

**ANALISIS SENTIMEN PERSPEKTIF MASYARAKAT INDONESIA  
PADA MEDIA SOSIAL TWITER TERHADAP DARK JOKES  
MENGUNAKAN METODE *NAÏVE BAYES CLASSIFIER*  
DENGAN *BOOSTING ADABOOST***

**SKRIPSI**



**DISUSUN OLEH:**

**MUHAMMAD FAIZAL ARIADI**

**H76216064**

**PROGRAM STUDI SISTEM INFORMASI**

**FAKULTAS SAINS DAN TEKNOLOGI**

**UNIVERSITAS ISLAM NEGERI SUNAN AMPEL**

**SURABAYA**

**2022**

## PERNYATAAN KEASLIAN

Saya yang bertanda tangan di bawah ini,

Nama : MUHAMMAD FAIZAL ARIADI.

NIM : H76216064

Program Studi : Sistem Informasi

Angkatan : 2016

Menyatakan bahwa saya tidak melakukan plagiat dalam penulisan skripsi saya yang berjudul “ANALISIS SENTIMEN PERSPEKTIF MASYARAKAT INDONESIA PADA MEDIA SOSIAL TWITTER TERHADAP DARK JOKES MENGGUNAKAN METODE NAÏVE BAYES CLASSIFIER DENGAN BOOSTING ADABOOST”. Apabila suatu saat nanti terbukti saya melakukan tindakan plagiat, maka saya bersedia menerima sanksi yang telah ditetapkan.

Demikian pernyataan keaslian ini saya buat dengan sebenar-benarnya.

Surabaya, 14 Januari 2022

Yang menandatangani,

The image shows a handwritten signature in black ink over a yellow revenue stamp. The stamp features the number '1000' in large red digits, the text 'STAMPED RUPAH' vertically on the left, and 'METERAI TEMPEL' on the right. A unique alphanumeric code '92FE3AJX566392973' is printed at the bottom of the stamp.

M. Faizal Ariadi

H76216064

## LEMBAR PERSETUJUAN PEMBIMBING

Skrpsi oleh

NAMA : MUHAMMAD FAIZAL ARIADI

NIM : H76216064

JUDUL : ANALISIS SENTIMEN PERSPEKTIF MASYARAKAT INDONESIA  
PADA MEDIA SOSIAL TWITER TERHADAP DARK JOKES  
MENGUNAKAN METODE *NAIVE BAYES CLASSIFIER* DENGAN  
*BOOSTING ADABOOST*

Ini telah diperiksa dan disetujui untuk diujikan

Surabaya, 14 Januari 2021

Dosen Pembimbing 1



(Mujib Ridwan, S.Kom., M.T)

NIP 198604272014031004

Dosen Pembimbing 2



(Ahmad Yusuf, M. Kom)

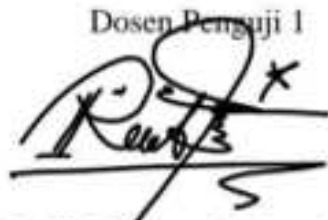
NIP 199001202014031003

## PENGESAHAN TIM PENGUJI SKRIPSI

Skripsi Muhammad Faizal Ariadi ini telah dipertahankan  
di depan tim penguji skripsi  
Surabaya, 10 Januari 2022

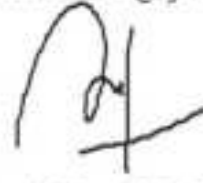
Mengesahkan,  
Dewan Penguji

Dosen Penguji 1



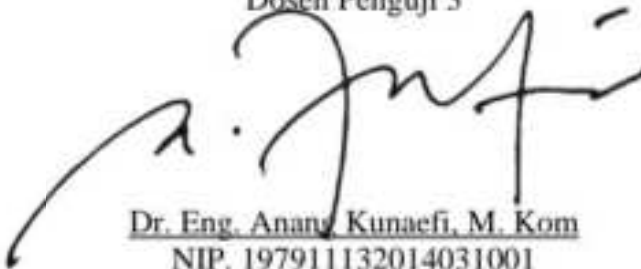
Mujib Ridwan, S.Kom., M.T  
NIP. 198604272014031004

Dosen Penguji 2



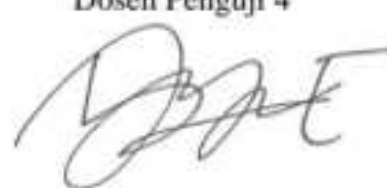
Ahmad Yusuf, M. Kom  
NIP. 199001202014031003

Dosen Penguji 3



Dr. Eng. Anang Kunaefi, M. Kom  
NIP. 197911132014031001

Dosen Penguji 4



Bayu Adhi Nugroho, M. Kom  
NIP. 197905182014031001

Mengetahui,

Dekan Fakultas Sains dan Teknologi

UIN Sunan Ampel Surabaya



Prof. Dr. H. Evi Fatimatur Rusydiyah, M.Ag  
NIP. 197312272005012003



**KEMENTERIAN AGAMA**  
**UNIVERSITAS ISLAM NEGERI SUNAN AMPEL SURABAYA**  
**PERPUSTAKAAN**

Jl. Jend. A. Yani 117 Surabaya 60237 Telp. 031-8431972 Fax. 031-8413300  
E-Mail: [perpus@uinsby.ac.id](mailto:perpus@uinsby.ac.id)

LEMBAR PERNYATAAN PERSETUJUAN PUBLIKASI  
KARYA ILMIAH UNTUK KEPENTINGAN AKADEMIS

Sebagai sivitas akademika UIN Sunan Ampel Surabaya, yang bertanda tangan di bawah ini, saya:

Nama : Muhammad Faizal Ariadi  
NIM : H76216064  
Fakultas/Jurusan : FAKULTAS SAINS DAN TEKNOLOGI/SISTEM INFORMASI  
E-mail address : [oreocoklat04@gmail.com](mailto:oreocoklat04@gmail.com)

Demi pengembangan ilmu pengetahuan, menyetujui untuk memberikan kepada Perpustakaan UIN Sunan Ampel Surabaya, Hak Bebas Royalti Non-Eksklusif atas karya ilmiah:

☒ Skripsi ☐ Tesis ☐ Desertasi ☐ Lain-lain (.....)

yang berjudul :

ANALISIS SENTIMEN PERSPEKTIF MASYARAKAT INDONESIA PADA

MEDIA SOSIAL TWITTER TERHADAP DARK JOKES MENGGUNAKAN

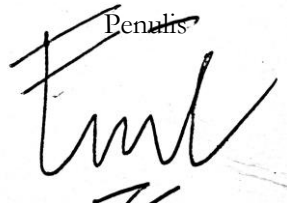
METODE NAÏVE BAYES CLASSIFIER DENGAN BOOSTING ADABOOST

beserta perangkat yang diperlukan (bila ada). Dengan Hak Bebas Royalti Non-Eksklusif ini Perpustakaan UIN Sunan Ampel Surabaya berhak menyimpan, mengalih-media/format-kan, mengelolanya dalam bentuk pangkalan data (database), mendistribusikannya, dan menampilkan/mempublikasikannya di Internet atau media lain secara **fulltext** untuk kepentingan akademis tanpa perlu meminta ijin dari saya selama tetap mencantumkan nama saya sebagai penulis/pencipta dan atau penerbit yang bersangkutan.

Saya bersedia untuk menanggung secara pribadi, tanpa melibatkan pihak Perpustakaan UIN Sunan Ampel Surabaya, segala bentuk tuntutan hukum yang timbul atas pelanggaran Hak Cipta dalam karya ilmiah saya ini.

Demikian pernyataan ini yang saya buat dengan sebenarnya.

Surabaya, 17 Januari 2022

Penulis  
  
(Muhammad Faizal Ariadi)















## DAFTAR GAMBAR

Gambar 3. 1 Diagram Alur Penelitian.....	23
Gambar 4. 1 Topik Terkait Dark Jokes (Google Trends) .....	29
Gambar 4. 2 Kode Tahapan Cleansing .....	32
Gambar 4. 3 Kode Tahapan Case Folding .....	33
Gambar 4. 4 Cuplikan Kode Tokenizing .....	34
Gambar 4. 5 Cuplikan Kode Normalization .....	35
Gambar 4. 6 Cuplikan Kode Filtering.....	36
Gambar 4. 7 Cuplikan Kode Stemming.....	37
Gambar 4. 8 Word Cloud Sentimen Positif .....	40
Gambar 4. 9 Word Cloud Sentimen Negatif.....	41
Gambar 4. 10 Word Cloud Sentimen Neutral .....	42
Gambar 4. 11 Cuplikan Kode Cross Validation .....	45
Gambar 4. 12 Grafik Hasil Cross Validation kondisi satu dan tiga.....	47
Gambar 4. 13 Pie Chart Sentimen Dark Jokes.....	48

## BAB 1

## PENDAHULUAN

## 1.1 Latar Belakang

Pada era modern sekarang ini, kemajuan dalam perkembangan informasi dan teknologi berkembang dengan cepat. Salah satu produk hasil perkembangan teknologi adalah internet. Di Indonesia sendiri pengguna internet semakin meningkat, dapat dilihat pada tahun 2020 dinyatakan bahwa sebanyak 175,4 juta masyarakat Indonesia yang menjadi pengguna aktif Internet. Berdasarkan total populasi masyarakat Indonesia yang berjumlah sebesar 272,1 juta jiwa, yang artinya 64% lebih dari setengah masyarakat Indonesia yang menggunakan Internet untuk mengakses dunia maya(We Are Social & Hootsuite, 2020).

Media sosial adalah salah satu fitur yang populer di internet saat ini, karena media sosial menyajikan layanan yang bisa diakses oleh pengguna untuk mengirimkan pesan secara bebas. Pesan yang ditulis oleh pengguna juga beraneka ragam, ada yang sisi positif bahkan juga yang negatif. Salah satu yang menjadi sorotan di media sosial adalah sebuah kritikan yang ditulis oleh pengguna internet. Media sosial yang sering digunakan oleh masyarakat ada berbagai macam seperti Twitter dan Facebook. Pengguna Twitter di Indonesia juga bertambah dengan pesat bahkan Indonesia menjadi peringkat 5 besar di bawah negara Inggris dan negara besar lainnya (Kementerian Komunikasi dan Informatika, 2020).

Semakin banyak pengguna media sosial maka semakin banyak juga berbagai macam komentar maupun kritikan yang diungkapkan oleh masyarakat tanpa ada batasan waktu. Masyarakat juga sudah mulai terbiasa menyampaikan sebuah opininya di media sosial. Opini yang disampaikan terkadang menuai pro dan kontra. Opini atau merupakan sebuah pendapat, pikiran, pandangan seseorang terhadap kebenaran yang relatif. Opini atau kritik dapat diutarakan melalui berbagai macam contohnya di media sosial. Di Indonesia sendiri semua kalangan masyarakat banyak menyampaikan kritiknya di media sosial, bahkan beberapa politisi juga

sering menyampaikan kritiknya. Penyampaian kritik maupun opini mempunyai khas masing-masing seperti kritikan langsung secara tajam atau ada juga secara halus, yang tidak kalah menarik ialah sebuah kritik yang dibalur dengan sedikit humor.

Manusia memiliki kebutuhan material dan non-material, kebutuhan material ialah sandang, pangan, dan sebagainya., sedangkan non-material seperti *gadget*, hiburan dan sebagainya., secara tidak langsung humor ialah salah satu kebutuhan manusia sebagai sarana hiburan maupun pelipur lara. Humor saat ini bukan hanya digunakan sebagai media hiburan saja bahkan humor bisa digunakan untuk wahana kritik sosial (Hidayah, 2020). Mengkritik sesuatu menggunakan humor adalah sebuah format baru yang sudah menjadi budaya yang maju sehingga humor dapat dijadikan sebuah wadah baru dalam penyampaian sebuah kritik sosial (Laraziza, 2014). Penyampaian kritik secara halus dalam bentuk komedi saat ini adalah sebuah komedi modern yang biasa disebut dengan *satire*. *Satire* ialah “a work or manner that blends a censori us attitude with humor and wit for improving human institutions or humanity”, secara tidak langsung *satire* saat ini bisa dijadikan karya atau media untuk mengkritik kejadian melalui perkataan maupun perbuatan dengan ditambahkannya bumbu humor. (Lázaro and Swift, 2001).

*Satire* saat ini tidak hanya berbentuk esai, dan drama melainkan banyak yang memakai *satire* dalam bentuk komedi seperti sebuah konsep komedi modern yang dibawakan diatas panggung dengan menyampaikan sebuah materi komedi yang mengandung sebuah pesan untuk ditertawakan lebih mudah ini dinamakan *Stand Up Comedy*. Indonesia sekarang ini sudah banyak ajang perlombaan *Stand Up Comedy* secara siaran TV maupun tidak, melalui ajang ini para *Stand Up Comedian* atau yang biasa disebut dengan komika membawakan materi kritik sosial yang berhubungan dengan Indonesia contohnya adalah deskriminasi dan disampaikan secara *satire*. Penyampaian kritik dengan metode *satire* ini dipelopori orang seorang komedian bernama Chris Rock yang berasal dari Amerika Serikat, di dalam setiap materi *Stand Up* dia menyampaikan sebuah kritik sosial yang pada saat itu sedang ramai diperbincangkan di Amerika Serikat soal maraknya kasus rasisme yang menimpa masyarakat berkulit hitam. Indonesia juga memiliki seorang

komedian yang memiliki materi komedi mengenai isu-isu sensitif terhadap kritik sosial Indonesia yaitu Pandji Pragiwaksono (Abdul, Mahbub and Ab, 2018).

Analisis sentimen adalah jenis penelitian dari *Text Mining* yang berfungsi untuk mengklasifikasikan teks berupa opini menjadi dua kelas negatif dan positif (Nurhuda, Widya Sihwi and Doewes, 2016). Analisis sentimen digunakan sebagai riset untuk menghitung opini sentiment maupun emosional yang diutarakan

secara tekstual (Zulfa and Winarko, 2017). Analisis sentimen bisa digunakan sebagai alat untuk meninjau media sosial saat ini, karena sangat memungkinkan untuk memvisualisasikan gambaran dari opini yang bersifat umum menjadi lebih luas dibalik dari maksud opini tersebut. Analisis sentimen di media sosial biasanya melihat sebuah sentimen dari emosi maupun kata yang disampaikan, umumnya ditujukan pada *brand* sebagai pengetahuan untuk melihat karakteristik konsumen, tingkat kepuasan, kualitas serta memanfaatkannya untuk strategi pemasaran (Zuhri and Alamsyah, 2017). Dapat dikatakan analisis sentimen jika digunakan tepat sasaran seperti bisnis, maka perubahan dinamika pasar dan kecocokan suatu *brand* untuk pelanggan akan terlihat. Jika analisis sentimen biasa digunakan untuk melihat kecocokan dari suatu *brand* maka analisis sentimen dapat juga digunakan untuk melihat kecocokan salah satu jenis komedi yaitu *dark jokes*.

Masalah yang dialami sekarang ini ialah dengan cara apa menentukan suatu kelas sentimen terhadap *dark jokes*. Klasifikasi sentimen mengenai *dark jokes* perlu diketahui karena dapat mengukur sejauh mana pandangan masyarakat terhadap *dark jokes*. Selain itu analisis sentimen ini dapat dimanfaatkan untuk masyarakat pengguna media sosial terutama pengguna Twitter dalam penggunaan *dark jokes*. Untuk memecahkan masalah yang sudah dijelaskan, maka solusi yang diusulkan yaitu menggunakan sistem analisis yang mengklasifikasikan dua sentimen negatif dan positif yang berkaitan dengan opini masyarakat mengenai sejauh mana pandangan positif masyarakat terhadap *dark jokes*. Metode klasifikasi ada berbagai macam yang bisa digunakan untuk mengklasifikasikan suatu *text mining* dari hasil analisis sentimen, salah satunya yaitu *Naïve Bayes Classifier*(NBC), metode ini dapat digunakan untuk menganalisa sebuah *brand* dengan tingkat keakuratan yang bagus dengan mengambil data opini dari forum kaskus (Zuhri and Alamsyah, 2017). Metode ini juga menghasilkan tingkat akurasi yang tinggi untuk menganalisa sentimen opini masyarakat dari Twitter (Cahyani, 2020). Dalam hasil penelitian yang dilakukan oleh (Syakuro, 2017) yang berjudul “ANALISIS SENTIMEN MASYARAKAT TERHADAP E-COMMERCE PADA MEDIA SOSIAL MENGGUNAKAN METODE NAÏVE BAYES CLASSIFIER (NBC) DENGAN SELEKSI FITUR INFORMATION GAIN (IG)” menyarankan jika metode yang digunakan yaitu NBC dapat digabungkan dengan metode *Adaboost* agar dapat



metode NBC dengan *Adaboost* untuk membandingkan  
dikombinasikan dengan NBC dan apakah dapat menaikkan  
goritma metode NBC.

**2.2.2. Rumusan Masalah**

1. Bagaimana membuat klasifikasi sentimen dari media sosial Twitter pada  
terkait dengan *dark jokes* menggunakan metode *Naïve Bayes*  
dan *Adaboost*?

2. Bagaimana perbandingan performa dari *Naïve Bayes Classifier* dan *Naïve Bayes Classifier* dengan  
metode *Adaboost* sebagai analisa sentimen *dark jokes*?

3. Bagaimana hasil analisis sentimen negatif dan positif terkait *dark jokes*?

**2.2.3. Batasan Masalah**

1. Batasan masalah dalam penelitian ini adalah:  
a. Data hanya dari media sosial Twitter.  
b. Data menggunakan kata kunci "*dark jokes, darkjokes, 1*"

## 1.2 Perumusan Masalah

1. Bagaimana membuat klasifikasi sentimen dari media sosial Twitter pada opini masyarakat terkait dengan *dark jokes* menggunakan metode *Naïve Bayes Classifier* dan *Adaboost*?
2. Bagaimana performa dari *Naïve Bayes Classifier* dan *Naïve Bayes Classifier* dengan optimasi *Adaboost* sebagai analisa sentimen *dark jokes*?
3. Bagaimana hasil analisis sentimen negatif dan positif terkait *dark jokes*?

### 1.3 Batasan Masalah

Adapun Batasan – Batasan masalah dalam penelitian ini adalah:

1. Pengambilan data hanya dari media sosial Twitter.
2. Pengambilan data menggunakan kata kunci “*dark jokes, darkjokes*, lelucon, komedi hitam, satir, sarkas”.
3. Data yang digunakan sebatas *tweet* yang berasal dari Indonesia.
4. Data yang digunakan berupa sebuah kata maupun kalimat.
5. Data yang diambil mulai dari tanggal 12-12-2020 sampai dengan 18-09-2021.
6. Tidak memproses emotikon.

## 1.4 Tujuan Penelitian

1. Mengetahui bagaimana cara implementasi *Naïve Bayes Classifier* & *Adaboost* dalam mengklasifikasikan sentimen terkait opini masyarakat terhadap *dark jokes*.

- n pihak yang berhubungan dalam penelitian ini, diantaranya:  
 lah satu bahan rujukan mengenai klasifikasi teks untuk pen  
 a.  
 nambah wawasan dan pengetahuan dalam implementasi meto  
*ive Bayes Classifier*.  
 nakan sebagai batasan acuan masyarakat Indonesia dalam berk  
 nggunaan *dark jokes* sebagai kritik di media sosial.  
 enambah wawasan dalam implementasi metode *Naïve*  
*& Adaboost*.

Hasil dari penelitian yang dilakukan ini diharapkan memiliki manfaat bagi berbagai macam pihak yang berhubungan dalam penelitian ini, diantaranya:

- 6

## BAB 2

### TINJAUAN PUSTAKA

## 2.1 Tinjauan Penelitian Terdahulu

Dalam memberikan pemahaman dan interpretasi keterkaitan dengan penelitian terdahulu yang memiliki relevansi dengan penelitian yang akan dilakukan, dapat dilihat lebih lanjut dalam Tabel 2.1 sebagai berikut:

Tabel 2. 1 Daftar Penelitian Terdahulu

No	Judul	Metode	Hasil
1	ANALISIS SENTIMEN MASYARAKAT TERHADAP BRAND SMARTFREN MENGGUNAKAN NAIVE BAYES CLASSIFIER DI FORUM KASKUS (Zuhri and Alamsyah, 2017)	<i>Naïve Bayes Classifier</i>	<ol style="list-style-type: none"> <li>Memiliki tujuan untuk mengetahui sejauh mana sentimen pengguna brand smartfren di indonesia.</li> <li>Hasil dari analisis sentimen dan <i>word cloud</i> dapat dijadikan acuan oleh perusahaan sebagai informasi untuk meningkatkan kualitas produk.</li> <li>Hasil klasifikasi mendapatkan nilai precision sebesar 98.42%, recall 98.40%, dan akurasi sebesar 98.40% yang diartikan sistem dapat mengklasifikasi dengan baik.</li> </ol>
2	ANALISIS SENTIMEN MASYARAKAT TERHADAP E-COMMERCE PADA MEDIA SOSIAL MENGGUNAKAN METODE <i>Naïve Bayes Classifier</i> (NBC) DENGAN SELEKSI FITUR <i>Information Gain</i> (IG) (Syakuro, 2017)	<i>Naïve Bayes Classifier</i> dan <i>Information Gain</i>	<ol style="list-style-type: none"> <li>Memiliki tujuan untuk menganalisa sentimen negatif dan positif di Twitter mengenai beberapa <i>e-commerce</i> yaitu lazada, tokopedia, dan bukalapak.</li> <li>Hasil dari penggabungan metode NBC dengan seleksi fitur <i>Information Gain</i> dapat meningkatkan batas <i>threshold</i> sebesar 2200 dengan waktu</li> </ol>

			proses menjadi 514 detik lebih baik.
3	INTEGRASI METODE INFORMATION GAIN UNTUK SELEKSI FITUR DAN ADABOOST UNTUK MENGURANGI BIAS PADA ANALISIS SENTIMEN REVIEW RESTORAN MENGGUNAKAN ALGORITMA NAÏVE BAYES (Utami, 2015)	<i>Naïve Bayes Classifier, Information Gain, &amp; Adaboost</i>	<ol style="list-style-type: none"> <li>1. Bertujuan untuk meningkatkan <i>accuracy</i> dan mengurangi bias pada sentiment analisis menggunakan NBC.</li> <li>2. Hasil klasifikasi sentiment menggunakan metode NBC dibandingkan dengan NBC disertai <i>Adaboost</i> mengalami peningkatan <i>accuracy</i> sebesar 29%.</li> </ol>
4	ANALISIS SENTIMEN DAN KLASIFIKASI KATEGORI TERHADAP TOKOH PUBLIK PADA TWITTER (Fathan Hidayatullah and Sn, 2014)	<i>Naïve Bayes Classifier</i>	<ol style="list-style-type: none"> <li>1. Memiliki tujuan klasifikasi sentiment terhadap tokoh publik yang dianggap layak untuk mengikuti pemilu.</li> <li>2. Hasil <i>accuracy</i> menggunakan pembobotan TF menghasilkan <i>accuracy</i> sebesar 79.91% dan TF – IDF sebesar 79.68%.</li> </ol>
5	PENERAPAN METODE ADABOOST UNTUK MENGOPTIMASI PREDIKSI PENYAKIT STROKE DENGAN ALGORITMA NAÏVE BAYES (Byna and Basit, 2020)	<i>Naïve Bayes Classifier &amp; Adaboost</i>	<ol style="list-style-type: none"> <li>1. Memiliki tujuan untuk mengoptimasikan hasil prediksi dari metode <i>Naïve Bayes Classifier</i>.</li> <li>2. Metode <i>Naïve Bayes Classifier</i> dapat di optimasikan dengan <i>Adaboost</i> dan meningkatkan <i>accuracy</i> dari penelitian ini sebesar 0.5%.</li> </ol>
6	SENTIMENT ANALYSIS OF SARCASM IN SPOKEN LANGUAGE (Angraini, 2017)	<i>Descriptive Analysis</i>	<ol style="list-style-type: none"> <li>1. Memiliki tujuan untuk mengetahui pengaruh penggunaan sarkasme dalam kehidupan.</li> <li>2. Hasil dari penelitian ini sebagian besar orang dapat menerima sarkasme dan dijadikan pelajaran. Dari 26 orang diwawancara, sebesar 65% sering menggunakan</li> </ol>

			sarkasme, 25% tidak menggunakannya karena dapat menyakiti orang.
7	COGNITIVE AND EMOTIONAL DEMANDS OF BLACK HUMOUR PROCESSING: THE ROLE OF INTELLIGENCE, AGGRESSIVENESS AND MOOD (Willinger <i>et al.</i> , 2017)	<i>Clustering</i>	<ol style="list-style-type: none"> <li>1. Memiliki tujuan untuk melihat populasi yang dapat mengerti tentang <i>black humor</i> dilihat dari, kapasitas intelektual, mood, dan agresivitasnya.</li> <li>2. Hasil dari penelitian dari 156 orang yang dapat menerima <i>dark jokes</i> dari komik “The Black Book” yang dapat memahaminya hanya orang yang memiliki jenjang pendidikan tinggi.</li> </ol>
8	FACTORS OF BLACK HUMOR POPULARITY (Gubanov, Gubanov and Rokotyanskaya, 2018)	<i>Descriptive Analysis</i>	<ol style="list-style-type: none"> <li>1. Memiliki tujuan untuk melihat factor apa yang membuat <i>black humor</i> populer dan apa yang memebedakannya dengan <i>humor</i> biasa.</li> <li>2. Hasil dari penelitiannya ialah <i>black humor</i> merupakan komedi yang menertawakan mereka yang terluka namun mereka paham bahwa itu haja skedar candaan. <i>Black Humor</i> saat ini dapat mengubah peristiwa hidup menjadi sebuah ide-ide yang abstrak.</li> </ol>
9	SENTIMEN ANALISIS KOMENTAR TOXIC PADA GRUP FACEBOOK GAME ONLINE MENGGUNAKAN KLASIFIKASI NAÏVE BAYES (Sidiq, Dermawan and Umaidah, 2020)	<i>Naïve Bayes Classifier</i>	<ol style="list-style-type: none"> <li>1. Memiliki tujuan untuk klasifikasi komentar toxic dan non-toxic di grup game online facebook.</li> <li>2. Metode <i>Naïve Bayes Classifier</i> dapat mengklasifikasikan toxic dan non-toxic dengan <i>accyuracy</i> sebesar 75%.</li> </ol>











podcastnya yang berjudul “Musuh Masyarakat” dimana kedua komedian tersebut membawakan materi *dark jokes* seperti “Selama ini mungkin cara kita memperlakukan pengemis itu istilahnya kayak ngasih umpan ke ikan, jadi setiap ada pengemis kasih uang, saya punya ide kalo kita kasih pengemis mending kasih kail pancingnya jadi kalo ada pengemis anda tempeleng pake tabloid lowongan kerja”(Gunawan, 2021).

*Pseudo labeling* merupakan metode *semi-supervised learning* yang difungsikan untuk labeling data dengan teknik yang dibuat dari data yang sudah terlabeli untuk melabeli data yang belum terlabel (Fudholi, DThomas H;Juwairi, 2019). Penggunaan metode ini memiliki cara kerja dari data yang sudah terlabel dan data yang belum terlabel maka didapatkan label yang sebenarnya, dengan cara kerja tersebut metode ini merupakan salah satu teknik *semi-supervised learning*.

Hasil pelabelan data dilakukan secara manual dan melibatkan campur tangan manusia. Dalam penelitian kali ini validasi manusia sangat penting dikarenakan data yang didapat ialah data yang belum terlabel dengan tepat. Sentimen analisis memiliki banyak masalah dalam pemrosesan datanya seperti kata yang ambigu, konteks dari suatu kalimat, typo, dan *slang word*.

Tahapan *pre-processing text* ini merupakan salah satu tahapan yang penting di *Text Mining*. Tahapan ini perlu dilakukan guna mengolah data berupa teks kedalam berbagai dokumen agar mempermudah proses klasifikasi. Umumnya dokumen yang nanti diproses memiliki teks yang tidak terstruktur menyebabkan kesulitan dalam proses klasifikasi sehingga perlu dilakukan tahap *pre-processing*. Alasan lainnya adalah ada beberapa data teks yang dimana data tersebut tidak merepresentasikan isi dari dokumen tersebut. Tahapan *pre-processing* ada beberapa tahap yaitu:

Tahapan ini digunakan sebagai pembersih spesial karakter yang biasa digunakan sebagai tanda baca seperti titik, koma, hashtag, karakter web,at, dan lain-lain.

## 2. Case Folding

Proses ini merupakan suatu proses yang berguna sebagai pengubah seluruh format huruf menjadi huruf kecil secara menyeluruh, hal ini dilakukan untuk menyamakan arti dalam sebuah kalimat (Manning, Raghavan and Schütze, 2008). Dalam penggunaan media sosial Twitter kebanyakan masyarakat menggunakan huruf kapital semua sebagai penekanan dari makna yang disampaikan.

### 3. Tokenizing

Proses selanjutnya yang dilakukan adalah *Tokenizing*. Proses ini merupakan proses untuk memisahkan teks dari paragraf maupun kalimat menjadi bagian kecil yang disebut dengan *token*, *token* tersebut akan dilanjutkan untuk proses analisis (Kalra and Aggarwal, 2018). Proses *Tokenizing* ini diperlukan untuk memudahkan dalam pembobotan suatu kata.

#### 4. Filtering

Kalimat yang sudah menjadi bagian kecil atau yang bisa disebut *token* dari proses *Tokenizing* akan dihilangkan kata-kata yang tidak memiliki makna seperti kata hubung, ini disebut dengan proses *Filtering* (Kalra and Aggarwal, 2018). Dalam proses *Filtering* ini menggunakan algoritma *stopword*.

## 5. Stemming

Tahapan ini adalah salah satu tahapan penting dalam *pre-processing*. *Stemming* merupakan proses menghilangkan *inflection* dalam suatu kata untuk dikembalikan kebentuk dasarnya sehingga kata yang memiliki imbuhan akan dihapus (Kalra and Aggarwal, 2018).

### 2.2.7 Pembobotan kata

Pembobotan kata dalam proses *Text Mining* bermaksud untuk memberikan nilai untuk term atau kata dari dokumen. Nilai yang diberikan pada kata tergantung dengan metode apa yang akan digunakan (Deolika, Kusri and Luthfi, 2019).

### 2.2.8 Naïve Bayes Classifier

$$P(A|B) = \frac{(P(B|A) * P(A))}{P(B)} \quad (1)$$

$P(B)$  : Probabilitas A

proses klasifikasi teks, maka pendekatan *Naïve Bayes Classification* yang mempunyai probabilitas tertinggi ( $C_{MAP}$ ) dengan persamaan yaitu :

$$C_{MAP} = p(c_j) p(h_i/c_j)$$

probabilitas maksimal  
probabilitas munculnya hipotesis dalam dokumen  
probabilitas munculnya hipotesis dalam kelas  
yang di klasifikasi

$$C_{MAP} = p(cj) p(hi/cj) \quad (3)$$

$C_{MAP}$  : Probabilitas maksimal  
 $p(c_j)$  : Probabilitas munculnya hipotesis dalam dokumen  
 $p(h_i|c_j)$  : Probabilitas munculnya hipotesis dalam kelas  
 $c_j$  : Kelas yang di klasifikasi

1. Menyiapkan data yang sudah melalui hasil pre-processing.
2. Menghitung probabilitas data jenis training.

3. Menghitung *term frequency* data training.
4. Menghitung probabilitas kelas dari data training.









### Tabel 2. 2 Tabel Confussion Matrix

Aktual \ Prediksi	Positif	Negatif
Positif	True Positif (TP)	False Positif (FP)
Negatif	False Negatif (FN)	True Negatif (TN)

Dari tabel 2.2 di atas sesuai kegunaannya yaitu menghitung *accuracy*, *recall*, dan *precision* dapat dirumuskan sebagai berikut:

### 1. Accuracy

Akurasi yaitu rasio perhitungan dari jumlah keseluruhan data yang benar dari total data.

$$Accuracy = \frac{TP + TN}{TP + FP + TN + FN} \quad (14)$$

## 2. Recall

Recall ialah rasio dari suatu prediksi benar negatif maupun positif dibandingkan dengan total data yang benar negatif maupun positif. Recall positif dapat disebut sebagai sensitivity, sedangkan specificity sebagai recall negatif.

$$\text{Sensitivity} = \frac{\text{TP}}{\text{TP} + \text{FN}} \quad (15)$$

$$\text{Specificity} = \frac{\text{TN}}{\text{TN} + \text{FP}} \quad (16)$$

### 3. Precision

Precision ialah perhitungan prediksi positif maupun negatif yang benar dibandingkan dengan total hasil prediksi.

$$\text{Precision positif} = \frac{TP}{TP + FP} \quad (17)$$

$$\text{Precision negatif} = \frac{TN}{TN + FN} \quad (18)$$

#### 4. *F-Measure*

F-Measure ialah perhitungan akurasi dari akurasi tes yang mengukur perbandingan rata-rata dari *precision* dan *recall*.

$$F - \text{Measure positif} = 2 \cdot \frac{\text{Precision Positif} \cdot \text{Recall Positif}}{\text{Precision Positif} + \text{Recall Positif}} \quad (19)$$

$$F - \text{Measure negatif} = 2 \cdot \frac{\text{Precision Negatif} \cdot \text{Recall Negatif}}{\text{Precision Negatif} + \text{Recall Negatif}} \quad (20)$$

## 2.3 Integrasi Keilmuan

Mengenai tema skripsi yang diambil dengan pandangan islam dengan cara mengintegrasikan ayat dari Al-Qur'an maupun hadits, penulis menggunakan ayat Al-Qur'an dan hadits berikut yang didapatkan dari hasil wawancara ustadz Azahri, beliau ialah ustadz sekaligus ketua bidang keagamaan di perumahan GKGA kedanyang RT 3. Wawancara yang dilaksanakan membahas mengenai pandangan tentang sentimen analis dan *dark jokes* menurut perspektif islam.

Hubungan integrasi penelitian ini dengan nilai islam tertuang dalam Al-Qur'an . Allah swt telah berfirman didalam Al-Qur'an surat Al-Hujurat ayat 6. Isi kandungan ayat tersebut menjelaskan bahwa segala sesuatu yang diterima haruslah diperiksa terlebih dahulu kebenarannya. Jika sesuatu tersebut tidak dipastikan kebenarannya, maka orang lain pun yang akan dirugikan.

يَأْتِيهَا الَّذِينَ ءَامَنُوا إِن جَاءَكُمْ فَاسِقٌ بِنَبَأٍ فَتَبَيَّنُوا أَن تُصِيبُوا قَوْمًا بِجَهْلَةٍ فَتُصْبِحُوا عَلَىٰ مَا  
فَعَلْتُمْ نَادِمِينَ

Artinya : “Hai orang-orang yang beriman, jika datang kepadamu orang fasik membawa suatu berita, maka periksalah dengan teliti agar kamu tidak menimpakan suatu musibah kepada suatu kaum tanpa mengetahui keadaannya yang menyebabkan kamu menyesal atas perbuatanmu itu”. (QS. Al-Hujurat: 6).

Pada dasarnya setiap individu haruslah dapat mencari sumber yang dapat dipercaya dan hal tersebut dapat dijadikan dasar dalam mendapatkan berita yang benar. Sedangkan langkah dan sikap yang diambil oleh kaum muslimin hendaknya agar tidak terburu-buru untuk bertindak akan tetapi dengan cara memeriksa kebenaran dari berita tersebut. Dikhawatirkan jika mempercayai suatu berita tanpa memeriksa kebenarannya, tindakan tersebut dianggap zhalim dan dapat merugikan suatu kaum. Sikap tersebut yang nantinya akan menyebabkan penyesalan akibat dari tidak mempertahankan keadilan dan kebenaran.

Penelitian ini juga membahas mengenai komedi atau humor yang berjenis dark jokes. Dalam Hadits yang diriwayatkan oleh Al-Tirmidzi (no. hadis: 2315), dan Abu Dawud (no. hadis: 4990) dari hadis Mu'awiyah bin Haidah radiyallahu 'anhu, bahwa Rasulullah shallallahu 'alaihi wasallam bersabda,

وَيْلٌ لِلَّذِي يُحَدِّثُ فَيَكْذِبُ لِيُضْحِكَ بِهِ الْقَوْمَ وَيْلٌ لَهُ وَيْلٌ لَهُ

“Celakalah orang yang berbicara kemudian ia berdusta agar manusia tertawa karenanya, maka celakalah ia, maka celakalah ia.”

Hadist diatas mengandung peringatan terhadap orang yang berdusta dalam bercanda. Pada dasarnya bercanda itu hukumnya mubah, namun yang menjadikannya haram dan dilarang adalah saat membuat orang tertawa akan tetapi mengandung informasi yang bohong dan mengada-ada. Informasi yang disampaikan harusnya bersifat ilmiah, memiliki sumber yang valid dan dapat dipertanggungjawabkan. Hal ini pula yang menjadikan penulis ingin mencari tahu lebih mengenai efek dari penggunaan dark jokes di media sosial Twitter.

## BAB 3

## METODOLOGI PENELITIAN

### 3.1 Jenis dan Sumber Data

Data dalam penelitian ini merupakan kumpulan dataset mengenai opini masyarakat yang berhubungan dengan *dark jokes* pada akhir tahun 2020 – 2021. Data didapatkan dari [www.api.twitter.com](https://www.api.twitter.com), situs tersebut menyediakan data yang dapat di publikasikan di Twitter seperti tweet, retweet, nama pengguna, maupun tanggal tweet. Untuk mendapatkan datanya dilakukan tahapan *crawling* data dari situs tersebut menggunakan *Python programming*. Dalam tahapan *crawling* ini ditentukan kata kunci yang ingin di cari dan dari negara mana *tweet* tersebut. Pada penelitian ini membutuhkan dataset yang berupa opini masyarakat Indonesia terkait dengan *dark jokes* maka dari itu kata kunci yang digunakan ialah “*dark jokes*” dan mengambil tweet dari Indonesia saja.

### 3.2 Metodologi Penelitian

Penelitian yang dilakukan ini akan menggunakan empat tahapan yang akan dilalui yaitu tahap awal penelitian, pemerosesan data, klasifikasi teks dan yang terakhir tahap analisis data. Empat langkah tersebut dapat ditinjau lebih jelas pada gambar berikut.



Pada gambar 3.1 terdapat empat proses inti dalam penelitiannya yaitu tahapan awal, data *processing*, klasifikasi teks, dan analisa hasil. Berikut akan di jelaskan proses-proses tersebut.

### 3.2.1 Tahap Penelitian Awal

Dalam tahapan awal ini memiliki tiga tahapan yang akan dilalui yakni sebagai berikut.

### 1. Perumusan Masalah

Tahap pertama ini ialah proses yang menjelaskan mengapa penelitian ini dilakukan, dari latar belakang tersebut dapat dibentuk menjadi rumusan masalah yang tertera pada bab pertama. Masalah yang diambil dalam penelitian ini ialah mengklasifikasi opini positif dan negatif masyarakat terhadap *dark jokes* menggunakan metode *Naïve Bayes Classifier*.

## 2. Studi Pustaka

Tahap kedua ialah proses studi literatur untuk menganalisis kebutuhan dari masalah yang dialami untuk mengimplementasikan algoritma apa yang akan diimplementasikan kedalam sebuah sistem, serta teori apa saja yang akan digunakan untuk penjabaran sistem ini.

### 3. Crawling Data

Tahap terakhir yakni pengumpulan data atau biasa disebut dengan *crawling* data. Untuk melakukan *crawling* data Twitter harus memiliki akun twitter developer agar dapat mengakses API dari Twitter itu sendiri. Setelah mendapatkan kode akses API Twitter langkah selanjutnya ialah melakukan proses *crawling* dengan *Python programming* dengan cara memasukkan kode akses yang sudah di dapatkan serta kata kunci yang akan dicari. Data Processing

Dalam tahapan ini data yang sudah didapatkan akan di olah lebih lanjut. Tahapan pengolahan ini terdiri dari beberapa tahapan, tahapan ini biasa disebut dengan data *preprocessing* namun sebelum dilakukan tahapan tersebut ada tahapan yang harus di lakukan yaitu menghapus duplikasi data, berikut adalah penjelasan setiap tahapan.

#### 4. Menghapus duplikasi data

Proses pertama yaitu penghapusan duplikasi data. Proses ini dilakukan karena data yang diambil merupakan seluruh teks opini yang ada di Twitter di dalam Twitter sendiri terdapat banyak data dari retweet masyarakat sehingga retweet tersebut akan menimbulkan banyaknya duplikasi data muncul sehingga perlu dilakukan penghapusan data tersebut.

## 5. Preprocessing

Proses kedua ini memiliki beberapa tahap lagi untuk mengolah data mentah menjadi data yang lebih bersih, *preprocessing* memiliki lima tahap yang digunakan untuk pemrosesannya sebagai berikut:

a. *Cleansing*

Tahapan awal *preprocessing* ini bertujuan untuk menghapus karakter atau simbol-simbol tanda baca dan menghapus alamat web maupun html yang tidak memiliki makna agar tidak mengganggu proses klasifikasi.

### b. Case Folding

Tahapan selanjutnya berfungsi sebagai teks yang di dapat dari Twitter memiliki huruf yang besar dan huruf kecil yang bercampuran atau tidak tersusun sesuai dengan cara penulisan, hal tersebut dapat mengganggu proses klasifikasi maka dari itu perlu di lakukan proses *case folding* ini untuk merubah format semua teks menjadi *lowercase*.

### c. Tokenizing

Pada tahapan ini teks data dari Twitter akan diberikan pemisah setiap kata sehingga berubah menjadi token-token atau potongan kata.

#### d. *Normalization*

Tahapan ini digunakan untuk membenarkan kata seperti kata yang memiliki arti sama namun memiliki gaya penulisan yang berbeda hal ini biasa di sebabkan karena adanya penyingkatan, *typo*, maupun bahasa yang biasa digunakan sehari hari atau bahasa gaul.

### e. *Filtering*

Kalimat yang sudah menjadi token-token ini akan diproses lagi untuk menghilangkan kata yang tidak memiliki makna seperti kata hubung. Untuk proses tersebut menggunakan algoritma *stoplist* yang sudah ada *library* NLTK,





## 1. Pembobotan Kata

Untuk pembobotan kata di penelitian ini menggunakan metode TF IDF. Dalam tahap ini setiap kata akan memiliki bobot masing-masing jadi tiap kata yang ada akan dirubah menjadi nilai. Proses ini menggunakan library *Python* yaitu TF IDF *Vectorizer*. Tabel 3.8 merupakan contoh perhitungan TF IDF secara manual.

Tabel 3. 1 Pembobotan Kata TF IDF

Token	TF	W (TF*IDF)				df	D /d f	IDF (log D/df )
	D1	D2	D3	D4				
Bisa	1 0.301	0 0	0 0	1 0.301	2	2	0.301	
Suka	1 0.301	0 0	1 0.301	0 0	2	2	0.301	
Sopan	0 0	1 0.602	0 0	0 0	1	4	0.602	
Takut	0 0	0 0	0 0	1 0.602	1	4	0.602	
Terima	0 0	0 0	0 0	1 0.602	1	4	0.602	

## 2. Naïve Bayes Classifier

Proses ini berguna sebagai proses klasifikasi data menggunakan metode *Naïve Bayes Classifier* dengan menghitung probabilitas teks Twitter untuk mendapatkan nilai probabilitas maximum ( $C_{MAP}$ ) (3) dari sentimen positif maupun negatif. Untuk tahapan klasifikasi lebih detail ada pada BAB II Sub Bab 2.2.8.

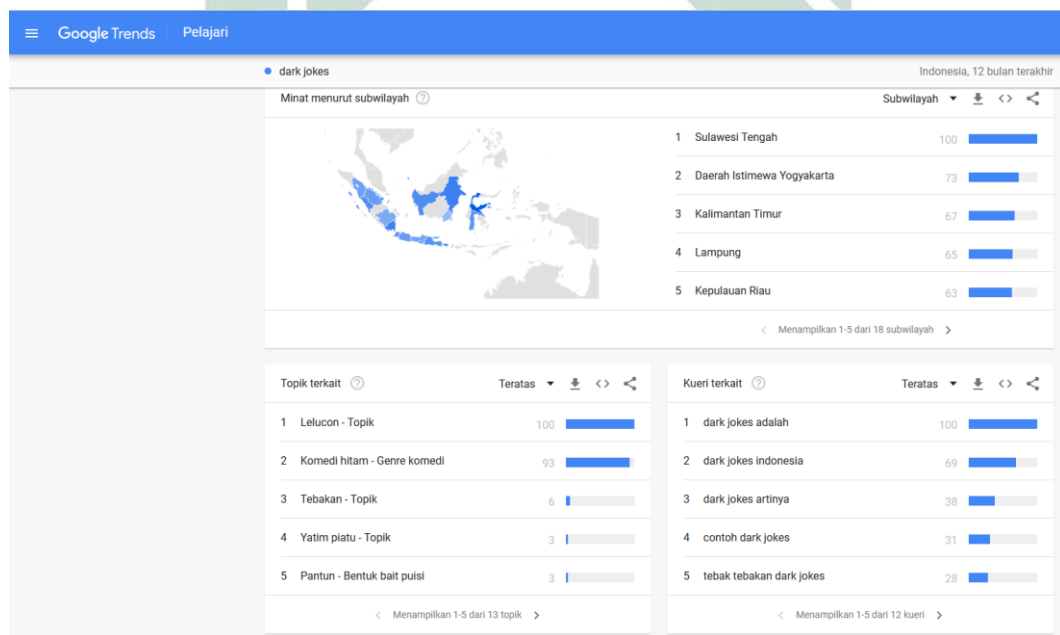


## HASIL DAN PEMBAHASAN

Pada bab ini akan dipaparkan proses dan hasil yang diperoleh dari penelitian ini mulai dari tahap awal hingga tahap akhir.

## 4.1 Crawling Data Twitter

Data dikumpulkan dengan melakukan pengambilan data dari Twitter berdasarkan kata kunci awal yang sudah diambil dari topik penelitian dan kata kunci yang terkait dengan topik penelitian yang didapat dari google trends. Beberapa topik terkait yang diambil dari google trends hanya 3 kata kunci teratas yang dapat dilihat pada Gambar 4.1.



Gambar 4. 1 Topik Terkait Dark Jokes (Google Trends)

Dari gambar 4.1 didapatkan kata kunci 3 teratas yang berhubungan dengan *dark jokes* yaitu “lelucon”, “komedi hitam”, dan “tebakan”. Maka kata kunci yang digunakan yaitu “*darkjokes*”, “*dark jokes*”, “satir”, “sarkas”, dan 3 kata kunci yang didapat dari Google Trends. Dalam proses crawling data ini menggunakan APIv2 Twitter dan Tweepy, hasil data yang didapat sebesar 21.636 data. Contoh data tweets yang didapatkan dari proses crawling data ditampilkan pada Tabel 4.1.











T6	gitu aja tersinggung semua bisa dibercandainkan kalian sih gak punya selera humor sjw kamu sok belain orang
T7	asli orang kalau dikritik alesannya gak ngerti dark jokes lo dark jokes kadang cuman bisa di pahami sama beberapa orang

#### 4.3.4 Normalization

Dalam tahapan ini digunakan bahasa pemrograman Python untuk memproses data mentah yang berhasil di dapat. Gambar 4.5 adalah cuplikan kode dari tahapan ini.

```
normalized_word = pd.read_excel("normalisasi.xlsx")

normalized_word_dict = {}

for index, row in normalized_word.iterrows():
    if row[0] not in normalized_word_dict:
        normalized_word_dict[row[0]] = row[1]

def normalized_term(document):
    return [normalized_word_dict[term] if term in normalized_word_dict else term for term in document]

data['tweet_normalized'] = data['tweet_tokens_wSW'].apply(normalized_term)

data['tweet_normalized'].head(10)
```

Gambar 4. 5 Cuplikan Kode *Normalization*

Tahapan ini ialah tahapan yang digunakan untuk membenarkan kata seperti kata yang memiliki arti sama namun memiliki gaya penulisan yang berbeda atau membenarkan penyingkatan kata kata. Tabel 4.5 ialah isi dari tahapan *normalization*.

Tabel 4. 5 Teks Twitter Hasil Normalization

T	Tweet
T1	ya   biarin   kalau   saya   suka   dark   jokes   ini   setidaknya   saya   bisa   sesuaikan   dengan   tingkat   dibawa   perasaan   kamu
T2	tidak   usah   alasan   menggunakan   dark   jokes   ya   kamu   yang   tidak   tahu   sopan   santun
T3	yang   suka   sama   dark   jokes   sini   saya   tampar
T4	punya   banyak   dark   jokes   yang   mau   dikeluarin   tapi   takut   tidak   ada   yang   bisa   terima
T5	jokes   mu   itu   tidak   ada   lucunya   loh   cuman   hati   nuranimu   saja   yang   gelap   dark
T6	begitu   saja   tersinggung   semua   bisa   dibercandainkan   kalian   sih   tidak   punya   selera   humor   sjw   kamu   sok   belain   orang

T7	asli   orang   kalau   dikritik   alasannya   tidak   ngerti   dark   jokes   lo   dark   jokes   kadang   cuman   bisa   di   pahami   sama   beberapa   orang
----	---

### 4.3.5 Filtering

Dalam tahapan ini digunakan bahasa pemrograman Python untuk memproses data mentah yang berhasil di dapat. Gambar 4.6 adalah cuplikan kode dari tahapan ini.

```
list_stopwords = stopwords.words('indonesian')

# ----- manually add stopword -----
# append additional stopword
list_stopwords.extend(["yg", "dg", "rt", "dgn", "ny", "d",
                        'amp', 'bian',
                        'nih', 'sih',
                        'si', 'tuh',
                        'aja', 'n', 't',
                        'nyg', 'pen', 'u', 'nan', 'loh', 'rt',
                        'yah'])

# convert list to dictionary
list_stopwords = set(list_stopwords)

#remove stopword pada list token
def stopwords_removal(words):
    return [word for word in words if word not in list_stopwords]

data['tweet_tokens_WSW'] = data['tweet_tokens'].apply(stopwords_removal)
```

Gambar 4. 6 Cuplikan Kode *Filtering*

Selanjutnya ialah proses *filtering* yakni melakukan pembuangan kata-kata yang tidak bermakna. Tabel 4.6 merupakan hasil dari proses *filtering*.

Tabel 4. 6 Teks Twitter Hasil Filtering

T	Tweet
T1	suka   setidaknya   bisa   sesuaikan   tingkat   dibawa   perasaan
T2	tidak   usah   alasan   menggunakan   tidak   tahu   sopan   santun
T3	suka   sama   sini   tampar



Tabel 4. 7 Teks Twitter Hasil Stemming

T	Tweet
T1	suka   tidak   bisa   sesuai   tingkat   bawa   rasa
T2	tidak   usah   alas   guna   tidak   tahu   sopan   santun
T3	suka   sama   sini   tampar
T4	punya   banyak   mau   keluar   tapi   takut   tidak   ada   bisa   terima
T5	tidak   ada   lucu   cuman   hati   nurani   saja   gelap
T6	begitu   saja   singgung   semua   bisa   canda   kalian   tidak   punya   selera   humor   bela   orang
T7	asli   orang   kritik   alasan   tidak   erti   kadang   cuma   bisa   paham   sama   berapa   orang

#### 4.4 Data Labeling

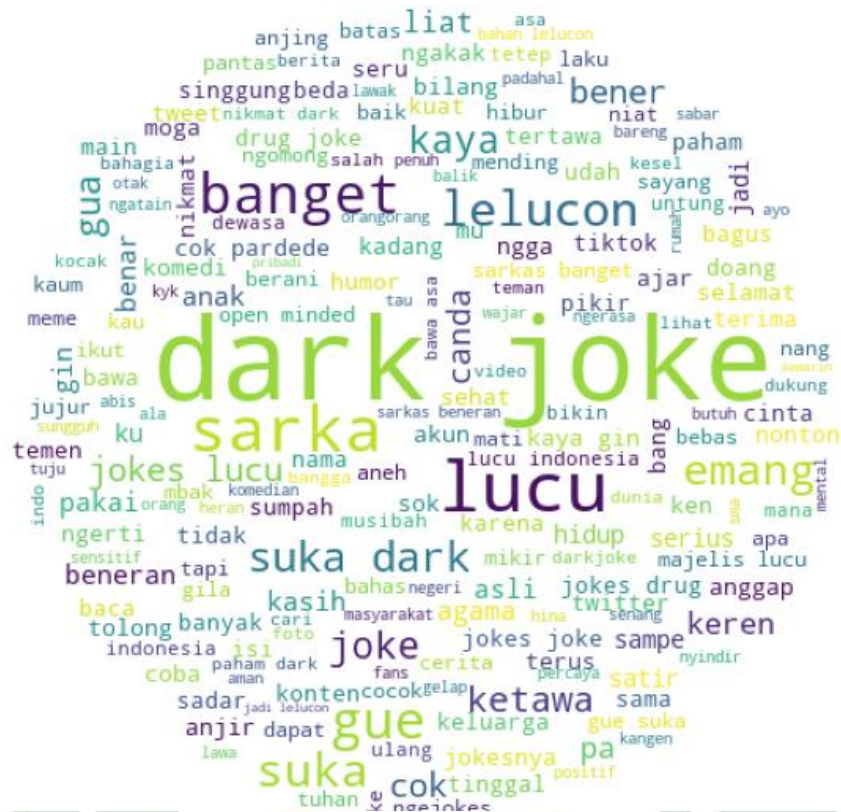
Penelitian ini menggunakan teknik *pseudo labeling* yaitu labeling yang menggunakan tiga model yaitu menggunakan label 1 *SentiStrength*, label 2 *SentiWord*, dan label 3 *TextBlob*. Indikator 1 sebagai positif, -1 sebagai negatif, 0 sebagai netral, dan X merupakan dataset yang tidak akan digunakan karena dari ketiga model menghasilkan label yang berbeda sehingga data menjadi ambigu dan bias. Dari 11.539 data yang berhasil dilabelkan menjadi 7.692 karena menghapus data yang berlabelkan X. Tabel 4.8 merupakan hasil dari *Pseudo Labeling* yang didapat.

Tabel 4. 8 Hasil Pseudo Labeling

T	Tweet	Label			Pseudo Label
		1	2	3	
T1	Ya biarin kalau gue suka dark jokes ini seenggaknya gue bisa sesuaikan dengan tingkat kebaperan lo	1	1	1	1







Pada gambar 4.8 merupakan hasil dari kumpulan kata yang ada pada sentimen positif. Kata yang dominan muncul ialah “*dark joke*” yang merupakan salah satu kata kunci yang digunakan, menjadikan kata tersebut paling sering muncul diikuti oleh kata “lucu” yang menyatakan bahwa pengguna Twitter terhibur dengan *dark jokes*.





pakar. Dari hasil pelabelan menggunakan *Pseudo Labeling* dan divalidasi dengan label pakar mendapatkan akurasi sebesar 70%, dari 200 data yang di validasi ada 160 data yang benar.

Tabel 4. 9 Validasi Data Labeling

T	Tweet	Label Pakar			Pseudo Label	B/S
		1	2	3		
T1	yang suka sama dark jokes sini saya tampar	1	-1	-1	-1	B
T2	txtdriatthaphan ajarin nge darkjokes	1	0	1	0	S
T3	Bayangin malaikat izrail gabut terus nyabut nyawa lu darkjokes darkjoke	1	1	1	1	B
T4	helloverrrr mastourbabe agakbete ntar dikasih dark jokes beneran bilanganya gapunya perasaan hati nuraninya dimana wkwwk	1	1	1	1	B
T5	ailuvsomatcha bukan apa sadar ga sih anak tiktok banyak yg toxic parah deh. ya ngatain fisik lah dikit dijadiin konten atau candaan	-1	-1	-1	1	S
T6	jawafess himbauan buat sjw kalau liat hal seperti ini masih ketawa tidak usah sok menghakimi dark Jokes wkwwk	-1	1	1	1	B
T7	sosmedkeras ini gak lucu sumpah. dark jokes yang salah tempat	1	0	1	-1	S
T8	ya biarin kalau gue suka dark jokes ini seenggaknya gue bisa	1	1	1	1	B

	sesuaikan dengan tingkat kebaperan lo					
T9	mothra guyonan murahan kaum dark jokes	-1	-1	-1	1	S





Tabel 4. 11 Hasil *Stratified Cross Validation* Kondisi Pertama

Fold	NBC				NBC dengan Adaboost			
	Accuracy	Precision	Recall	F1	Accuracy	Precision	Recall	F1
1	0,592	0,668	0,592	0,574	0,447	0,537	0,447	0,367
2	0,605	0,683	0,605	0,573	0,452	0,365	0,452	0,365
3	0,738	0,766	0,738	0,742	0,566	0,762	0,566	0,556
4	0,8	0,807	0,8	0,802	0,747	0,818	0,747	0,74
5	0,819	0,828	0,819	0,819	0,783	0,832	0,783	0,781
6	0,872	0,875	0,872	0,873	0,822	0,855	0,822	0,821
7	0,88	0,885	0,88	0,879	0,858	0,877	0,858	0,858
8	0,791	0,795	0,791	0,792	0,738	0,817	0,738	0,728
9	0,783	0,789	0,783	0,782	0,75	0,782	0,75	0,744
10	0,694	0,697	0,694	0,682	0,322	0,421	0,322	0,338
Rata - Rata	0,757	0,779	0,757	0,751	0,648	0,706	0,648	0,629

## 2. *Cross Validation* data kondisi kedua

Kondisi kedua menggunakan dataset asli yang tidak dikondisikan yaitu 3.698 positif, 1.279 netral, dan 2.715 negatif dengan total data 7.692 data. Hasil dari *cross validation* pada kondisi ini dapat dilihat pada Tabel 4.12.

Tabel 4. 12 Hasil *Stratified Cross Validation* Kondisi Kedua

Fold	NBC				NBC dengan Adaboost			
	Accuracy	Precision	Recall	F1	Accuracy	Precision	Recall	F1
1	0,648	0,733	0,648	0,585	0,663	0,701	0,663	0,654
2	0,661	0,731	0,661	0,604	0,716	0,758	0,716	0,69
3	0,607	0,703	0,607	0,537	0,676	0,72	0,676	0,664
4	0,65	0,722	0,65	0,593	0,707	0,739	0,707	0,68
5	0,655	0,731	0,655	0,605	0,684	0,723	0,684	0,645
6	0,663	0,747	0,663	0,609	0,713	0,745	0,713	0,707
7	0,668	0,754	0,668	0,609	0,713	0,753	0,713	0,703
8	0,605	0,697	0,605	0,543	0,55	0,554	0,55	0,54
9	0,577	0,628	0,577	0,503	0,54	0,631	0,54	0,511
10	0,622	0,701	0,622	0,562	0,49	0,735	0,49	0,499
Rata - Rata	0,635	0,714	0,635	0,575	0,645	0,705	0,645	0,629

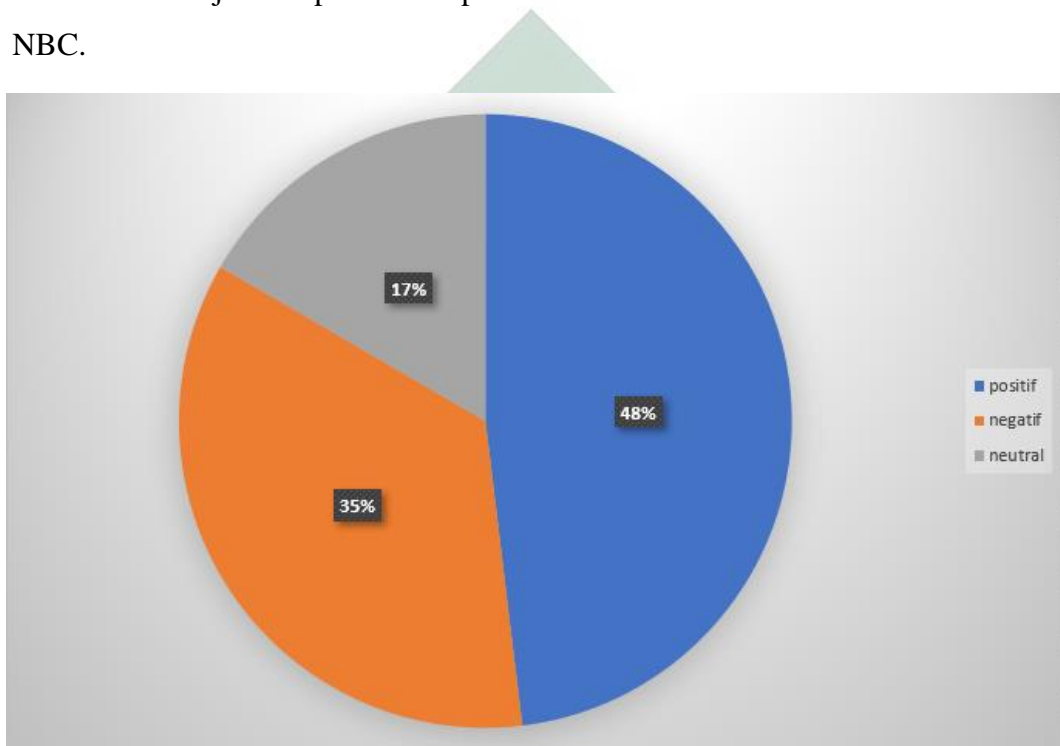




## 4.8 Analisis Hasil Sentimen

Dari 7.692 data yang terkumpul hasil klasifikasi sentimen didapatkan 2.715 label negatif, 1.279 label neutral, dan 3.698 label positif. Berdasarkan dari hasil dari klasifikasi sentimen tersebut, disimpulkan bahwa pandangan masyarakat terhadap *dark jokes* lebih cenderung pada sentimen positif.

Pada gambar 4.5 ditampilkan hasil pelabelan menggunakan *pie chart*. *Chart* tersebut menunjukkan presentase perbedaan antara label sentimen hasil metode NBC.



Gambar 4. 13 *Pie Chart* Sentimen *Dark Jokes*

Sesuai hasil yang didapat dari gambar 4.5, *tweet* yang memiliki label positif lebih dominan daripada 2 label yang lain meskipun perbedaannya tidak terlalu jauh dengan label negatif. Hal ini disebabkan masyarakat pengguna Twitter memandang *dark jokes* lebih kearah yang positif. Dari banyaknya label positif terhadap *dark jokes* bisa dibilang masyarakat Twitter juga sudah cerdas dalam memandang *dark jokes*, faktor pendorong masyarakat semakin cerdas karena untuk memahami suatu *dark jokes* dikarenakan untuk memahami suatu hal tabu memang membutuhkan kecerdasan yang lebih (Willinger *et al.*, 2017).

Sesuai hasil analisis penelitian ini, bisa disarankan untuk para pelaku stand up comedian atau para masyarakat yang ingin membuat atau mengutarakan *dark jokes*









- Ford, T. E., Richardson, K. and Petit, W. E. (2015) 'Disparagement humor and prejudice: Contemporary theory and research', *Humor*, 28(2), pp. 171–186. doi: 10.1515/humor-2015-0017.
- Freund, Y. (2001) 'An Adaptive Version of the Boost by Majority Algorithm', *Machine Learning*, 43(3), pp. 293–318. doi: 10.1023/A:1010852229904.
- Fudholi, D;Thomas H;Juwairi, K. (2019) 'Pemanfaatan\_Teknik Media Jamur'.
- Gubanov, N. N., Gubanov, N. I. and Rokotyanskaya, L. (2018) 'Factors of Black Humor Popularity', 283(Cesses), pp. 379–383. doi: 10.2991/cesses-18.2018.85.
- Gunawan, E. F. (2021) 'Komodifikasi Kemiskinan pada Konten Pocast Dark Joke Coki Pardede dan Tretan Muslim'.
- Hearst, M. A. (1999) 'Untangling text data mining', pp. 3–10. doi: 10.3115/1034678.1034679.
- Hidayah, I. N. . (2020) 'Strategi Penciptaan Humor oleh Sadana Agung pada Stand Up Indonesia Season 6 Kompas TV Kajian: Anatomy of Humor Asa Berger'.
- Hu, Wei and Hu, Weiming (2005) 'Network-based intrusion detection using adaboost algorithm', *Proceedings - 2005 IEEE/WIC/ACM International Conference on Web Intelligence, WI 2005*, 2005, pp. 712–717. doi: 10.1109/WI.2005.107.
- Hutapea, T. & I. (2018) 'Penerapan Algoritma Modified K-Nearest Neighbour Pada Pengklasifikasian Penyakit Kejiwaan Skizofrenia', *Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer*, 2(10), pp. 3957–3961.
- Kalra, V. and Aggarwal, R. (2018) 'Importance of Text Data Preprocessing & Implementation in RapidMiner', *Proceedings of the First International Conference on Information Technology and Knowledge Management*, 14(January), pp. 71–75. doi: 10.15439/2017km46.
- Kementerian Komunikasi dan Informatika (2020). Available at: [https://kominfo.go.id/content/detail/2366/indonesia-peringkat-lima-pengguna-twitter/0/sorotan\\_media](https://kominfo.go.id/content/detail/2366/indonesia-peringkat-lima-pengguna-twitter/0/sorotan_media) (Accessed: 1 October 2020).
- Kusrorong, Neutrino Sae B. Sina, D. R. R. (2019) 'KAJIAN MACHINE



4) ‘KRITIK SOSIAL DALAM KOMEDI (Studi Kasus Stand-Up  
Mangan Mahasiswa Fakultas Ilmu Sosial Universitas Negeri Jakarta)’,  
AL DALAM ROMAN MOMO KARYA MICHAEL ENDE (ANALISIS  
ASTRA) SKRIPSI Diajukan.

5) 'TEXT MINING UNTUK KLASIFIKASI KONTEN ABSTRAK BAHASA INGGRIS MENGGUNAKAN METODE REDUKSI DIMENSI DENGAN NAIVE BAYES', 151(2), pp. 10–17.

and Swift, J. (2001) 'British Satire', *English*, (1736), pp. 569–572.

., Raghavan, P. and Schütze, H. (2008) *Introduction to Information*  
Cambridge University Press. Available at:  
[google.co.id/books?id=t1PoSh4uwVcC](https://books.google.co.id/books?id=t1PoSh4uwVcC).

*al.* (2018) ‘How to evaluate sentiment classifiers for Twitter time-series data’. *PLoS ONE*. doi: 10.1371/journal.pone.0194317.

Widya Sihwi, S. and Doewes, A. (2016) 'Analisis Sentimen terhadap Calon Presiden Indonesia 2014 berdasarkan Opini dari digunakan Metode Naive Bayes Classifier', *Jurnal Teknologi & Informatika*. doi: 10.20961/its.v2i2.630.

N. (2020) 'Pengembangan sistem prediksi daftar ulang calon baru menggunakan metode ADABOOST'. Available at: [insby.ac.id/42063/](https://ojs.umsida.ac.id/index.php/insby.ac.id/42063/).

Farida, Y. and Puspitasari, D. (2016) 'Klasifikasi Menggunakan  
id Bayesian-Neural Network (Studi Kasus: Identifikasi Virus  
*Jurnal Matematika 'MANTIK'*, 1(2), p. 38. doi:  
tik.2016.1.2.38-43.

Dermawan, B. A. and Umaidah, Y. (2020) 'Sentimen Analisis  
kic pada Grup Facebook Game Online Menggunakan Klasifikasi  
, *Jurnal Informatika Universitas Pamulang*, 5(3), p. 356. doi:  
rmatika.v5i3.6571.



