

## **BAB III**

### **METODOLOGI PENELITIAN**

#### **3.1 Jenis Penelitian**

Jenis penelitian yang dilaksanakan merupakan penelitian kuantitatif, yaitu penelitian yang pengumpulan datanya melalui pengambilan komentar-komentar oleh seseorang terhadap film di youtube yang di simpan di notepad. Selain itu juga mengambil dari data komentar yang sudah dikumpulkan oleh peneliti sebelumnya. Kemudian data teks tersebut diolah menjadi bentuk angka sehingga dapat diolah.

#### **3.2 Instrumen Penelitian**

Dalam penelitian ini, dibutuhkan beberapa komponen peralatan, yaitu:

##### **3.2.1 Kebutuhan *Software***

Adapun perangkat lunak yang dibutuhkan dalam penelitian ini adalah sebagai berikut :

a. Sistem Operasi

Sistem operasi yang digunakan dalam penelitian ini adalah Windows 7.

b. *Hypertext Preprocessor (PHP)*

Merupakan bahasa pemrograman yang akan digunakan untuk mengimplementasi hasil dari rancangan yang sudah dibuat.

c. Notepad++

Tools yang digunakan untuk editor bahasa pemrograman PHP.

d. MySQL

*Software* ini digunakan untuk menyimpan database dari sistem yang akan dibangun.

e. Ms. Word

*Software* ini digunakan untuk membuat laporan hasil penelitian.

### 3.2.2 Kebutuhan *Hardware*

Selain kebutuhan *software*, diperlukan pula *hardware* yang harus dipenuhi agar penelitian ini berjalan dengan lancar. Adapun *hardware* yang digunakan dalam penelitian ini adalah sebagai berikut:

a. *Personal Computer* atau laptop dengan spesifikasi :

Prosesor : Dual core

Sistem Operasi : Windows 7

RAM : 2 GB

b. Printer, digunakan untuk mencetak hasil penelitian ke dalam bentuk *hardcopy*.

### 3.3 Pengumpulan Data

Untuk mendapatkan data yang nantinya akan digunakan dalam penelitian ini, penulis mendapatkannya dari peneliti terdahulu yang sudah mengelompokkan beberapa komentar ke dalam kelompok positif dan negatif. Selain itu juga penulis

mengumpulkan sendiri dari komentar-komentar yang diberikan oleh pengguna situs youtube. Berikut adalah beberapa contoh data yang penulis ambil dari situs youtube dengan kategori film aksi:

Tabel 3. 1 Data Komentar Review Film

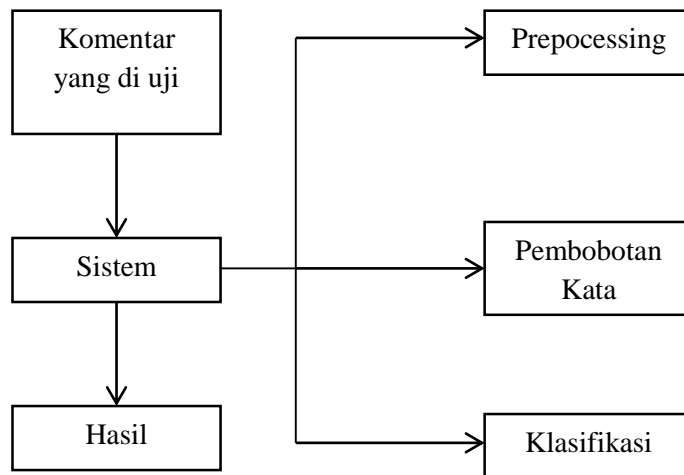
No	Tanggal	User	Komentar
1	24-12-2014	Icy Ball	<p>Many people complain about these things with the movie:</p> <p>Too much destruction.</p> <p>Killing off Zod.</p> <p>The more serious take on the title character.</p> <p>The unoriginal story.</p> <p>People are still forgetting about all of this:</p> <p>The Avengers had a lot of destruction.</p> <p>Two-Face, Bane, and Ra's Al Ghul were killed off.</p> <p>The Dark Knight Trilogy had a more serious take on Batman.</p> <p>Many superhero movies have similar stories, like Batman Begins and the Avengers.</p>
2	24-12-2014	Grkpektis	<p>I don't give a fuck what you retarded Superman and DC fanboys say Man of Steel was one of the worst movies I have ever seen, fuck this POS movie. No I am not a Marvel fanboy I gave up on them when they started sucking China's dick in IM3</p>
3	06-01-2015	Johnny Valencia	<p>For those who are great fans of the franchise this movie could considered as an insult; they completely changed the whole "old superman</p>

			<p>perception" to make way a new and in many levels improved character. This is what studios are getting us all used to. the constant remaking of these films (superman, x-men, spiderman) are made with the one and only purpose of filling up studios pockets. By the way, how many times are we gonna keep seeing New york get crumbled into wreckage in these movies. Give us a break for crying out loud.</p>
4	23-12-2014	Craven Lestat	<p>Holy crap this movie is the exact plot of the movie "The Warriors" I mean everything from the leader of the strongest group bringing all other gangs together, to his death and then it getting blamed on a specific gang,then they escape blah blah.Funniest is that the sirens part is again stolen from "The warriors" where the female gang called the Lizzies does the same thing and lures them and tries to kill them.This movie has a ridiculously stolen plot which is why I could not enjoy it.Want to see a good ninja movie Check out pretty much any movie with Sho Kosugi in it like "Enter the ninja" etc.Btw the "Warriors" is a classic I recommend you watch it and even Rockstar games who makes The Grand theft auto series made a game after it and it was a great game.Yes you can watch it online on Youtube but I will not link because am not promoting anyones channels but just find it and enjoy.</p>

4	06-12-2014	Ella Madison	I thinks it's a good movie. I really like this movie.
5	25-12-2014	Roxanne Lipz	It's sad that alot of racist white people use other cultures as slaves and can't do it themselves no wonder why the ISIS Muslim want to kill them smh
6	31-12-2014	Shad Xevyn	This is NOT Transporter anything and Jason Statham is NOT even in the movie !!! Somebody is trying to RUIN Youtube and they're doing a good job of it.
7	25-12-2014	Ricky Concepcion	This is NOT Transporter anything and Jason Statham is NOT even in the movie !!! Somebody is trying to RUIN Youtube and they're doing a good job of it.
8	05-01-2015	Cinta Shyna	I cannt say anything jst awesome
9	06-11-2014	Knowledge Power	How is it possible actor Micheal White is not a big budget mega action star? He is much better than the Rock and bring a certain intensity to the screen that in my view is authentic.
10	05-01-2015	Neosomato Hypergenes is	MAN that was some crazy far out stuff like its really out of this world man crazy and wierd dude wicked strange and mind bending far out crazy

### 3.4 Metode yang Diusulkan

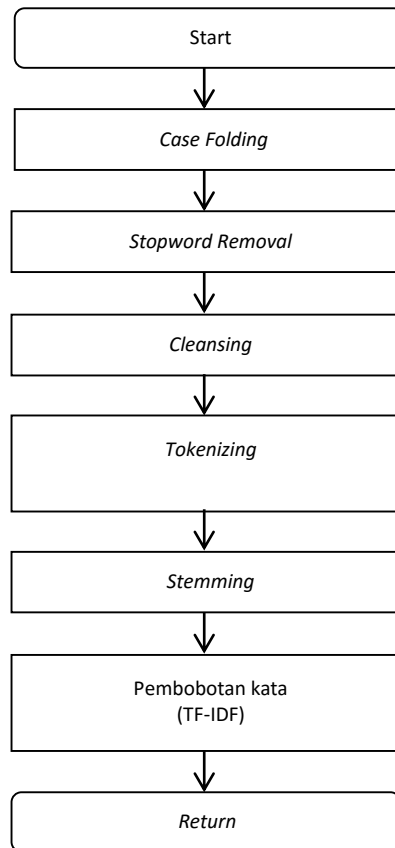
Proses *text mining* secara umum memiliki tahapan yaitu *preprocessing text* kemudian dilakukan pembobotan kata dan kemudian diolah menggunakan algoritma yang dipakai dalam kasus ini menggunakan algoritma *K-NN*. *Preprocessing text* juga terdiri dari beberapa tahapan yaitu *cleansing*, *parsing*, *tokenizing*, *stopword removal*, *stemming*, dan pembobotan kata. Alur prosesnya jika digambarkan adalah sebagai berikut:



Gambar 3. 1 Rancangan Arsitektur Sistem

#### 3.4.1 *Preprocessing*

Karena sistem tidak bisa membaca dokumen teks dikarenakan strukturnya tidak teratur maka diperlukanlah tahapan *preprocessing* yaitu merubah dari teks menjadi sebuah angka yang terstruktur sehingga dapat dikenali oleh sistem. *Preprocessing* ini terdiri dari beberapa langkah yaitu *cleansing*, *parsing*, *tokenizing*, *stopword removal*, *stemming* dan pembobotan kata. Adapun diagram alirnya adalah sebagai berikut:



Gambar 3. 2 Diagram Alir *Preprocessing*

Langkah pertama yang dilakukan adalah *case folding*, yaitu merubah kalimat tersebut ke dalam bentuk huruf kecil seluruhnya, mengganti huruf kapital menjadi huruf kecil agar seragam. Kemudian setelah kalimat tersebut seragam maka selanjutnya adalah melakukan penyaringan kata-kata yang tidak bermakna. Kata-kata yang tidak bermakna ini akan memakan banyak memori dan pemborosan waktu proses. Kata-kata yang disaring nantinya yaitu kata hubung misalnya tahap ini dinamakan *stopword removal*. Pada tahap *cleansing* akan dilakukan pembersihan dokumen dari tanda baca dan simbol-simbol. Tanda baca dan simbol-simbol itu akan dideteksi dan digantikan oleh spasi. Kemudian dilakukan *tokenizing*, dokumen yang sudah dibersihkan dari tanda

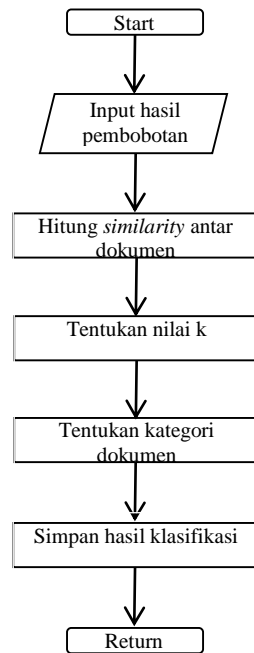
baca dan simbol-simbol nantinya akan dipecah menjadi kata, pemecahan ini berdasarkan oleh spasi. Langkah selanjutnya yaitu melakukan *stemming*, yaitu merubah kata-kata tersebut menjadi kata dasar. Proses ini akan menghilangkan imbuhan kata, yakni awalan, sisipan, awalan-akhiran.

Karena komputer tidak dapat membaca dokumen teks, maka perlulah representasi dari teks menjadi angka, tahap ini yaitu pembobotan kata. Tiap kata yang muncul dalam dokumen tadi akan di beri bobot tergantung dari frekuensi kemunculannya dalam tiap dokumen. Bobot ini akan disimpan dan digunakan untuk langkah perhitungan menggunakan algoritma *K-NN*.

#### **3.4.2 Klasifikasi Menggunakan *K-Nearest Neighbor (K-NN)***

Pada tahap klasifikasi ini yaitu mengambil hasil dari pembobotan kata. Hasil dari pembobotan kata tersebut selanjutnya akan dihitung nilai similaritasnya atau kemiripan antara dokumen uji dengan dokumen latih menggunakan rumus (5). Setelah didapatkan nilai similaritasnya tentukan nilai  $k$ , ambil hasil similaritas tersebut sesuai nilai  $k$  di mulai dari nilai similaritas yang paling tinggi. Setelah mendapatkan hasil similaritas sejumlah nilai  $k$ , tentukan hasil klasifikasi menggunakan rumus (6) atau (7). Berikut adalah diagram alirnya:





Gambar 3. 3 Diagram Alir Proses K-NN

### 3.4.3 Manualisasi Sistem

Penghitungan manual berikut agar memberikan sedikit gambaran tentang perancangan sistem yang akan di bangun. Berikut adalah contoh manualisasinya:

a. Input dokumen

Tabel 3. 2 Input Dokumen

ID	Komentar	Sentimen
1	Movie/book terrible young readers read messages fuckin stupid ruining younger generations girls.	Positif
2	Stupid deceiver tittle!!! Stupid liar viewers, waste time!!!!	Negatif
3	Nice! love. credits trailer	Positif
4	Nice video,!, Joss Whedon, eagerly awaiting real woman film! wait. :) past time	Positif

5	Stupidest movie	Negatif
X	Its very nice this film ;)	?????

b. *Case Folding*

Pada tahap ini semua teks dirubah ke huruf kecil

Tabel 3. 3 Proses *Case Folding*

ID	Komentar	Sentimen
1	movie/book terrible young readers read messages fuckin stupid ruining younger generations girls.	Positif
2	stupid deceiver tittle!!! stupid liar viewers, waste time!!!!	Negatif
3	nice! love. credits trailer	Positif
4	nice video,!, joss whedon, eagerly awaiting real woman film! wait. :) past time	Positif
5	stupidest movie	Negatif
X	its very nice this film ;)	?????

c. *Stopword Removal*

Yaitu membersihkan dokumen dari kata yang tidak memiliki makna, misal kata hubung dan sebagainya.

Tabel 3. 4 Proses *Stopward Removal*

ID	Komentar	Sentimen
1	movie/book terrible young readers read messages fuckin stupid ruining younger generations girls.	Positif
2	stupid deceiver tittle!!! stupid liar viewers, waste time!!!!	Negatif
3	nice! love. credits trailer	Positif

4	nice video,!, joss whedon, eagerly awaiting real woman film! wait. :) past time	Positif
5	stupidest movie	Negatif
X	nice film ;)	???????

d. *Cleansing*

Yaitu membersihkan dokumen dari simbol-simbol dan tanda baca

Tabel 3. 5 Proses *Cleansing*

ID	Komentar	Sentimen
1	movie book terrible young readers read messages fuckin stupid ruining younger generations girls	Positif
2	stupid deceiver tittle stupid liar viewers waste time	Negatif
3	nice love credits trailer	Positif
4	nice video joss whedon eagerly awaiting real woman film wait past time	Positif
5	stupidest movie	Negatif
X	nice film	???????

e. *Tokenizing*

Setelah itu memecah dari kalimat menjadi kata-kata kecil

Tabel 3. 6 Proses *Tokenizing*

ID	Komentar	Sentimen
1	movie // book // terrible // young // readers // read // messages // fuckin // stupid // ruining // younger // generations // girls	Positif
2	stupid // deceiver // tittle // stupid // liar // viewers //	Negatif

	waste // time	
3	nice // love // credits // trailer	Positif
4	nice // video // joss // whedon // eagerly // awaiting // real // woman // film // wait // past // time	Positif
5	stupidest // movie	Negatif
X	nice // film	???????

f. *Stemming*

Setelah menjadi kata-kata kecil kemudian merubah kata tersebut menjadi kata dasar.

Tabel 3. 7 Hasil *Stemming*

ID	Komentar	Sentimen
1	movi // book // terribl // young // reader // read // messag // fuckin // stupid // ruin // younger // gener // girl	Positif
2	stupid // deceiv // tittl // stupid // liar // viewer // wast // time	Negatif
3	nice // love // credit // trailer	Positif
4	nice // video // joss // whedon // eagerli // await // real // woman // film // wait // past // time	Positif
5	stupidest // movi	Negatif
X	nice // film	???????

g. *Pembobotan Kata*

Pembobotan kata dimulai dari mencari nilai TF yaitu menghitung jumlah kata yang muncul pada suatu dokumen. Contohnya kata nice muncul satu kali pada D1, D4, dan D5. Dari nilai TF itu akan didapatkan nilai DF yang

didasarkan pada jumlah kata muncul pada semua dokumen. Kata nice muncul pada D1, D4, dan D5 maka didapatkan nilai DF adalah 3.

Tabel 3. 8 Hasil Pembobotan Kata

Informasi TF dokumen								
No	Kata	D1	D2	D3	D4	D5	Dx	DF
1	movi	1	0	0	0	1	0	2
2	book	1	0	0	0	0	0	1
3	terribl	1	0	0	0	0	0	1
4	young	1	0	0	0	0	0	1
5	reader	1	0	0	0	0	0	1
6	read	1	0	0	0	0	0	1
7	messag	1	0	0	0	0	0	1
8	fuck	1	0	0	0	0	0	1
9	stupid	1	2	0	0	0	0	2
10	ruin	1	0	0	0	0	0	1
11	younger	1	0	0	0	0	0	1
12	gener	1	0	0	0	0	0	1
13	girl	1	0	0	0	0	0	1
14	deceiv	0	1	0	0	0	0	1
15	tittl	0	1	0	0	0	0	1
16	liar	0	1	0	0	0	0	1
17	viewer	0	1	0	0	0	0	1
18	wast	0	1	0	0	0	0	1
19	time	0	1	0	1	0	0	2
20	nice	0	0	1	1	0	1	3
21	love	0	0	1	0	0	0	1
22	credit	0	0	1	0	0	0	1
23	trailer	0	0	1	0	0	0	1
24	video	0	0	0	1	0	0	1
25	joss	0	0	0	1	0	0	1

26	whedon	0	0	0	1	0	0	1
27	eagerli	0	0	0	1	0	0	1
28	await	0	0	0	1	0	0	1
29	real	0	0	0	1	0	0	1
30	woman	0	0	0	1	0	0	1
31	film	0	0	0	1	0	1	1
32	wait	0	0	0	1	0	0	1
33	stupidest	0	0	0	0	1	0	1

Dari nilai DF itu nanti akan dicari lagi nilai IDF yaitu dengan menggunakan rumus (1):

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

Misal dihitung IDF kata movi:

$$IDF(movi) = \log\left(\frac{6}{2}\right)$$

$$IDF(movi) = \log 3$$

$$IDF(movi) = 0,477$$

Dari nilai IDF tersebut akan dicari lagi nilai TF-IDF yaitu perkalian antara TF dengan IDF. Hasilnya adalah sebagai berikut:

Tabel 3. 9 Hasil TF-IDF

TF-IDF							
No	Kata	D1	D2	D3	D4	D5	Dx
1	movi	0,477	0	0	0	0,477	0
2	book	0,778	0	0	0	0	0
3	terribl	0,778	0	0	0	0	0
4	young	0,778	0	0	0	0	0

5	reader	0,778	0	0	0	0	0
6	read	0,778	0	0	0	0	0
7	messag	0,778	0	0	0	0	0
8	fuck	0,778	0	0	0	0	0
9	stupid	0,477	0,954	0	0	0	0
10	ruin	0,778	0	0	0	0	0
11	younger	0,778	0	0	0	0	0
12	gener	0,778	0	0	0	0	0
13	girl	0,778	0	0	0	0	0
14	deceiv	0	0,778	0	0	0	0
15	tittl	0	0,778	0	0	0	0
16	liar	0	0,778	0	0	0	0
17	viewer	0	0,778	0	0	0	0
18	wast	0	0,778	0	0	0	0
19	time	0	0,477	0	0,477	0	0
20	nice	0	0	0,301	0,301	0	0,301
21	love	0	0	0,778	0	0	0
22	credit	0	0	0,778	0	0	0
23	trailer	0	0	0,778	0	0	0
24	video	0	0	0	0,778	0	0
25	joss	0	0	0	0,778	0	0
26	whedon	0	0	0	0,778	0	0
27	eagerli	0	0	0	0,778	0	0
28	await	0	0	0	0,778	0	0
29	real	0	0	0	0,778	0	0
30	woman	0	0	0	0,778	0	0
31	film	0	0	0	0,477	0	0,477
32	wait	0	0	0	0,778	0	0
33	stupidest	0	0	0	0	0,778	0

Kemudian dari nilai TF-IDF itu masih perlu dinormalisasi dengan menggunakan rumus (4).

Contoh normalisasi  $w = \text{movi}$  pada  $d1$ :

$$\text{TF-IDF}(w, d) = \frac{\text{TF-IDF}(w, d)}{\sqrt{\sum_{w=1}^n \text{TF-IDF}(w, d)^2}}$$

$$\text{TF-IDF}(\text{movi}, d1) = \frac{\text{TF-IDF}(\text{movi}, d1)}{\sqrt{\sum_{w=1}^n \text{TF-IDF}(w, d1)^2}}$$

$$\text{TF-IDF}(\text{movi}, d1)$$

$$= \frac{0,477}{\sqrt{0,477^2 + 0,778^2 + 0,778^2 + 0,778^2 + 0,778^2 + 0,778^2 + 0,778^2 + 0,477^2 + 0,778^2 + 0,778^2 + 0,778^2 + 0,778^2 + 0,778^2}}$$

$$\text{TF-IDF}(\text{movi}, d1) = \frac{0,477}{\sqrt{6,655}}$$

$$\text{TF-IDF}(\text{movi}, d1) = \frac{0,477}{2,667}$$

$$\text{TF-IDF}(\text{movi}, d1) = 0,179$$

Hasilnya seluruhnya adalah sebagai berikut:

Tabel 3. 10 Normalisasi TF-IDF

TF-IDF Normalisasi							
No	Kata	D1	D2	D3	D4	D5	Dx
1	movi	0,187	0	0	0	0,522	0
2	book	0,305	0	0	0	0	0
3	terribl	0,305	0	0	0	0	0
4	young	0,305	0	0	0	0	0
5	reader	0,305	0	0	0	0	0
6	read	0,305	0	0	0	0	0
7	messag	0,305	0	0	0	0	0
8	fuck	0,305	0	0	0	0	0
9	stupid	0,187	0,468	0	0	0	0
10	ruin	0,305	0	0	0	0	0
11	younger	0,305	0	0	0	0	0
12	gener	0,305	0	0	0	0	0
13	girl	0,305	0	0	0	0	0
14	deceiv	0	0,381	0	0	0	0
15	tittl	0	0,381	0	0	0	0
16	liar	0	0,381	0	0	0	0



17	viewer	0	0,381	0	0	0	0
18	wast	0	0,381	0	0	0	0
19	time	0	0,234	0	0,206	0	0
20	nice	0	0	0,218	0,130	0	0,534
21	love	0	0	0,564	0	0	0
22	credit	0	0	0,564	0	0	0
23	trailer	0	0	0,564	0	0	0
24	video	0	0	0	0,335	0	0
25	joss	0	0	0	0,335	0	0
26	whedon	0	0	0	0,335	0	0
27	eagerli	0	0	0	0,335	0	0
28	await	0	0	0	0,335	0	0
29	real	0	0	0	0,335	0	0
30	woman	0	0	0	0,335	0	0
31	film	0	0	0	0,206	0	0,846
32	wait	0	0	0	0,335	0	0
33	stupidest	0	0	0	0	0,852	0

Langkah selanjutnya adalah menghitung *Cosine Similarity* dengan menggunakan rumus (5) kemudian hasil tersebut diurutkan.

Misalkan menghitung *cosine similarity* antara  $d_2$  dengan  $d_x$ :

$$\text{cosSim}(d_j, q_k) = \frac{\sum_{i=1}^n (td_{ij} \times tq_{ik})}{\sqrt{\sum_{i=1}^n td_{ij}^2 \times \sum_{i=1}^n tq_{ik}^2}}$$

$$\text{cosSim}(d_3, d_x) = \frac{0,218 \times 0,534}{\sqrt{(0,218^2 + 0,564^2 + 0,564^2 + 0,564^2) \times (0,534^2 + 0,846^2)}}$$

$$\text{cosSim}(d_3, d_x) = \frac{0,116}{\sqrt{(0,047 + 0,318 + 0,318 + 0,318) \times (0,285 + 0,716)}}$$

$$\text{cosSim}(d_3, d_x) = \frac{0,116}{\sqrt{(1,001) \times (1,001)}}$$

$$\text{cosSim}(d_3, d_x) = 0,116$$

Hasilnya keseluruhan adalah sebagai berikut:

Tabel 3. 11 Hasil *Cosine Similarity*

	Sim(x,di)	Kategori
D4	0,713	Positif
D3	0,116	Positif
D1	0	Negatif
D2	0	Negatif
D5	0	Negatif

#### h. Klasifikasi

Pada perhitungan klasifikasi menggunakan *K-NN* misal ditentukan nilai *k-values* nya adalah 3 ( $k=3$ ), sesuai perhitungan dari rumus (7) yaitu menghitung nilai *cosine similarity* paling tinggi.

Probabilitas terhadap sentimen positif:

$$p(x, c_m) = \sum_{l=1}^m SIM(X, d_j) \in c_m$$

$$p(x, positif) = 0,713 + 0,116$$

$$p(x, positif) = 0,829$$

Probabilitas terhadap sentimen negatif:

$$p(x, c_m) = \sum_{l=1}^m SIM(X, d_j) \in c_m$$

$$p(x, negatif) = 0$$

Dari hasil diatas didapatkan bahwa probabilitas dx terhadap sentimen positif lebih besar dibandingkan sentimen negatif. Hasil akhir adalah sebagai berikut:

Tabel 3. 12 Hasil Klasifikasi

ID	Komentar	Sentimen
1	movie/book terrible young readers read messages fuckin stupid ruining younger generations girls.	Positif
2	stupid deceiver tittle!!! stupid liar viewers, waste time!!!!	Negatif
3	nice! love. credits trailer	Positif
4	nice video,!, joss whedon, eagerly awaiting real woman film! wait. :) past time	Positif
5	stupidest movie	Negatif
X	its very nice this film ;)	Positif