

## BAB III

### METODE PENELITIAN

#### 3.1. Metode Pengumpulan Data

Data yang digunakan pada penelitian ini merupakan data sentimen dari pengguna aplikasi android yang memberikan komentarnya pada fasilitas *user review* pada website googleplay yang kemudian dipisahkan menjadi sentimen positif dan sentimen negatif. Sentimen positif yaitu sentimen dimana pengguna memberikan komentarnya dengan memberikan penilaian secara positif atau memberikan keterangan bahwa aplikasi ini baik atau mempunyai kepuasan untuk menginstal aplikasi tersebut. Sedangkan sentimen negatif merupakan sentimen sebaliknya dari sentimen positif yaitu pengguna memberikan komentarnya dengan memberikan ketidakpuasan terhadap aplikasi tersebut.

Untuk memperoleh data tersebut peneliti mengambil data dari *user review* di <https://play.google.com> pada bulan Desember 2012 – Januari 2013 dengan memilih secara acak aplikasi yang memiliki review dari user dengan bahasa Indonesia baik aplikasi berbayar maupun aplikasi gratis. Data yang diambil berjumlah 2.000 data sentimen dengan 1.000 data sentimen positif dan 1.000 data sentimen negatif.

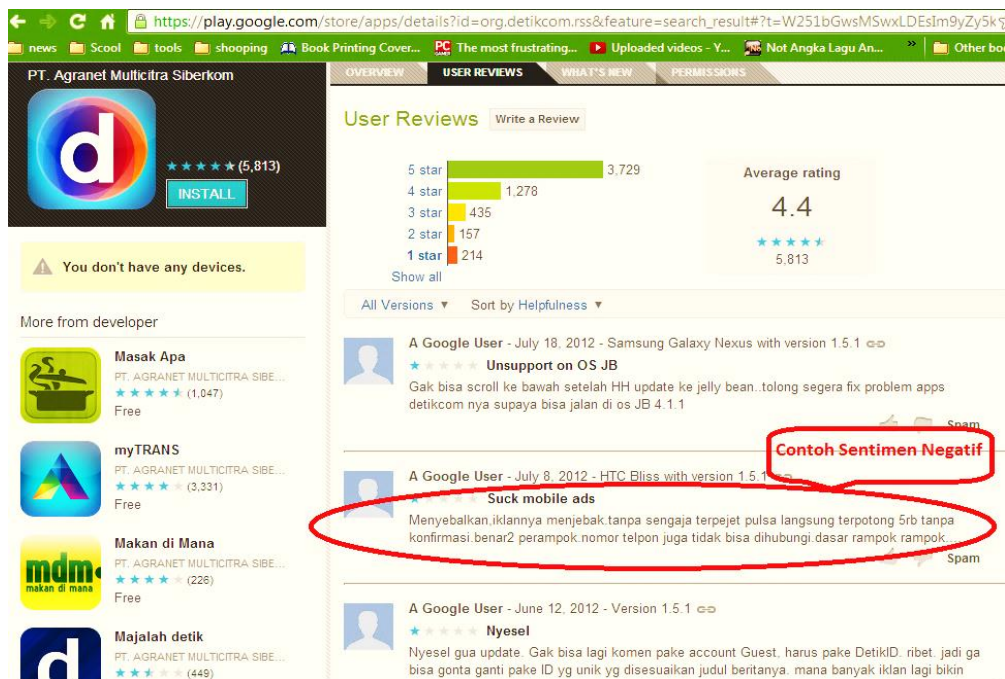
The screenshot shows the Google Play Store interface for the 'detikcom' app. The app is developed by PT. Agranet Multicitra Siberkom and has a 4.4 average rating based on 5,813 reviews. The 'User Reviews' section is visible, showing a bar chart of star ratings and a list of reviews. A red circle highlights a specific review from 'A Google User' dated September 6, 2012, with a 5-star rating. The review text is: 'Lebih enak liat di app ini dibanding di twitter. Appnya ga gitu berat dan ga muluk2. Sederhana tapi memuaskan. Excellent!'. A red box labeled 'Contoh Sentimen Positif' points to this review.

Star Rating	Count
5 star	3,729
4 star	1,278
3 star	435
2 star	157
1 star	214

Average rating: 4.4

Example Review: **5 stars** Mantap  
Lebih enak liat di app ini dibanding di twitter. Appnya ga gitu berat dan ga muluk2. Sederhana tapi memuaskan. Excellent!

Gambar 3. 1 Contoh sentimen positif berbahasa indonesia



Gambar 3. 2 Contoh sentimen negatif berbahasa indonesia

### 3.2. Metode Pemrosesan Awal Data

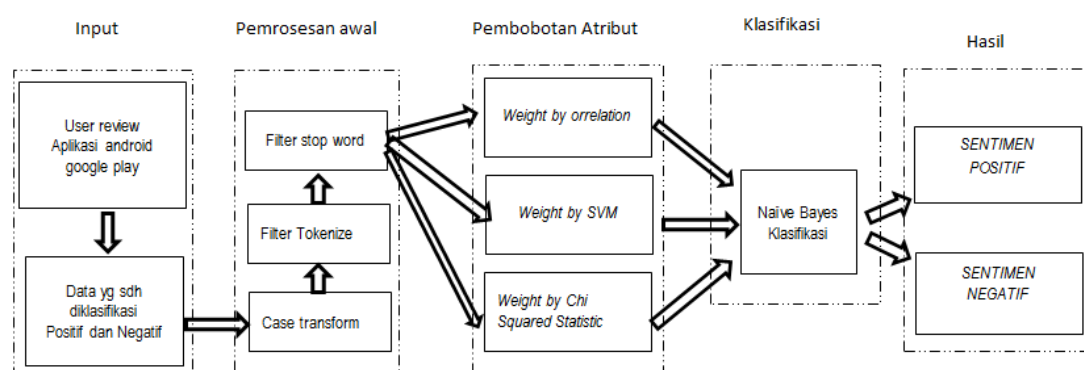
Guna memudahkan dalam mengelola data maka data perlu kita berikan analisa sentimen secara manual dengan membaca maksud dari kalimat yang ada dalam sentimen tersebut, sehingga dapat diberikan penilaian bahwa sentimen tersebut merupakan setimen negatif atau positif.

Setelah itu data kita pisahkan antara sentimen positif dan negatif masing masing kedalam folder yang berbeda, yaitu untuk sentimen positif kita masukkan ke folder positif sedangkan sentimen negatif kita masukkan ke folder negatif hal ini dilakukan guna memudahkan saat pengambilan, penambahan, pengurangan maupun saat pemrosesan terhadap data tersebut.

Seperti telah dijelaskan sebelumnya maka masing – masing folder positif dan negatif berisi 1.000 file berformat \*.txt yang berisi sentimen yang terhadap aplikasi android berbahasa Indonesia yang sudah dipisahkan atara positif dan negatif sesuai foldernya masing – masing.

### 3.3. Model Yang Diusulkan

Pada penelitian ini peneliti mengusulkan suatu metode dalam analisis sentimen guna mengetahui sentimen user dalam aplikasi android, dengan menggunakan metode klasifikasi naïve bayes serta menggunakan pemrosesan data awal dengan beberapa tahapan diantaranya ekstraksi teks dan melakukan pengujian perbandingan unjuk kerja dari metode pembobotan *weight by correlation*, *weight by chi squared statistic* dan *weight by SVM* terhadap akurasi klasifikasi sentimen pengguna aplikasi android dengan metode klasifikasi naïve bayes .



Gambar 3. 3 Model Yang Diusulkan

### 3.4. Experimen Dan Pengujian

*Tools* yang digunakan dalam experimen ini adalah Rapidminer yang di update dengan penambahan *plugin text processing* yang telah memiliki fitur pemrosesan teks diataranya:

#### 3.4.1. Transform Cases

Dengan fitur *transform cases* kita dapat secara otomatis mengubah semua huruf pada teks menjadi huruf kecil semua atau menjadi huruf kapital semua, pada penelitian ini semua huruf dirubah kedalam huruf kecil karena mayoritas teks berupa tulisan opini yang sebagian besar merupakan huruf kecil semua.

### 3.4.2. Filter Stop Word (Indonesia)

Dengan fitur ini maka teks sebelum di klasifikasikan di hilangkan dulu teks yang tidak berhubungan dengan analisa sentimen sehingga dimensi teks akan berkurang tanpa mengurangi isi sentimen dari teks tersebut.

Fiter stopword bahasa indonesia ini penulis ambil dari internet yang dibuat oleh Wang Pidong seorang Ph.D dari National University Singapore dengan penulis menambahkan beberapa kata yang memiliki arti sama dengan kata – kata yang sudah ada dalam daftar stopwords tersebut.

### 3.4.3. Filter Tokenize

Tahapan ini juga menghilangkan karakter-karakter tertentu seperti tanda baca serta memfilter berdasarkan panjang teks.

Untuk metode pengujian atau evaluasi dilakukan pengujian terhadap model - model yang diteliti untuk mendapatkan informasi model diusulkan. Evaluasi dan validasi menggunakan metode sebagai berikut:

### 3.4.4. Mengubah Teks Menjadi Matrix

Untuk dapat diolah atau di proses maka data awal yang berupa kalimat setelah dilakukan pemrosesan awal data akan menjadi suatu atribut berupa teks, dan untuk dapat diterapkan kedalam algoritma selanjutnya maka dibutuhkan adanya transformasi data dari teks kedalam sebuah matrix yang berisi numerik.

Pada penelitian ini penulis menggunakan proses pembentukan vector kalimat dengan menggunakan TF-IDF (*term frequency-inverse document frequency*) Matrix yang dirumuskan sebagai berikut:

$$TF-IDF = TF * IDF = TF * \log(n/df)$$

Dimana: tf = frekuensi teks  
df = frekuensi dokument  
n = jumlah dokumen

contoh perhitungan matriks TF-IDF

jika ada tabel atribut sebagai berikut :

Tabel 3. 1 Tabel Atribut contoh Perhitungan TF-IDF

	<b>Dok 1</b>	<b>Dok 2</b>	<b>Dok 3</b>	<b>df</b>
Aplikasi	6	0	7	2
Bagus	9	2	3	3
Buruk	10	4	0	2

Maka perhitungna matrix TF-IDF nya

Tabel 3. 2 Tabel Perhitungan TF-IDF

	<b>Dok 1</b>	<b>Dok 2</b>	<b>Dok 3</b>
Aplikasi	$6 \cdot \log(3/2)$	0	$7 \cdot \log(3/2)$
Bagus	$9 \cdot \log(3/3)$	$2 \cdot \log(3/3)$	$3 \cdot \log(3/3)$
Buruk	$10 \cdot \log(3/2)$	$4 \cdot \log(3/2)$	0

Dan hasil matrix TF-IDF nya adalah sebagai berikut

Tabel 3. 3 Tabel Matrix TF-IDF

	<b>Dok 1</b>	<b>Dok 2</b>	<b>Dok 3</b>
Aplikasi	1.06	0.00	1.23
Bagus	0.00	0.00	0.00
Buruk	1.76	0.70	0.00

### 3.4.5. Pemilihan Atribut

Pemilihan atribut dilakukan sebelum proses klasifikasi terhadap sentimen atau *dataset*, dimaksudkan untuk mengurangi banyaknya jumlah atribut yang ada pada sentimen dengan tanpa mengurangi isi dari sentimen tersebut. Pada eksperimen ini penulis mencoba mengambil beberapa kombinasi pengambilan atribut dengan didasarkan dari adanya pembobotan yang dilakukan terhadap atribut tersebut, dengan menggunakan parameter mengambil sebanyak p% dari jumlah tribut diambil dari

atribut yang mempunyai bobot lebih tinggi menjadi prioritas untuk diklasifikasikan kedalam metode naïve bayes.

Dalam penelitian ini penulis melakukan beberapa kali kombinasi berbagai variasi pemilihan atribut berdasarkan bobot atribut tersebut diantaranya:

- a.  $p = 5\%$  atribut diambil sebanyak 5% dari total atribut
- b.  $p = 10\%$  atribut diambil sebanyak 10% dari total atribut
- c.  $p = 15\%$  atribut diambil sebanyak 15% dari total atribut
- d.  $p = 20\%$  atribut diambil sebanyak 20% dari total atribut
- e.  $p = 25\%$  atribut diambil sebanyak 25% dari total atribut
- f.  $p = 30\%$  atribut diambil sebanyak 30% dari total atribut
- g.  $p = 35\%$  atribut diambil sebanyak 35% dari total atribut
- h.  $p = 40\%$  atribut diambil sebanyak 40% dari total atribut
- i.  $p = 45\%$  atribut diambil sebanyak 45% dari total atribut
- j.  $p = 50\%$  atribut diambil sebanyak 50% dari total atribut
- k.  $p = 55\%$  atribut diambil sebanyak 55% dari total atribut
- l.  $p = 60\%$  atribut diambil sebanyak 60% dari total atribut
- m.  $p = 65\%$  atribut diambil sebanyak 65% dari total atribut
- n.  $p = 70\%$  atribut diambil sebanyak 70% dari total atribut
- o.  $p = 75\%$  atribut diambil sebanyak 75% dari total atribut
- p.  $p = 80\%$  atribut diambil sebanyak 80% dari total atribut
- q.  $p = 85\%$  atribut diambil sebanyak 85% dari total atribut
- r.  $p = 90\%$  atribut diambil sebanyak 90% dari total atribut
- s.  $p = 95\%$  atribut diambil sebanyak 95% dari total atribut
- t.  $p = 100\%$  atribut diambil sebanyak 100% dari total atribut

dengan pengambilan atribut berdasarkan bobot yang diperoleh dari masing – masing pembobotan secara urut dari tingkat bobot paling tinggi menuju bobot paling rendah.

#### **3.4.6. Klasifikasi Sentimen**

Pengkalsifikasian data dilakukan terhadap *dataset* berdasarkan atribut yang dimiliki oleh *dataset* review pengguna aplikasi android tersebut. Pada penelitian ini penulis melakukan berberapa kali eksperimen guna mengetahui pebedaan hasil klasifikasi terhadap metode pembobotan yang dilakukan terhadap atribut dari *dataset* dengan menggunakan algoritma naïve bayes.

Pada tahapan pertama dilakukan eksperimen tanpa memperhatikan pembobotan atribut, dengan memproses semua atribut kedalam algoritma naïve bayes. Kemudian dilakukan pengujian dengan mempertimbangkan pemilihan atribut berdasarkan pemilihan atribut sesuai variasi pemilihan atribut dari  $p = 5\%$  sampai dengan  $p =$

100% yang semuanya dilakukan untuk semua pembobotan atribut dan dilakukan dengan metode kalsifikasi yang sama yaitu naïve bayes.

**3.4.7. Cross validation**

*Cross validation* adalah pengujian standar yang dilakukan untuk memprediksi *error rate*. Data *training* dibagi secara acak ke dalam beberapa bagian dengan perbandingan yang sama kemudian *error rate* dihitung bagian demi bagian, selanjutnya hitung rata-rata seluruh *error rate* untuk mendapatkan *error rate* secara keseluruhan.

Pada penelitian ini penulis menggunakan metode pengujian *10 fold cross validation* yang akan mengulang pengujian sebanyak 10 kali dan hasil pengukuran adalah nilai rata-rata dari 10 kali pengujian, karena hasil dari berbagai percobaan yang ekstensif dan pembuktian teoritis, menunjukkan bahwa *10 fold cross validation* adalah pilihan terbaik untuk mendapatkan hasil validasi yang akurat.

Uji ke	Dataset									
1	█									
2		█								
3			█							
4				█						
5					█					
6						█				
7							█			
8								█		
9									█	
10										█

Gambar 3. 4 skema 10 fold cross validation

Hitam = k-subset (data testing)  
 Putih = data training

**3.4.8. Confusion matrix**

Metode ini menggunakan tabel matriks seperti pada Tabel 3.1 , jika data set hanya terdiri dari dua kelas, kelas yang satu dianggap sebagai positif dan yang lainnya negatif [15]

Tabel 3. 4 Model Confusion Matrix

Kenyataan	Diprediksi sebagai	
	+	-
+	True positives	False positives
-	False negatives	true negatives

*True positives* adalah jumlah *record* positif yang diklasifikasikan sebagai positif, *false positives* adalah jumlah *record* negatif yang diklasifikasikan sebagai positif, *false negatives* adalah jumlah *record* positif yang diklasifikasikan sebagai negatif, *true negatives* adalah jumlah *record* negatif yang diklasifikasikan sebagai negative, kemudian masukkan data uji. Setelah data uji dimasukkan ke dalam *confusion matrix*, hitung nilai - nilai yang telah dimasukkan tersebut untuk dihitung jumlah *sensitivity (recall)*, *specificity*, *precision* dan *accuracy*. *Sensitivity* digunakan untuk membandingkan jumlah *true positives* terhadap jumlah *record* yang positif sedangkan *specificity* adalah perbandingan jumlah *true negatives* terhadap jumlah *record* yang negatif. Untuk menghitung akurasi digunakan persamaan di bawah [16]:

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \quad (7)$$

Keterangan:

TP = jumlah *true positives*

TN = jumlah *true negatives*

FP = jumlah *false positives*

FN = jumlah *false negatives*

*Precision* adalah rasio jumlah parameter yang benar yang termasuk ke dalam *true positive* dengan jumlah parameter yang benar yang termasuk ke dalam *true positive* dan jumlah parameter yang salah tetapi benar (*false positive*). *Recall rate* adalah rasio jumlah parameter yang benar *true positive* dengan jumlah parameter yang benar *true positive* dan jumlah parameter yang *false negativ*).

Adapun rumus untuk menghitung *precision* dan *recall*, sebagai berikut:

$$Precision = \frac{TP}{TP+FP} \quad (8)$$

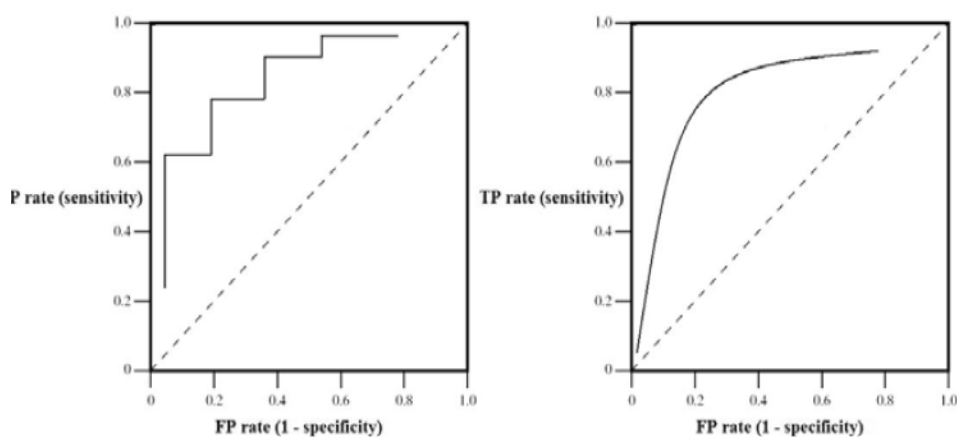
$$Recall = \frac{TP}{TP+FN} \quad (9)$$



Nilai ROC (*Receiver Operating Characteristics*) Curve sering digunakan untuk menilai hasil dari prediksi berupa grafik. Berikut penjelasan singkat tentang ROC Kurva ROC pertama kali dikembangkan oleh teknik elektro dan radar dalam perang dunia II untuk mendeteksi objek musuh dalam benteng pertahanan [7] (contohnya: cerita pearl harbor yang diserang tahun 1941, atau masalah operator penerima radar di inggris)

- ROC sudah lama digunakan dalam teori deteksi sinyal.
- ROC sering digunakan dalam penelitian kesehatan.
- ROC juga sering digunakan dalam penelitian *machine learning* dan *data mining* (salah satu pendahulu yang menggunakan ROC dalam *machine learning* adalah Spackman, seseorang yang mendemonstrasikan nilai dari kurva ROC dalam evaluasi dan perbandingan algoritma)
- Dalam permasalahan klasifikasi, ROC merupakan dasar kinerja dari teknik untuk visualisasi, pengorganisasian dan pemilihan klasifikasi

Dalam permasalahan klasifikasi digunakan keputusan 2 class (klasifikasi biner), salah satu objek digambarkan satu elemen yang saling berpasangan yaitu positive atau negative. Model klasifikasi yang lain yaitu dengan penamaan label pada class. Kurva ROC juga dikenal sebagai grafik ROC yaitu grafik 2 dimensi yang mana TP rate adalah plot untuk Y-axis dan FP rate adalah plot untuk X-axis. Grafik ROC (discrete dan continuous) digambarkan dibawah ini :



Gambar 3. 5 Grafik ROC (Discrete dan Continuous)

Pada ruang ROC dipisah oleh garis diagonal hasil ROC dikategorikan ke dalam “*good classification*” jika poin berada di atas garis diagonal begitu juga sebaliknya

dikategorikan “*poor classification*” jika poin berada di bawah garis diagonal. Dapat disimpulkan bahwa satu poin dalam ruang ROC lebih baik daripada jika satu poin ke utara-selatan dari persegi (jika TP rate lebih tinggi dan FP rate lebih rendah atau kedua-duanya). Kurva ROC adalah alat dua dimensi yang digunakan untuk menilai kinerja klasifikasi. ROC sering digunakan untuk perbandingan model klasifikasi. Berikut kategori klasifikasi untuk mencari akurasi dengan menggunakan AUC (*Area Under Curve*) :

Tabel 3. 5 Kategori Klasifikasi dengan menggunakan hasil AUC

<b>Nilai AUC</b>	<b>Kategori</b>
0.90 - 1.00	excellent classification
0.80 - 0.90	good classification
0.70 - 0.80	fair classification
0.60 - 0.70	poor classification
0.50 - 0.60	failure classification