
| P(cina Embarrass) | $= (0+1)/(6+29) = 0,0285714285714286$ |
| P(asing Embarrass) | $= (0+1)/(6+29) = 0,0285714285714286$ |
| P(kondensat Embarrass) | $= (0+1)/(6+29) = 0,0285714285714286$ |
| P(biar Embarrass) | $= (0+1)/(6+29) = 0,0285714285714286$ |
| P(kabur Embarrass) | $= (0+1)/(6+29) = 0,0285714285714286$ |
| P(korupsi Physically Threatened) | $= (0+1)/(7+29) = 0,0277777777777778$ |
| P(super Physically Threatened) | $= (0+1)/(7+29) = 0,0277777777777778$ |
| P(kakap Physically Threatened) | $= (0+1)/(7+29) = 0,0277777777777778$ |
| P(etnis Physically Threatened) | $= (0+1)/(7+29) = 0,0277777777777778$ |
| P(cina Physically Threatened) | $= (0+1)/(7+29) = 0,0277777777777778$ |
| P(asing Physically Threatened) | $= (0+1)/(7+29) = 0,0277777777777778$ |
| P(kondensat Physically Threatened) | $= (0+1)/(7+29) = 0,0277777777777778$ |
| P(biar Physically Threatened) | $= (0+1)/(7+29) = 0,0277777777777778$ |
| P(kabur Physically Threatened) | $= (0+1)/(7+29) = 0,0277777777777778$ |
| P(korupsi Sexually Harassed) | $= (0+1)/(6+29) = 0,0285714285714286$ |
| P(super Sexually Harassed) | $= (0+1)/(6+29) = 0,0285714285714286$ |

| | |
|---------------------------------|--------------------------------------|
| P(kakap Sexually Harassed) | = $(0+1)/(6+29)= 0,0285714285714286$ |
| P(etnis Sexually Harassed) | = $(0+1)/(6+29)= 0,0285714285714286$ |
| P(cina Sexually Harassed) | = $(0+1)/(6+29)= 0,0285714285714286$ |
| P(asing Sexually Harassed) | = $(0+1)/(6+29)= 0,0285714285714286$ |
| P(kondensat Sexually Harassed) | = $(0+1)/(6+29)= 0,0285714285714286$ |
| P(biar Sexually Harassed) | = $(0+1)/(6+29)= 0,0285714285714286$ |
| P(kabur Sexually Harassed) | = $(0+1)/(6+29)= 0,0285714285714286$ |
| P(korupsi Neutral) | = $(0+1)/(6+29)= 0,0285714285714286$ |
| P(super Neutral) | = $(0+1)/(6+29)= 0,0285714285714286$ |
| P(kakap Neutral) | = $(0+1)/(6+29)= 0,0285714285714286$ |
| P(etnis Neutral) | = $(0+1)/(6+29)= 0,0285714285714286$ |
| P(cina Neutral) | = $(0+1)/(6+29)= 0,0285714285714286$ |
| P(asing Neutral) | = $(0+1)/(6+29)= 0,0285714285714286$ |
| P(kondensat Neutral) | = $(0+1)/(6+29)= 0,0285714285714286$ |
| P(biar Neutral) | = $(0+1)/(6+29)= 0,0285714285714286$ |
| P(kabur Neutral) | = $(0+1)/(6+29)= 0,0285714285714286$ |

Tahapan selanjutnya dilakukan perhitungan probabilitas kelas dengan menggunakan rumus persamaan 2.3

$$P(R/E/P/S/N | d) = P(R/E/P/S/N) * \prod_i P(a_i | R/E/P/S/N)$$

Keterangan :

$P(R/E/P/S/N | d)$ = Peluang kemunculan kata-kata pada sebuah kelas/kategori dengan dokumen data uji, d adalah dokumen data baru

$\prod_i P(a_i | R/E/P/S/N)$ = Peluang kemunculan a_i pada kelas/kategori, a_i adalah kata-kata baru yang akan dilakukuan pengujian.

$$P(\text{testing}|R) = P(R) \times P(\text{korupsi}|R) \times P(\text{super}|R) \times P(\text{kakap}|R) \times P(\text{etnis}|R) \times P(\text{cina}|R) \times P(\text{asing}|R) \times P(\text{kondensat}|R) \times P(\text{biar}|R) \times P(\text{kabur}|R)$$

$$P(\text{testing}|R) = 0.2 \times 0,0588235294117647 \times 0,0294117647058824 \times 0,0294117647058824 \times 0,0294117647058824 \times 0,0294117647058824 \times 0,0294117647058824 \times 0,0294117647058824 \times 0,0294117647058824$$

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.

$$\begin{aligned}
 &= 8,8037E-08 \\
 P(\text{testing}|E) &= P(E) \times (P(\text{korupsi}|E) \times P(\text{super}|E) \times P(\text{kakap}|E) \times \\
 &\quad P(\text{etnis}|E) \times P(\text{cina}|E) \times P(\text{asing}|E) \times P(\text{kondensat}|E) \times \\
 &\quad P(\text{biar}|E) \times P(\text{kabur}|E)) \\
 P(\text{testing}|E) &= 0.2 \times 0,0285714285714286 \times 0,0285714285714286 \times \\
 &\quad 0,0285714285714286 \times 0,0285714285714286 \times \\
 &\quad 0,0285714285714286 \times 0,0285714285714286 \\
 &= 1,08798E-09 \\
 P(\text{testing}|P) &= P(P) \times (P(\text{korupsi}|P) \times P(\text{super}|P) \times P(\text{kakap}|P) \times \\
 &\quad P(\text{etnis}|P) \times P(\text{cina}|P) \times P(\text{asing}|P) \times P(\text{kondensat}|P) \times \\
 &\quad P(\text{biar}|P) \times P(\text{kabur}|P)) \\
 P(\text{testing}|P) &= 0.2 \times 0,0277777777777778 \times 0,0277777777777778 \times \\
 &\quad 0,0277777777777778 \times 0,0277777777777778 \times \\
 &\quad 0,0277777777777778 \times 0,0277777777777778 \times \\
 &\quad 0,0277777777777778 \\
 &= 2,55219E-11 \\
 P(\text{testing}|S) &= P(S) \times (P(\text{korupsi}|S) \times P(\text{super}|S) \times P(\text{kakap}|S) \times \\
 &\quad P(\text{etnis}|S) \times P(\text{cina}|S) \times P(\text{asing}|S) \times P(\text{kondensat}|S) \times \\
 &\quad P(\text{biar}|S) \times P(\text{kabur}|S)) \\
 P(\text{testing}|S) &= 0.2 \times 0,0285714285714286 \times 0,0285714285714286 \times \\
 &\quad 0,0285714285714286 \times 0,0285714285714286 \times \\
 &\quad 0,0285714285714286 \times 0,0285714285714286 \\
 &= 1,08798E-09 \\
 P(\text{testing}|N) &= P(N) \times (P(\text{korupsi}|N) \times P(\text{super}|N) \times P(\text{kakap}|N) \times \\
 &\quad P(\text{etnis}|N) \times P(\text{cina}|N) \times P(\text{asing}|N) \times P(\text{kondensat}|N) \times \\
 &\quad P(\text{biar}|N) \times P(\text{kabur}|N)) \\
 P(\text{testing}|N) &= 0.2 \times 0,0285714285714286 \times 0,0285714285714286 \times \\
 &\quad 0,0285714285714286 \times 0,0285714285714286 \times \\
 &\quad 0,0285714285714286 \times 0,0285714285714286 \\
 &= 1,08798E-09
 \end{aligned}$$

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.

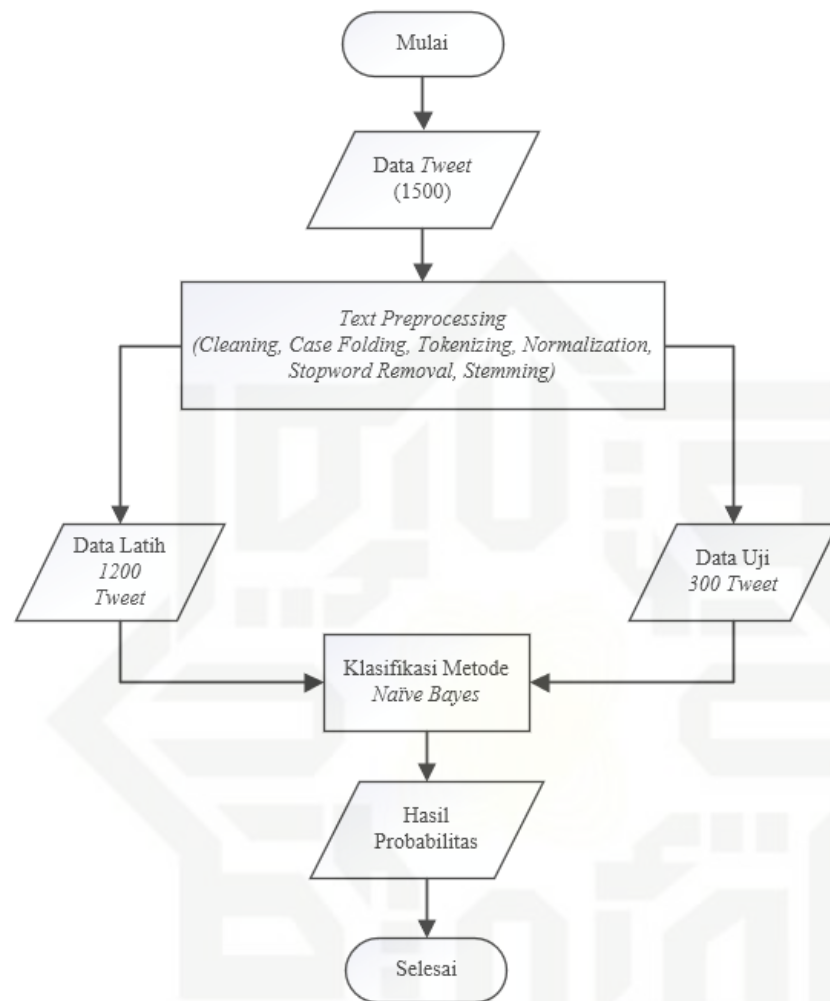
Nilai probabilitas tertinggi adalah kategori *Racist* yaitu sebesar 8,8037E-08 sehingga *tweet* tersebut diklasifikasikan ke dalam pelecehan kelas *Racist*.

1.7 Analisa Rancangan

Analisa fungsional sistem akan menjelaskan mengenai rancangan sistem yang akan dibuat menggunakan *flowchart*. Sedangkan perancangan sistem analisa klasifikasi *tweet* pelecehan *online* pada Twitter dirancang berdasarkan tahapan analisa yang telah dilakukan sebelumnya.

1.7.1 Rancangan *Flowchart*

Flowchart digunakan untuk menggambarkan secara keseluruhan dari proses suatu sistem dan juga untuk merencanakan tahapan suatu kegiatan. Dalam penerapan *machine learning*, penggunaan *flowchart* merupakan alat yang tepat untuk menggambarkan *machine learning* yang akan dibuat. Adapun *flowchart* klasifikasi *tweet* pelecehan *online* pada Twitter dapat dilihat pada Gambar 4.7 sebagai berikut:



Gambar 4. 7 Rancangan *Flowchart* Klasifikasi Pelecehan Online

1.7.2 Rancangan Basis Data

Tujuan dari perancangan stuktur ini yaitu untuk menentukan nama *field*, *type field*, lebar *field*, dan keterangan dari *field* tersebut yang ada pada setiap *file*. Basis data pada sistem ini berisikan satu tabel digunakan untuk menampung *tweet* yang akan digunakan. Rancangan basis data dapat dilihat pada Tabel 4.12 berikut :

Tabel 4. 12 Rancangan Tabel *Tweet*

| Nama <i>field</i> | <i>Type</i> dan <i>length</i> | <i>Primary Key</i> | <i>Null</i> | <i>Default</i> |
|-------------------|-------------------------------|--------------------|-------------|----------------|
| id | int (50) | <i>Primary key</i> | NO | - |
| user_name | varchar (150) | - | NO | - |
| screen_name | varchar (150) | - | NO | - |

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.

| Nama field | Type dan lengt | Primary Key | Null | Default |
|-------------------|-----------------------|--------------------|-------------|----------------|
| tweet | varchar (255) | - | NO | - |
| tweetid | bigint (20) | - | NO | - |
| lang | varchar (10) | - | YA | - |
| location | varchar (255) | - | YA | - |
| date_tweeted | datetime | - | NO | - |
| preprocess | enum | - | NO | - |
| class | varchar (50) | - | NO | - |

