

KOMPUTASI HUMOR, STUDI KASUS BERITA HOAKS

HUMOR COMPUTATION, CASE STUDY HOAX NEWS



OMAR WAHID

2018130038

PEMBIMBING I : Dr. Eng. Yuyun, S.Kom., M.T

PEMBIMBING II : Adnan, ST.,M.T.,PhD

PROGRAM PASCA SARJANA

STMIK HANDAYANI

MAKASSAR

2021

HUMOR COMPUTATION,CASE STUDY HOAX NEWS

TESIS

Diajukan Sebagai Salah Satu Syarat Mencapai Gelar Magister Komputer
Program Studi Pasca Sarjana Sistem Komputer

Disusun dan Diajukan Oleh :

OMAR WAHID
2018130038

PRODI MAGISTER SISTEM KOMPUTER
SEKOLAH TINGGI MANAJEMEN INFORMATIKA DAN KOMPUTER
(STMIK) HANDAYANI
MAKASSAR
2021

TESIS

HUMOR COMPUTATION, CASE STUDY HOAX NEWS

Disusun dan diajukan oleh :

OMAR WAHID

2018130038

Telah dipertahankan di depan Panitia Ujian Tesis

Pada Tanggal 03 Oktober 2020

Dan dinyatakan telah memenuhi syarat

Menyetujui
Komisi Penasehat



Dr. Eng. Yuyun, S.Kom., M.T
Ketua



Adnan, ST., M.T., PhD
Anggota

Ketua Program Studi
Sistem Komputer

Direktur Program Pascasarjana
STMIK Handayani

Prof. Dr. Ir. Andani Ahmad., M.T



Dr. Eng. Yuyun, S.Kom., M.T



TANDA PERSETUJUAN PERBAIKAN
UJIAN AKHIR MAGISTER

Pada hari **Sabtu** tanggal **03 Oktober 2020** telah dilaksanakan Ujian Akhir mahasiswa a.n

Nama Mahasiswa : Omar Wahid
Nomor Pokok : 2018130038
Program Pendidikan : S2 (Magister) Program Studi Sistem Komputer
Judul Penelitian : **Humor Computation, Case Study Hoax News**

Hasil ujian menyepakati bahwa sebelum penggandaan tesis, yang bersangkutan harus menyempurnakan tesisnya sesuai saran dan masukan yang muncul pada ujian tersebut.

Hasil penyempurnaan tesis tersebut ditunjukkan kepada Panitia Ujian Akhir, dan dinyatakan selesai jika Panitia Ujian Akhir menandatangani persetujuan di bawah ini

Panitia Ujian Akhir

Tanda Tangan

Ketua : Dr. Eng. Yuyun, S.Kom., M.T
Sekertaris : Adnan, ST.,M.T.,PhD
Anggota : 1. Dr.Eng. Hazriani, M.T
2. Dr. Eng. Ir. Zulfajri B.Hasanuddin, M.Eng
3. Dr. Imran Taufiq, ST., M.Si

Mengetahui
Ketua Program Studi,

Prof. Dr. Ir. Andani Achmad, M.T
NIP. 19601231 198703 1 022

PERNYATAAN KEASLIAN TESIS

Yang bertanda tangan di bawah ini :

Nama : **Omar Wahid**
NIM : 2018130038
Program Studi : Sistem Komputer

Menyatakan dengan sebenar-benarnya bahwa tesis yang berjudul : **“Humor Computation, Case Study Hoax News”** adalah benar-benar hasil karya sendiri. Dalam naskah tesis ini tidak terdapat karya ilmiah yang pernah diajukan/ditulis/diterbitkan sebelumnya, kecuali yang secara tertulis dalam naskah ini dan disebutkan sumber kutipan dan daftar Pustaka.

Apabila dikemudian hari ternyata di dalam naskah tesis ini adalah hasil karya orang lain maka saya bersedia menerima sanksi apapun sesuai dengan peraturan yang berlaku.

Makassar, 03 Oktober 2020

Yang menyatakan,



Omar Wahid

PRAKATA

Puji Syukur kami panjatkan kehadirat ALLAH SWT atas rahmat dan hidayah-NYA kepada penulis sehingga dapat menyelesaikan tesis ini, yang menjadi salah satu persyaratan untuk menyelesaikan jenjang Pendidikan Strata Dua (S2) Magister Komputer pada Sekolah Tinggi Manajemen Informatika dan Komputer (STMIK) Handayani Makassar.

Alhamdulillah selama menyelesaikan tesis ini penulis telah memperoleh banyak bantuan, petunjuk dan saran dari berbagai pihak. Penulis menyadari sepenuhnya bahwa dengan bantuan dan partisipasi tersebut tidak mungkin tesis ini dapat terselesaikan dengan baik. Untuk itu penulis mempersembahkan ucapan terima kasih teristimewa dan doa kepada kedua orang tua penulis yang telah berpulang Keharibaan Allah SWT, Ibunda yang terkasih **Nukria** dan Ayahanda **La Ode Abdul Wahid Ado** yang semasa hidupnya selalu memberikan suport kepada penulis dalam menempuh pendidikan. Dan tak lupa pula penulis ucapkan banyak-banyak terima kasih kepada istri tercinta saya **Ade Srianik** dan anak anak saya **Athar Ziyad Ado, Tiara Adreena Farzana Omar** dan **Rakha Wahid Ado** atas dukungannya yang menjadi semangat dan selalu memberikan energi positif kepada penulis disetiap tatapan, canda dan tawanya.

Penulis juga mengucapkan terima kasih yang sebesar-besarnya kepada :

1. Dr. Eng. Yuyun, S,Kom., M.T selaku Direktur pascasarjana STMIK Handayani Makassar dan sekaligus sebagai pembimbing I penulis yang selalu memberikan arahan, motivasi, ilmu, bimbingan sehingga penelitian ini dapat diselesaikan dengan baik.
2. Bapak Adnan, ST.,M.T.,PhD, selaku pembimbing II penulis yang telah meluangkan waktu kepada penulis untuk membimbing dan berkonsultasi tentang materi tesis ini.
3. Prof. Dr. Ir. Andani Ahmad., M.T selaku ketua program studi Pasca Sarjana STMIK Handayani Makassar yang selama ini telah banyak membantu penulis dan memberikan ilmu pengetahuan sehingga dapat menyelesaikan studi program Pasca Sarjana di kampus STMIK Handayani Makassar.
4. Bapak/Ibu Dosen dan Staf Pascasarjana STMIK Handayani Makassar yang telah banyak membantu proses Akademik mulai dari perkuliahan hingga saat ini.
5. Saudaraku **Moldefya Wahid Ado, Arsul Risal Wahid Ado, Rafdidiar Wahid Ado, Risa Wahid Ado, Fara Hamdana Wahid Ado** dan semua keluargaku tercinta yang selalu mendoakan dan memberi dukungan sehingga penulis sampai pada tahap akhir.
6. Teman-teman seperjuangan Pasca Sarjana kelas Kendari, Ilin Sukma, Nur Islamudin, Faizal Aris, Nilam Kusumawati, Asmira,

Andi Iwan, Aris Susanto, Salam, La Ode Bakrim, La Ija, Jufer, yang selalu memberikan dukungan dalam segala bentuk motivasi dan Alhamdulillah penulis sudah sampai di tahap akhir.

Namun penulis menyadari bahwa tesis ini belum dapat penulis buat secara sempurna dengan berbagai kekurangan yang ada, oleh karena itu saran maupun kritik yang bersifat membangun akan penulis terima untuk penyempurnaan tesis ini.

Semoga materi dalam tesis ini dapat bermanfaat bagi pengembangan ilmu pengetahuan khususnya dibidang Informatika Komputer dan dapat bermanfaat bagi kita semua terutama bagi penulis pribadi, Amin

Makassar, 03 Oktober 2020

Penulis

ABSTRAK

OMAR WAHID. *Humor computation, study case hoax news* (dibimbing oleh Dr. Eng. Yuyun, S.Kom.,MT dan Adnan, ST.,M.T.,PhD).

Hoax adalah informasi berbahaya yang menyesatkan persepsi manusia. Dengan penyebarannya yang begitu pesat dikhawatirkan dapat menimbulkan keresahan dan permusuhan ditengah masyarakat. Tujuan dari penelitian ini adalah membuat sebuah pemodelan untuk merubah berita hoaks menjadi humor. Pemodelan ini menggunakan teknik web scrapping dalam pengambilan data hoaks dan corpus humor, kemudian menerapkan tahapan *text mining* yaitu tahapan *preprocessing* yang meliputi *tokenizing*, *stopwords*, dan *stemming*. Selanjutnya dilakukan tahapan pembobotan kata (TF-Idf) dan menghitung tingkat kemiripan antar dua dokumen dengan menggunakan metode *cosine similarity*. Data yang digunakan pada penelitian ini adalah 50 data hoaks dan 182 corpus data humor untuk dilakukan pencocokkan data menggunakan algoritma *cosine similarity*. Hasil pengujian *Normalized Discounted Cumulative Gain* (nDCG) didapatkan skor tertinggi yaitu 0,696. Serta ditemukan 7 (tujuh) data hoaks yang tidak memiliki relevansi dengan corpus data humor.

Kata kunci : Hoaks, humor, *web scrapping*, *preprocessing*, *tf-idf*, *cosine similarity*.

ABSTRACT

OMAR WAHID. *Humor computation, study case hoax news* (supervised by Dr. Eng. Yuyun, S.Kom.,MT dan Adnan, ST.,M.T.,PhD).

Hoaxes are dangerous information that misleads human perceptions. It is feared that its rapid spread could cause unrest and hostility in the community. The purpose of this research is to create a model to turn hoax news into humor. This modeling uses web scraping techniques to collect hoax data and humor corpus, then applies the text mining stage, namely the preprocessing stage which includes tokenizing, stop words, and stemming. Furthermore, the word weighting step (TF-Idf) is carried out and calculating the level of similarity between two documents using the cosine similarity method. The data used in this study were 50 data hoaxes and 182 corpus humor data for matching the data using cosine similarity algorithm. The test results of the Normalized Discounted Cumulative Gain (nDCG) obtained the highest score, namely 0.696. Also found 7 (seven) hoax data that have no relevance to the humor corpus data.

Keywords : Hoax, humor, preprocessing, tf-idf, *cosine similarity*.

DAFTAR ISI

Halaman Judul	i
Halaman Pengajuan.....	ii
Halaman Pengesahan.....	iii
Halaman Persetujuan	iv
Lembar Pernyataan Keaslian	v
Prakata	vi
Abstrak	ix
Abstract	x
Daftar Isi	xi
Daftar Gambar	xiii
Daftar Tabel	xv
BAB I PENDAHULUAN	1
A. Latar Belakang Masalah	1
B. Rumusan Masalah.....	5
C. Tujuan Penelitian	5
D. Kegunaan Penelitian.....	6
E. Batasan Penelitian.....	6
F. Organsisasi/Sistematika	6
BAB II TINJAUAN PUSTAKA.....	8
A. Computation	8
B. Definisi Humor	9
C. Definisi <i>Hoax</i>	13
D. Tinjauan Umum Media Sosial	15
E. <i>Web Scraping</i>	16
F. <i>Text Mining</i>	17

G. Pembobotan Kata (<i>Tf-Idf</i>)	19
H. <i>Cosine Similarity</i>	20
I. <i>Discounted Cumulative Gain (DCG)</i>	21
J. <i>Machine Learning</i>	22
K. Road Map Penelitian.....	22
L. Kerangka Pikir	24
BAB III METODOLOGI PENELITIAN	25
A. Tahapan Penelitian	25
B. Waktu dan Lokasi Penelitian.....	26
C. Jenis Penelitian.....	26
D. Sumber Data.....	27
E. Instrumen Penelitian	27
F. Perancangan Sistem.....	27
BAB IV HASIL DAN PEMBAHASAN	39
A. Hasil Proses <i>Preprocessing</i>	39
B. Model Berita Hoaks Menjadi Humor	54
C. Gambaran umum sistem.....	57
D. Pengujian Sistem <i>Discounted Comulative Gain (DCG)</i>	60
E. Pengujian Tingkat Kelucuan	64
BAB V PENUTUP	73
A. Kesimpulann	73
B. Saran	74
DAFTAR PUSTAKA	75

DAFTAR GAMBAR

Gambar 2.1. Tahap <i>Tokenizing</i>	17
Gambar 2.2. Tahap Filtering	17
Gambar 2.3. Tahap Stemming	18
Gambar 2.4. Tahap Tagging	18
Gambar 2.5. Tahap Text Mining	19
Gambar 2.6. Kerangka Pikir	24
Gambar 3.1 Tahapan Penelitian	25
Gambar 3.2 <i>Flowchart Sistem</i>	28
Gambar 3.3. <i>Flowchart preproses</i>	30
Gambar 3.4 Tampilan <i>scraping</i> data hoaks	31
Gambar 3.5. Alur <i>Web Scraping</i>	33
Gambar 3.6. <i>Flowchart filtering</i>	33
Gambar 3.7. <i>Flowchart tokenizing</i>	34
Gambar 3.8. <i>Flowchart stopwords</i>	35
Gambar 3.9. <i>Flowchart stemming</i>	36
Gambar 3.10. <i>Flowchart tf-idf</i>	37
Gambar 3.11. <i>Flowchart cosine similarity</i>	38
Gambar 4.1. Contoh perhitungan <i>Tf-Idf</i> kalimat lucu 1	46
Gambar 4.2. Contoh perhitungan <i>Tf-Idf</i> kalimat lucu 2	47
Gambar 4.3. Contoh perhitungan <i>Tf-Idf</i> kalimat lucu 3	48
Gambar 4.4. Contoh perhitungan <i>Tf-Idf</i> kalimat lucu 4	49

DAFTAR TABEL

Tabel 2.1. Road Map Penelitian	22
Tabel 3.1. Daftar link berita hoaks	32
Tabel 4.1. Contoh hasil proses <i>filtering</i> berita hoaks	40
Tabel 4.2 Contoh hasil proses <i>filtering</i> kalimat lucu	40
Tabel 4.3 Contoh hasil proses <i>tokenizing</i> berita hoaks	41
Tabel 4.4 Contoh hasil proses <i>tokenizing</i> kalimat lucu	41
Tabel 4.5 Contoh hasil proses <i>stopwords</i> berita hoaks.....	42
Tabel 4.6 Contoh hasil proses <i>stopwords</i> kalimat lucu	42
Tabel 4.7 Contoh hasil proses <i>stemming</i> berita hoaks	43
Tabel 4.8 Contoh hasil proses <i>stemming</i> berita lucu	43
Tabel 4.9 Contoh hasil <i>Tf-Idf</i>	44
Tabel 4.10 <i>Hoax to Humor</i>	57
Tabel 4.11 Hasil <i>Discounted Cumulative Gain (DCG)</i>	61
Tabel 4.12 Tabulasi tingkat kelucuan dikaitkan dengan umur	67
Tabel 4.13 Tabulasi tingkat kelucuan dikaitkan dengan jenis kelamin	68
Tabel 4.14 Tabulasi tingkat kelucuan dikaitkan dengan pekerjaan	69
Tabel 4.15 Tabulasi tingkat kelucuan dikaitkan dengan pendidikan terakhir	70

BAB I

PENDAHULUAN

A. Latar Belakang Masalah

Perkembangan teknologi informasi memberikan kemudahan pada masyarakat dalam menjalin komunikasi dan mendapatkan informasi. Hal ini didukung dengan adanya internet sebagai media pencari informasi yang canggih. Pencarian informasi melalui internet sendiri sudah menjadi trend kekinian di tengah masyarakat. Pada sisi lain penyebaran informasi di internet yang nyaris tanpa kendali mempunyai masalah tersendiri, tidak semua informasi yang disebarakan berupa fakta atau sering disebut hoaks. Hoaks adalah informasi berbahaya yang sengaja di buat untuk menyesatkan persepsi masyarakat, hoaks biasanya bermuatan provokasi, SARA bahkan bersifat menakut-nakuti sehingga dapat menimbulkan kecemasan karena disajikan seolah-olah sebagai kebenaran. (Rasywir & Purwarianti, 2015).

Dampak berita *hoax* dapat langsung menyerang pemikiran masyarakat tanpa disadari secara langsung. Tingkat pengetahuan dan pemahaman tentang berita hoaks akan dapat mempengaruhi pemikiran pembacanya. (Vukovi et al., 2009).

Fenomena berita hoaks di Indonesia begitu masif yang seolah tanpa henti cukup merepotkan pemerintah. Kementerian Komunikasi dan Informatika bahkan membentuk satuan tugas khusus, yang diberi nama

Drone 9 untuk memantau konten di Internet, termasuk berita palsu (*hoax*) (*Kementerian Komunikasi dan Informatika, 2018*).

Pada saat Pemilihan Umum Presiden dan Legislatif atau Pemilihan Kepala Daerah berlangsung kemunculan berita *hoax* semakin banyak. Ini dapat dilihat saat Pilkada DKI Jakarta 2017. Saat itu, *hoax* banyak beredar di masyarakat. Dewan Pers Indonesia menilai *hoax* telah memasuki tahap serius. Selama akhir 2016 hingga April 2017, munculnya *hoax* semakin intensif, menyerang para kandidat yang bersaing. Sebuah survei yang dilakukan oleh Masyarakat Telematika Indonesia (MASTEL) sepuluh hari sebelum putaran pertama pemilu Jakarta, menemukan bahwa distribusi konten tipuan sangat tinggi. Sebanyak 44,3% responden menyatakan bahwa mereka menerima konten tipuan setiap hari dan 17,2% dari mereka menerimanya lebih dari sekali per hari (Utami, 2019).

Penyebaran berita *hoax* saat ini cukup tinggi di tengah merebaknya Covid-19. Identifikasi Kementerian Komunikasi dan Informatika melalui Tim Ais tanggal 16 Maret 2020, ditemukan sebanyak 242 berita *hoaks* dan disinformasi tentang virus corona (Covid-19). Konten-konten berita *hoaks* itu tersebar di pesan instan, media sosial maupun website. Menkominfo menilai penyebaran berita *hoaks* di saat merebaknya Covid-19 dapat membuat kepanikan dan ketakutan di tengah masyarakat (Agus Tri Haryanto, 2020).

Pesatnya penyebaran berita *hoax* dikhawatirkan dapat menimbulkan keresahan, kebencian dan permusuhan di tengah

masyarakat. Jika tidak cepat ditanggulangi, kegaduhan yang ditimbulkan dari berita hoaks di dunia maya melalui *platform* sosial media dapat berimbas ke dunia nyata. Pembahasan yang terjadi di media sosial berpotensi dapat menanamkan pemahaman publik mengenai suatu hal dalam kehidupan berbangsa dan bernegara. Kegaduhan di media sosial dapat berimbas dalam kehidupan nyata karena di anggap dapat membentuk pemahaman dan memaknai sesuatu yang dapat mempengaruhi pendapat sosial kita. Kegaduhan akibat berita hoaks di media sosial semacam itu seringkali menggunakan sentimen identitas yang berakhir pada hujatan dan kebencian sehingga dapat melemahkan semangat kebhinekaan yang menjadi pondasi dalam bermasyarakat, berbangsa dan bernegara. Sehingga konsep kebhinekaan yang menjadi pondasi kita mengalami perubahan yang diakibatkan opini-opini yang terbentuk oleh media sosial (Juliswara, 2017).

Tidak hanya dapat menimbulkan keresahan dan permusuhan berita palsu (*hoax*) juga bisa berdampak buruk pada kesehatan mental, dalam sebuah studi, para psikolog sepakat bahwa berita *hoax* bisa memberikan dampak buruk pada kesehatan mental, seperti *post-traumatic stress syndrome* (PTSD), menimbulkan kecemasan sampai kekerasan. Psikolog percaya, orang yang terpapar berita hoax juga bisa membutuhkan terapi, karena diselimuti kecemasan, stres, dan merasa kesepian karena berita palsu. Psikolog meyakini, berita hoax dihadirkan untuk memanipulasi

banyak orang. Sebab, berita palsu bisa memanfaatkan kelompok orang yang takut, dan mengambil keuntungan ketakutan itu (Wisnubrata, 2019).

Melihat dampak negatif berita palsu (*hoax*) sangat buruk, maka penulis merasa perlu membuat suatu penelitian yang dapat merubah berita *hoax* menjadi sebuah humor atau lelucon. Dimana humor dapat membuat orang tertawa yang menggambarkan kesenangan atau kebahagiaan seseorang. Sehingga dengan humor diharapkan dapat meminimalisir kegaduhan yang ditimbulkan oleh maraknya berita palsu (*hoax*) yang sangat meresahkan.

Humor juga berdampak positif bagi kesehatan, diantaranya bisa mengurangi rasa sakit, mengoptimalkan fungsi otak, membuat rileks, meningkatkan sistem kekebalan tubuh, mengurangi stres, mengurangi rasa takut dan baik untuk pernafasan dan pencernaan. Selain itu, humor juga bermanfaat bagi hubungan sosial. Orang yang humoris akan lebih mudah terlibat dalam pergaulan karena dapat membuat pemikiran positif, selain itu juga dianggap sebagai teman yang menyenangkan sehingga lebih mudah bergaul dan menciptakan hubungan sosial yang baik dengan orang lain. (Istiningtyas, 2014).

Penelitian ini menggunakan beberapa referensi dari penelitian sebelumnya diantaranya jurnal yang berjudul "*Computational Humor*", menggunakan metode semantik (Kim Binsted, 2006). Dan "*Getting Serious about the Development of Computational Humor*", menggunakan

penerapan *artificial intelligence* dengan metode *semantik* (Oliviero Stock and Carlo Strapparava, 2003).

Berdasarkan uraian studi literatur yang telah dilakukan maka penelitian yang dilakukan adalah "***Computation Humor, Case Study Hoax News***".

B. Rumusan Masalah

Berdasarkan permasalahan yang telah diuraikan pada latar belakang, maka rumusan masalah yang akan di bahas oleh penulis adalah :

1. "Bagaimana penerapan algoritma *cosine similarity* untuk merubah text hoaks menjadi humor?"
2. "Bagaimana pengujian *Discounted Comulative Gain* (DCG) untuk mengukur kualitas hasil rangking dokumen?"
3. "Bagaimana pengujian tingkat kelucuan sistem dengan hasil survei?"

C. Tujuan Penelitian

Adapun tujuan yang ingin dicapai dari penelitian ini adalah :

1. Untuk penerapan algoritma *cosine similarity* untuk merubah text *hoax* menjadi humor.
2. Untuk mengukur kualitas hasil rangking dokumen dengan pengujian *Discounted Comulative Gain* (DCG).
3. Untuk menguji tingkat kelucuan sistem dengan hasil survei.

D. Kegunaan Penelitian

Dengan penerapan algoritma *cosine similarity* ini diharapkan dapat merubah berita *hoax* menjadi sebuah humor atau lelucon sehingga dapat meningkatkan kondisi kesehatan menjadi lebih sehat dan kegaduhan yang diakibatkan oleh berita *hoax* dapat diminimalisir.

E. Batasan Penelitian

Dalam penelitian ini perlu adanya pembahasan yang terbatas sehingga permasalahan tidak melebar. Adapun batasan masalah pada penerapan algoritma *cosine similarity* merubah berita *hoax* menjadi humor adalah berita hoaks dan corpus data humor yang berbentuk teks berbahasa Indonesia.

F. Organisasi/Sistematika

Sistematika penulisan pada penelitian ini, yaitu :

Bab I Pendahuluan, berisi latar belakang masalah, rumusan masalah, tujuan penelitian, kegunaan penelitian, batasan penelitian serta sistematika penulisan.

Bab II Tinjauan Pustaka, berisi tentang teori komputasi, definisi humor, *hoaks*, tinjauan umum media sosial, web scraping, text mining, pembobotan kata (tf-idf), *cosine similarity*, *discounted cumulative gain*, *machine learning*, road map penelitian, kerangka konseptual.

Bab III Metodologi Penelitian, berisi tahapan penelitian, waktu dan lokasi, jenis penelitian, sumber data, instrumen penelitian..

Bab IV Hasil dan Pembahasan, berisi tentang arsitektur sistem, tahap *preprocessing*, pembobotan kata (*Tf-Idf*), *cosine similarity*, model berita hoaks menjadi humor, dan gambaran umum sistem.

Bab V Penutup, berisi tentang kesimpulan dan saran penelitian.

BAB II

TINJAUAN PUSTAKA

A. *Computation*

1. Teori komputasi

Komputasi adalah suatu model yang membahas apakah dan bagaimanakah suatu masalah pada ilmu komputer dan matematika dapat diselesaikan menggunakan algoritma. Komputabilitas dan kompleksitas adalah bahasan utama dari bidang ilmu ini dalam hubungannya dengan formalisme komputasi. (Deadydiedy, 2015).

Komputasi juga dapat disebut sebagai sebuah komputer fisik, yang dibahas pada ilmu teori fisik (*Physic of Computation*). Sistem fisik yang dimaksud yaitu komputer digital, komputer quantum, komputer yang dapat menganalisa DNA, dan komputer molekular..

Teori komputasi terbagi atas tiga, yaitu :

1. Teori Otomata (*automata theory*)
2. Teori Komputabilitas (*computability theory*)
3. Teori Kompleksitas (*computational complexity theory*)

2. Komputasi Modern

Komputasi modern dapat disebut sebuah konsep sistem yang dapat menyimpan instruksi-instruksi yang diterima dalam sebuah memory. Sehingga dapat dikatakan komputer merupakan sebuah komputasi modern, karena saat ini kita lebih sering melakukan komputasi

menggunakan komputer. Prinsip kerja komputasi modern mencari solusi dari permasalahan yang ditemukan dengan melakukan perhitungan.

Adapun perhitungan itu meliputi:

1. Akurasi
2. Kecepatan
3. Problem Volume Besar
4. Modelling
5. Kompleksitas

Komputasi modern digolongkan ke dalam tiga jenis, yaitu:

1. *Mobile Computing* adalah perkembangan teknologi di bidang komunikasi menggunakan jaringan tanpa kabel serta fleksibel bisa digunakan dimana saja. contohnya smartphone, GPS, dll.
2. *Grid Computing* adalah teknologi komputer yang mengatasi masalah komputasi skala besar.
3. *Cloud Computing* adalah model komputasi yang terukur dinamis dan sumber daya virtual yang menyediakan layanan IT berbasis model dalam internet.

Proses *mobile computing* bergantung pada penggunaanya, sedangkan *grid computing* apakah pengguna mendapatkan server atau tidak dan *cloud computing* prosesnya membutuhkan jaringan.

B. Definisi Humor

Menurut Ensiklopedia Indonesia, kata humor yang berasal dari bahasa Yunani mempunyai arti getah. Humor dapat disebut dengan sikap

riang dalam kehidupan sehari-hari. (Didiek Rahmanadji, 2009) humor merupakan rasa atau gejala yang menjadi perangsang kita untuk tertawa atau hasrat secara mental untuk tertawa, humor berwujud rasa, kepekaan di dalam diri individu (*sense of humor*) yang merupakan gejala atau hasil ciptaan didalam ataupun diluar diri individu. Dampak yang diakibatkan oleh humor adalah dorongan yang dapat menimbulkan sikap tersenyum hingga tertawa terbahak-bahak.

Setiap individu mempunyai selera humor yang tidak dapat dihilangkan. Humor bisa menjadi penghibur hati seseorang, bisa pula sebagai media pendidikan dan sebagai ajang dalam mengungkapkan sikap kritik sosial dalam masyarakat. Humor adalah ungkapan seseorang yang mampu membuat suasana lebih rileks dan menarik karena mampu menghidupkan suasana. Sehingga dapat dikatakan bahwa humor mempunyai peranan yang vital dalam kehidupan manusia. Cerita yang mengandung humor, sebagian besar bercerita tentang kelucuan yang disebabkan kecerdikan, kebodohan, kemalangan keberuntungan pemeran utamanya.

Hal-hal yang dapat membuat orang tertawa karena humor adalah sebagai berikut:

- a. Sesuatu yang norak, atau lebih buruk yang diutarakan kepada orang lain, tetapi tidak mengatakannya langsung kepada yang bersangkutan, sehingga melahirkan keinginan untuk tertawa bagi pendengarnya.

- b. Ada sesuatu yang janggal yang diinginkan oleh satu individu dari individu lainnya, sehingga menimbulkan berbagai bentuk reaksi dari orang yang mengharapkan berupa ekspresi wajah dengan sorot mata yang serius atau dengan ucaan yang spontan terucapa sesuai dengan keadaan saat itu.
- c. Karena tingginya beban dan tekanan psikis sehingga individu cenderung ingin menyampingkan tekanan tersebut dengan humor.

FUNGSI HUMOR

1. Kesehatan Fisik

Dampak tertawa pada sistem kekebalan tubuh, memperlihatkan beberapa manfaat pada kesehatan, seperti jumlah dan kemampuan sel imun yang dapat membunuh sel-sel yang membahayakan tubuh meningkat, bertambahnya antibodi *IgA (imunoglobulin A)* yang dapat membunuh infeksi pada saluran napas atas, menambah kemampuan tubuh dalam memanfaatkan oksigen, serta meningkatkan aktivitas-aktivitas sel pembunuh di dalam tubuh.

Tertawa juga dapat melatih otot-otot jantung, serta relaksasi otot, melancarkan peredaran darah, mengurangi hormon-hormon yang dapat menyebabkan stres dll. (Franzini, 2001)

2. Kebahagiaan Psikologis

Banyak hal terjadi dalam kehidupan yang dapat membuat tingkat stres menjadi tinggi, antara lain terjadinya bencana, masalah yang terjadi dalam keluarga atau orang lain, tekanan dan beban kerja

kerja yang tinggi, masalah ekonomi, dan lain sebagainya yang bisa berdampak buruk bagi kesehatan mental dan fisik individu, serta dapat melahirkan hal-hal negatif seperti emosi yang tidak stabil, turunnya kemampuan berpikir otak, dan permasalahan perilaku. Namun, tingkat stres individu satu dan individu yang lain berbeda-beda tergantung cara memahami suatu kejadian dan cara menyikapinya. Emosi negatif akan sirna jika perasaan bahagia, semakin banyak perasaan bahagia maka akan dapat menimbulkan banyak emosi positif. Dengan tertawa yang ditimbulkan oleh humor perasaan akan selalu ceria dan berenergi, mengurangi stres, kecemasan, ketegangan, dan tidak mudah tersinggung. Sehingga semakin sering orang berhumor maka perasaan bahagia akan sering timbul.

3. Hubungan Sosial

Humor merupakan fenomena sosial, berhumor sambil tertawa sering kali terjadi saat bercengkrama dengan teman-teman atau orang lain. Sekalipun tertawa saat sendirian pada saat menonton film yang bermuatan humor atau membaca cerita komedi, hal itu tergolong *pseudo-social* karena tetap menampilkan atau membahas orang lain secara virtual. Pada konteks sosial umumnya humor dianggap sebagai sebuah permainan, dan seperti permainan sebagaimana biasanya sangat jarang dianggap serius dalam berhumor, ungkapan dan aksi semata-mata hanya main-main. Manfaat humor dalam

dalam hubungan sosial pun cukup banyak. Orang yang humoris akan lebih mudah terlibat dalam pergaulan karena dapat membuat pemikiran positif, selain itu juga dianggap sebagai teman yang menyenangkan sehingga lebih mudah bergaul dan menciptakan hubungan sosial yang baik dengan orang lain. (Istiningtyas, 2014).

C. Definisi Hoaks

Hoaks adalah upaya untuk memperdayai atau mempermainkan pembacanya atau pendengarnya untuk mempercayai suatu informasi/berita yang salah atau palsu sehingga dianggap sebagai kebenaran.

Tujuan Hoaks untuk menggiring opini publik, dan membentuk persepsi tentang sesuatu yang salah menjadi suatu kebenaran serta untuk mengukur kecerdasan dan ketelitian media sosial dan internet. Umumnya ada beberapa ragam tujuan disebarkannya berita hoaks antara lain sebagai bahan candaan, persaingan yang tidak sehat untuk menjatuhkan pesaingnya (*black campaign*), promosi barang palsu, ataupun mengajak melakukan amalan-amalan baik yang dalilnya belum jelas sehingga dapat melahirkan paham-paham radikal. Tetapi banyak penerima berita hoaks yang tanpa menyaring terlebih dahulu dari sumber yang terpercaya ikut menyebarkan berita hoaks tersebut sehingga berita hoaks ini sangat cepat tersebar luas.

Jenis-jenis Informasi Hoax

1. *Fake news*: Berita yang tidak sesuai isinya dengan berita yang asli. Tujuan dari berita ini adalah membuat atau meniru suatu berita dan memasukkan informasi bohong dalam berita itu. Semakin tidak sesuai dengan isi dan tujuan berita aslinya penulis berita hoaks tersebut merasa semakin baik.
2. *Clickbait* : *link* yang disimpan di dalam suatu situs dan letaknya sangat strategis dengan tujuan untuk menarik perhatian orang masuk ke situs lainnya. Judul dalam link tersebut dibuat sedemikian rupa sehingga dapat memikat pembacanya tetapi isinya sesuai dengan berita aslinya.
3. *Confirmation bias* : menfasirkan kejadian yang terjadi sebagai penguatan dari kepercayaan masyarakat yang sudah ada.
4. *Misinformation* : suatu Informasi tidak akurat atau salah, dengan tujuan untuk menipu.
5. *Satire* : Sebuah ungkapan dalam bentuk teks yang dikemas secara humor dan ironi serta hal yang dibesar-besarkan untuk mengulas kejadian yang sedang ramai diperbincangkan.
6. *Post-truth* : Opini publik yang dibentuk berdasarkan kejadian yang mana emosi lebih dominan daripada fakta.
7. Propaganda : Membentuk opini publik dengan mengedarkan informasi yang belum jelas kebenarannya dan bahkan berisi kebohongan. (Rahadi, 2017).

D. Tinjauan Umum Media Sosial

1. Definisi Media Sosial

Media Sosial adalah interaksi sosial menggunakan media online. Dengan media sosial siapapun dapat membuat, mengedit sekaligus menyebarkan sendiri konten berita, artikel, promosi, gambar/foto, dan video. Media sosial lebih fleksibel, cakupannya luas, cepat, interaktif, dan lebih efektif dan efisien.

2. Sejarah Media Sosial

Ide dasar media sosial adalah bagaimana orang-orang dari seluruh dunia dapat saling terhubung satu sama lain. Pada tahun 1978, secara tidak langsung konsep media sosial sudah ada, selain penggunaan telepon yang terhubung dengan modem, penggunaan surat elektronik sudah ada pada masa itu dengan menggunakan sistem papan buletin untuk saling terhubung dengan orang. (Nurudin, 2012).

Munculnya situs *GeoCities* pada tahun 1995 menjadi tonggak dasar dikenalnya *website*. Situs *GeoCities* merupakan media *website* yang dapat diakses serta datanya dapat disimpan. Pada tahun 1997, *SixDegrees* dan *Classmates* juga hadir, dimana *SixDegrees* hadir dengan konsep situs jaringan sosial untuk menghubungkan orang tanpa harus kenal atau berteman terlebih dahulu. Sedangkan *Classmates* yang lebih dulu didirikan hanya pada hubungan antar kenalan atau teman. Karena cakupannya lebih luas dan lebih canggih, sehingga media sosial pertemanan pertama di dunia yang dikenal adalah *SixDegrees*.

Selanjutnya lahir situs yang disebut *Blogger* pada tahun 1999 yang penggunanya dapat merancang halaman situsnya sendiri. Jejaring sosial ini dapat berisi pendapat tentang berbagai hal, baik persoalan pribadi, sosial maupun hal-hal yang bersifat politik. Tahun 2000 lahir jejaring sosial yang sangat fenomenal yaitu *Frienster*. Jejaring sosial ini mendorong pengembangan situs-situs jejaring sosial lainnya seperti *MySpace* (2003), *LinkedIn* (2003), *Facebook* (2004), dan *Twitter* (2006).

E. Web Scraping

Web scraping adalah teknik mengambil atau memperoleh informasi secara otomatis dari sebuah website. Tujuan web scraping adalah untuk menelusuri suatu informasi tertentu, kemudian informasi tersebut di ekstraksi, dan menyimpan ke dalam halaman web yang baru. Beberapa web scraping lebih berfokus untuk mengubah data menjadi lebih terstruktur.

Umumnya proses pengambilan informasi dari sebuah website dalam bahasa markup berupa HTML atau XHTML, dan menganalisis informasi tersebut untuk di ambil bagian tertentu dari halaman tersebut sesuai dengan kebutuhan yang akan digunakan. Web scraping biasa juga disebut screen scraping. Web scraping berbeda dengan data mining karena data mining lebih berkonsentrasi untuk mempelajari pola semantik atau kecenderungan dari sejumlah besar data yang telah ditemukan.

F. Text Mining

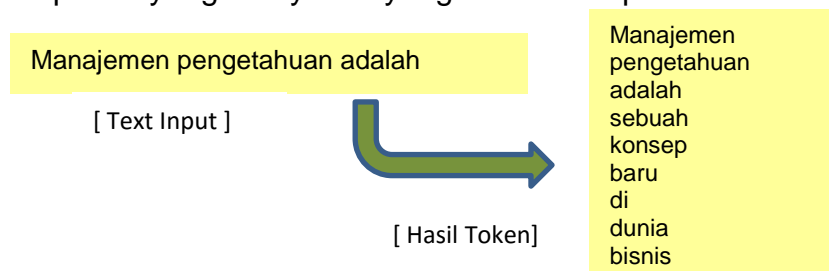
Text Mining memiliki definisi menambang data yang berupa teks dimana sumber data biasanya didapatkan dari dokumen, dan tujuannya adalah mencari kata-kata yang dapat mewakili isi dari dokumen sehingga dapat dilakukan analisa keterhubungan antar dokumen.

Tahapan Text Mining

Tahapan yang dilakukan secara umum adalah :

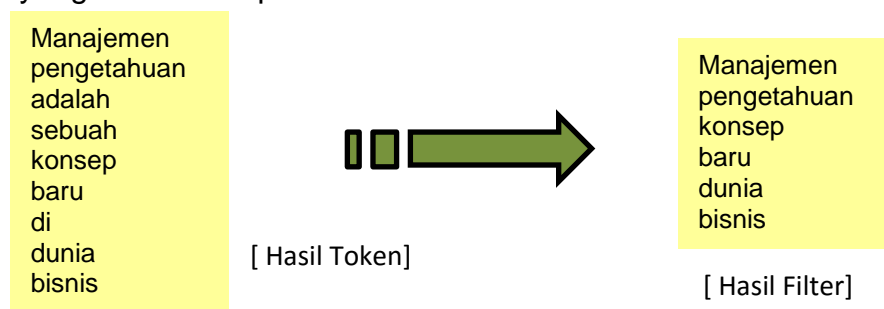
1. *Tokenizing*

Tokenizing adalah tahap pemotongan *string* input berdasarkan tiap kata yang menyusun yang bisa dilihat pada Gambar 2.1 :



2. **Gambar 2.1.** Tahap *Tokenizing* (J. Mooney, 2000)

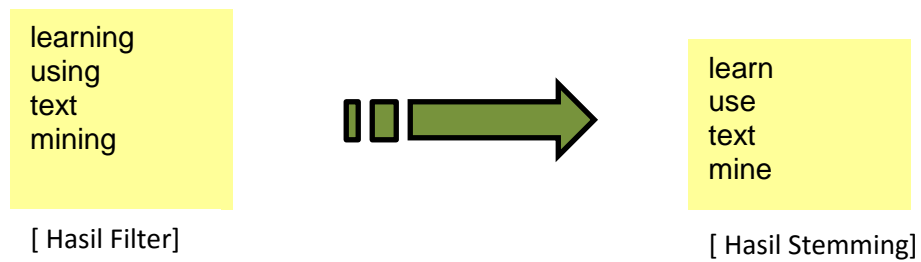
Filtering adalah tahap mengambil kata-kata penting dari hasil tokenizing. Bisa menggunakan algoritma *stop list* (membuang kata yang kurang penting) atau *word list* (menyimpan kata penting) yang bisa dilihat pada Gambar 2.2 :



Gambar 2.2. Tahap *Filtering* (J. Mooney, 2000)

3. Stemming

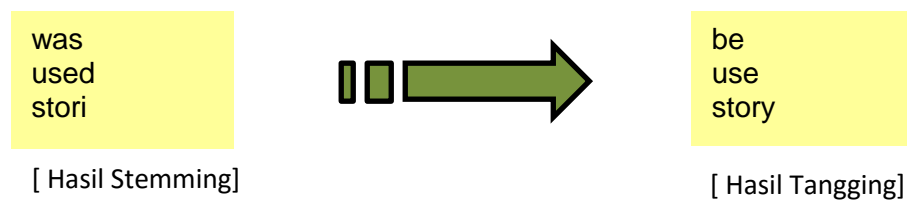
Stemming adalah tahap mencari root kata dari tiap kata hasil *filtering* yang bisa dilihat pada Gambar 2.3 :



Gambar 2.3. Tahap Stemming (*J. Mooney, 2000*)

4. Tagging

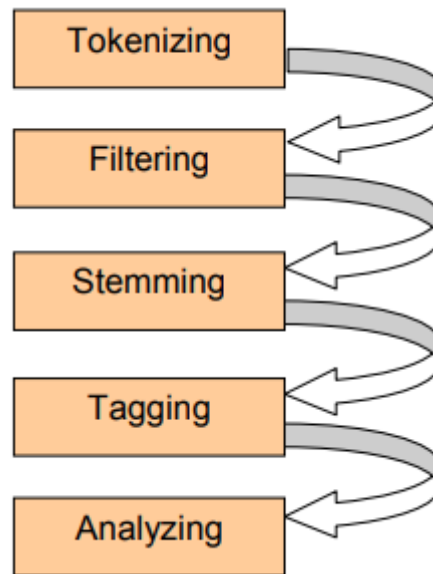
Tagging adalah tahap mencari bentuk awal/root dari tiap kata lampau atau kata hasil *stemming* yang bisa dilihat pada Gambar 2.4



Gambar 2.4. Tahap Tagging (*J. Mooney, 2000*)

5. Analyzing

Analyzing adalah tahap penentuan seberapa jauh keterhubungan antar kata-kata antar dokumen yang ada yang bisa dilihat pada Gambar 2.5.



Gambar 2.5. Tahap *Text Mining* (J. Mooney, 2000)

G. Pembobotan Kata (*Tf-Idf*)

Metode ini digunakan untuk melakukan perhitungan bobot kata yang paling sering digunakan pada saat pencarian informasi. Metode TF-IDF lebih mudah, efisien dan hasil yang akurat.

Metode Term Frequency-Inverse Document Frequency (TF-IDF) adalah suatu metode algoritma untuk mendapatkan bobot hubungan suatu kata (term) dalam suatu dokumen dan korpus dengan perhitungan menggunakan rumus. Proses ini digunakan untuk menggambarkan seberapa penting sebuah kata di dalam sebuah dokumen terhadap dokumen dalam korpus. Di dalam dokumen semakin banyak frekuensi kemunculan kata maka bobot kata semakin besar serta menunjukkan pentingnya kata tersebut di dalam sebuah dokumen dan jika kata tersebut muncul dalam banyak dokumen maka bobot akan semakin kecil.

H. *Cosine Similarity*

Metode Cosine Similarity merupakan suatu algoritma untuk mengukur kemiripan antara dua buah dokumen dalam satu dimensi dengan melakukan perkalian dua buah dokumen sehingga diperoleh nilai cosinus. Perbandingan dua buah dokumen dikatakan mirip apabila hasil nilai cosine similarity mempunyai nilai 1. (Melita et al., 2018)

Perangkingan dokumen yang diperoleh dari perhitungan algoritma cosine similarity berdasarkan dengan tingkat kemiripan dokumen dengan korpus. Dokumen dapat dipastikan mirip dengan korpus apabila rangking kemiripannya paling tinggi.

Dalam pemberian bobot, metode cosine similarity tidak terpengaruh dengan panjang dan pendeknya sebuah dokumen, sehingga proses pencarian kemiripan antar item dapat dipastikan sesuai dengan keyword yang dihasilkan dan ini menjadi suatu kelebihan utama dari metode cosine similarity.

Berikut adalah rumus *cosine similarity*

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

keterangan :

$\text{cosSim}(d_j, q_k)$ = tingkat kesamaan dokumen dengan query tertentu

td_{ij} = term ke-i dalam vektor untuk dokumen ke-j

tq_{ik} = term ke-i dalam vektor untuk query ke-k

n = jumlah term yang unik dalam data set

I. **Discounted Cumulative Gain (DCG)**

Metode *Discounted Cumulative Gain (DCG)* adalah suatu metode mencari posisi dokumen dalam daftar hasil untuk ukuran kualitas peringkat dokumen. Adapun rumus persamaan DCG adalah sebagai berikut :

$$DCG_p = \sum_{i=1}^p \frac{rel\ i}{\log_2(i + 1)}$$

Setelah ditemukan ukuran kualitas dokumen, selanjutnya mencari perwakilan dokumen didalam daftar database yang relevan sesuai dengan perankingannya. Adapun persamaan dari IDCG adalah sebagai berikut :

$$IDCG_p = \sum_{i=1}^{REL} \frac{rel\ i}{\log_2(i + 1)}$$

Selanjutnya mencari nilai *Normalized Discounted Cumulative Gain* (nDCG) untuk memperoleh perhitungan rata-rata kinerja algoritma yang digunakan pada mesin pencari. Pada perhitungan normalisasi dengan membandingkan hasil DCG dan IDCG perolehan nilai 1 mempunyai posisi peringkat paling mendekati. Persamaan nDCG , sebagai berikut:

$$nDCG = \frac{DCG_p}{IDCG_p}$$

J. **Machine Learning**

Penerapan ilmu komputer saat ini sangat membantu aktivitas manusia saat ini, salah cabangnya yaitu machine learning yang mempunyai kaitan erat dengan *Artificial Intelligent* (Kecerdasan Buatan).

Dimana konsep machine learning menitik beratkan pembangunan algoritma untuk otomatisasi program komputer dalam melakukan tugasnya tanpa perintah dari user.

Banyaknya pemanfaatan teknologi dalam dunia industri saat ini yang mengungkap konsep *machine learning* baik dalam maupun luar negeri serta banyaknya pelajar dan mahasiswa yang mengambil konsentrasi pada bidang ini, menandakan bahwa cabang ilmu ini sangat diminati. (Robert, 2017).

K. Road Map Penelitian

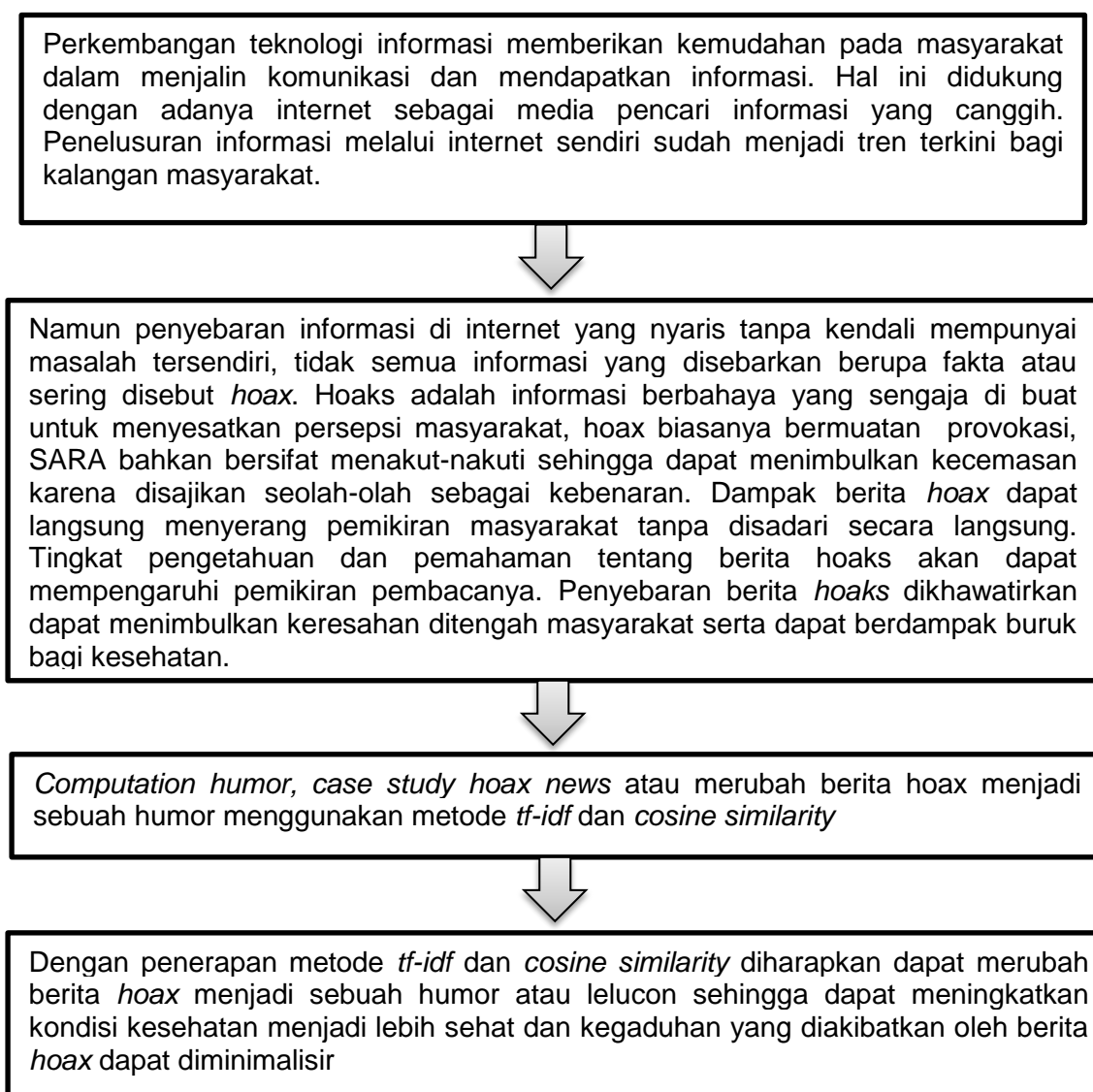
Beberapa penelitian terkait dengan penelitian ini antara lain:

Tabel 2.1. Road Map Penelitian

No	Judul	Penulis	Tahun	Metode	Hasil
1	An Intelligent Automatic Hoax Detection System	Marin Vuković, dkk	2009	Artificial Neural Network	73,86%
2	Penerapan Metode Term Frequency Inverse Document Frequency (Tf-Idf) Dan Cosine Similarity Pada Sistem Temu Kembali Informasi Untuk Mengetahui Syarah Hadits Berbasis Web (Studi Kasus: Syarah Umdatil Ahkam)	Ria Melita, Victor Amrizal, Hendra Bayu Suseno, Taslimun Dirjam	2018	Tf-Idf	Stemming Model 90,93% confusion matrix 100%, recall 88,7%, accuracy 88,73%, dan error rate 11,27%.
3	A Survey on Text Mining in Social Networks	Rizwana Irfan, dkk	2004	Supervised dan Unseperviased	-
4	Evaluating Machine Learning Algorithms for Fake News Detection	Shlok Gilda	2017	<ul style="list-style-type: none"> • Support Vector Machines • Stochastic Gradient Descent • Gradient Boosting • Bounded Decision Trees • Random Forests 	Stochastic Gradient Descent 77,2%
5	Information Extraction From Text Messages Using Data Mining Techniques	Sartaj Ahmad & Rishabh Varma	2018	Support Vector Machines dan KNN	87%
6	Study of Hoax News Detection Using Naïve Bayes Classifier in Indonesia Language	Inggrid Yanuar, dkk	2017	Naïve Bayes dan SVM	78,6%
7	Hoax in Modern Politics: The Meaning of Hoax in	Pratiwi Utami	2019	-	-

	Indonesian Politics and Democracy				
8	Computational Humor	Kim Binsted	2006	Semantik	-
9	Getting Serious about the Development of Computational Humor	Oliviero Stock & Carlo Strapparava	2003	Penerapan Artificial Intelligence dengan metode semantik	-
10	Sejarah, Teori, Jenis, dan Fungsi Humor	Didiek Rahmanadji	2007	-	-
11	Humor Dalam Kajian Psikologi Islam	Listya Istiningtyas	2014	-	-

L. Kerangka Pikir



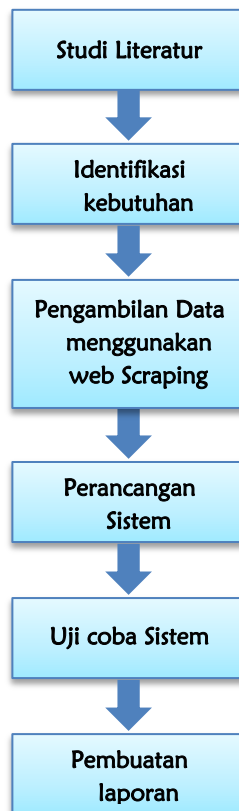
Gambar 2.6. Kerangka Pikir

BAB III

METODOLOGI PENELITIAN

A. Tahapan Penelitian

Tahapan penelitian yang dilakukan, yaitu sebagai berikut :



Gambar 3.1. Tahapan Penelitian

1. Studi literatur dilakukan dengan mencari penelitian-penelitian terkait *computation humor*, *hoaks*, *text mining*, *algoritma tf-idf* dan *cosine similarity*, baik dari buku, jurnal maupun internet.
2. Identifikasi kebutuhan dilakukan penetapan berbagai kebutuhan penelitian untuk menunjang kegiatan penelitian.

3. Pengambilan data dilakukan dengan mengambil data dari situs kominfo.go.id untuk berita hoaks dan corpus data humor diambil dari situs woazy.com dengan menggunakan metode *web scraping*.
4. Perancangan sistem. Data yang telah diambil kemudian diimplementasikan pada algoritma sistem. Penelitian ini menggunakan algoritma *tf-idf* dan *cosine similarity*.
5. Uji coba sistem dilakukan setelah sistem selesai dirancang, maka pengujian akurasi sistem dapat dilakukan dengan melihat output dari hasil implementasi algoritma yang digunakan.
6. Pembuatan laporan. Setelah melewati semua tahapan, proses akhir adalah menuliskan laporan penelitian menyeluruh sebagai bahan publikasi dan penyusunan naskah tugas akhir magister

B. Waktu dan Lokasi Penelitian

1. Waktu Penelitian

Waktu penelitian dilaksanakan selama kurang lebih 6 bulan, dimulai pada bulan Maret sampai September 2020.

2. Lokasi Penelitian

Penelitian ini dilakukan di kota kendari.

C. Jenis Penelitian

Jenis penelitian ini merupakan penelitian eksperimental yang bersifat analisis sehingga dari ruang lingkup masalah dapat dilakukan

dengan metode studi pustaka (*library research*), metode pengumpulan data (*field research*) dan perancangan sistem serta analisis.

D. Sumber Data

Tahapan Pengambilan data melalui proses *web scraping* dengan mengambil link berita hoaks pada situs kominfo.go.id dan link corpus data humor pada situs woazy.com. Data hoaks yang digunakan sebanyak 50 dan corpus data humor sebanyak 182.

E. Instrumen Penelitian

1. Software

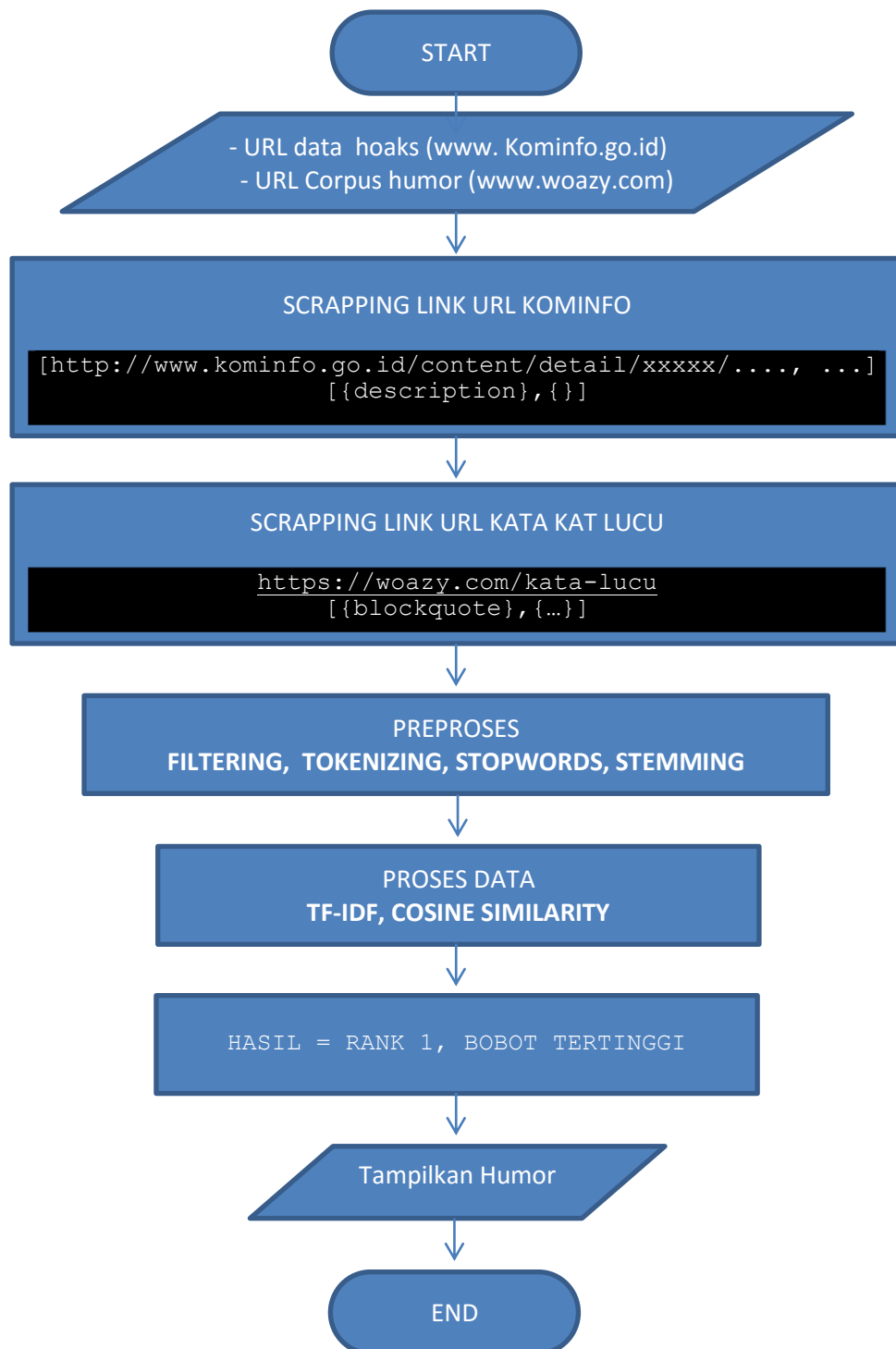
- a. Windows 10
- b. Python
- c. Microsoft Excel

2. Hardware

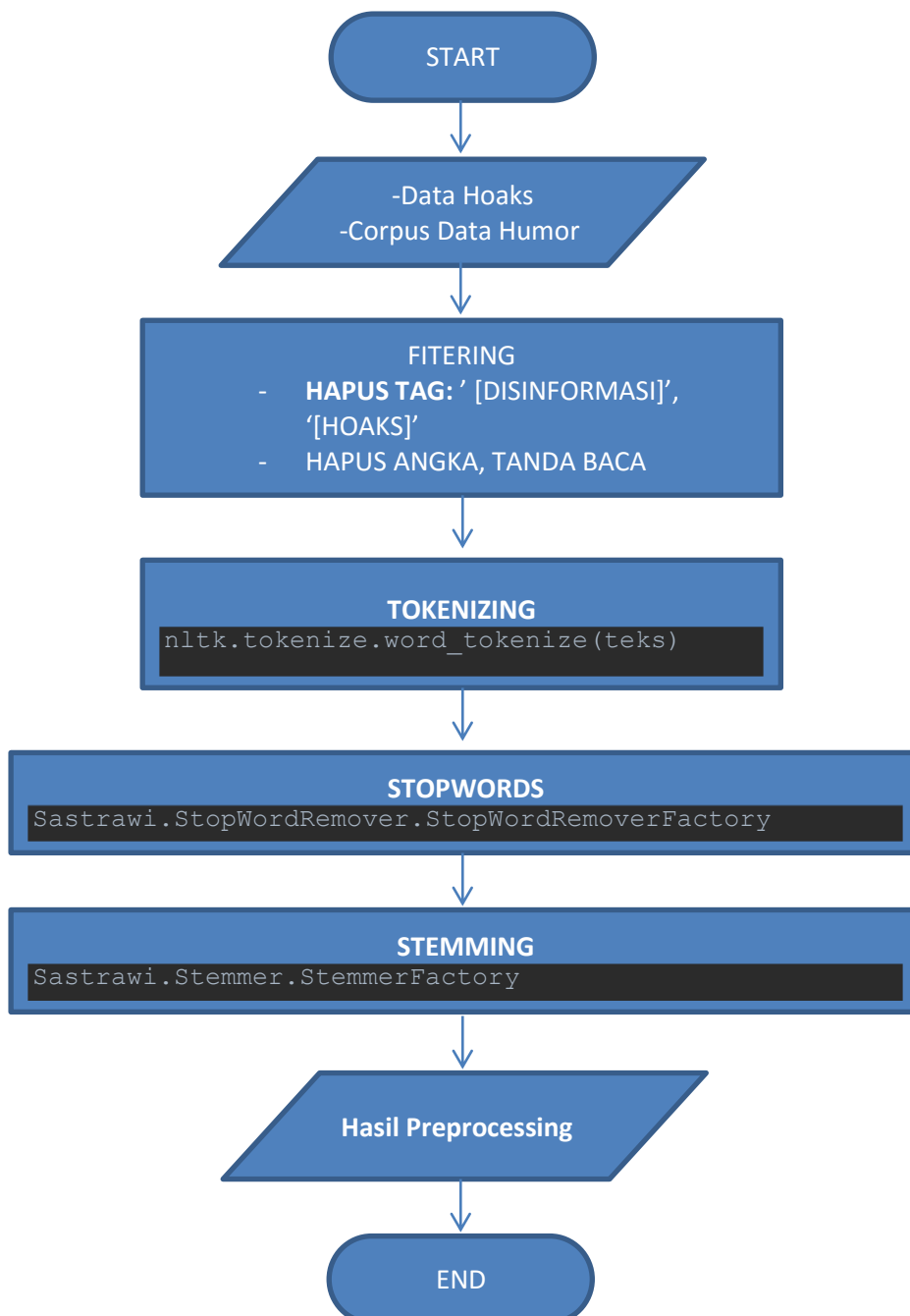
Laptop Lenovo AMD 80E1 dengan Prosesor AMD A8-6410 APU with AMD Radeon R5 Graphics , RAM 4GB.

F. Perancangan Sistem

Arsitektur sistem *computation humor* ditunjukkan pada gambar 3.2

Gambar 3.2. *Flowchart* Sistem

Data hoax yang digunakan berasal dari *link* berita yang disiapkan dalam website kominfo sedangkan data kalimat lucu diambil dari website woazy.com. Data hoax dan data kalimat lucu diperoleh melalui metode *web scraping*. Setelah kedua data diperoleh proses selanjutnya adalah tahap *preprocessing* dengan tujuan mengubah bentuk data yang belum terstruktur menjadi data yang terstruktur sesuai dengan kebutuhan. Selanjutnya adalah tahap pembobotan kata pada kalimat hoax dan kalimat lucu dengan menggunakan *TF-IDF*. Setelah melakukan pembobotan kata, langkah selanjutnya mencari kemiripan dokumen dengan korpus dengan menggunakan rumus *cosine similarity*. Tahap akhir dari sistem adalah perangkingan data berdasarkan hasil perhitungan *cosine similarity*.



Gambar 3.3. *Flowchart* preproses

Pada gambar 3.3 menjelaskan tentang tahapan proses *preprocessing* yaitu *filtering*, *tokenizing*, *stopwors*, *stemming*.

1. Data Masukan

Data masukan merupakan data yang diambil dari link berita hoaks di situs kominfo dan corpus data humor yang diambil dari situs woazy.com. melalui proses *web scraping* sebagaimana tampilannya pada Gambar 3.4.

no	title	description	link	pubdate	kategori	title_remove tag	title_ritering	title_tr	
0	1	Makan Telur Rebus Tengah Malam Dapat Mencegah ... WhatsApp y...	Telah beredar sebuah pesan berantai WhatsApp y...	http://www.kominfo.go.id/content/detail/28879/...	Mon, 24 Aug 2020 17:33:41 +0700	DISINFORMASI	Makan Telur Rebus Tengah Malam Dapat Mencegah ...	makan telur rebus tengah malam dapat mencegah ...	[makan, rebus, le malam, dapat,
1	2	Akun WhatsApp Mengatasnamakan Sekda Kabupaten ...	Beredar akun WhatsApp menggunakan foto profil ...	http://www.kominfo.go.id/content/detail/28878/...	Mon, 24 Aug 2020 17:22:46 +0700	HOAKS	Akun WhatsApp Mengatasnamakan Sekda Kabupaten ...	akun whatsapp mengatasnamakan sekda kabupaten ...	[akun, what mengatasnamakan sekda, ka
2	3	Penggunaan Obat Trombolisis di RESPON Tidak Bis...	Beredar pesan berantai pada media sosial yang ...	http://www.kominfo.go.id/content/detail/28877/...	Mon, 24 Aug 2020 17:31:51 +0700	DISINFORMASI	Penggunaan Obat Trombolisis di RESPON Tidak Bis...	penggunaan obat trombolisis di rpson tidak bis...	[penggunaan, tromboli rspor
3	4	Gedung Kejaksaan Terbakar, Menko Marves Luhut ...	Telah beredar unggahan di media sosial berisi ...	http://www.kominfo.go.id/content/detail/28876/...	Mon, 24 Aug 2020 17:31:05 +0700	HOAKS	Gedung Kejaksaan Terbakar, Menko Marves Luhut ...	gedung kejaksaan terbakar menko marves luhut m...	[ge kejak terbakar, rr marv
4	5	Demokrat: Habib Rizieq ada di Arab, kenapa tak...	Beredar postingan di media sosial Facebook yan...	http://www.kominfo.go.id/content/detail/28875/...	Mon, 24 Aug 2020 17:30:19 +0700	DISINFORMASI	Demokrat: Habib Rizieq ada di Arab, kenapa tak...	demokrat habib rizieq ada di arab kenapa tak d...	[demokrat, l rizieq, ada, di, ke
5	6	Minum Kopi Dengan Jeruk Nipis Dapat Meredakan ...	Telah beredar informasi di media sosial yang m...	http://www.kominfo.go.id/content/detail/28874/...	Mon, 24 Aug 2020 10:28:51 +0700	DISINFORMASI	Minum Kopi Dengan Jeruk Nipis Dapat Meredakan ...	minum kopi dengan jeruk nipis dapat meredakan ...	[minum dengan, nipis, dapat,
6	7	Foto Penampakan Tiga Orang yang Mengganlung An...	Beredar sebuah foto di media Facebook yang me...	http://www.kominfo.go.id/content/detail/28873/...	Mon, 24 Aug 2020 10:27:29 +0700	DISINFORMASI	Foto Penampakan Tiga Orang yang Mengganlung An...	foto penampakan tiga orang yang mengganlung an...	[foto, penamp tiga, orang, meng
7	8	Jokowi Sudah Gagal Dan Tidak Bisa Jadi Preside...	Telah beredar melalui media sosial Facebook Ka...	http://www.kominfo.go.id/content/detail/28872/...	Mon, 24 Aug 2020 10:26:37 +0700	DISINFORMASI	Jokowi Sudah Gagal Dan Tidak Bisa Jadi Preside...	jokowi sudah gagal dan tidak bisa jadi preside...	[jokowi, s gagal, dan, bisa,]
8	9	Berkas - Arsip Perkara Hukum Hangus Dalam Keba...	Telah beredar di media sosial Twitter sebuah p...	http://www.kominfo.go.id/content/detail/28871/...	Mon, 24 Aug 2020 10:25:45 +0700	DISINFORMASI	Berkas - Arsip Perkara Hukum Hangus Dalam Keba...	berkas arsip perkara hukum hangus dalam kebak...	[berkas, perkara, h hangus, dal
9	10	Dokumen WHO Terkait Anjuran Tidak Menggunakan ...	Beredar sebuah dokumen di media sosial yang di...	http://www.kominfo.go.id/content/detail/28854/...	Sun, 23 Aug 2020 20:00:45 +0700	HOAKS	Dokumen WHO Terkait Anjuran Tidak Menggunakan ...	dokumen who terkait anjuran tidak menggunakan ...	[dokumen terkait, an tidak, men

Gambar 3.4. Tampilan *scraping* data hoaks

Adapun distribusi daftar *headline Link* berita hoaks pada *website* kominfo di *update* secara berkala, dan dalam setiap penyajiannya hanya menampilkan 10 link berita hoaks. Untuk data masukan berita hoax adalah *title* atau judul. Data masukkan corpus data humor yang di ambil pada situs www.woazy.com sebanyak 182 kalimat lucu.

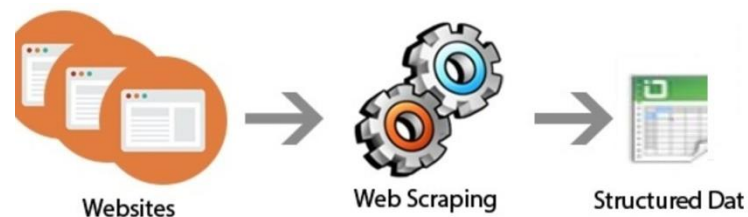
Daftar link berita hoaks tanggal 24 Agustus 2020 diperlihatkan pada tabel 3.1.

Tabel 3.1. Daftar link berita hoaks

No	Link
1	http://www.kominfo.go.id/content/detail/28879/disinformasi-makan-telur-rebus-tengah-malam-dapat-mencegah-covid-19/0/laporan_isu_hoaks
2	http://www.kominfo.go.id/content/detail/28878/hoaks-akun-whatsapp-mengatasnamakan-sekda-kabupaten-kepulauan-selayar/0/laporan_isu_hoaks
3	http://www.kominfo.go.id/content/detail/28877/disinformasi-penggunaan-obat-trombolisis-di-rspon-tidak-bisa-lebih-dari-4-jam-setelah-kejadian-stroke/0/laporan_isu_hoaks
4	http://www.kominfo.go.id/content/detail/28876/hoaks-gedung-kejaksaan-terbakar-menko-marves-luhut-minta-pemutihan-sejumlah-koruptor/0/laporan_isu_hoaks
5	http://www.kominfo.go.id/content/detail/28875/disinformasi-demokrat-habib-rizieq-ada-di-arab-kenapa-tak-dirikan-khilafah-di-sana-beraninya-cuma-di-indonesia-saja-dasar-provokator/0/laporan_isu_hoaks
6	http://www.kominfo.go.id/content/detail/28874/disinformasi-minum-kopi-dengan-jeruk-nipis-dapat-meredakan-migrain/0/laporan_isu_hoaks
7	http://www.kominfo.go.id/content/detail/28873/disinformasi-foto-penampakan-tiga-orang-yang-menggantung-anggota-tni-di-pohon/0/laporan_isu_hoaks
8	http://www.kominfo.go.id/content/detail/28872/disinformasi-jokowi-sudah-gagal-dan-tidak-bisa-jadi-presiden-lagi-jutaan-rakyat-akan-memaksa-angkat-kaki/0/laporan_isu_hoaks
9	http://www.kominfo.go.id/content/detail/28871/disinformasi-berkas-arsip-perkara-hukum-hangus-dalam-kebakaran-gedung-kejaksaan-agung/0/laporan_isu_hoaks
10	http://www.kominfo.go.id/content/detail/28854/hoaks-dokumen-who-terkait-anjuran-tidak-menggunakan-masker-selama-pandemi-covid-19/0/laporan_isu_hoaks

2. Web Scraping

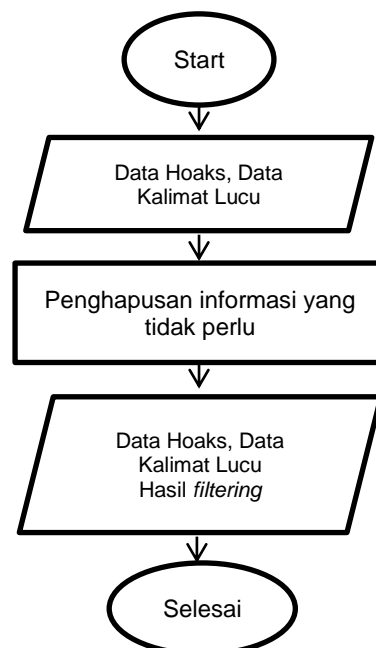
Proses *Web Scraping* bertujuan untuk mengambil sebuah dokumen dari internet, berupa halaman-halaman web dalam bahasa markup seperti HTML atau XHTML, dan menganalisis dokumen tersebut untuk diambil data tertentu dari halaman tersebut untuk digunakan.



Gambar 3.5. Alur *Web Scraping* (www.dev.to, 2019)

3. Filtering

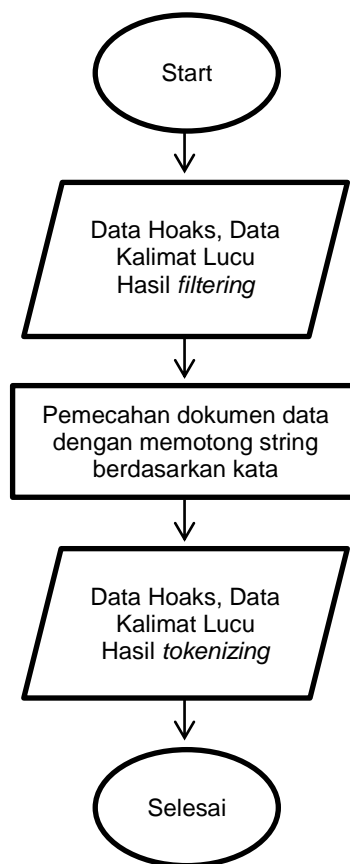
Filtering bertujuan untuk mengambil isi dari berita tanpa terikat dengan *noise* yang dapat berpengaruh dengan proses *preprocessing* sehingga tidak merusak bobot dari suatu dokumen. Alur dari Proses *filtering* diperlihatkan pada Gambar 3.6.



Gambar 3.6 *Flowchart filtering*

4. *Tokenizing*

Tokenizing berfungsi untuk membagi teks yang dapat berupa kalimat, paragraf atau dokumen, menjadi *token-token*/bagian-bagian tertentu. Pada penelitian ini, *tokenizing* memecah teks yang berupa dokumen berdasarkan kata. Alur dari Proses ini dapat dilihat pada Gambar 3.7.

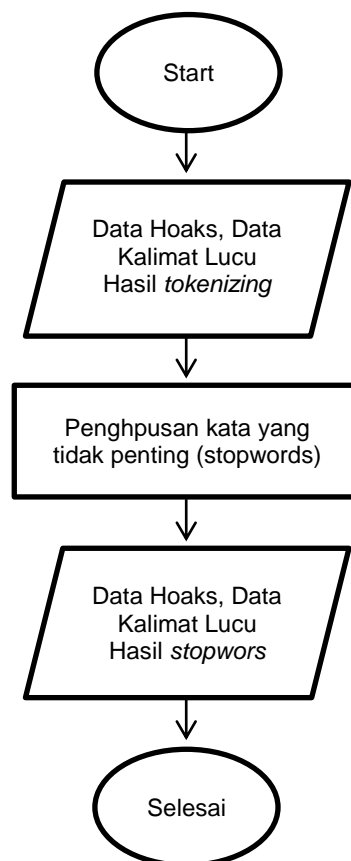


Gambar 3.7. *Flowchart tokenizing*

5. *Stopwords*

Stop-word yaitu penghilangan kata tidak penting (*stopword*) pada deskripsi melalui pengecekan kata-kata hasil token deskripsi apakah termasuk di dalam daftar kata tidak penting (*stop-word*) atau tidak.

Jika termasuk di dalam *stop-word* maka kata-kata tersebut akan di-*remove* dari deskripsi sehingga kata-kata yang tersisa di dalam deskripsi di anggap sebagai kata-kata penting. Proses *stopwords* ini mempunyai alur yang digambarkan pada gambar 3.8 sebagai berikut.

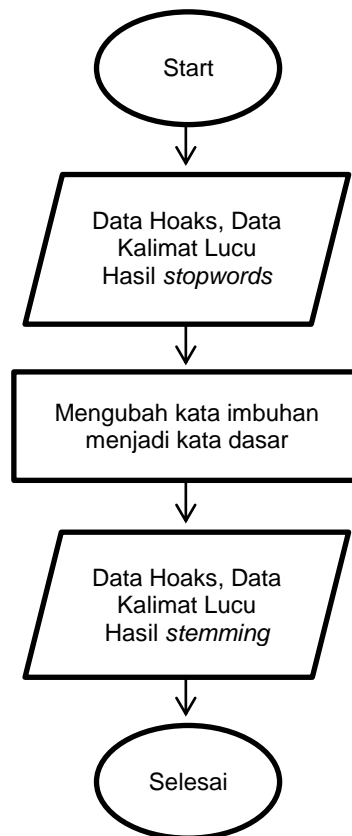


Gambar 3.8. *Flowchart stopwords*

6. *Stemming*

Stemming adalah tahap mencari root kata dari tiap kata hasil *stopwords*. Pada tahap ini dilakukan proses pengembalian berbagai bentuk kata imbuhan ke dalam suatu representasi yang sama dengan menghilangkan imbuhan seperti diantaranya “yang”, “di”, “ke”,

“me”, “meng”, “kan” menjadi bentuk kata dasarnya (stem). Alur stemming dapat dilihat pada gambar 3.9.

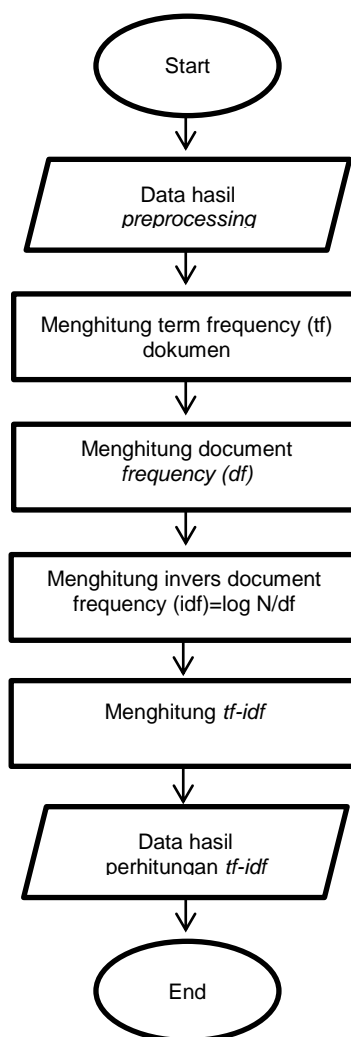


Gambar 3.9. *Flowchart stemming*

7. Pembobotan *tf-idf*

Pada text mining, term frequency – inverse document frequency (TF-IDF) bertujuan untuk memberikan bobot hubungan suatu kata (term) terhadap dokumen. Di mana term merupakan kata dari hasil proses preprocessing tf menyatakan jumlah berapa banyak keberadaan suatu term (t) dalam satu dokumen (d), df merupakan perhitungan berapa banyaknya dokumen dimana suatu term (t) muncul dan idf berfungsi

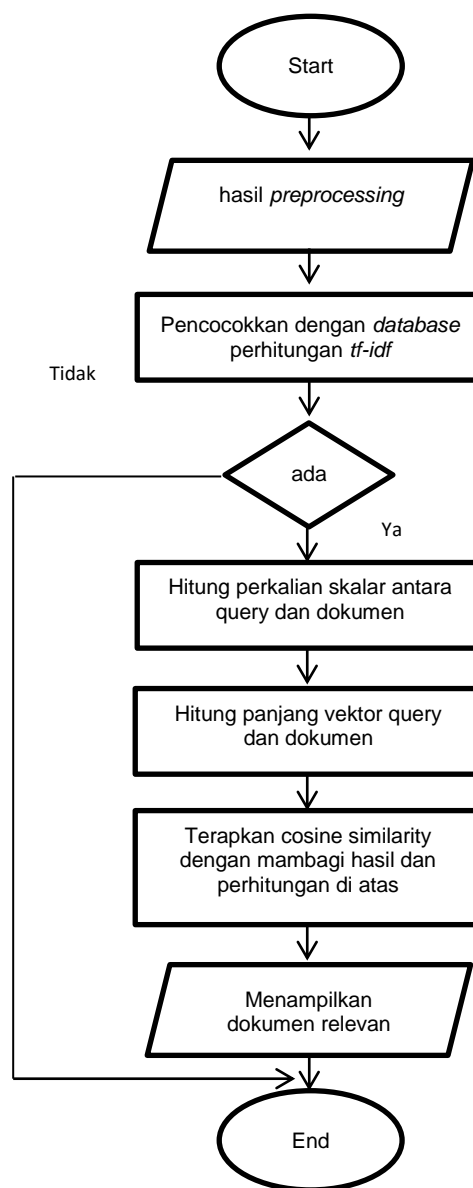
mengurangi bobot suatu term jika kemunculannya banyak tersebar di seluruh koleksi dokumen. Hal ini dikarenakan terkadang suatu term muncul di hampir sebagian besar dokumen mengakibatkan proses pencarian term unik terganggu. Proses pembobotan kata TF-IDF diperlihatkan pada Gambar 3.10.



Gambar 3.10. *Flowchart tf-idf*

8. Cosine similarity

Setelah proses *tf-idf* dengan memberikan bobot dokumen terhadap korpus, langkah terakhir yang dilakukan mencari kemiripan dokumen dengan korpus dengan menggunakan rumus *cosine similarity*. Proses ini ditunjukkan pada gambar 3.11.



Gambar 3.11. Flowchart cosine similarity

BAB IV

HASIL DAN PEMBAHASAN

Bab ini membahas tentang hasil tahapan *preprocessing* melalui proses *tokenizing*, *stopword*, dan *stemming* dilanjutkan dengan penerapan metode *TF-IDF* dan *cosine similarity*.

A. Hasil Proses *Preprocessing*

1. *Filtering*

Proses ini menghilangkan isi dokumen yang tidak dibutuhkan, data yang di ambil dari situs www.kominfo.go.id, bagian yang diberi kotak warna merah merupakan bagian yang tidak diperlukan dalam proses *computation hoax* yang disebut *noise*, dapat dilihat pada tabel 4.1. Untuk proses *filtering* corpus data humor di ambil dari situs woazy.com, ditunjukkan pada tabel 4.2.

```
def removetag(teks):
    t = teks.replace('[DISINFORMASI]', '')
    t = t.replace('[HOAKS]', '')
    t = t.replace('"', '')
    t = t.replace("'", '')
    return t.strip()

def filtering(teks):
    lteks = teks.lower() # lower case
    oteks = re.sub(r"\d+", "", lteks) # hilangkan angka
    tteks = oteks.translate(str.maketrans("", "", string.punctuation))
    #hilangkan tanda baca
    return tteks.strip()

title_hoax = hoaxs.title[idx]
t_filter = filtering(title_hoax)

katalucu = kata2lucu.kalimat[idxk]
k_filter = filtering(katalucu)
```

Source kode tahap *filtering*

Tabel 4.1. Contoh hasil proses filtering berita hoaks

Hasil <i>web scraping</i>	Hasil <i>filtering</i>
[HOAKS] Gedung Kejaksaan Terbakar, Menko Marves Luhut Minta Pemutihan' Sejumlah Koruptor	gedung kejaksaan terbakar menko marves luhut minta pemutihan sejumlah koruptor

Dapat dilihat pada tabel 4.1, bahwa data hasil web scraping yaitu “[HOAKS]” telah dihilangkan pada proses *filtering*.

Tabel 4.2 Contoh hasil proses *filtering* corpus data humor

Hasil <i>web scraping</i>	Hasil <i>filtering</i>
“kebakaran terjadi di kamar mayat terbakar melarikan diri dan bingung yang baca berita ini”	kebakaran terjadi di kamar mayat terbakar melarikan diri dan bingung yang baca berita ini

2. *Tokenizing*

Pada proses teks yang masih berbentuk kalimat dipecah menjadi kata-kata yang dapat dilihat pada tabel 4.3 dan 4.4 hasil *tokenizing* dengan menggunakan pemisahan tanda petik di setiap kata serta tanda koma.

```
def tokenisasi(teks):
    return nltk.tokenize.word_tokenize(teks)

title_hoax = hoaxs.title[idx]
t_tokens = tokenisasi(t_filter)

katalucu = kata2lucu.kalimat[idxk]
k_tokens = tokenisasi(k_filter)
```

Source kode tahap *tokenizing*

Tabel 4.3. Contoh hasil proses *tokenizing* berita hoaks

Hasil <i>filtering</i>	Hasil <i>tokenizing</i>
gedung kejaksaan terbakar menko marves luhut minta pemutihan sejumlah koruptor	'gedung', 'kejaksaan', 'terbakar', 'menko', 'marves', 'luhut', 'minta', 'pemutihan', 'sejumlah', 'koruptor'

Tabel 4.4. Contoh hasil proses *tokenizing* corpus data humor

Hasil <i>filtering</i>	Hasil <i>tokenizing</i>
kebakaran terjadi di kamar mayat terbakar melarikan diri dan bingung yang baca berita ini	'kebakaran', 'terjadi', 'di', 'kamar', 'mayat', 'terbakar', 'melarikan', 'diri', 'dan', 'bingung', 'yang', 'baca', 'berita', 'ini'

3. *Stopwords*

Kata-kata yang dihapus dalam proses stopword adalah kata-kata yang telah ada di dalam stop list. Pada penelitian ini, stop list yang digunakan yaitu corpus sastraawi. Ditunjukkan pada Tabel 4.5 dan 4.6.

```

stop_factory_word = StopWordRemoverFactory().get_stop_words()
                    # load default stopword
more_stopword_word = ['daring', 'online', 'rp', 't', 'ri']
                    # menambahkan stopword
data_stop_word = stop_factory_word + more_stopword_word
                    # menggabungkan stopword
dictionary = ArrayDictionary(data_stop_word)
strRem = StopWordRemover(dictionary)

title_hoax = hoaxs.title[idx]
t_stopwords = filterSastrawi(' '.join(t_tokens)).strip()

katalucu = kata2lucu.kalimat[idkx]
k_stopwords = filterSastrawi(' '.join(k_tokens)).strip()

```

Source kode tahap *stopwords*

Tabel 4.5. Contoh hasil proses *stopwords* berita hoaks

Hasil <i>tokenizing</i>	Hasil <i>stopwords</i>
'gedung', 'kejaksaan', 'terbakar', 'menko', 'marves', 'luhut', 'minta', 'pemutihan', 'sejumlah', 'koruptor'	gedung kejaksaan terbakar menko marves luhut minta pemutihan sejumlah koruptor

Tabel 4.6. Contoh hasil proses *stopwords* corpus data humor

Hasil <i>tokenizing</i>	Hasil <i>stopwords</i>
'kebakaran', 'terjadi', 'di', 'kamar', 'mayat', 'terbakar', 'melarikan', 'diri', 'dan', 'bingung', 'yang', 'baca', 'berita', 'ini'	kebakaran terjadi kamar mayat terbakar melarikan diri bingung baca berita

4. *Stemming*

Pada proses ini kata-kata berimbuhan yaitu “kebakaran”, “kejaksaan”, diubah ke bentuk dasar dengan menghilangkan imbuhanannya sehingga menghasilkan bentuk yaitu “bakar”, “jaksa” pada hasil *stemming*. Proses ini dapat di lihat pada tabel 4.7 dan 4.8.

```
factoryStemer = StemmerFactory()
stemmerTeks = factoryStemer.create_stemmer()
def stemmer(teks):
    hasil = stemmerTeks.stem(teks)
    return tokenisasi(hasil)

title_hoax = hoaxs.title[idx]
t_stemm = stemmer(t_stopwords)

katalucu = kata2lucu.kalimat[idkx]
k_stemm = stemmer(k_stopwords)
```

Source kode tahap *stemming*

Tabel 4.7. Contoh hasil proses *stemming* berita hoaks

Hasil <i>stopwords</i>	Hasil <i>stemming</i>
gedung kejaksaan terbakar menko marves luhut minta pemutihan sejumlah koruptor	'gedung', 'jaksa', 'bakar', 'menko', 'marves', 'luhut', 'minta', 'mutih', 'jumlah', 'koruptor'

Tabel 4.8. Contoh hasil proses *stemming* corpus data humor

Hasil <i>stopwords</i>	Hasil <i>stemming</i>
kebakaran terjadi kamar mayat terbakar melarikan diri bingung baca berita	'bakar', 'jadi', 'kamar', 'mayat', 'bakar', 'lari', 'diri', 'bingung', 'baca', 'berita'

5. Pembobotan kata (*TF-IDF*)

Tahap pembobotan kata dengan menggunakan metode term frequency inverse document frequency (*TF-IDF*), dimana daftar term hasil *stemming* dilakukan perhitungan untuk mengetahui bobot perkata dengan menghitung jumlah term frequency dokumen (*tf*) terlebih dahulu, kemudian menghitung nilai jumlah dokumen yang memiliki term (*df*), dan selanjutnya menghitung nilai *idf* dengan rumus $\log=N/df$, langkah terakhir adalah menentukan bobot kata dengan mengalikan *TF* dan *IDF* dengan rumus $W_{dt}=TF_{dt} \times IDF_t$. Hasil dari proses perhitungan ini disimpan dalam microsoft excel.

Pembobotan (*tf-idf*) pada sistem *computation* ini adalah dengan menghitung setiap *title* berita hoax terhadap masing-masing korpus humr. Jadi satu *title* hoaks akan di proses sebanyak korpus data humor yaitu 182.

Hasil pembobotan tf-idf pada gambar 4.1 didapat empat (4) corpus data humor, untuk data kalimat hoaks “Gedung Kejaksaan Terbakar, Menko Marves Luhut Minta 'Pemutihan' Sejumlah Koruptor”. Keempat corpus data humor yang mempunyai bobot adalah :

1. kebakaran terjadi di kamar mayat terbakar melarikan diri dan bingung yang baca berita ini
2. jika kamu ingin bersinar seperti matahari pertama kamu harus terbakar sepertinya
3. jika kamu ingin bersinar seperti matahari hal pertama yang harus kamu lakukan adalah terbakar sepertinya.
4. kamu berada disudut jalan tersipu malu wajah kamu berseri dgn penuh segan kamu ulurkan tanganmu tanpa ragu seraya berkata minta sedekahnya

Proses perhitungan pembobotan corpus data humor pertama dapat dilihat pada gambar 4.1.

Doc 1 (Hoax) : Gedung Kejaksaan Terbakar, Menko Marves Luhut Minta 'Pemutihan' Sejumlah Koruptor

Doc 2 (Humor) : kebakaran terjadi di kamar mayat terbakar melarikan diri dan bingung yang baca berita ini

Stemm Doc 1 : ['gedung', 'jaksa', 'bakar', 'menko', 'marves', 'luhut', 'minta', 'mutih', 'jumlah', 'koruptor']

Stemm Doc 2 : ['bakar', 'jadi', 'kamar', 'mayat', 'bakar', 'lari', 'diri', 'bingung', 'baca', 'berita']

Dict Word List : ['baca', 'bakar', 'berita', 'bingung', 'diri', 'gedung', 'jadi', 'jaksa', 'jumlah', 'kamar', 'koruptor', 'lari', 'luhut', 'marves', 'mayat', 'menko', 'minta', 'mutih']

	term	d1	d2	df	n_per_df	idf	idf_plus_1	wdt_d1	wdt_d2
0	baca	0	1	1	2.0	0.30103	1.30103	0.00000	1.30103
1	bakar	1	2	2	1.0	0.00000	1.00000	1.00000	2.00000
2	berita	0	1	1	2.0	0.30103	1.30103	0.00000	1.30103
3	bingung	0	1	1	2.0	0.30103	1.30103	0.00000	1.30103
4	diri	0	1	1	2.0	0.30103	1.30103	0.00000	1.30103
5	gedung	1	0	1	2.0	0.30103	1.30103	1.30103	0.00000
6	jadi	0	1	1	2.0	0.30103	1.30103	0.00000	1.30103
7	jaksa	1	0	1	2.0	0.30103	1.30103	1.30103	0.00000
8	jumlah	1	0	1	2.0	0.30103	1.30103	1.30103	0.00000
9	kamar	0	1	1	2.0	0.30103	1.30103	0.00000	1.30103
10	koruptor	1	0	1	2.0	0.30103	1.30103	1.30103	0.00000
11	lari	0	1	1	2.0	0.30103	1.30103	0.00000	1.30103
12	luhut	1	0	1	2.0	0.30103	1.30103	1.30103	0.00000
13	marves	1	0	1	2.0	0.30103	1.30103	1.30103	0.00000
14	mayat	0	1	1	2.0	0.30103	1.30103	0.00000	1.30103
15	menko	1	0	1	2.0	0.30103	1.30103	1.30103	0.00000
16	minta	1	0	1	2.0	0.30103	1.30103	1.30103	0.00000
17	mutih	1	0	1	2.0	0.30103	1.30103	1.30103	0.00000

Gambar 4.1. Contoh perhitungan *Tf-Idf* kalimat lucu 1

Proses perhitungan pembobotan corpus data humor dua :

Doc 1 (Hoaks) : Gedung Kejaksaan Terbakar, Menko Marves Luhut
Minta 'Pemutihan' Sejumlah Koruptor

Doc 2 (Humor) : jika kamu ingin bersinar seperti matahari pertama
kamu harus terbakar sepertinya

Stemm Doc 1 : ['gedung', 'jaksa', 'bakar', 'menko', 'marves', 'luhut',
'minta', 'mutih', 'jumlah', 'koruptor']

Stemm Doc 2 : ['kamu', 'sinar', 'matahari', 'pertama', 'kamu', 'bakar',
'seperti']

Dict Word List : ['bakar', 'gedung', 'jaksa', 'jumlah', 'kamu', 'koruptor',
'luhut', 'marves', 'matahari', 'menko', 'minta', 'mutih',
'pertama', 'seperti', 'sinar']

	term	d1	d2	df	n_per_df	idf	idf_plus_1	wdt_d1	wdt_d2
0	bakar	1	1	2	1.0	0.00000	1.00000	1.00000	1.00000
1	gedung	1	0	1	2.0	0.30103	1.30103	1.30103	0.00000
2	jaksa	1	0	1	2.0	0.30103	1.30103	1.30103	0.00000
3	jumlah	1	0	1	2.0	0.30103	1.30103	1.30103	0.00000
4	kamu	0	2	1	2.0	0.30103	1.30103	0.00000	2.60206
5	koruptor	1	0	1	2.0	0.30103	1.30103	1.30103	0.00000
6	luhut	1	0	1	2.0	0.30103	1.30103	1.30103	0.00000
7	marves	1	0	1	2.0	0.30103	1.30103	1.30103	0.00000
8	matahari	0	1	1	2.0	0.30103	1.30103	0.00000	1.30103
9	menko	1	0	1	2.0	0.30103	1.30103	1.30103	0.00000
10	minta	1	0	1	2.0	0.30103	1.30103	1.30103	0.00000
11	mutih	1	0	1	2.0	0.30103	1.30103	1.30103	0.00000
12	pertama	0	1	1	2.0	0.30103	1.30103	0.00000	1.30103
13	seperti	0	1	1	2.0	0.30103	1.30103	0.00000	1.30103
14	sinar	0	1	1	2.0	0.30103	1.30103	0.00000	1.30103

Gambar 4.2. Contoh perhitungan *Tf-Idf* kalimat 2

Proses perhitungan pembobotan corpus data humor ketiga :

Doc 1 (Hoaks) : Gedung Kejaksaan Terbakar, Menko Marves Luhut
Minta 'Pemutihan' Sejumlah Koruptor

Doc 2 (Humor) : jika kamu ingin bersinar seperti matahari hal pertama
yang harus kamu lakukan adalah terbakar sepertinya

Stemm Doc 1 : ['gedung', 'jaksa', 'bakar', 'menko', 'marves', 'luhut',
'minta', 'mutih', 'jumlah', 'koruptor']

Stemm Doc 2 : ['kamu', 'sinar', 'matahari', 'pertama', 'harus', 'kamu',
'laku', 'bakar', 'seperti']

Dict Word List : ['bakar', 'gedung', 'harus', 'jaksa', 'jumlah', 'kamu',
'koruptor', 'laku', 'luhut', 'marves', 'matahari', 'menko',
'minta', 'mutih', 'pertama', 'seperti', 'sinar']

	term	d1	d2	df	n_per_df	idf	idf_plus_1	wdt_d1	wdt_d2
0	bakar	1	1	2	1.0	0.00000	1.00000	1.00000	1.00000
1	gedung	1	0	1	2.0	0.30103	1.30103	1.30103	0.00000
2	harus	0	1	1	2.0	0.30103	1.30103	0.00000	1.30103
3	jaksa	1	0	1	2.0	0.30103	1.30103	1.30103	0.00000
4	jumlah	1	0	1	2.0	0.30103	1.30103	1.30103	0.00000
5	kamu	0	2	1	2.0	0.30103	1.30103	0.00000	2.60206
6	koruptor	1	0	1	2.0	0.30103	1.30103	1.30103	0.00000
7	laku	0	1	1	2.0	0.30103	1.30103	0.00000	1.30103
8	luhut	1	0	1	2.0	0.30103	1.30103	1.30103	0.00000
9	marves	1	0	1	2.0	0.30103	1.30103	1.30103	0.00000
10	matahari	0	1	1	2.0	0.30103	1.30103	0.00000	1.30103
11	menko	1	0	1	2.0	0.30103	1.30103	1.30103	0.00000
12	minta	1	0	1	2.0	0.30103	1.30103	1.30103	0.00000
13	mutih	1	0	1	2.0	0.30103	1.30103	1.30103	0.00000
14	pertama	0	1	1	2.0	0.30103	1.30103	0.00000	1.30103
15	seperti	0	1	1	2.0	0.30103	1.30103	0.00000	1.30103
16	sinar	0	1	1	2.0	0.30103	1.30103	0.00000	1.30103

Gambar 4.3. Contoh perhitungan *Tf-Idf* kalimat 3

Proses perhitungan pembobotan corpus data humor keempat :

Doc 1 (Hoaks) : Gedung Kejaksaan Terbakar, Menko Marves Luhut
Minta 'Pemutihan' Sejumlah Koruptor

Doc 2 (Humor) : kamu berada disudut jalan tersipu malu wajah kamu
berseri dgn penuh segan kamu ulurkan tanganmu
tanpa ragu seraya berkata minta sedekahnya

Stemm Doc 1 : ['gedung', 'jaksa', 'bakar', 'menko', 'marves', 'luhut',
'minta', 'mutih', 'jumlah', 'koruptor']

Stemm Doc 2 : ['kmu', 'ada', 'sudut', 'jalan', 'sipu', 'maluwajah', 'kmu',
'seri', 'dgn', 'penuh', 'segan', 'kmu', 'ulur', 'tangan',
'ragu', 'kata', 'minta', 'sedekah']

Dict Word List : ['ada', 'bakar', 'dgn', 'gedung', 'jaksa', 'jalan', 'jumlah',
'kata', 'kmu', 'koruptor', 'luhut', 'maluwajah', 'marves',
'menko', 'minta', 'mutih', 'penuh', 'ragu', 'sedekah',
'segan', 'seri', 'sipu', 'sudut', 'tangan', 'ulur']

	term	d1	d2	df	n_per_df	idf	idf_plus_1	wdt_d1	wdt_d2
0	ada	0	1	1	2.0	0.30103	1.30103	0.00000	1.30103
1	bakar	1	0	1	2.0	0.30103	1.30103	1.30103	0.00000
2	dgn	0	1	1	2.0	0.30103	1.30103	0.00000	1.30103
3	gedung	1	0	1	2.0	0.30103	1.30103	1.30103	0.00000
4	jaksa	1	0	1	2.0	0.30103	1.30103	1.30103	0.00000
5	jalan	0	1	1	2.0	0.30103	1.30103	0.00000	1.30103
6	jumlah	1	0	1	2.0	0.30103	1.30103	1.30103	0.00000
7	kata	0	1	1	2.0	0.30103	1.30103	0.00000	1.30103
8	kmu	0	3	1	2.0	0.30103	1.30103	0.00000	3.90309
9	koruptor	1	0	1	2.0	0.30103	1.30103	1.30103	0.00000
10	luhut	1	0	1	2.0	0.30103	1.30103	1.30103	0.00000
11	maluwajah	0	1	1	2.0	0.30103	1.30103	0.00000	1.30103
12	marves	1	0	1	2.0	0.30103	1.30103	1.30103	0.00000
13	menko	1	0	1	2.0	0.30103	1.30103	1.30103	0.00000
14	minta	1	1	2	1.0	0.00000	1.00000	1.00000	1.00000
15	mutih	1	0	1	2.0	0.30103	1.30103	1.30103	0.00000
16	penuh	0	1	1	2.0	0.30103	1.30103	0.00000	1.30103
17	ragu	0	1	1	2.0	0.30103	1.30103	0.00000	1.30103
18	sedekah	0	1	1	2.0	0.30103	1.30103	0.00000	1.30103
19	segan	0	1	1	2.0	0.30103	1.30103	0.00000	1.30103
20	seri	0	1	1	2.0	0.30103	1.30103	0.00000	1.30103
21	sipu	0	1	1	2.0	0.30103	1.30103	0.00000	1.30103
22	sudut	0	1	1	2.0	0.30103	1.30103	0.00000	1.30103
23	tangan	0	1	1	2.0	0.30103	1.30103	0.00000	1.30103
24	ulur	0	1	1	2.0	0.30103	1.30103	0.00000	1.30103

Gambar 4.4. Contoh perhitungan *Tf-Idf* kalimat lucu 4

6. Cosine similarity

Setelah proses *tf-idf*, langkah terakhir yang dilakukan mencari kemiripan dokumen dengan korpus dengan menggunakan algoritma *cosine similarity*.

Berikut perhitungan *cosine similarity* dengan modul *sklearn* untuk keempat dokumen di atas :

- Hasil perhitungan *cosine similarity* kalimat lucu pertama :

```
# cosine similarity dengan modul sklearn
from sklearn.metrics.pairwise import cosine_similarity
cosim = cosine_similarity([wdt_d1, wdt_d2])
print("Hasil Cosim Modul Sklearn =", round(cosim[0,1],3))
round(cosim[0,1],3)
```

Hasil Cosim Modul Sklearn = 0.119

Gambar 4.5. Contoh hasil *cosine similarity* kalimat 1

Untuk proses perhitungan *cosine similarity* kalimat lucu pertama, dapat dilihat pada gambar 4.6.

```
In [23]: # cosine similarity rumus manual
#Tujuan zip () adalah untuk memetakan indeks serupa dari beberapa container sehingga dapat
a_vect = wdt_d1
b_vect = wdt_d2
print(wdt_d1)
panjang_vector_d1 = sum(av * av for av in a_vect) ** 0.5 # exponent dari penjumlahan seti
print("v d1:", panjang_vector_d1)
panjang_vector_d2 = sum(bv * bv for bv in b_vect) ** 0.5 # exponent dari penjumlahan seti
print("v d2:", panjang_vector_d2)
dot = sum(av * bv for av, bv in zip(a_vect, b_vect)) # penjumlahan dari setiap perkalian
print('dot:', dot)
cosine = dot / (panjang_vector_d1 * panjang_vector_d2) # jumlah perkalian setiap bobot k
round(cosine,3)

[0.0, 1.0, 0.0, 0.0, 0.0, 1.3010299956639813, 0.0, 1.3010299956639813, 1.3010299956639813,
299956639813, 1.3010299956639813, 0.0, 1.3010299956639813, 1.3010299956639813, 1.3010299956
v d1: 4.029157659679846
v d2: 4.188249323636232
dot: 2.0

Out[23]: 0.119
```

Gambar 4.6. Contoh proses hasil *cosine similarity* kalimat 1

Pada gambar 4.6 di atas, hasil perhitungan *cosine similarity* untuk kalimat pertama adalah 0,119

- Proses perhitungan *cosine similarity* kalimat lucu kedua :

```
# cosine similiraty dengan modul sklearn
from sklearn.metrics.pairwise import cosine_similarity
cosim = cosine_similarity([wdt_d1, wdt_d2])
print("Hasil Cosim Modul Sklearn =", round(cosim[0,1],3))
round(cosim[0,1],3)
```

Hasil Cosim Modul Sklearn = 0.065

Gambar 4.7. Contoh hasil *cosine similarity* kalimat 2

Untuk proses perhitungan *cosine similarity* kalimat lucu kedua

```
In [26]: # cosine similarity rumus manual
#Tujuan zip () adalah untuk memetakan indeks serupa dari beberapa container sehingga c
a_vect = wdt_d1
b_vect = wdt_d2
print(wdt_d1)
panjang_vector_d1 = sum(av * av for av in a_vect) ** 0.5 # exponent dari penjumlahan
print("v d1:", panjang_vector_d1)
panjang_vector_d2 = sum(bv * bv for bv in b_vect) ** 0.5 # exponent dari penjumlahan
print("v d2:", panjang_vector_d2)
dot = sum(av * bv for av, bv in zip(a_vect, b_vect)) # penjumlahan dari setiap perk
print('dot:', dot)
cosine = dot / (panjang_vector_d1 * panjang_vector_d2) # jumlah perkalian setiap bc
round(cosine,3)

[1.0, 1.3010299956639813, 1.3010299956639813, 1.3010299956639813, 0.0, 1.3010299956639813, 0.0, 1.3010299956639813, 1.3010299956639813, 1.3010299956639813, 0.0, 0.0, 0.0]
v d1: 4.029157659679846
v d2: 3.8133230124052373
dot: 1.0
```

Out[26]: 0.065

Gambar 4.8. Contoh proses hasil *cosine similarity* kalimat 2

Pada gambar di atas hasil perhitungan *cosine similarity* untuk kalimat kedua adalah 0.065

- Proses perhitungan *cosine similarity* kalimat lucu ketiga :

```
# cosine similarity dengan modul sklearn
from sklearn.metrics.pairwise import cosine_similarity
cosim = cosine_similarity([wdt_d1, wdt_d2])
print("Hasil Cosim Modul Sklearn =", round(cosim[0,1],3))
round(cosim[0,1],3)
```

Hasil Cosim Modul Sklearn = 0.059

Gambar 4.9. Contoh hasil *cosine similarity* kalimat lucu 3

Untuk proses perhitungan *cosine similarity* kalimat lucu ketiga, dapat dilihat pada gambar 4.10.

```
In [29]: # cosine similarity rumus manual
#Tujuan zip () adalah untuk memetakan indeks serupa dari beberapa container sehingga
a_vect = wdt_d1
b_vect = wdt_d2
print(wdt_d1)
panjang_vector_d1 = sum(av * av for av in a_vect) ** 0.5 # exponent dari penjumlah
print("v d1:", panjang_vector_d1)
panjang_vector_d2 = sum(bv * bv for bv in b_vect) ** 0.5 # exponent dari penjumlah
print("v d2:", panjang_vector_d2)
dot = sum(av * bv for av, bv in zip(a_vect, b_vect)) # penjumlahan dari setiap pe
print('dot:', dot)
cosine = dot / (panjang_vector_d1 * panjang_vector_d2) # jumlah perkalian setiap
round(cosine,3)
<
[1.0, 1.3010299956639813, 0.0, 1.3010299956639813, 1.3010299956639813, 0.0, 1.301029
299956639813, 0.0, 1.3010299956639813, 1.3010299956639813, 1.3010299956639813, 0.0,
v d1: 4.029157659679846
v d2: 4.234004073707793
dot: 1.0
```

Out[29]: 0.059

Gambar 4.10. Contoh proses hasil *cosine similarity* kalimat 3

Hasil perhitungan *cosine similarity* untuk kalimat ketiga adalah

0.059

- Proses perhitungan *cosine similarity* kalimat lucu keempat :

```
# cosine similiraty dengan modul sklearn
from sklearn.metrics.pairwise import cosine_similarity
cosim = cosine_similarity([wdt_d1, wdt_d2])
print("Hasil Cosim Modul Sklearn =", round (cosim[0,1],3))
round(cosim[0,1],3)
```

Hasil Cosim Modul Sklearn = 0.039

Gambar 4.11. Contoh hasil *cosine similarity* kalimat lucu 4

Proses perhitungan *cosine similarity* kalimat lucu keempat, dapat dilihat pada gambar 4.12.

```
In [32]: # cosine similarity rumus manual
#Tujuan zip () adalah untuk memetakan indeks serupa dari beberapa container sehi
a_vect = wdt_d1
b_vect = wdt_d2
print(wdt_d1)
panjang_vector_d1 = sum(av * av for av in a_vect) ** 0.5 # exponent dari penju
print("v d1:", panjang_vector_d1)
panjang_vector_d2 = sum(bv * bv for bv in b_vect) ** 0.5 # exponent dari penju
print("v d2:", panjang_vector_d2)
dot = sum(av * bv for av, bv in zip(a_vect, b_vect)) # penjumlahan dari setic
print('dot:', dot)
cosine = dot / (panjang_vector_d1 * panjang_vector_d2) # jumlah perkalian set
round (cosine,3)

[0.0, 1.3010299956639813, 0.0, 1.3010299956639813, 1.3010299956639813, 0.0, 1.30
1.3010299956639813, 0.0, 1.3010299956639813, 1.3010299956639813, 1.0, 1.30102999
0, 0.0]
v d1: 4.029157659679846
v d2: 6.319146947270702
dot: 1.0

Out[32]: 0.039
```

Gambar 4.12. Contoh proses hasil *cosine similarity* kalimat lucu 4

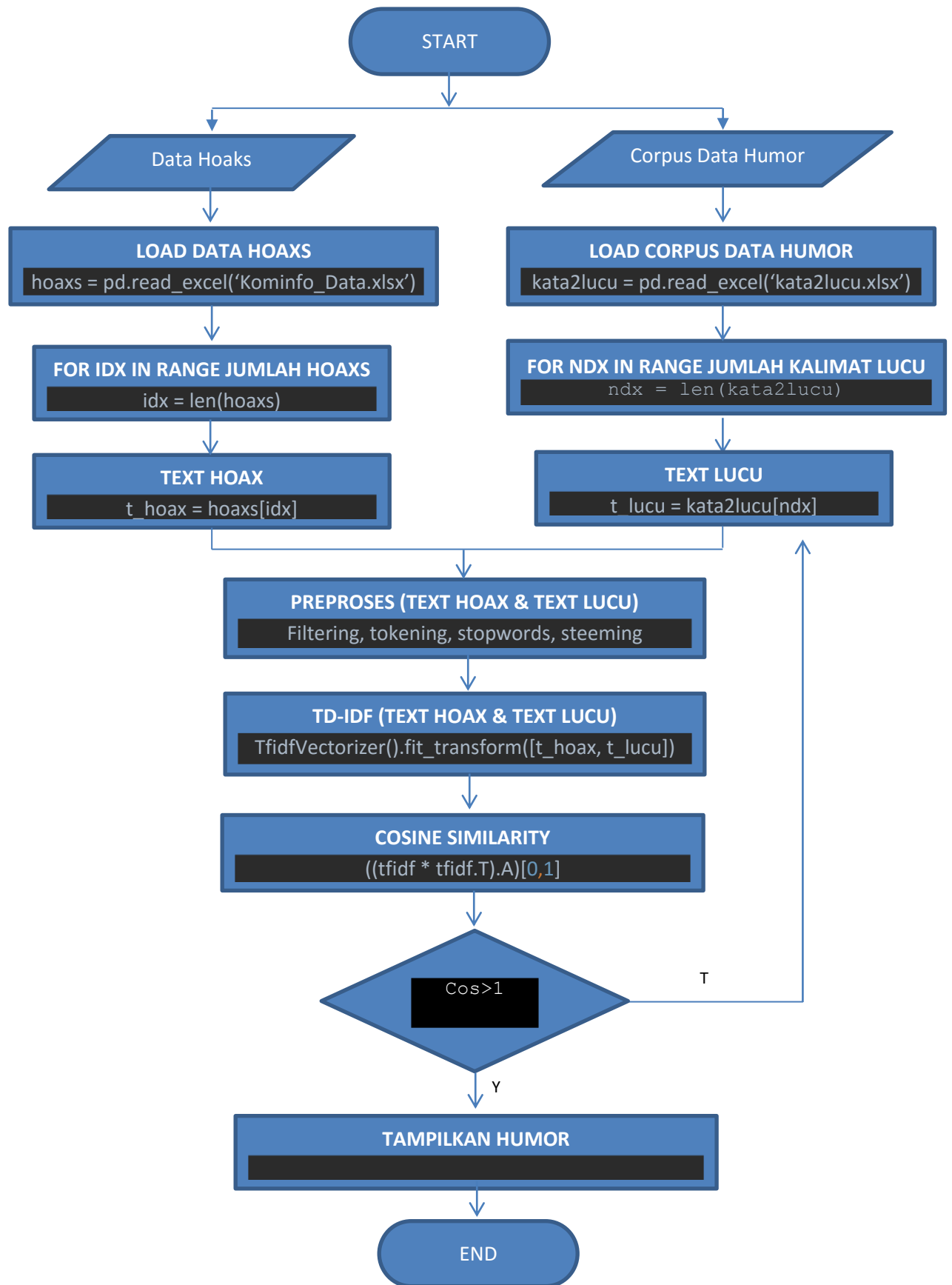
Hasil perhitungan *cosine similarity* untuk kalimat keempat adalah 0.039

Berdasarkan hasil perhitungan *cosine similarity*, dari keempat kalimat lucu di atas yaitu kalimat pertama adalah 0,119, kalimat kedua 0.065, kalimat ketiga adalah 0.059 dan kalimat lucu keempat adalah 0.039,

dari hasil perhitungan di atas dapat ditarik kesimpulan bahwa kalimat lucu pertama yang paling terdekat dengan kalimat hoaks yaitu 'kebakaran terjadi di kamar mayat terbakar melarikan diri dan bingung yang baca berita ini' . Hal tersebut dikarenakan aturan *cosine similarity* adalah semakin besar bobot nilai cosinus (maksimal 1) maka tingkat semakin relevan dokumen dengan yang dibandingkan.

B. Model Berita Hoaks Menjadi Humor

Secara keseluruhan proses computation humor dapat dilihat pada gambar 4.13. *Flowchart hoax to humor*

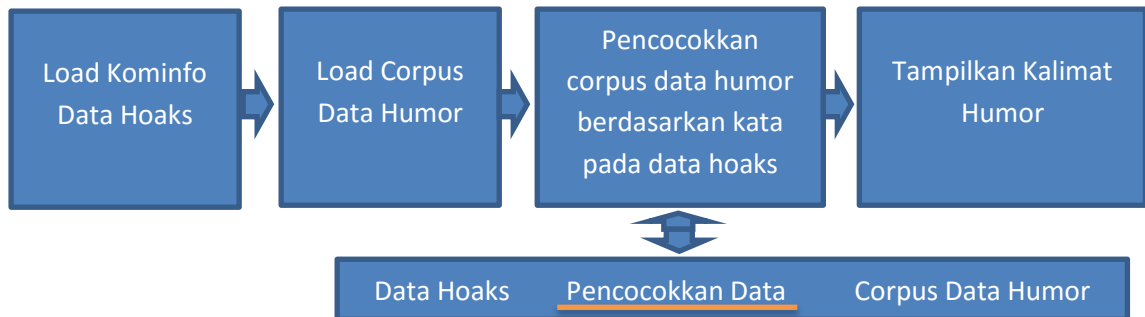


Gambar 4.13. Flowchart Hoax to Humor

Flowchart pada gambar 4.13 menggambarkan alur proses komputasi berita hoaks menjadi humor.

1. Proses memuat data hoaks dan data kalimat lucu yang tersimpan di dalam *microsoft excel* yang telah diperoleh melalui proses *web scrapping*.
2. Setelah kedua data ditemukan, selanjutnya adalah tahapan *preprocessing* yaitu *filtering*, *tokenizing*, *stopwords* dan *stemming*.
3. Proses pembobotan (*Tf-Idf*), pada tahapan ini data hoaks akan dihubungkan dengan setiap data kalimat lucu. Data kalimat lucu yang diperoleh sebanyak 182. Sehingga untuk pemberian bobot, satu data hoaks akan diproses sesuai dengan jumlah kalimat lucu.
4. Proses *cosine similarity*, setelah proses pembobotan (*tf-idf*) selanjutnya mencari kemiripan antar dokumen dengan korpus dengan menggunakan algoritma *cosine similarity*. Semakin tinggi nilai *cosine similarity* korpus kalimat lucu maka semakin mirip dengan data hoaks tersebut.
5. Setelah proses *cosine similarity*, tahap akhir adalah menampilkan kalimat humor.

C. Gambaran umum sistem



Gambar 4.14. Gambaran Umum Sistem

Pada gambar 4.14 menjelaskan tentang gambaran umum proses kerja sistem. Dimana data hoaks dan corpus data lucu di panggil, kemudian dilakukan pencocokan corpus data humor berdasarkan kata pada data hoaks, setelah proses pencocokan kata maka output dari sistem adalah tampilan kalimat humor.

Output dari sistem computation humor dapat dilihat pada tabel 4

Tabel 4.10. *Hoax to Humor*

No	Hoaks	Humors_rank1
1	Makan Telur Rebus Tengah Malam Dapat Mencegah Covid-19	menahan gelak tawa itu tidak sehat gelak tawa akan turun ke bawah dan membasahi celana anda – p valery sayang... tadi pagi aku tidak bisa makan karena merindukan kamu siangny juga tidak makan karena memikirkan kamu dan malamnya aku tidak bisa tidur karena kelaparan
2	Akun WhatsApp Mengatasnamakan Sekda Kabupaten Kepulauan Selayar	
3	Penggunaan Obat Trombolisis di RSPON Tidak Bisa Lebih dari 4 Jam Setelah Kejadian Stroke	bermimpilah tentang apa yang ingin yang kamu impikan jadilah seperti yang kamu inginkan karena bisa jadi itu memang hanya mimpi
4	Gedung Kejaksaan Terbakar, Menko Marves Luhut Minta 'Pemutihan' Sejumlah Koruptor	kebakaran terjadi di kamar mayat terbakar melarikan diri dan bingung yang baca berita ini
5	Demokrat: Habib Rizieq ada di Arab, kenapa tak dirikan khilafah di sana? Beraninya Cuma di Indonesia Saja, Dasar Provokator	sebenarnya yang melahirkan uang adalah mpok indonesia bank indonesia cuma menanam benih

6	Minum Kopi Dengan Jeruk Nipis Dapat Meredakan Migrain	kalo pagi enaknya minum kopi sama susu kopinya diminum susunya di pegang
7	Foto Penampakan Tiga Orang yang Menggantungkan Anggota TNI di Pohon	orang bijak taat pajak orang jahat makan pajak orang gila engga di ajak
8	Jokowi Sudah Gagal Dan Tidak Bisa Jadi Presiden Lagi, Jutaan Rakyat Akan Memaksa Angkat Kaki	diduga karena tidak tahan lagi dengan bau kaki pemiliknya sepasang sepatu di lirikan ke rumah sakit
9	Berkas - Arsip Perkara Hukum Hangus Dalam Kebakaran Gedung Kejaksaan Agung	kebakaran terjadi di kamar mayat terbakar melarikan diri dan bingung yang baca berita ini
10	Dokumen WHO Terkait Anjuran Tidak Menggunakan Masker Selama Pandemi Covid-19	aku punya kemampuan tidur selama setahun tapi cuman bisa sekali yaitu di tanggal desember
11	Angka Perceraian Melonjak Drastis selama Pandemi Covid-19	aku punya kemampuan tidur selama setahun tapi cuman bisa sekali yaitu di tanggal desember
12	Semua Penduduk Indonesia akan Dipajaki	sebenarnya yang melahirkan uang adalah mpok indonesia bank indonesia cuma menanam benih
13	Ustaz Abdul Somad Sebut Facebook Haram	jika aada orang yang menyebutmu jelek maka jangan berputus asa karena belum tentu orang tersebut berkata bohong
14	Chadwick Boseman Tewas Dibunuh dengan Racun	mengetahui ibunya akan memotong ayam seorang anak melaporkan pembunuhan berencana pada kepolisian
15	Akun WhatsApp Mengatasnamakan Sekda Palembang Ratu Dewa	
16	Korban Luka Akibat Bentrokan di Ciledug	seorang pemuda mengalami luka ringan setelah jatuh cinta
17	Artikel Berjudul Gli Italiani Chiamano Anies Baswedan Troppo Stupido Come Il Governatore di Jakarta	
18	Tank Tiongkok Lindas Demonstran Hongkong	
19	Sistem Ganjil Genap Ditiadakan Mulai Senin Selama PSBB	aku punya kemampuan tidur selama setahun tapi cuman bisa sekali yaitu di tanggal desember
20	Tentara Merah RRC Berkamufase Jadi Banser dan Ansor	carilah istri yang saleha suka lupa pake beha agar jadi sakinah sekali kena ingin nambah jadilah suami yang siaga siap antem pakai gaya apa saja wanita cantik bukan menjadi jaminan kehidupan kita menyenangkan apalagi yang jelek
21	Pernyataan Jokowi Pertumbuhan Ekonomi Indonesia Terjun Bebas Tidak Diberi Utang Lagi	sebenarnya yang melahirkan uang adalah mpok indonesia bank indonesia cuma menanam benih
22	Formulir Online Pendaftaran Kartu Indonesia Pintar	sebenarnya yang melahirkan uang adalah mpok indonesia bank indonesia cuma menanam benih
23	Gaji Tambahan dari Jokowi Berupa Mata Uang Yen	sahabat terbaik adalah uang karena uang dapat mendatangkan sahabat
24	35 Guru Meninggal Akibat Covid-19 di Surabaya	gajah mati meninggalkan gading harimau mati meninggalkan belang manusia mati meninggalkan utang

25	Puluhan Orang India Berjoget Merayakan Mendagri India Positif Covid-19	orang bijak taat pajak orang jahat makan pajak orang gila engga di ajak
26	Foto Sri Mulyani Berjilbab Kode pilkada sudah dekat Atau tagihan sudah jatuh tempo	seorang pemuda mengalami luka ringan setelah jatuh cinta
27	Foto Para Pekerja China akan Melakukan Cuti	seberat apapun pekerjaan apabila tidak dikerjakan maka akan terasa ringan
28	Foto Uang 212 Bergambar Perempuan Bercadar	sahabat terbaik adalah uang karena uang dapat mendatangkan sahabat
29	Instruksi KemenBUMN Angkat Advisor Gaji Rp 40 Juta	
30	Akhirnya Jokowi Selesai Buat Tol Laut	belajar buat kita pintar pintar buat kita sukses sukses buat kita kaya kaya buat kita sombong sombong di benci oleh tuhan makanya kita ga usah belajar
31	Melihat Papua dan Aceh Sejahtera Timor Leste Ingin Gabung dengan Indonesia	cobalah melihat pada sisi yg tak terlihat jika tetap tak terlihat cobalah nyalakan lampu
32	Putra Sulung Presiden Gibran akan Membebaskan Palestina dari Israel	
33	Ratusan Warga di Tiap Kecamatan Kota Serang Positif Covid-19	diduga sebagai penyebab kemacetan polisi tidur dibangunkan oleh warga
34	Panggilan Interview Karyawan PT. Garuda Indonesia 08 September 2020	sebenarnya yang melahirkan uang adalah mpok indonesia bank indonesia cuma menanam benih
35	Nama Wakil Bupati Pasuruan Masuk Tim Sukses Pemenangan Gus Ipul	kebahagiaan tersedia bagi mereka yang menang tetapi lebih baik bahagia diatas penderitaan orang lain dari pada menang
36	Kakeknya Arteria Dahlan PKI	
37	Akun Facebook Arthur Waroka Minta Uang Hingga Pulsa ke Warga	sahabat terbaik adalah uang karena uang dapat mendatangkan sahabat
38	Pesan Singkat Hadiah Rp 175 Juta dari Shopee	setelah jabotabek pemerintah akan membuat program kota bersaudara three sister cities yaitu ngawi – cengkareng – ciomas yang apabila disingkat menjadi ngacengmas
39	Beberapa Negara Beli Alat Tes Virus Corona sejak 2017 dan 2018	katanya uang nggak bisa membeli kebahagiaan tapi cewecewe habis belanja kok bahagia bener
40	Anies Terima Piagam DKI Jadi Provinsi Covid-19 Tertinggi	carilah istri yang saleha suka lupa pake beha agar jadi sakinah sekali kena ingin nambah jadilah suami yang siaga siap antem pakai gaya apa saja wanita cantik bukan menjadi jaminan kehidupan kita menyenangkan apalagi yang jelek
41	Donald Trump Sebut Jokowi Presiden Terbaik di Dunia	jika aada orang yang menyebutmu jelek maka jangan berputus asa karena belum tentu orang tersebut berkata bohong
42	Qari Internasional Asal Papua	hatihati dengan sate madura karena dagingnya berasal dari ayam dan kambing mati
43	Tangkapan Layar Obrolan WhatsApp Christiany Eugenia Paruntu (CEP)	cintailah tetangga tetapi jangan sampai tertangkap basah

44	Wisma Atlet Mencekam dan Semua Ruangan Penuh Pasien Covid-19	uang tidak cuma semua nya terus ada mastercard dan visa
45	Imbauan Warga Jakarta Tidak Keluar Rumah karena Akan Dilakukan Rapid Test di Mall dan Pasar	ketika dilanda kesulitan keuangan... jangan pernah menyerah ... benturkan lah kepala anda ditembok atau batu maka akan keluar duit ...
46	Rapid Test Perjalanan Dicabut	malu bertanya sesat di jalan terlalu banyak tanya malumaluin
47	Ledakan Kasus Corona di Kudus Mencapai 2.210 Kasus	saat anda gagal mencapai impian anda jangan pernah berhenti untuk terus mencoba sampai akhirnya tak ada lagi kekuatan untuk mencobanya
48	RI Larang WNA Masuk Saat Corona kecuali TKA China	sebenarnya ane udah bosan hidup atas di dunia ini tapi kalau masuk neraka lagi ntar ketemu lagi sama enteente semua
49	Foto Telah ditemukan situs peninggalan Prasejarah dimasa Raja Firaun	gajah mati meninggalkan gading harimau mati meninggalkan belang manusia mati meninggalkan utang
50	CEO Netflix Ditangkap dengan Tuduhan Pornografi Anak	cintailah tetangga tetapi jangan sampai tertangkap basah

D. Pengujian Sistem *Discounted Comulative Gain (DCG)*

Tahap pengujian sistem menggunakan metode *Discounted Comulative Gain* untuk menilai kualitas hasil ranking dokumen sehingga algoritma yang diterapkan pada sistem dapat diketahui efektivitasnya.

Langkah pertama melakukan perhitungan nilai berdasarkan posisi dokumen di hasil pencarian (DCG_p) (Järvelin & Kekäläinen, 2002).

Menggunakan persamaan :

$$DCG_p = \sum_{i=1}^p \frac{rel\ i}{\log_2(i + 1)}$$

Dengan rel adalah nilai urutan dari posisi. Nilai urutan dinyatakan dalam skala linear dari nol sampai dengan tiga, nol artinya sangat tidak relevan dan tiga artinya sangat relevan. Selanjutnya menghitung nilai dokumen yang mirip dalam database ($IDCG_p$), dengan rumus:

$$IDCGp = \sum_{i=1}^{REL} \frac{rel\ i}{\log_2(i + 1)}$$

Hasil dari DCG akan dinormalisasikan menggunakan perhitungan *Normalized Discounted Cumulative Gain (nDCG)* untuk mendapatkan representase peringkat algoritma yang digunakan pada mesin pencari, menggunakan persamaan :

$$nDCG = \frac{DCGp}{IDCGp}$$

Tabel 4.11 Hasil *Discounted Cumulative Gain (DCG)*

No	hoaxs	dcg_score	idcg_score	ndcg_score
1	Makan Telur Rebus Tengah Malam Dapat Mencegah Covid-19	0,672	2,562	0,262
2	Akun WhatsApp Mengatasnamakan Sekda Kabupaten Kepulauan Selayar	0	0	
3	Penggunaan Obat Trombolisis di RSPON Tidak Bisa Lebih dari 4 Jam Setelah Kejadian Stroke	5,993	10,527	0,569
4	Gedung Kejaksaan Terbakar, Menko Marves Luhut Minta 'Pemutihan' Sejumlah Koruptor	0,672	2,562	0,262
5	Demokrat: Habib Rizieq ada di Arab, kenapa tak dirikan khilafah di sana? Beraninya Cuma di Indonesia Saja, Dasar Provokator	5,871	9,951	0,59
6	Minum Kopi Dengan Jeruk Nipis Dapat Meredakan Migrain	0,672	2,562	0,262
7	Foto Penampakan Tiga Orang yang Menggantung Anggota TNI di Pohon	7,327	10,527	0,696
8	Jokowi Sudah Gagal Dan Tidak Bisa Jadi Presiden Lagi, Jutaan Rakyat Akan Memaksa Angkat Kaki	6,153	9,951	0,618
9	Berkas - Arsip Perkara Hukum Hangus Dalam Kebakaran Gedung Kejaksaan Agung	0,508	2,131	0,238
10	Dokumen WHO Terkait Anjuran Tidak Menggunakan Masker Selama Pandemi Covid-19	0,672	2,562	0,262

11	Angka Perceraian Melonjak Drastis selama Pandemi Covid-19	0,508	2,131	0,238
12	Semua Penduduk Indonesia akan Dipajaki	1,622	4,544	0,357
13	Ustaz Abdul Somad Sebut Facebook Haram	0,342	1,631	0,21
14	Chadwick Boseman Tewas Dibunuh dengan Racun	0,342	1,631	0,21
15	Akun WhatsApp Mengatasnamakan Sekda Palembang Ratu Dewa	0	0	
16	Korban Luka Akibat Bentrokan di Ciledug	0,508	2,131	0,238
17	Artikel Berjudul Gli Italiani Chiamano Anies Baswedan Troppo Stupido Come Il Governatore di Jakarta	0	0	
18	Tank Tiongkok Lindas Demonstran Hongkong	0	0	
19	Sistem Ganjil Genap Ditiadakan Mulai Senin Selama PSBB	0,672	2,562	0,262
20	Tentara Merah RRC Berkamufase Jadi Banser dan Ansor	3,136	7,04	0,445
21	Pernyataan Jokowi Pertumbuhan Ekonomi Indonesia Terjun Bebas Tidak Diberi Utang Lagi	2,054	4,823	0,426
22	Formulir Online Pendaftaran Kartu Indonesia Pintar	0,993	3,305	0,3
23	Gaji Tambahan dari Jokowi Berupa Mata Uang Yen	3,255	6,581	0,495
24	35 Guru Meninggal Akibat Covid-19 di Surabaya	0,672	2,562	0,262
25	Puluhan Orang India Berjoget Merayakan Mendagri India Positif Covid-19	7,08	10,337	0,685
26	Foto Sri Mulyani Berjilbab Kode pilkada sudah dekat Atau tagihan sudah jatuh tempo	1,467	4,254	0,345
27	Foto Para Pekerja China akan Melakukan Cuti	2,458	5,093	0,483
28	Foto Uang 212 Bergambar Perempuan Bercadar	2,336	5,355	0,436
29	Instruksi KemenBUMN Angkat Advisor Gaji Rp 40 Juta	0	0	
30	Akhirnya Jokowi Selesai Buat Tol Laut	0,993	3,305	0,3
31	Melihat Papua dan Aceh Sejahtera Timor Leste Ingin Gabung dengan Indonesia	1,777	4,823	0,368
32	Putra Sulung Presiden Gibran akan Membebaskan Palestina dari Israel	0	0	

33	Ratusan Warga di Tiap Kecamatan Kota Serang Positif Covid-19	0,833	2,948	0,283
34	Panggilan Interview Karyawan PT. Garuda Indonesia 08 September 2020	0,833	2,948	0,283
35	Nama Wakil Bupati Pasuruan Masuk Tim Sukses Pemenangan Gus Ipul	1,777	4,823	0,368
36	Kakeknya Arteria Dahlan PKI	0	0	
37	Akun Facebook Arthur Waroka Minta Uang Hingga Pulsa ke Warga	3,119	6,346	0,491
38	Pesan Singkat Hadiah Rp 175 Juta dari Shopee	0,174	1	0,174
39	Beberapa Negara Beli Alat Tes Virus Corona sejak 2017 dan 2018	0,508	2,131	0,238
40	Anies Terima Piagam DKI Jadi Provinsi Covid-19 Tertinggi	3,285	7,265	0,452
41	Donald Trump Sebut Jokowi Presiden Terbaik di Dunia	3,201	5,611	0,57
42	Qari Internasional Asal Papua	0,174	1	0,174
43	Tangkapan Layar Obrolan WhatsApp Christiany Eugenia Paruntu (CEP)	0,174	1	0,174
44	Wisma Atlet Mencekam dan Semua Ruang Penuh Pasien Covid-19	0,833	2,948	0,283
45	Imbauan Warga Jakarta Tidak Keluar Rumah karena Akan Dilakukan Rapid Test di Mall dan Pasar	4,397	6,813	0,645
46	Rapid Test Perjalanan Dicabut	0,342	1,631	0,21
47	Ledakan Kasus Corona di Kudus Mencapai 2.210 Kasus	0,342	1,631	0,21
48	RI Larang WNA Masuk Saat Corona kecuali TKA China	0,508	2,131	0,238
49	Foto Telah ditemukan situs peninggalan Prasejarah dimasa Raja Firaun	2,619	5,861	0,447
50	CEO Netflix Ditangkap dengan Tuduhan Pornografi Anak	1,379	3,305	0,417

Pada tabel 4.11, merupakan hasil perhitungan pengujian *Discounted Cumulative Gain*. Dari 50 data hoaks dan 182 corpus kalimat lucu, nilai tertinggi ditemukan pada data hoax nomor 7, dengan nilai *DCG* 7,327, *IDCG* 10,527, *nDCG* 0,696, dan ditemukan ada 7 (tujuh) data hoax

yang tidak memiliki relevansi dengan corpus data humor yaitu pada data hoaks nomor 2, 15, 17, 18, 29, 32 dan 36.

E. Pengujian Tingkat Kelucuan

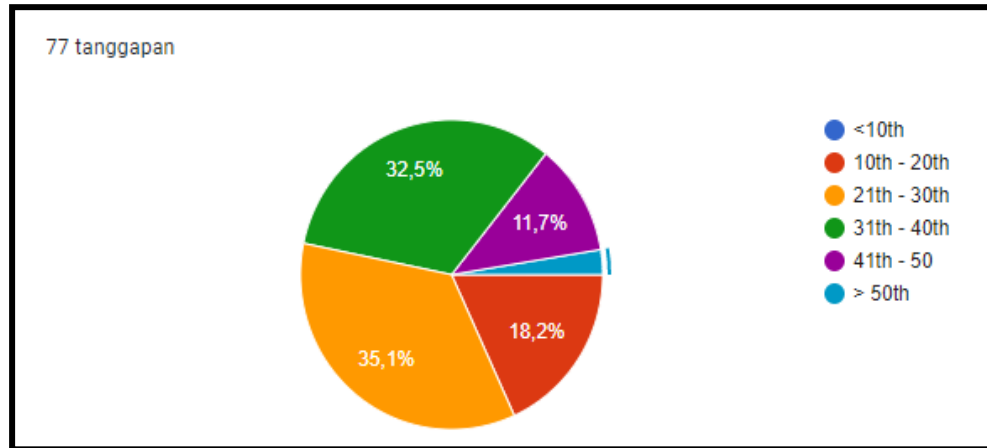
Pengujian tingkat kelucuan dilakukan dengan survey secara online kepada para pengguna media sosial menggunakan *google form*. Data responden pada penelitian ini, diperoleh dengan membagikan alamat *link form* (<https://forms.gle/Byw817av7TMchES67>) di *facebook* dan grup-grup *whatsApp* peneliti untuk diketahui dan dikunjungi oleh responden. Penyebaran kuisisioner di mulai pada tanggal 17 September sampai dengan 22 September 2020.

Aspek-aspek yang ditanyakan dalam kuisisioner adalah: 1). Profil responden yang terdiri dari nama, jenis kelamin, umur, pekerjaan, dan pendidikan terakhir. 2). Tanggapan responden tentang tingkat kelucuan dari perubahan berita hoaks ke humor. Untuk pilihan jawabannya terdiri dari 4 pilihan, yaitu : 1 = Tidak lucu, 2 = Biasa, 3 = Lucu dan 4 = Sangat lucu.

Seluruh anggota populasi akan diambil sebagai responden. Karena pada penelitian *eksplanatoris* dilakukan pendekatan kasus maka besarnya sampel penelitian adalah sebesar jumlah kuisisioner yang di isi, kembali, dan layak untuk diolah (Batubara, 2016). Pengguna media sosial yang mengisi kuisisioner ini sebanyak 77 responden.

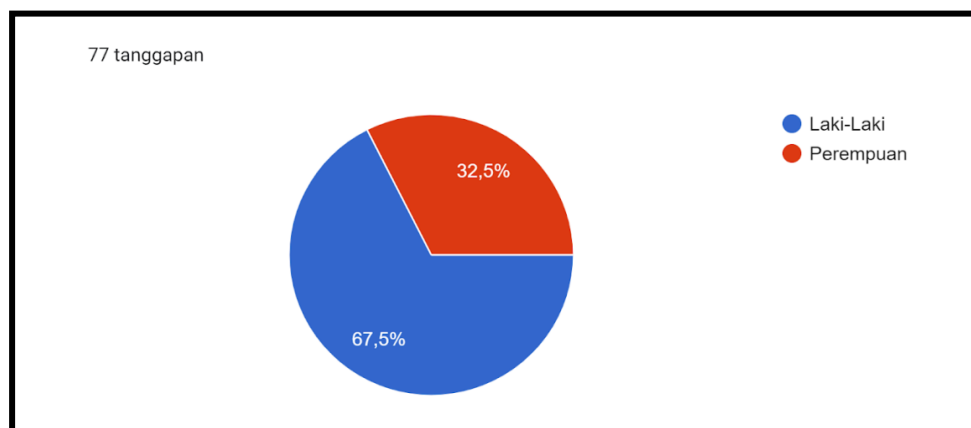
1. Profil Responden

a. Umur dan Jenis Kelamin



Gambar 4.15 Umur Responden

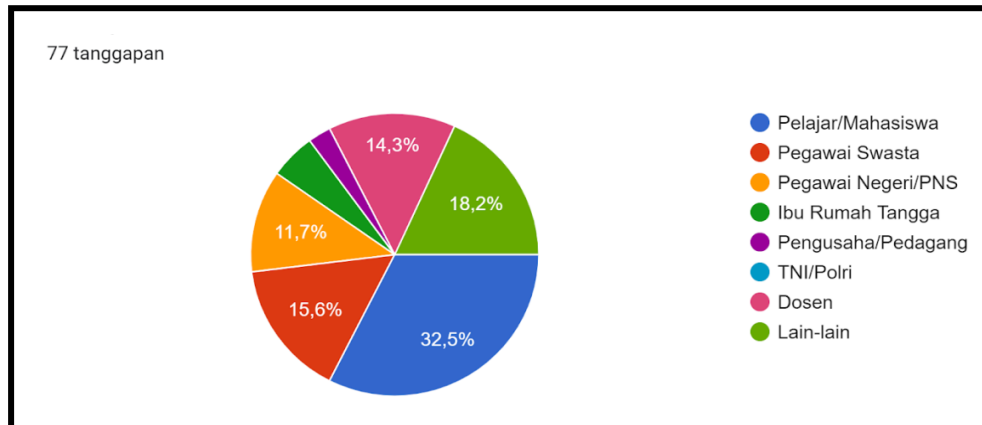
Berdasarkan grafik pada gambar 4.15, terlihat bahwa responden dalam penelitian ini didominasi oleh responden dengan segmen usia 21-30 tahun (35,1%) dan usia 31-40 tahun (32,5%), diikuti dengan segmen usia 10-20 tahun (18,2%), 41-50 tahun (11,7%) dan urutan terakhir usia di atas 50 tahun (2,6%).



Gambar 4.16 Jenis Kelamin Responden

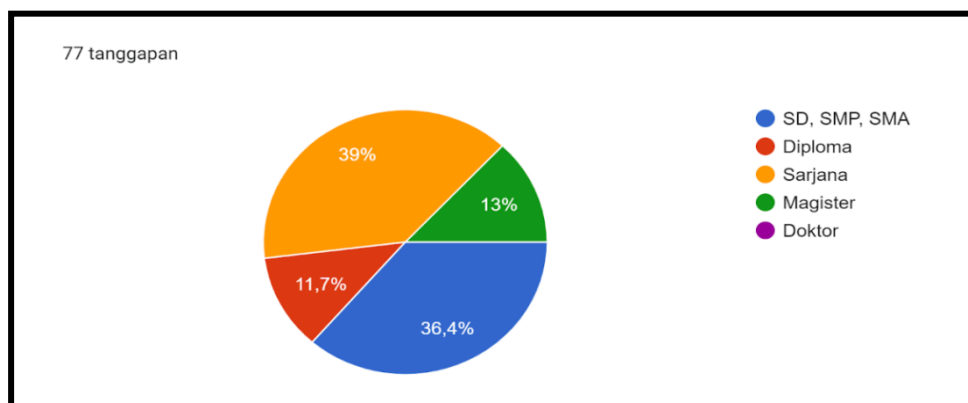
Grafik pada gambar 4.16, dapat dilihat bahwa jenis kelamin didominasi oleh responden laki-laki sebanyak 52 orang (67%) dan responden perempuan sebanyak 25 orang (32,5%).

b. Pekerjaan dan Pendidikan Terakhir



Gambar 4.17 Pekerjaan Responden

Berdasarkan grafik pada gambar 4.17, terlihat bahwa pekerjaan responden didominasi oleh pelajar/mahasiswa sebesar 32,5%, kemudian pekerjaan lain-lain sebesar 18,2%, diikuti dengan pegawai swasta sebesar 15,6%, dosen sebesar 14,3%, pegawai negeri sipil sebesar 11,7%, ibu rumah tangga sebesar 5,2% dan urutan terakhir pengusaha/pedagang sebesar 2,6%.



Gambar 4.18 Pendidikan Terakhir Responden

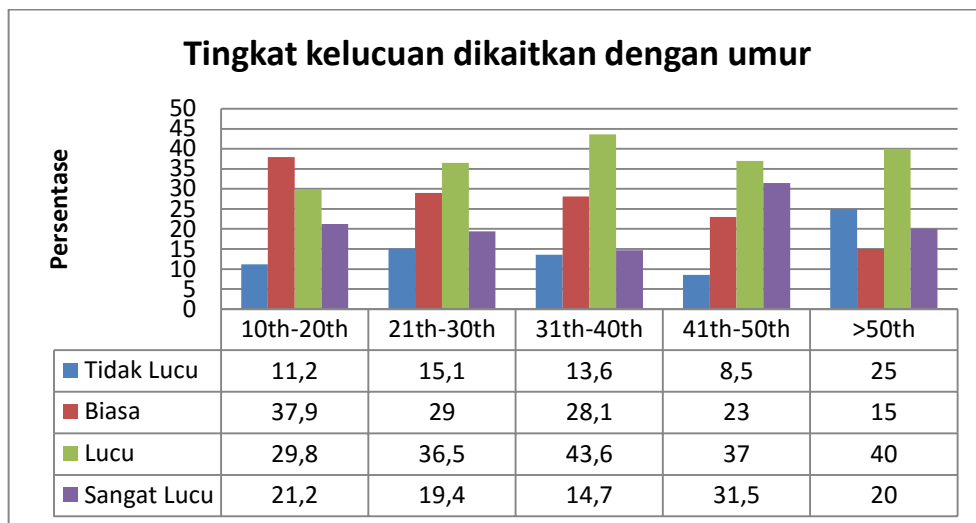
Jika di lihat dari pendidikan terakhir responden pada penelitian ini, sarjana sebanyak 30 orang (39%), di ikuti SD, SMP, SMA sebanyak 28 orang (36,4%), magister 10 orang (13%), urutan terakhir pendidikan responden adalah diploma sebesar 9 orang (11,7%).

2. Tabulasi

Tabel 4.12 Tabulasi tingkat kelucuan dikaitkan dengan umur

UMUR	Respon nden	Tingkat Kelucuan				Dalam Persentase			
		<i>Tidak Lucu</i>	<i>Biasa</i>	<i>Lucu</i>	<i>Sangat Lucu</i>	<i>Tidak Lucu</i>	<i>Biasa</i>	<i>Lucu</i>	<i>Sangat Lucu</i>
10th-20th	14	1,6	5,3	4,2	3,0	11,2	37,9	29,8	21,2
21th-30th	27	4,1	7,8	9,9	5,2	15,1	29,0	36,5	19,4
31th-40th	25	3,4	7,0	10,9	3,7	13,6	28,1	43,6	14,7
41th-50th	9	0,8	2,1	3,3	2,8	8,5	23,0	37,0	31,5
>50th	2	0,5	0,3	0,8	0,4	25,0	15,0	40,0	20,0
Total	77	10,3	22,5	29,1	15,1				

Berdasarkan tabel di atas, tingkat kelucuan pada hasil penelitian ini responden untuk umur 10 – 20 tahun lebih banyak memilih “Biasa” (37,9%) dan responden umur 21 tahun sampai 50 tahun ke atas lebih banyak memilih “Lucu”. Grafik tingkat kelucuan dikaitkan dengan umur dapat di lihat pada gambar di bawah ini

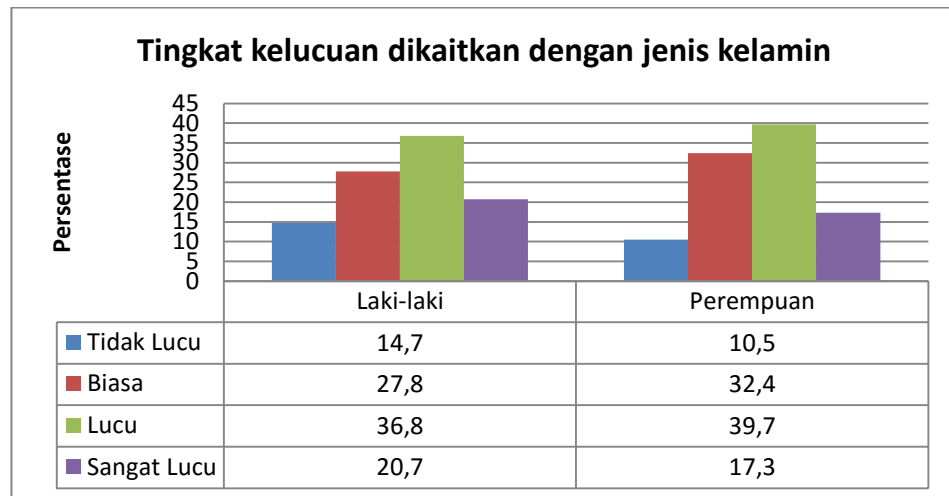


Gambar 4.19 Grafik Tingkat Kelucuan Dikaitkan dengan Umur

Tabel 4.13 Tabulasi tingkat kelucuan dikaitkan dengan jenis kelamin

Jenis Kelamin	Respon	Tingkat Kelucuan				Dalam Persentase			
		<i>Tidak Lucu</i>	<i>Biasa</i>	<i>Lucu</i>	<i>Sangat Lucu</i>	<i>Tidak Lucu</i>	<i>Biasa</i>	<i>Lucu</i>	<i>Sangat Lucu</i>
Laki-Laki	52	7,7	14,4	19,1	10,8	14,7	27,8	36,8	20,7
Perempuan	25	2,6	8,1	9,9	4,3	10,5	32,4	39,7	17,3
Rata-rata	77	10,3	22,5	29,1	15,1				

Jika dilihat dari jenis kelaminnya, responden laki-laki dan perempuan lebih banyak memilih “Lucu” yaitu sebesar 36,8% dan 39,7%. Jenis kelamin mempengaruhi Tingkat kepekaan humor (*sense of humor*). Dalam mencerna atau menangkap suatu humor perempuan membutuhkan waktu sedikit lebih lama dibanding pria tetapi tidak mempengaruhi dalam menikmati humor.. Grafik tingkat kelucuan dikaitkan dengan jenis kelamin dapat di lihat pada gambar 4.20.

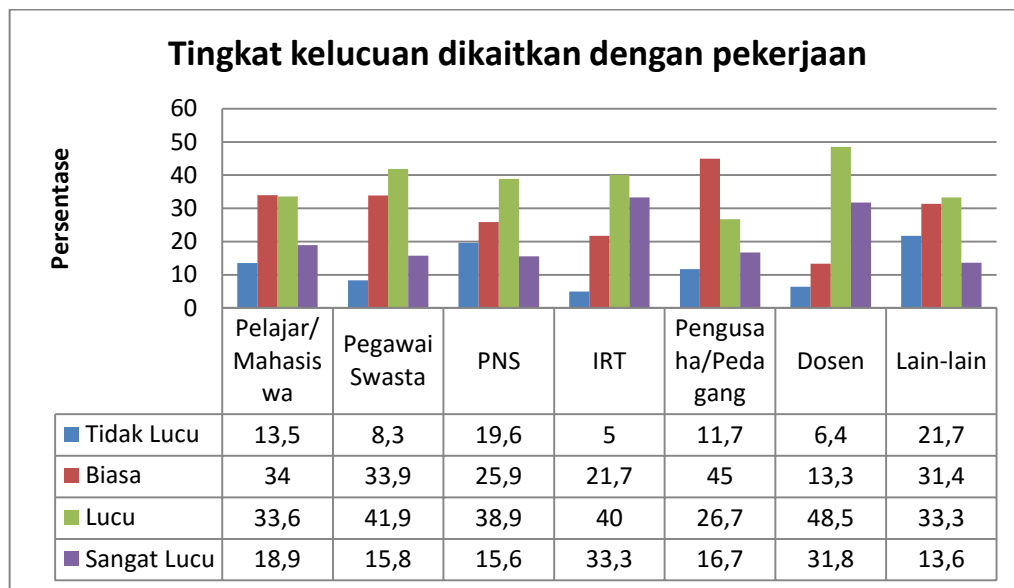


Gambar 4.20 Grafik Tingkat Kelucuan Dikaitkan dengan Jenis Kelamin

Tabel 4.14 Tabulasi tingkat kelucuan dikaitkan dengan pekerjaan

Pekerjaan	Respon nden	Tingkat Kelucuan				Dalam Persentase			
		<i>Tidak Lucu</i>	<i>Biasa</i>	<i>Lucu</i>	<i>Sangat Lucu</i>	<i>Tidak Lucu</i>	<i>Biasa</i>	<i>Lucu</i>	<i>Sangat Lucu</i>
Pelajar/Mahasiswa	25	3,4	8,5	8,4	4,7	13,5	34,0	33,6	18,9
Pegawai Swasta	12	1,0	4,1	5,0	1,9	8,3	33,9	41,9	15,8
PNS	9	1,8	2,3	3,5	1,4	19,6	25,9	38,9	15,6
IRT	4	0,2	0,9	1,6	1,3	5,0	21,7	40,0	33,3
Pengusaha/Pedagang	2	0,2	0,9	0,5	0,3	11,7	45,0	26,7	16,7
Dosen	11	0,7	1,5	5,3	3,5	6,4	13,3	48,5	31,8
Lain-lain	14	3,0	4,4	4,7	1,9	21,7	31,4	33,3	13,6
<i>Rata-rata</i>	77	10,3	22,5	29,1	15,1				

Bila di lihat dari tabel 4.14, responden dengan pekerjaan pelajar/mahasiswa dan pengusaha/pedagang lebih banyak memilih "Biasa". Untuk pegawai swasta, PNS, ibu rumah tangga, dosen dan lain-lain lebih banyak memilih "Lucu".

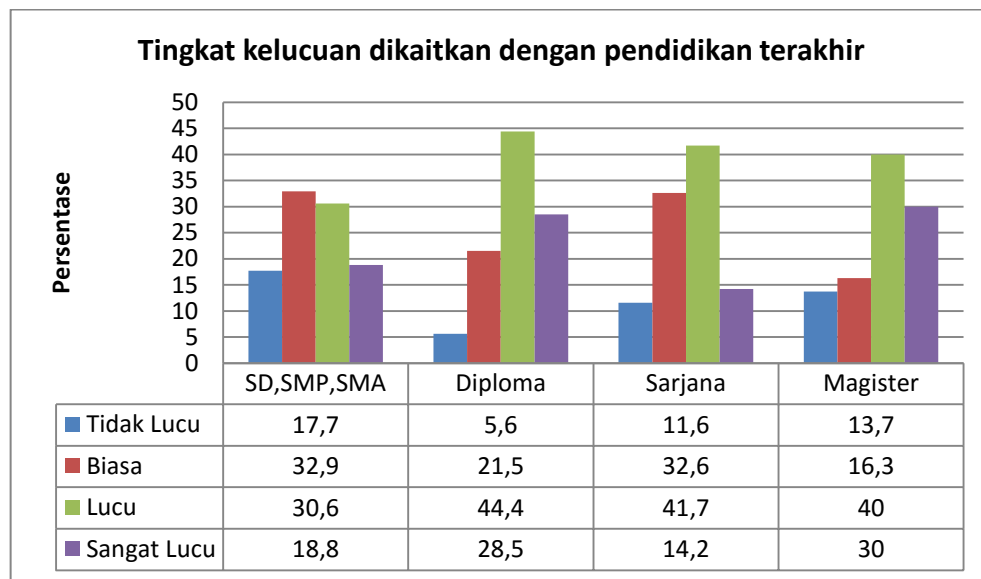


Gambar 4.21 Grafik Tingkat Kelucuan Dikaitkan dengan Pekerjaan

Tabel 4.15 Tabulasi tingkat kelucuan dikaitkan dengan pendidikan terakhir

Pendidikan Terakhir	Respon	Tingkat Kelucuan				Dalam Persentase			
		<i>Tidak Lucu</i>	<i>Biasa</i>	<i>Lucu</i>	<i>Sangat Lucu</i>	<i>Tidak Lucu</i>	<i>Biasa</i>	<i>Lucu</i>	<i>Sangat Lucu</i>
SD,SMP,SMA	28	5,0	9,2	8,6	5,3	17,7	32,9	30,6	18,8
Diploma	9	0,5	1,9	4,0	2,6	5,6	21,5	44,4	28,5
Sarjana	30	3,5	9,8	12,5	4,3	11,6	32,6	41,7	14,2
Magister	10	1,4	1,6	4,0	3,0	13,7	16,3	40,0	30,0
Total	77	10,3	22,5	29,1	15,1				

Berdasarkan tabel 4.15, dapat di lihat pendidikan terakhir SD, SMP, SMA lebih banyak memilih “Biasa”, sedangkan untuk diploma, sarjana dan magister lebih banyak memilih “Lucu”.



Gambar 4.22 Grafik Tingkat Kelucuan Dikaitkan dengan Pendidikan Terakhir

3. Analisa Hasil Pengujian Tingkat Kelucuan

Salah satu faktor yang mempengaruhi *sense of humor* adalah pengetahuan seseorang pada subjek humor yang dilontarkan. Ketika seseorang memahami subjek humor yang di bahas, maka orang tersebut akan tertawa, tetapi materi humor tidak akan membuat individu tertawa apabila individu tersebut tidak memahami materi humor tersebut. (Utomo, 2009).

Menurut (Pratama, 2019) umur di bawah 20 tahun kurang tertarik dengan berita hoaks. Berdasarkan tabulasi hasil pengujian tingkat kelucuan dan dikaitkan dengan referensi dari penelitian yang telah dikemukakan di atas, dapat disimpulkan bahwa kurangnya pengetahuan tentang berita hoaks terhadap responden umur di bawah 20 tahun mempengaruhi pilihan responden terhadap materi yang disajikan pada

penelitian ini. Selain itu responden umur di bawah 20 tahun lebih mudah memahami humor dalam bentuk gambar atau video dibandingkan dengan teks disebabkan karena kemampuan memahami dan menganalisa teks. Hal ini dapat dikaitkan dengan pekerjaan dan pendidikan terakhir responden, pelajar/mahasiswa lebih banyak yang memilih biasa. Seseorang yang masih berstatus pelajar/mahasiswa biasanya masih berada di umur 20 tahun ke bawah.

4. Keterbatasan Pengujian Tingkat Kelucuan

Selama proses pengujian tingkat kelucuan yang dilakukan dengan survei online terdapat keterbatasan dalam mengamati kondisi lapangan secara nyata. Pada saat responden mengisi kuisioner, peneliti tidak dapat mengawasi dan mendampingi. Oleh sebab itu, bisa saja terjadi bias karena dalam mengisi pernyataan-pernyataan yang ada responden kurang memahami maksud dari survei tersebut.

BAB V

PENUTUP

A. Kesimpulan

Berdasarkan hasil penelitian dapat disimpulkan bahwa :

1. *Computation humor, case study news hoax* dapat dilakukan dengan menggunakan algoritma *cosine similarity* dengan terlebih dahulu melalui tahap *preprocessing*.

Hasil Perhitungan *cosine similarity* pada data hoaks “Gedung Kejaksaan Terbakar, Menko Marves Luhut Minta 'Pemutihan' Sejumlah Koruptor” adalah 0,119, dengan corpus humor, “kebakaran terjadi di kamar mayat terbakar melarikan diri dan bingung yang baca berita ini”.

2. Hasil pengujian sistem menggunakan metode *Discounted Cumulative Gain* ditemukan nilai tertinggi pada data hoax nomor 7, dengan nilai kualitas hasil ranking dokumen (*DCG*) sebesar 7,327, nilai dokumen yang relevan dalam database (*IDCG*) sebesar 10,527, dan *nDCG* sebesar 0,696 serta ditemukan ada 6 (enam) data hoaks yaitu nomor 2, 17, 18, 29, 32 dan 36, tidak memiliki relevansi dengan corpus data humor.
3. Hasil analisis data pengujian tingkat kelucuan menggunakan survei online, berdasarkan umur diperoleh umur 10 – 20 tahun lebih banyak

memilih Biasa, dan responden umur 21 tahun sampai 50 tahun ke atas lebih banyak memilih Lucu. Untuk analisis data berdasarkan jenis kelamin, responden laki-laki dan perempuan lebih banyak yang memilih Lucu. Jika dikaitkan dengan pekerjaan responden, pelajar/mahasiswa dan pengusaha/pedagang lebih banyak memilih Biasa sedangkan pegawai swasta, PNS, ibu rumah tangga, dosen dan lain-lain lebih banyak memilih Lucu. Untuk hasil analisis pendidikan terakhir responden, SD, SMP, SMA lebih banyak memilih Biasa, sedangkan untuk diploma, sarjana dan magister lebih banyak memilih lucu.

B. Saran

Saran untuk pengembangan penelitian *computation humor, case study news hoax* yaitu :

1. Pada penelitian ini berita hoaks yang diperoleh dengan proses *web scrapping*, hanya mengambil judul (*title*) berita hoaks sebagai sumber data hoaks. Maka saran untuk pengembangan sistem, sumber data hoaks dapat menggunakan isi berita hoaks
2. *Computation humor* dikembangkan dengan penambahan corpus data humor, sebagaimana pada penelitian ini hanya sebanyak 182 corpus data humor.
3. Berita Hoaks disandingkan dengan informasi yg baik sehingga menjadi sesuatu yang bermanfaat.

DAFTAR PUSTAKA

- Agus Tri Haryanto. (2020). *Kominfo Sergap 242 Konten Hoax Corona*.
<https://inet.detik.com/law-and-policy/d-4942767/kominfo-sergap-242-konten-hoax-corona>
- Batubara, H. H. (2016). Di Prodi Pgmi Uniska Muhammad Arsyad Al Banjari. *Universitas Islam Kalimantan MAB*, 8(1), 40–50.
- Deadydiedy. (2015). *KOMPUTASI & TEORI KOMPUTASI | deadydiedy*.
<https://deadydiedy.wordpress.com/2015/04/05/komputasi-teori-komputasi/>
- Didiek Rahmanadji. (2009). Sejarah, teori, jenis, dan fungsi humor. In *Jurnal bahasa dan seni* (Vol. 35, Issue 2).
- Franzini, L. R. (2001). Humor in therapy: The case for training therapists in its uses and risks. In *Journal of General Psychology* (Vol. 128, Issue 2). Taylor & Francis Group.
<https://doi.org/10.1080/00221300109598906>
- Istiningtyas, L. (2014). Humor Dalam Kajian Psikologi Islam. In *Jurnal Ilmu Agama* (Vol. 15, Issue 1).
- Järvelin, K., & Kekäläinen, J. (2002). Cumulated gain-based evaluation of IR techniques. In *ACM Transactions on Information Systems* (Vol. 20, Issue 4). <https://doi.org/10.1145/582415.582418>
- Juliswara, V. (2017). Mengembangkan Model Literasi Media yang Berkebhinnekaan dalam Menganalisis Informasi Berita Palsu (Hoax) di Media Sosial. In *Jurnal Pemikiran Sosiologi* (Vol. 4, Issue 2). <https://doi.org/10.22146/jps.v4i2.28586>
- Kementerian Komunikasi dan Informatika. (2018). https://www.kominfo.go.id/content/detail/12292/kenalan-dengan-cyber-drone-9-polisi-internet-indonesia/0/sorotan_media
- Melita, R., Amrizal, V., Suseno, H. B., Dirjam, T., Studi, P., Informatika, T., & Sains, F. (2018). (*Tf-Idf*) Dan *Cosine Similarity* Pada Sistem Temu Kembali Informasi Untuk Mengetahui Syarah Hadits Berbasis Web (*Studi Kasus : Syarah Umdatil Ahkam*) (Vol. 11, Issue 2).
- Pratama, M. I. (2019). Efek Menonton Tayangan Video Ciri – Ciri Berita Hoaks Terhadap Tingkat Pengetahuan Remaja Tentang Berita Hoaks. In *Jurnal Ilmu Sosial dan Ilmu Politik* (Vol. 8, Issue 5).
- Rahadi, D. R. (2017). Perilaku Pengguna Dan Informasi Hoax Di Media

- Sosial. In *Jurnal Manajemen Dan Kewirausahaan* (Vol. 5, Issue 1).
<https://doi.org/10.26905/jmdk.v5i1.1342>
- Rasywir, E., & Purwarianti, A. (2015). Import citahttp://Eksperimen pada Sistem Klasifikasi Berita Hoax Berbahasa Indonesia Berbasis Pembelajaran Mesin. In *Jurnal Cybermatika* (Vol. 3, Issue 2).
<https://www.mendeley.com/import/>
- Robert. (2017). *Konsep dan Implementasi Machine Learning*.
<http://www.catatanrobert.com/konsep-dan-implementasi-machine-learning/>
- Utami, P. (2019). Hoax in Modern Politics. In *Jurnal Ilmu Sosial dan Ilmu Politik* (Vol. 22, Issue 2). <https://doi.org/10.22146/jsp.34614>
- Vukovi, M., Pripuzi, K., & Belani, H. (2009). *An Intelligent Automatic Hoax Detection System **.
- Wisnubrata. (2019). *Dampak Buruk Berita Hoax pada Kesehatan Mental, Ini Penjelasannya Halaman all - Kompas.com*.
<https://lifestyle.kompas.com/read/2019/10/08/120209420/dampak-buruk-berita-hoax-pada-kesehatan-mental-ini-penjelasannya?page=all>
- www.inixindojogja.co.id. (2017). *Mengenal Machine Learning*.
<https://inixindojogja.co.id/mengenal-machine-learning/>