

**SISTEM PENENTUAN KONTRIBUSI TOKOH POLITIK
MENGUNAKAN *K-NEAREST NEIGHBOR***

SKRIPSI

**Oleh:
MUHAMAD NIZAR ZULMI
NIM. 12650037**



**JURUSAN TEKNIK INFORMATIKA
FAKULTAS SAINS DAN TEKNOLOGI
UNIVERSITAS ISLAM NEGERI MAULANA MALIK IBRAHIM
MALANG
2019**

**SISTEM PENENTUAN KONTRIBUSI TOKOH POLITIK
MENGUNAKAN *K-NEAREST NEIGHBOR***

SKRIPSI

**Diajukan Kepada:
Universitas Islam Negeri Maulana Malik Ibrahim Malang
Untuk Memenuhi Salah Satu Persyaratan Dalam
Memperoleh Gelar Sarjana Komputer (S.Kom)**

**Oleh:
MUHAMAD NIZAR ZULMI
NIM. 12650037**

**JURUSAN TEKNIK INFORMATIKA
FAKULTAS SAINS DAN TEKNOLOGI
UNIVERSITAS ISLAM NEGERI MAULANA MALIK IBRAHIM
MALANG
2019**

HALAMAN PERSETUJUAN
SISTEM PENENTUAN KONTRIBUSI TOKOH POLITIK
MENGUNAKAN *K-NEAREST NEIGHBOR*


SKRIPSI

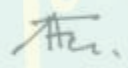
Oleh:
MUHAMAD NIZAR ZULMI
NIM. 12650037

Telah Diperiksa dan Disetujui untuk Diuji
Tanggal: 10 Juni 2018

Dosen Pembimbing I

Dosen Pembimbing II


Dr. Cahyo Crisdian
NIP. 19740424200901 1 008


Fatchurrochman, M. Kom
NIP. 19700731200501 1 002

Mengetahui,
Ketua Jurusan Teknik Informatika
Fakultas Sains dan Teknologi
Universitas Islam Negeri Maulana Malik Ibrahim Malang




Dr. Cahyo Crisdian
NIP. 19740424 200901 1 008

LEMBAR PENGESAHAN

SISTEM PENENTUAN KONTRIBUSI TOKOH POLITIK
MENGUNAKAN *K-NEAREST NEIGHBOR*

SKRIPSI

Oleh :
MUHAMAD NIZAR ZULMI
NIM. 12650037

Telah Dipertahankan di Depan Dewan Penguji Skripsi
dan Dinyatakan Diterima Sebagai Salah Satu Persyaratan
Untuk Memperoleh Gelar Sarjana Komputer (S.Kom)
Tanggal Juni 2019

Susunan Dewan Penguji	Tanda Tangan
Penguji Utama : <u>Roro Inda Melani, S.Kom., M.Sc</u> NIP. 19780925 200501 2 008	()
Ketua Penguji : <u>Dr. Ir. M. Amin Hariyadi, M.T</u> NIP. 19670118 200501 1 001	()
Sekretaris Penguji : <u>Dr. Cahyo Crys dian</u> NIP. 19740424 200901 1 008	()
Anggota Penguji : <u>Fatchurrochman, M. Kom</u> NIP. 19700731200501 1 002	()

Mengesahkan,
Ketua Jurusan Teknik Informatika
Fakultas Sains dan Teknologi
Universitas Islam Negeri Maulana Malik Ibrahim Malang



Dr. Cahyo Crys dian
NIP. 19740424 200901 1 008

**PERNYATAAN
ORISINALITAS PENELITIAN**

Saya yang bertanda tangan di bawah ini:

Nama : Muhamad Nizar Zulmi
NIM : 12650037
Fakultas / Jurusan : Sains dan Teknologi / Teknik Informatika
Judul Penelitian : Sistem Penentuan Kontribusi Tokoh Politik
Menggunakan *K-Nearest Neighbor*

Menyatakan dengan sebenar-benarnya bahwa hasil penelitian saya ini tidak terdapat unsur-unsur penjiplakan karya penelitian atau karya ilmiah yang pernah dilakukan atau dibuat oleh orang lain, kecuali yang secara tertulis dikutip dalam naskah ini dan disebutkan dalam sumber kutipan dan daftar pustaka.

Apabila ternyata hasil penelitian ini terbukti terdapat unsur-unsur jiplakan, maka saya bersedia untuk mempertanggung jawabkan, serta diproses sesuai peraturan yang berlaku.

Malang, 10 Juni 2018

Penulis



Muhamad Nizar Zulmi
NIM. 12650037

MOTTO

إِنْ أَعْصَمْتُمْ لِنَفْسِكُمْ

“Jika kalian berbuat baik, sesungguhnya kalian berbuat baik bagi diri kalian sendiri” (QS. Al-Isra:7)



KATA PENGANTAR

Assalamu 'alaikum Warahmatullahi Wabarakatuh

Alhamdulillah segala puji dan syukur kehadirat Allah SWT atas berkah, rahmat serta hidayah-Nya, sholawat serta salam penulis tujukan kepada baginda Rasulullah Shalallahu 'alaihi wasalam sebagai pembawa rahmat bagi seluruh alam ini yang senantiasa dilimpahkan kepada penulis, sehingga dapat menyelesaikan skripsi ini dengan judul “Sistem Penentuan Kontribusi Tokoh Politik Menggunakan *K-Nearest Neighbor*”

Dalam menyelesaikan skripsi ini banyak hambatan dan kekurangan yang penulis hadapi namun pada akhirnya dapat terlampaui dengan adanya bimbingan, dukungan dan bantuan dari berbagai pihak. Untuk itu pada kesempatan ini penulis menyampaikan terimakasih kepada:

1. Bapak Kuwat dan Ibu Subekti selaku orang tua yang memberikan dukungan terbaik kepada penulis.
2. Prof. Dr. Abdul Haris, M.Ag selaku Rektor Universitas Islam Negeri Maulana Malik Ibrahim Malang
3. Bapak Dr. Cahyo Crysdiyan selaku Ketua Jurusan Teknik Informatika dan juga pembimbing I, yang senantiasa memberikan waktu, memberikan semangat, dan motivasi.
4. Bapak Fatchurrochman M.Kom selaku pembimbing II, yang telah memberikan arahan kepada penulis dalam penyusunan skripsi ini.

5. Segenap Dosen Teknik Informatika yang telah memberikan bimbingan keilmuan dengan begitu sabar selama masa studi.
6. Teman – teman seperjuangan Teknik Informatika dari berbagai angkatan yang tidak dapat saya sebutkan satu persatu.
7. Teman teman dan saudara UKM Seni Religius juga Orda Triscom yang telah membentuk pribadi ku menjadi seperti sekarang.

Dalam skripsi ini, penulis menyadari sepenuhnya bahwa dalam penulisan skripsi ini masih jauh dari kesempurnaan karena keterbatasan kemampuan yang penulis miliki. Dengan segenap kerendahan hati, penulis mengharapkan kritik dan saran yang membangun penulisan skripsi ini agar dapat memberikan manfaat khususnya bagi penulis sendiri dan umumnya kepada seluruh pembaca skripsi ini. Semoga skripsi ini berguna bagi perkembangan ilmu pengetahuan.

Wassalamu 'alaikum Warahmatullahi Wabarakatuh.

Malang, 10 Juni 2018

Muhamad Nizar Zulmi

DAFTAR ISI

HALAMAN JUDUL	i
HALAMAN PENGAJUAN	ii
HALAMAN PERSETUJUAN	iii
HALAMAN PENGESAHAN	iv
HALAMAN PERNYATAAN	v
MOTTO	vi
KATA PENGANTAR	viii
DAFTAR ISI	ix
DAFTAR GAMBAR	xi
DAFTAR TABEL	xii
ABSTRAK	xiii
BAB I PENDAHULUAN	1
1.1 Latar Belakang Masalah	1
1.2 Pertanyaan Masalah	4
1.3 Tujuan Penelitian	4
1.4 Batasan Masalah	4
1.5 Manfaat Penelitian	5
1.6 Sistematika Penulisan	5
BAB II STUDI PUSTAKA	7
2.1 Kontribusi	7
2.2 Text Mining	8
2.2.1 Text Preprocessing	9
2.2.2 Feature Selection	9
2.3 K-Nearest Neighbor	11
2.4 Penelitian Terkait	12
BAB III DESAIN DAN IMPLEMENTASI	17
3.1 Sistem Desain	17
3.1.1 Preprocessing	17
3.1.2 Feature Selection	18
3.1.3 Pembobotan	23

3.1.4 Similarity.....	24
3.1.5 Evaluasi dan Validasi.....	25
3.2 Data.....	26
3.2.1 Data Dokumen.....	27
3.2.2 Data Stopword.....	28
3.2.3 Data Kata Dasar.....	28
3.2.4 Data Kata Kunci.....	29
3.2.5 Data Narasumber.....	36
3.3 Analisis Sistem.....	37
3.3.1 Stemming.....	37
3.3.2 TF-IDF.....	38
3.3.3 K-Nearest Neighbor.....	39
3.3.3.1 Menentukan Nilai K.....	40
3.3.3.2 Menghitung Euclidean Distance.....	41
3.3.4 Akurasi.....	44
BAB IV UJI COBA DAN PEMBAHASAN.....	46
4.1 Langkah Uji Coba.....	46
4.2 Hasil Uji Coba.....	47
4.3 Pembahasan.....	50
4.4 Implementasi.....	51
4.4.1 Implementasi Sistem.....	55
4.5 Integrasi Islam.....	59
BAB V KESIMPULAN DAN SARAN.....	61
5.1 Kesimpulan.....	61
5.2 Saran.....	61
DAFTAR PUSTAKA	
LAMPIRAN	

DAFTAR GAMBAR

Gambar 3.1 Desain Sistem.....	17
Gambar 3.2 Flowchart Preprocessing	18
Gambar 3.3 Flowchart Stopword Removal.....	19
Gambar 3.4 Flowchart Stemming	21
Gambar 3.5 Flowchart Pembobotan.....	23
Gambar 3.6 Flowchart Similarity.....	24
Gambar 3.7 Flowchart Evaluasi dan Validasi.....	26
Gambar 3.8 Dokumen Berita Online	27
Gambar 4.1 Tampilan Halaman Utama	52
Gambar 4.2 Tampilan List Dokumen.....	52
Gambar 4.3 Tampilan Kontribusi Tokoh.....	53
Gambar 4.4 Tampilan Pencarian Dokumen.....	54
Gambar 4.5 Tampilan Pencarian Kontribusi Tokoh	54

DAFTAR TABEL

Tabel 3.1 Tabel Contoh Stopword	28
Tabel 3.2 Tabel Contoh Kata Dasar	29
Tabel 3.3 Tabel Kata Kunci Kategori Pendidikan	30
Tabel 3.4 Tabel Kata Kunci Kategori Pembangunan.....	31
Tabel 3.5 Tabel Kata Kunci Kategori Sosial	32
Tabel 3.6 Tabel Kata Kunci Kategori Hukum	33
Tabel 3.7 Tabel Kata Kunci Kategori Teknologi.....	34
Tabel 4.1 Tabel Hasil Penentuan Kontribusi	48
Tabel 4.2 Tabel Hasil Penentuan Kontribusi Lanjutan	49



ABSTRAK

Zulmi, M. Nizar. 2019. Sistem Penentuan Kontribusi Tokoh Politik menggunakan K-Nearest Neighbor. Skripsi. Jurusan Teknik Informatika Fakultas Sains dan Teknologi Universitas Islam Negeri Maulana Malik Ibrahim Malang.

Pembimbing: (I) Dr. Cahyo Crysdian dan (II) Fatchurrohman M.Kom

Kata Kunci: pembobotan kata, TF-IDF, *K-Nearest Neighbor*

Dalam hal penentuan jenis dokumen sudah menjadi hal yang umum ditemukan dalam komputer. salah satu metode yang digunakan yaitu *K-Nearest Neighbor* dengan menggunakan pembobotan TF-IDF. Penelitian ini menggunakan dokumen dari media *online* yang digunakan sebagai data training dan data testing. tujuan yang diperoleh yaitu untuk menentukan bidang kontribusi dari tokoh politik di Indonesia. bidang kontribusi meliputi pendidikan, pembangunan, sosial, teknologi, dan hukum. untuk menentukan jenis dokumen dalam data testing diperlukan term unik untuk mewakili dari tiap-tiap bidang kontribusi. dokumen yang relevan dari tokoh akan dilakukan penentuan bidang kontribusi dan pengurutan sesuai jumlah prioritas. hasil yang didapatkan dilakukan perbandingan dengan hasil dari narasumber. setelah dilakukan pengujian didapati akurasi sebesar 60% untuk kesesuaian hasil dari sistem dengan hasil dari narasumber. dengan adanya sistem penentuan kontribusi tokoh politik diharapkan bisa mengetahui kontribusi dari tokoh-tokoh yang ada di Indonesia. dikarenakan sebaik-baiknya manusia adalah yang bermanfaat terhadap manusia lain.

ABSTRACT

Zulmi, M. Nizar. 2019. System for Determining the Contribution of Political Leaders using K-Nearest Neighbor. Thesis. History Informatics Engineering Faculty of Science and Technology State Islamic University Maulana Malik Ibrahim Malang.

Advisor: (I) Dr. Cahyo Crysdiand and (II) Fatchurrohman M.Kom

Keywords: *term weighting, TF-IDF, K-Nearest Neighbor*

In terms of determining the type of document has become a common thing found in computers. one of the methods used is K-Nearest Neighbor using TF-IDF weighting. This study uses documents from online media that are used as training data and testing data. the objectives obtained were to determine the field of contribution of political leaders in Indonesia. fields of contribution include education, development, social, technology, and law. to determine the type of document in the data testing unique terms are needed to represent each of the contribution fields. the relevant documents from the figures will be determined in the field of contributions and sorting according to the number of priorities. the results obtained were compared with the results of the speakers. after testing, found an accuracy of 60% for the suitability of the results of the system with the results of the speakers. with the existence of a system for determining the contribution of political figures it is hoped that they can know the contributions of figures in Indonesia. because the best of human beings is beneficial to other humans.

الملخص

زلمي ، م. نزار. 2019 م. نظام تحديد مساهمة قادة السياسيين باستخدام *K-Nearest Neighbor*. البحث العلمي. قسم الهندسة المعلوماتية كلية العلوم والتكنولوجيا بجامعة الإسلامية الحكومية مولانا مالك إبراهيم مالانج.
المشرف: (1) الدكتور جهيوكريسدیان و (2) فتح الرحمن الماجستير

الكلمات الرئيسية: ترجيح الكلمة ، *TF-IDF* ، *K-Nearest Neighbor*

أصبح تحديد نوع البيانات شيء شائع في أجهزة الكمبيوتر. إحدى الطرق المستخدمة هي *K-Nearest Neighbour* التي تستخدم ترجيح *TF-IDF*. تستخدم هذه الدراسة البيانات من وسائط عبر الإنترنت تستخدم كبيانات التدريب وبيانات الاختبار. تهدف هذا البحث إلى تحديد مجال مساهمة قادة السياسيين في إندونيسيا. وهذه مجالات المساهمة تشمل على التعليمية، والتنمية، والاجتماعية، والتكنولوجيا، والقانونية. ولتحديد نوع البيانات عند الاختبار ، احتاج إلى مصطلحات فريدة لتمثيل كل مجال من مجالات المساهمة. وأما البيانات فمأخوذة من القادة المناسبة بمجال المساهمة ويتم بترتيبه حسب عدد الأولويات. تمت هذه البيانات بمقارنة النتائج المحسولة مع نتائج من عينة البحث. بعد الاختبار ، وجدت دقة 60 ٪ مدى ملاءمة نتائج النظام مع نتائج من عينة البحث. مع وجود نظام تحديد مساهمة قادة السياسيين، معرفة مساهمات السياسيين في إندونيسيا. لأن خير الناس أنفعهم لغيره.

BAB I

PENDAHULUAN

1.1 Latar Belakang Masalah

Pada era globalisasi seperti saat ini tentunya kita dituntut untuk mengimbangi perkembangan teknologi dan informasi. Teknologi sebagai alat penunjang pengelolaan sumber daya dan informasi sebagai petunjuk terhadap pengelolaan yang dilakukan. Dalam perkembangan tersebut tentunya diperlukan integrasi antara teknologi dan informasi sehingga mendapatkan hasil yang maksimal.

Salah satu bentuk integrasi antara teknologi dan informasi yaitu media *online*. Menurut pendapat dari Asep M. Romli dalam bukunya yang berjudul *Panduan Mengelola Media Online* terbit pada tahun 2012 disebutkan bahwa media *online* yaitu media massa atau umum yang disajikan secara *online* pada situs media internet. Media *online* mudah diakses karena menggunakan internet yang bisa mencakup seluruh dunia. Di Indonesia pun perkembangan media *online* semakin menunjukkan peningkatan.

Dari Survei yang telah dilaksanakan oleh Asosiasi Penyelenggara Jaringan Internet Indonesia atau disebut (APJII) pada tahun 2016 mengungkap bahwa 132,7 juta penduduk wilayah Indonesia telah terkoneksi dengan internet dari total penduduk Negara Indonesia yaitu sebanyak 256,2 juta orang. Dan juga dilansir dari situs *alexacom* diakses pada 5 Februari 2017 disebutkan bahwa situs *tribunnews.com* dan *detik.com* menduduki urutan keempat dan kelima yang paling

banyak dikunjungi di Indonesia. Tentunya hal tersebut membuktikan bahwa media *online* sudah menjadi kebutuhan utama bagi masyarakat pengguna internet di Indonesia.

Dalam media *online* tentu fokus utamanya adalah penyampaian pesan kepada pembaca terhadap suatu objek yang dibahas. Objek tersebut bisa berupa perorangan, kelompok atau yang lain. Khususnya perorangan seperti seorang tokoh yang dijadikan objek, tentunya berita tersebut mengarahkan pembaca untuk menyampaikan hal-hal positif, negatif atau netral terkait tokoh tersebut. Dari hal tersebut diaksudkan agar pembaca bisa menarik kesimpulan atas berita yang diberikan.

Untuk menentukan inti dari suatu berita kita harus membaca dan memahami secara seksama. tentunya harus bersabar jika berita yang disuguhkan terlampaui panjang. Dari hal itulah untuk mempermudah memahami inti dari berita tersebut maka penulis berinovasi untuk membuat aplikasi analisis kontribusi tokoh berdasarkan dokumen media *online*. bertujuan untuk memudahkan pembaca mengetahui kontribusi tokoh yang dibahas berdasarkan dokumen. Dan berikut ayat Al-Qur'an yang menerangkan tentang pentingnya untuk memberikan kontribusi yang baik dan bermanfaat kepada sesama. Allah Subhanahu Wa Ta'ala berfirman :

وَلِكُلِّ وِجْهَةٌ هُوَ مُوَلِّيٰهَا فَاسْتَبِقُوا الْخَيْرَاتِ ۚ اَيْنَ مَا تَكُونُوا يَأْتِ بِكُمْ اللّٰهُ جَمِيعًا ۚ اِنَّ

اللّٰهُ عَلٰى كُلِّ شَيْءٍ قَدِيْرٌ

Terjemahan : “Dan bagi tiap-tiap umat ada kiblatnya (sendiri) yang ia menghadap kepadanya. Maka berlomba-lombalah (dalam membuat) kebaikan. Di mana saja kamu berada pasti Allah akan mengumpulkan kamu sekalian (pada hari kiamat). Sesungguhnya Allah Maha Kuasa atas segala sesuatu.” (Qur’an Surat Al-Baqarah ayat 148)

Tafsir Jalalayn : “(Dan bagi masing-masing) maksudnya masing-masing umat (ada arah dan tujuan) maksudnya kiblat (tempat ia menghadapkan wajahnya) di waktu shalatnya. Menurut suatu qiraat bukan 'muwalliihaa' tetapi 'muwallaahaa' yang berarti majikan atau yang menguasainya, (maka berlomba-lombalah berbuat kebaikan) yakni segera menaati dan menerimanya. (Di mana saja kamu berada, pastilah Allah akan mengumpulkan kamu semua) yakni di hari kiamat, lalu dibalas-Nya amal perbuatanmu. (Sesungguhnya Allah Maha Kuasa atas segala sesuatu)”.

Dari ayat Al-Quran tersebut dapat diketahui bahwa untuk menjadi manusia yang terbaik adalah dengan memberikan kontribusi yang baik dan bermanfaat terhadap sesama manusia maupun kepada lingkungan. Sejalan dengan hal tersebut, maka tokoh-tokoh dalam bidang politik harus bisa mewujudkan hal tersebut, yaitu memberikan kontribusi yang bermanfaat terhadap masyarakat.

Untuk membantu mengetahui kontribusi tokoh maka diperlukan sebuah metode untuk menganalisa. Dari beberapa penelitian mengenai klasifikasi teks metode yang umum digunakan yaitu *Naive Bayes Classifier* dan *K-Nearest Neighbour*. Dari kedua metode tersebut yang sudah banyak digunakan adalah metode *Naive Bayes Classifier*, maka dalam penelitian ini menggunakan metode *K-Nearest Neighbour* untuk penentuan kontribusi tokoh berdasarkan dokumen.

1.2 Pertanyaan Masalah

Berdasarkan latar belakang tersebut maka rumusan masalah yang akan diselesaikan adalah sebagai berikut :

1. Berapakah Akurasi Penentuan Kontribusi Tokoh Politik menggunakan *K-Nearest Neighbour*
2. Berapakah *Error Rate* Penentuan Kontribusi Tokoh Politik menggunakan *K-Nearest Neighbour*

1.3 Tujuan Penelitian

Adapun tujuan dalam penelitian ini adalah sebagai berikut :

1. Mengukur Akurasi Penentuan Kontribusi Tokoh Politik menggunakan *K-Nearest Neighbour*
2. Mengukur *Error Rate* Penentuan Kontribusi Tokoh Politik menggunakan *K-Nearest Neighbour*

1.4 Batasan Masalah

Untuk menghindari meluasnya pembahasan dalam penelitian, maka perlunya batasan masalah. Adapun batasan masalah dalam penelitian ini adalah sebagai berikut :

1. Dokumen yang digunakan adalah dokumen berbahasa Indonesia
2. Dokumen diambil dari situs www.kompas.com dan situs www.republika.co.id dalam rentan waktu antara 1 januari 2018 hingga 31 maret 2018.
3. Bidang kontribusi yaitu hukum, sosial, teknologi, pendidikan dan pembangunan.

1.5 Manfaat Penelitian

Penelitian ini diharapkan dapat memberikan alternatif bagi masyarakat untuk membantu menemukan kontribusi dari tokoh tokoh politik di Indonesia. Kontribusi tersebut didapatkan dari media *online* berbahasa Indonesia.

1.6 Sistematika Penulisan

Untuk memahami laporan penelitian ini, maka materi-materi yang tertera pada dikelompokkan menjadi beberapa sub bab dengan sistematika penulisan sebagai berikut :

BAB I PENDAHULUAN

Berisi tentang latar belakang, perumusan masalah, tujuan dan manfaat penelitian, batasan penelitian, dan sistematika penulisan.

BAB II STUDI PUSTAKA

Bab ini berisikan teori yang berupa pengertian dan definisi yang diambil dari kutipan buku yang berkaitan dengan penyusunan laporan skripsi serta beberapa literature review yang berhubungan dengan penelitian.

BAB III DESAIN DAN IMPLEMENTASI

Bab ini berisikan tentang perancangan sistem, desain sistem yang digunakan dalam mempersiapkan konsep penelitian

BAB IV UJICOBA DAN PEMBAHASAN

Bab ini menjelaskan tentang hasil dari ujicoba yang telah dilakukan dan disertakan dengan pembahasan mengenai hasil dari ujicoba serta integrasi Islam.

BAB V KESIMPULAN DAN SARAN

Bab ini berisi kesimpulan dan saran yang berkaitan dengan analisa dan optimalisasi sistem berdasarkan dari hasil ujicoba dan pembahasan.

DAFTAR PUSTAKA

LAMPIRAN



BAB II

STUDI PUSTAKA

2.1 Kontribusi

Kontribusi yaitu berasal dari bahasa Inggris yaitu *contribute*, *contribution*, yang mempunyai makna keikutsertaan, keterlibatan, pengaruh, melibatkan diri maupun sumbangan. sehingga kontribusi dapat berupa materi atau juga sebuah perbuatan. Sebagai contoh bersifat materi misalnya yaitu seorang yang memberikan bantuan kepada orang lain demi kebaikan sesama. Kontribusi yang mempunyai arti sebagai tindakan seperti berupa perbuatan yang dilakukan oleh individu yang menyebabkan dampak baik positif maupun negatif terhadap pihak lain. seperti melakukan kerja bakti di lingkungan untuk meningkatkan suasana aman dan nyaman di lingkungan tempat tinggal sehingga dapat memberikan dampak positif bagi masyarakat.

Dengan berkontribusi maka seseorang telah berusaha meningkatkan efektivitas dan rasa sosial dalam hidupnya. Hal tersebut dilaksanakan yaitu dengan meningkatkan perannya, hal yang menjadi bidang keahliannya, agar lebih tepat sesuai dengan kemampuan. Kontribusi juga dapat dilakukan dalam berbagai bidang seperti bidang ideologi, kepemimpinan, ekonomi, kebudayaan, dan lainnya (Bakti, 2012).

Berdasarkan pengertian kontribusi yang dikemukakan di atas maka dapat diartikan bahwa kontribusi adalah keterlibatan dalam hal materi atau tindakan yang akan menimbulkan dampak positif maupun negatif terhadap individu

tersebut maupun orang lain dan berpengaruh terhadap reputasi serta kompetensi terhadap individu tersebut.

2.2 Text Mining

BAB I *Text Mining* adalah suatu bidang dalam *data mining*. *Text mining* yaitu proses yang bertujuan untuk mendapatkan informasi seperti pengguna dengan melakukan interaksi terhadap sekumpulan dokumen menggunakan metode analisis, yaitu merupakan suatu sistem dalam *data mining* yang salah satu komponennya adalah klasifikasi. Fungsi text mining adalah untuk memperoleh suatu informasi yang bermanfaat dari sekumpulan data dokumen. Oleh karena itulah, data yang difungsikan pada proses *text mining* adalah sekumpulan data yang mempunyai bentuk tidak terstruktur atau semi terstruktur. Tujuan khusus dari *text mining* yaitu pengelompokan teks dan pengkategorisasian teks (Feldman dan Sanger, 2007).

BAB II *Text mining* adalah tumpuan pada percobaan yang termasuk baru. *Text mining* memberikan kemudahan dan solusi terhadap masalah penelitian analisa, klasifikasi dan pencarian teks yang tidak terstruktur dalam kuantitas yang banyak. Dalam proses pencarian solusi, *text mining* menggunakan dan mengembangkan beberapa metode dari bidang lain, seperti *Machine Learning*, *Linguistic*, *Data Mining*, *Statistik*, *Information Retrieval*, dan *Matematik*. *Natural Language Processing*, dan *Visualization*. Penelitian *text mining* antara lain seperti ekstraksi dan juga penyimpanan teks, kemudian *preprocessing* konten teks, pengumpulan data statistic *indexing*, dan analisa konten.

Pada intinya proses yang dilakukan pada *text mining* banyak menggunakan dari penelitian *data mining* akan tetapi yang membedakan adalah pola yang

dipilih pada *text mining* diambil dari kumpulan bahasa alami yang formatnya tidak terstruktur sedangkan dalam *data mining* pola yang digunakan dari *database* yang terstruktur (Han & Kamber, 2006). Tahapan dari *text mining* secara umum adalah *text preprocessing* dan *feature selection*.

2.2.1 Text Preprocessing

Tahap dari *text preprocessing* adalah proses pertama dari *text mining*. Pada tahapan ini adalah melakukan semua fungsi, dan proses untuk mempersiapkan data yang digunakan pada proses *knowledge discovery* sistem *text mining*. Kegiatan yang dilakukan adalah *to Lower Case*, yaitu kegiatan yang bertujuan untuk mengubah karakter huruf menjadi karakter huruf kecil dan *Tokenizing* yaitu proses dimana pemecah deskripsi yang pada awalnya berupa kalimat-kalimat menjadi kata-kata kemudian menghilangkan delimiter delimiter seperti tanda titik (.), koma (,), spasi dan karakter angka yang ada pada kata tersebut (Weiss, 2005).

2.2.2 Feature Selection

Tahapan pada seleksi fitur (*feature selection*) adalah bertujuan untuk mengurangi kepadatan dari suatu kalimat, atau menghapus kata-kata yang pada akhirnya nanti tidak digunakan atau tidak menggambarkan sebagai kunci dari dokumen sehingga proses dari pengelompokan data bisa menjadi lebih efektif dan akurat. Proses yang dilakukan adalah menghilangkan *stopword* (*stopword removal*) dan *stemming* terhadap kata yang memiliki imbuhan. *Stopword* adalah kumpulan kata yang bukan merupakan kata unik dari suatu dokumen. Contohnya “di”, “oleh”, “pada”, “sebuah”, “karena” dan lain sebagainya. Sebelum proses *stopword removal* dilaksanakan, maka perlu dibuat daftar *stopword* (*stoplist*). Pada suatu kata yang mengandung *stoplist* maka kata-kata tersebut akan

dihilangkan dari proses sehingga term yang tersisa di dalam proses dianggap sebagai term yang menjadi ciri isi dari suatu dokumen atau *keywords*.

Pada proses selanjutnya setelah melewati proses *stopword removal* proses selanjutnya adalah yaitu proses *stemming*. *Stemming* yaitu proses dimana penguraian bentuk (*variants*) dari suatu term menjadi bentuk kata dasarnya (*stem*) (Tala, 2003). Fungsi dari proses *stemming* yaitu menghilangkan imbuhan-imbuhan berupa prefiks, sufiks, maupun konfiks yang terdapat pada suatu kata. Jika pada imbuhan tersebut tidak dihapus maka setiap satu kata dasar akan disimpan dengan berbagai macam bentuk yang berbeda sesuai dengan imbuhan yang digunakan maka hal tersebut akan menambah berat pada beban *database*.

Jika pada proses tersebut telak dilaksanakan maka proses menghilangkan imbuhan-imbuhan yang terdapat pada setiap kata dasar, menyebabkan satu kata dasar akan disimpan sekali walaupun kata dasar tersebut pada sumber data sudah berubah dari bentuk aslinya dan mendapatkan berbagai macam imbuhan. Dikarenakan bahasa Indonesia mempunyai aturan morfologi maka pada proses *stemming* dilaksanakan berdasarkan pada aturan morfologi bahasa Indonesia. Berdasarkan penelitian terkait, terdapat beberapa algoritma *stemming* yang bisa digunakan untuk *stemming* pada kata yang berbahasa Indonesia diantaranya seperti algoritma algoritma *Porter stemmer bahasa Indonesia*, *Confix-Stripping*, algoritma *Arifin dan Sutiono*, dan algoritma *Idris*. Algoritma *Confix-Stripping* adalah suatu algoritma atau metode yang akurat dalam proses *stemming* bahasa Indonesia (Agusta, 2009).

2.3 K-Nearest Neighbor

K-Nearest Neighbor (KNN) yaitu proses dari metode klasifikasi yang dilakukan berdasarkan data pembelajaran yang sudah terklasifikasikan sebelumnya. Dalam Tahapan ini seperti dalam hal *supervised learning*, dimana hasil dari *dinstance* yang diklasifikasikan berdasarkan mayoritas dari kedekatan suatu jarak dari kategori yang digunakan dalam metode *K-Nearest Neighbor*. Diberikan pada titik *query* yaitu akan ditemukan sejumlah k obyek atau (titik *training*) yang memiliki kedekatan dengan titik *query* pilihan. Proses klasifikasi menggunakan hasil terbanyak diantara klasifikasi dengan k obyek dimana algoritma *K-Nearest Neighbor* menggunakan klasifikasi kedekatan ketetanggaan sebagai prediksi dari *query instance* yang dicari. Jauh atau dekatnya tetangga biasanya dihitung dengan berdasarkan fungsi *Euclidean Distance*.

$$D(a, b) = \sqrt{\sum_{k=1}^d (a_k - b_k)^2}$$

Fungsi $D(a, b)$ yaitu jarak skalar dari dua vektor data a dan b yang merupakan bentuk matrik berukuran d dimensi (Nugraheni, 2013).

Pada tahapan *training*, proses dari algoritma ini akan melaksanakan penyimpanan pada vektor-vektor dan melakukan klasifikasi pada sampel data *training*. Pada tahap ini klasifikasi terhadap ciri-ciri yang sama akan disimpan dan sebagai dasar untuk *testing* data (yang klasifikasinya tidak diketahui). Jarak dari proses yang akan dicari terhadap seluruh proses *training sample* dihitung dan jumlah k buah yang memiliki kedekatan terbesar akan diambil. Titik yang baru maka klasifikasinya termasuk pada klasifikasi terbanyak dari titik tersebut.

Dalam penentuan nilai *Euclidean Distance* pada algoritma KNN adalah sebagai berikut:

1. Memilih parameter k (jumlah tetangga paling dekat).
2. Menentukan kuadrat jarak *Eucliden* terhadap data *training* yang diberikan.
3. Hasil diurutkan secara *ascending*
4. Memilih sesuai kategori y (Klasifikasi *Nearest Neighbor* berdasarkan nilai k)
5. Dengan menggunakan kategori *Nearest Neighbor* yang paling mayoritas maka dapat diprediksikan kategori objek .

Nilai k yang berpengaruh tergantung pada jumlah data. Pada penelitian terkait, nilai k yang tinggi akan mengurangi *noise* pada proses klasifikasi, tetapi dengan membuat batas pada setiap klasifikasi maka akan membuat hasilnya semakin kabur. Nilai k yang optimal dapat diperoleh dengan optimasi parameter, yaitu menggunakan *cross-validation*. Pada kasus khusus dimana suatu klasifikasi diprediksikan dengan berdasarkan data *training* yang memiliki kedekatan (dengan kata lain, $k = 1$) disebut sebagai algoritma *Nearest Neighbor*. Keberhasilan dan keakuratan algoritma KNN dipengaruhi oleh fitur-fitur yang tidak relevan atau seperti jika pada bobot fitur tersebut tidak sesuai dengan relevansinya terhadap klasifikasi. Penelitian terhadap algoritma ini mayoritas membahas bagaimana memilih dan memberi bobot terhadap fitur agar hasil klasifikasi menjadi lebih optimal (Ayistiya, 2014).

2.4 Penelitian Terkait

Dari berbagai referensi terdapat beberapa penelitian yang terkait terhadap penelitian ini yaitu penelitian pertama oleh Riandri Anggono menjelaskan bahwa penelitian tersebut menggunakan metode *Naive Bayes* dan *K-Nearest Neighbor* dalam suatu proses klasifikasi teks yang mempunyai kata homograf. Tidak hanya itu dilakukan juga analisis pada performa hasil dari klasifikasi dua metode ini. Parameter performansi yang digunakan adalah *precision*, *recall*, dan *f-measure*. Pada akhir pengujian metode *Naive Bayes* menghasilkan nilai rata-rata *f-measure* yang lebih tinggi dibanding *K-Nearest Neighbor* (Anggono, 2009).

Pada penelitian kedua oleh Defri Rosdiansyah yaitu melakukan klasifikasi orientasi sentimen dalam 3 jenis yaitu positif, negatif dan netral menggunakan kombinasi *K-Nearest Neighbor* (KNN) dan pendekatan *lexicon*. Topik penelitian tersebut yaitu menentukan analisis sentimen dari kasus politik, khususnya sentimen publik terhadap tokoh politik pada kandidat calon presiden Republik Indonesia 2014. Algoritma KNN pada penelitian tersebut akan bertindak sebagai algoritma yang akan melakukan klasifikasi orientasi sentimen sementara. Sedangkan pendekatan *lexicon* untuk dapat melakukan identifikasi kalimat netral pada tahapan *preprocessing query* juga melakukan evaluasi kalimat negasi. Sistem tersebut telah dapat menentukan orientasi sentimen pada media Twitter dengan tingkat akurasi *K Nearest Neighbor* adalah 70%, *Lexicon* 77%, Kombinasi A 78% dan Kombinasi B 82 % (Rosdiansyah, 2014).

Penelitian ketiga oleh Deny Chandra disebutkan bahwa perkembangan pada media digital telah tumbuh dengan masif. Diperkirakan 80% media digital tersebut dalam bentuk yang tidak terstruktur. Dengan tingginya volume dokumen

teks ini dipicu dengan aktivitas berbagai sumber dalam berita tersebut. Kebutuhan analisa pada text mining diperlukan untuk menangani teks yang tidak terstruktur. Untuk pengelompokan jenis berita, mayoritas peneliti yang berusaha untuk melakukan klasifikasi terhadap berita ini melakukannya secara otomatis, yaitu salah contohnya adalah dengan menggunakan klasifikasi *Naive Bayes*. Pada penelitian tersebut selain menggunakan metode *Naive Bayes*, peneliti juga menggunakan metode dengan fitur *N-Gram*. Penggunaan *N-Gram* terbukti mampu melakukan penambahan pada jenis kata sebelum memasuki proses *stemming*. Dengan memiliki tambahan jenis kata ini maka membantu proses pada klasifikasi *Naive Bayes* menjadi lebih optimal. dan juga klasifikasi *Naive Bayes* dengan tambahan fitur *N-Gram* mendapatkan hasil akurasi maksimalnya adalah 78.66% untuk data uji berita ekonomi, news, edukasi, kesehatan, olahraga, entertainment, dan lain-lain dalam Bahasa Indonesia (Chandra, 2016).

Pada penelitian keempat oleh Ahmad Fathan dijelaskan bahwa analisa dengan melaksanakan klasifikasi *tweet* yang mempunyai konten sentimen masyarakat tentang tokoh tertentu. Pada metode klasifikasi yang digunakan dalam penelitian tersebut adalah menggunakan *Naive Bayes Classifier*. *Naive Bayes Classifier* dikombinasikan dengan fitur tambahan untuk dapat mendeteksi negasi dan pembobotan menggunakan term frequency juga TF-IDF. Klasifikasi pada *tweet* yang telah dilaksanakan pada penelitian ini diperoleh berdasarkan kombinasi antara kelas sentimen dan kelas kategori. Dengan klasifikasi sentimen mempunyai opsi dari positif dan negatif sedangkan pada klasifikasi kategori terdiri dari kapabilitas, integritas, dan juga akseptabilitas. Hasil dari pengujian dapat diketahui bahwa akurasi dengan *term frequency* memberikan hasil akurasi

yang lebih optimal daripada akurasi dengan fitur TF-IDF. Metode *Support Vector Machine* menghasilkan akurasi performansi lebih baik dari pada metode *Naive Bayes* baik dalam klasifikasi sentimen maupun dalam klasifikasi kategori. Namun, secara keseluruhan dapat diketahui bahwa pada penggunaan metode *Support Vector Machine* dan *Naive Bayes* memiliki performansi yang cukup baik untuk melakukan klasifikasi *tweet* (Fathan, 2014).

Pada penelitian kelima oleh Handoko disebutkan bahwa proses klasifikasi yang digunakan pada suatu metode feature selection yaitu *Term Frequency-Inverse Document Frequency* (TF-IDF) akan memilih beberapa kata unik dari keseluruhan kata yang ada. Selain itu klasifikasi juga melalui beberapa tahapan yang lain, yaitu tokenisasi, pembuangan *stopword*, dan *stemming*. Metode yang digunakan untuk mengklasifikasikan sekumpulan teks yang belum terdefinisi ini adalah metode *Naive Bayes Classifier* dan *K-Nearest Neighbor*, yang mana akan diperbandingkan tingkat presisinya pada penelitian ini. Dalam penelitian tersebut, nilai *feature selection* yang tinggi akan membuat hasil klasifikasi dengan menggunakan metode *Naive Bayes Classifier* memiliki tingkat presisi yang lebih tinggi daripada metode *K-Nearest Neighbor*. Selain itu untuk jumlah dokumen pelatihan yang banyak akan membuat klasifikasi dengan metode KNN lebih baik daripada dengan menggunakan metode NBC (Handoko, 2008).

Penelitian keenam yang dilakukan oleh Ricky Imanuel dapat diketahui bahwa untuk menentukan prediksi dari tingkat pengunduran diri mahasiswa dalam jumlah yang banyak sulit jika dilakukan secara manual, maka dibutuhkan sebuah algoritma yang dapat mengelompokkan prediksi dari tingkat pengunduran diri mahasiswa secara otomatis dan mudah menggunakan komputer. Dalam

memprediksi digunakan sistem pendukung keputusan berupa *prototype* dan dalam menganalisis menggunakan Metode *K-Nearest Neighbor* karena memiliki kriteria kerja mencari jarak terpendek antara data yang akan dievaluasi dengan *k* terdekat dalam data pelatihannya. Maka dari hasil pengujian menggunakan *prototype* sistem pendukung keputusan dan dibandingkan secara manual yang menggunakan metode Algoritma *K-Nearest Neighbor* dengan 4 variabel yaitu IPK, Pekerjaan orang tua, jurusan dan semester adalah mendapatkan kesesuaian 79% (Imanuel, 2014).

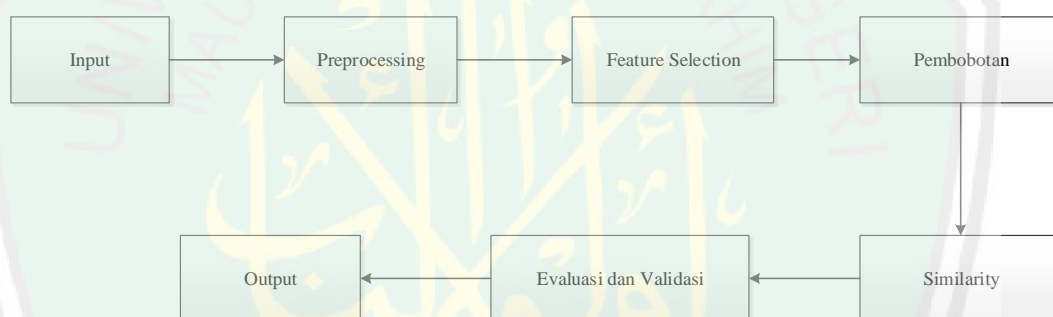
Dari beberapa penelitian yang terkait, maka pada penelitian ini melakukan penentuan kontribusi tokoh politik dengan menggunakan metode *K-Nearest Neighbor*. Metode *K-Nearest Neighbor* dipilih karena untuk menganalisa dokumen dalam jumlah banyak metode ini lebih baik daripada metode *Naive Bayes Classifier*, dengan menggunakan nilai *k* yang tepat maka akan mendapatkan hasil validasi yang lebih baik. Data yang digunakan dari penelitian ini bersumber dari media *online* dan kategori dari dokumen tersebut yaitu mengenai bidang ekonomi, pendidikan dan pembangunan. Model dan pembobotan dokumen dihitung dengan menggunakan *Term Frequency/TF* serta *Term Frequency-Inverse Document Frequency /TF-IDF*.

BAB III

DESAIN DAN IMPLEMENTASI

3.1 Sistem Desain

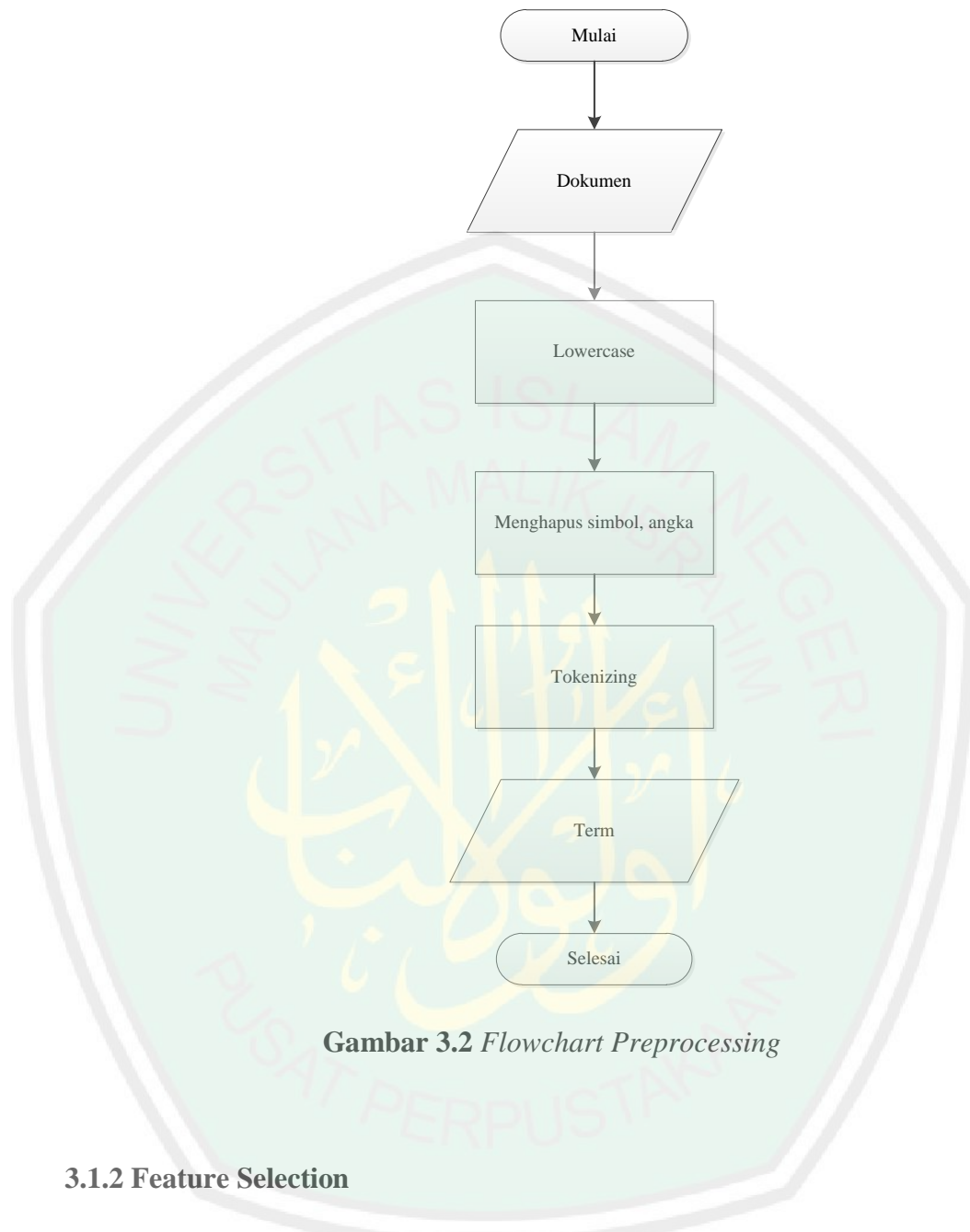
Untuk memberikan penjelasan terhadap rancang bangun dari sistem, maka diperlukan sistem desain agar bisa menggambarkan rancangan dari bagian *input*, proses sampai *output* dari sistem tersebut. dalam penelitian ini rancang bangun dari sistem desain ditunjukkan dengan Gambar 3.1



Gambar 3.1 Desain Sistem

3.1.1 Preprocessing

Pada tahap *preprocessing* ini yaitu proses dimana dokumen akan dibuat *Lowercase*, kemudian diproses *Tokenizing* untuk memisahkan tiap *term* dan menghapus tanda penghubung, angka, serta simbol. Hal ini dimaksudkan agar *term* tidak terlampaui banyak, dan diambil sesuai term yang mendukung penelitian. Berikut *flowchart* dari *Preprocessing* sesuai dengan Gambar 3.2

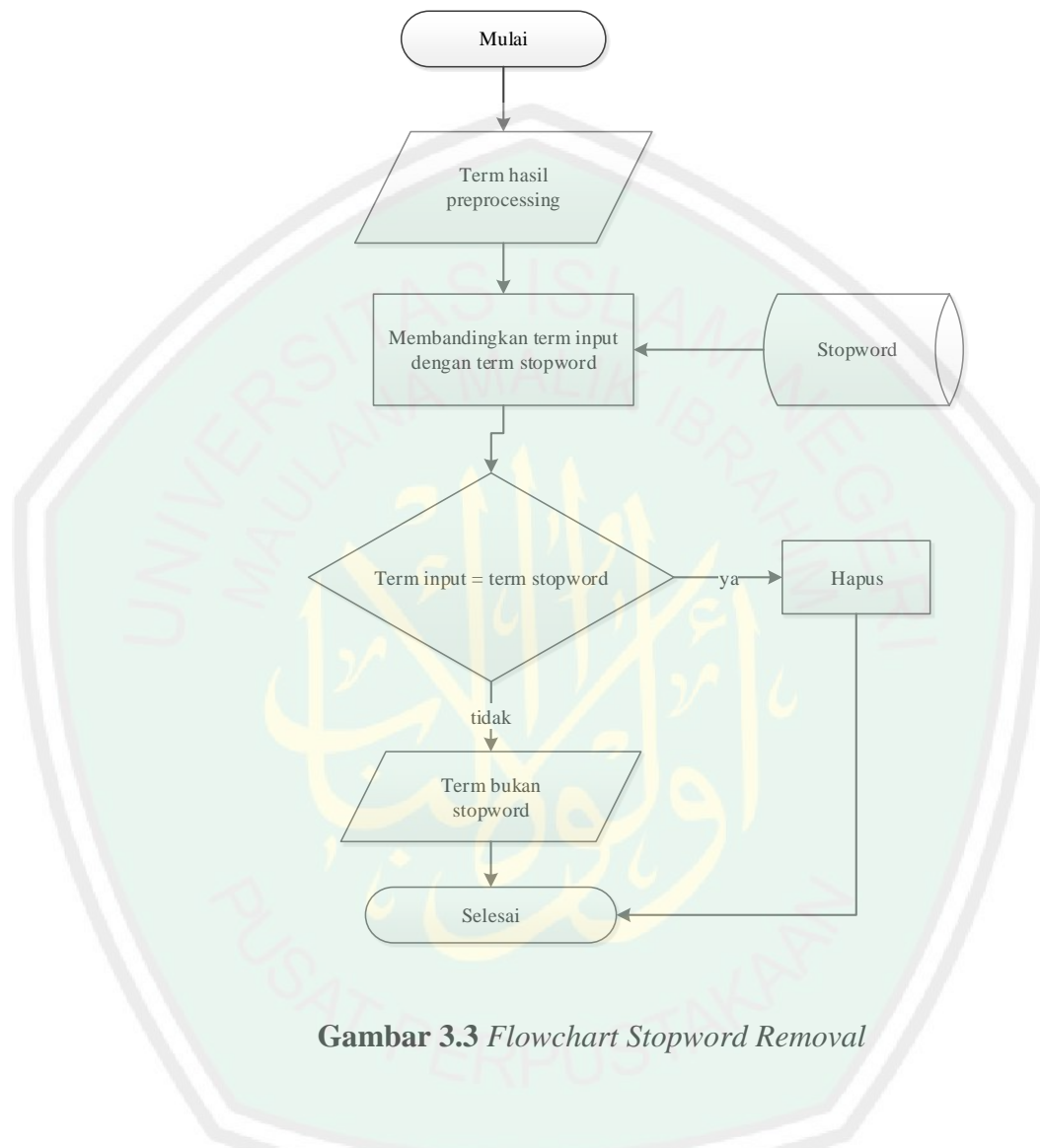


Gambar 3.2 Flowchart Preprocessing

3.1.2 Feature Selection

Pada tahap *feature selection* atau seleksi fitur yaitu dilakukan tahapan *Stopword Removal/ filtering* yaitu menghapus term/kata yang ada dalam *database stopwords*, yaitu kata yang tidak digunakan dalam penelitian dan jumlahnya banyak dalam dokumen. Tahap selanjutnya *stemming*, yaitu merubah kata/term menjadi kata dasar, jika pada data term sesuai dengan kata dasar yang ada dalam

database maka tidak perlu dilakukan perubahan menjadi kata dasar. Dalam proses *feature selection* ditunjukkan dengan Gambar 3.3

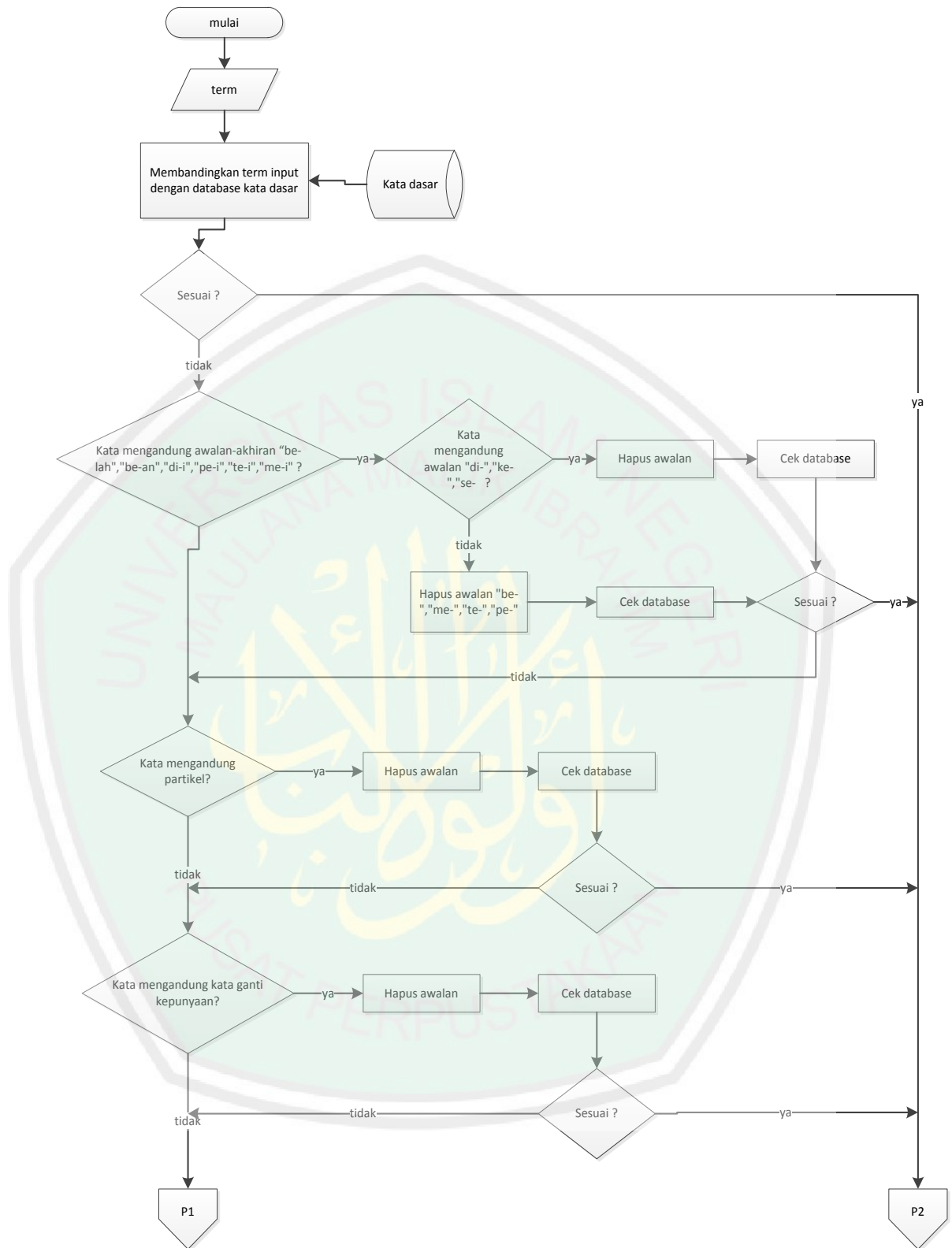


Gambar 3.3 Flowchart Stopword Removal

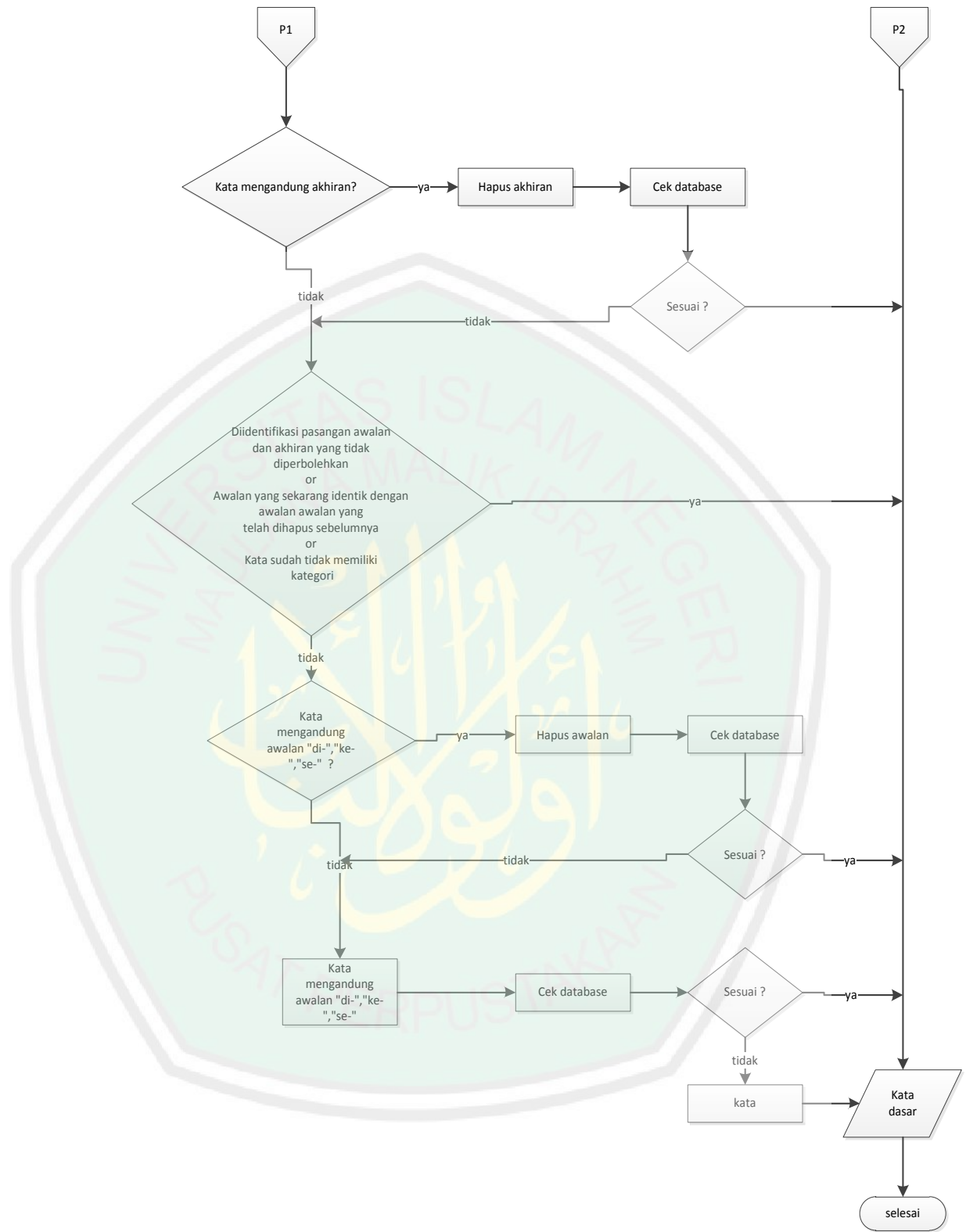
Pada proses *stemming* yaitu untuk membuat *term* dari dokumen yang diinputkan menjadi kata dasar yang sesuai. Hal ini dimaksudkan untuk mempermudah proses identifikasi dari *term* yang akan dilakukan proses pembobotan dan juga proses testing pada tahapan selanjutnya. Gambaran mengenai proses *stemming* ditunjukkan dengan Gambar 3.4 (Zaman, 2014).

Dalam gambar tersebut dijelaskan proses stemming dari awal hingga akhir dengan penunjukan posisi, arah flowchart dengan jelas guna memudahkan pemahaman penerapannya.





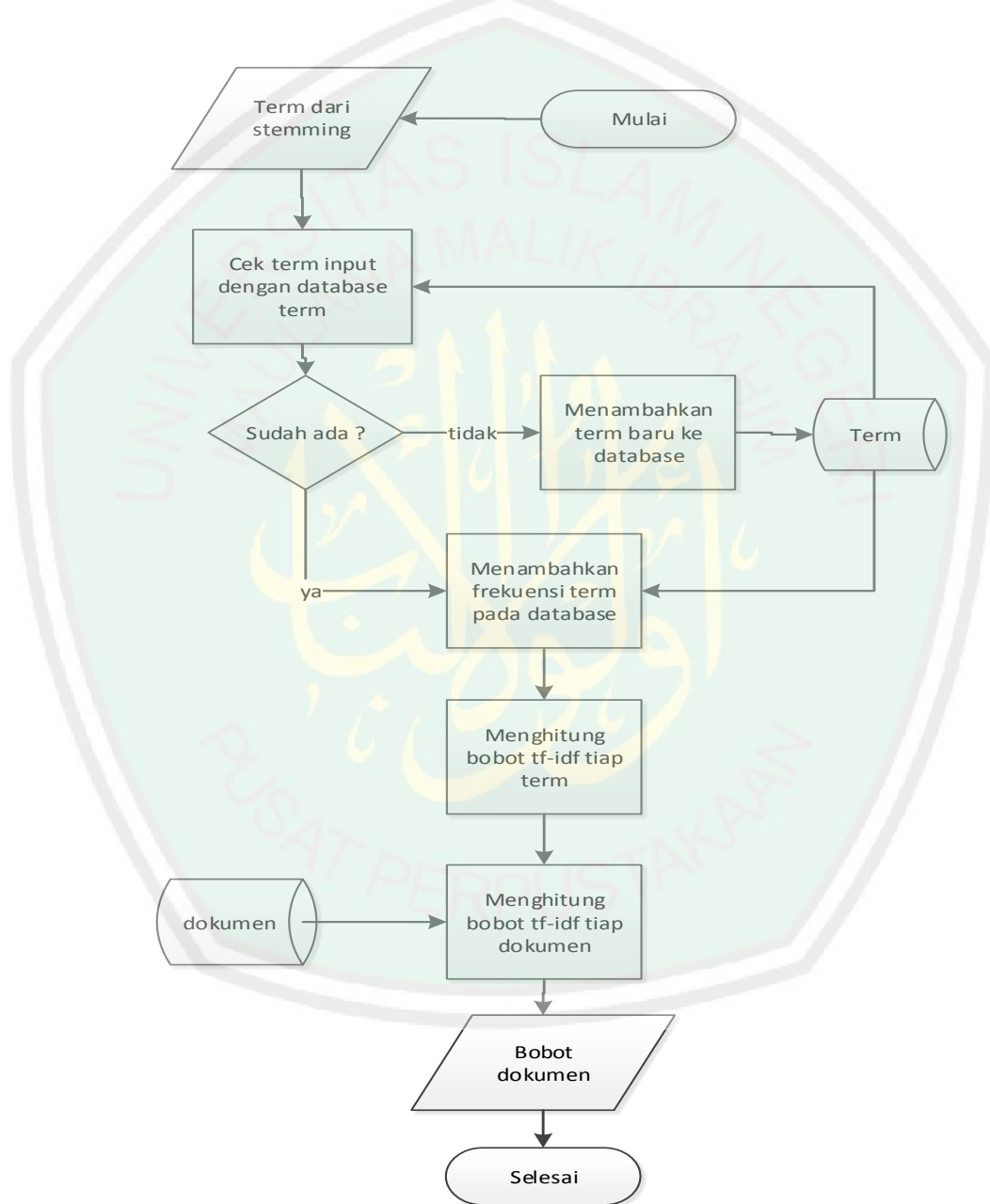
Gambar 3.4 Flowchart Stemming



Gambar 3.4 Flowchart Stemming (Lanjutan)

3.1.3 Pembobotan

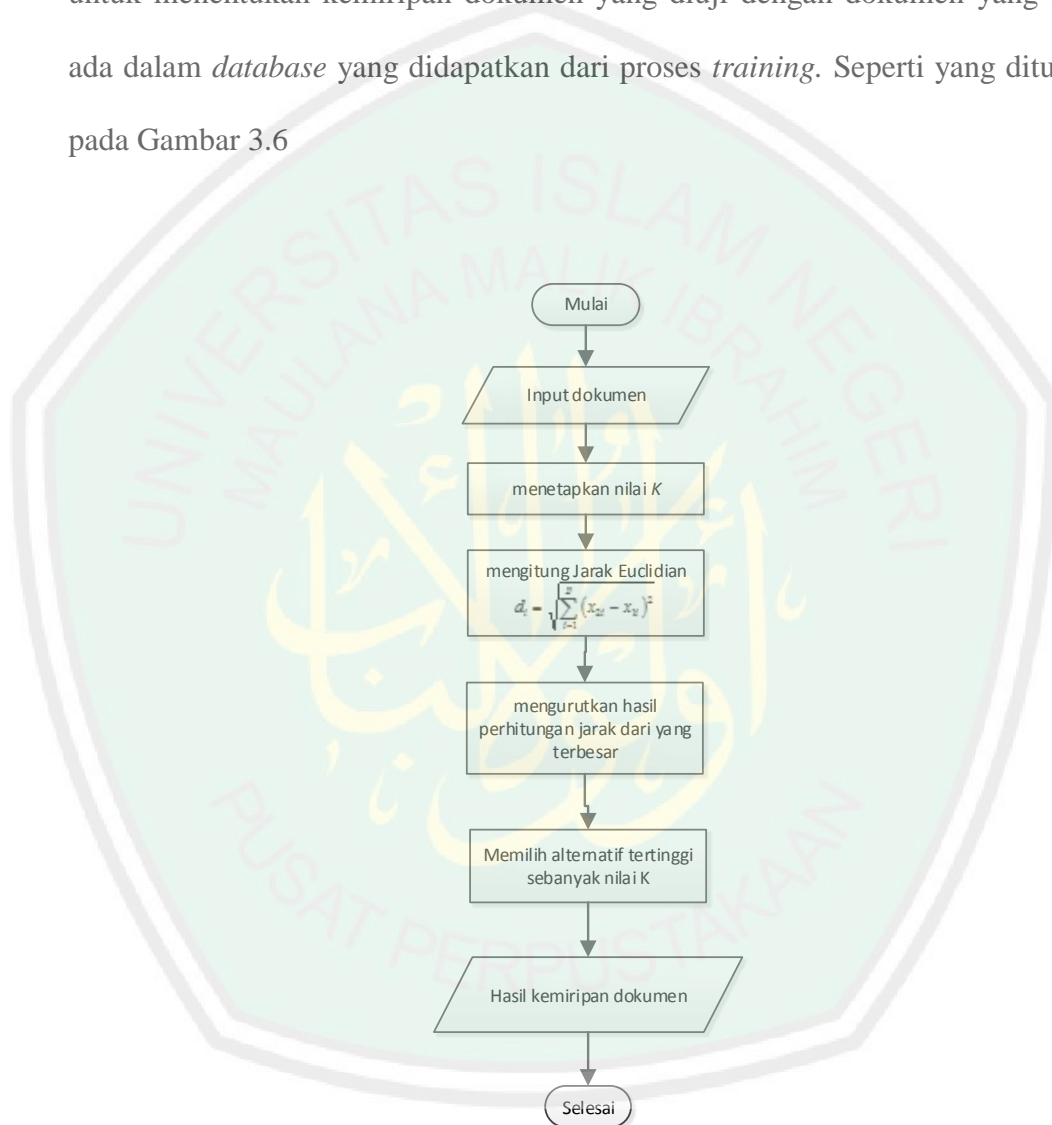
Pada tahap pembobotan yaitu memberikan nilai jumlah frekuensi kemunculan dari term/kata yang diinputkan. Frekuensi yang dimaksudkan adalah berapa kali kemunculan *term* pada dokumen yang diinputkan. Dalam tahap pembobotan ini menggunakan TF-IDF, sesuai yang ditunjukkan pada Gambar 3.5



Gambar 3.5 Flowchart Pembobotan

3.1.4 Similarity

Tahap dari proses *similarity* dokumen ini menggunakan metode *K Nearest Neighbor*. Dalam tahap ini yaitu berfungsi sebagai *testing* dari data *term* dan dokumen yang didapatkan dari proses sebelumnya. Dalam proses ini bertujuan untuk menentukan kemiripan dokumen yang diuji dengan dokumen yang sudah ada dalam *database* yang didapatkan dari proses *training*. Seperti yang ditunjukkan pada Gambar 3.6

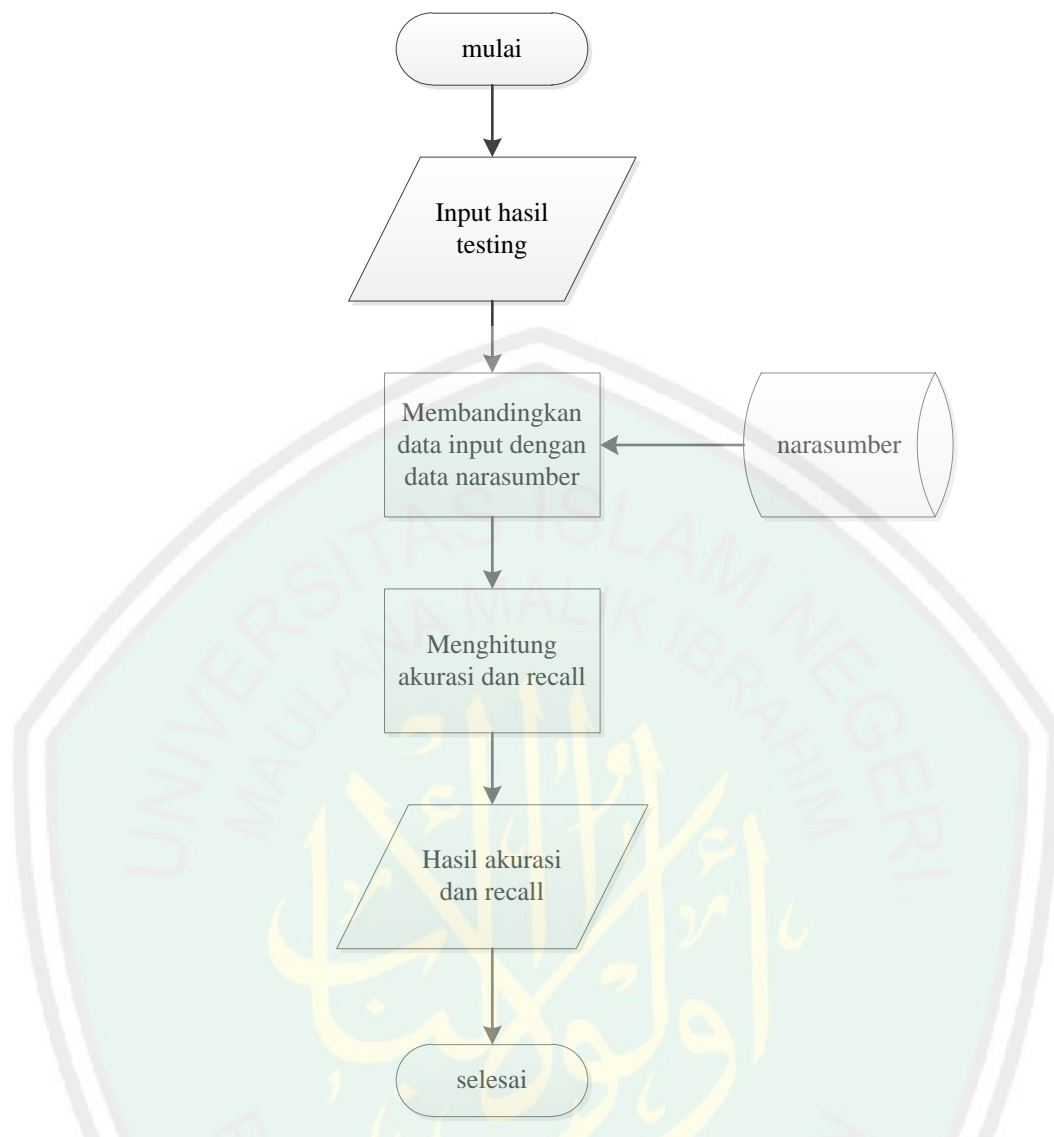


Gambar 3.6 Flowchart Similarity

3.1.5 Evaluasi dan Validasi

Pada tahap evaluasi yaitu melakukan review kembali terhadap hasil dari *testing*. Proses ini bertujuan untuk melakukan evaluasi dari hasil penilaian dari proses *similarity*, apakah sudah mendapatkan kemiripan dokumen, atau masih terdapat kesalahan. Dan juga pada tahap ini bermaksud untuk memberikan tambahan informasi agar fungsi dari sistem lebih berkembang. Hal yang dievaluasi yaitu meliputi jumlah dokumen, kuantitas, jenis dokumen, bahasa yang digunakan, metode, dan hasil dari testing digunakan sebagai indikator dalam pengembangan sistem untuk menjadi lebih optimal dan akurat.

Pada tahap validasi yaitu melakukan verifikasi terhadap hasil dari uji coba untuk menentukan akurasi dan *recall* dari proses *testing* yang sudah dilakukan. Dari hasil yang sudah didapatkan dari proses sebelumnya akan dibandingkan dengan hasil data atau kesimpulan yang didapatkan dari narasumber. Fungsi akurasi yaitu untuk memastikan apakah kesimpulan dari testing dengan data dari narasumber sesuai atau tidak. Dan juga hasil kesimpulan tersebut akan diproses *recall*/temu kembali sehingga dapat mengetahui analisa dari sistem yang sedang dibuat. Proses dalam menentukan evaluasi dan validasi digambarkan sesuai dengan Gambar 3.7



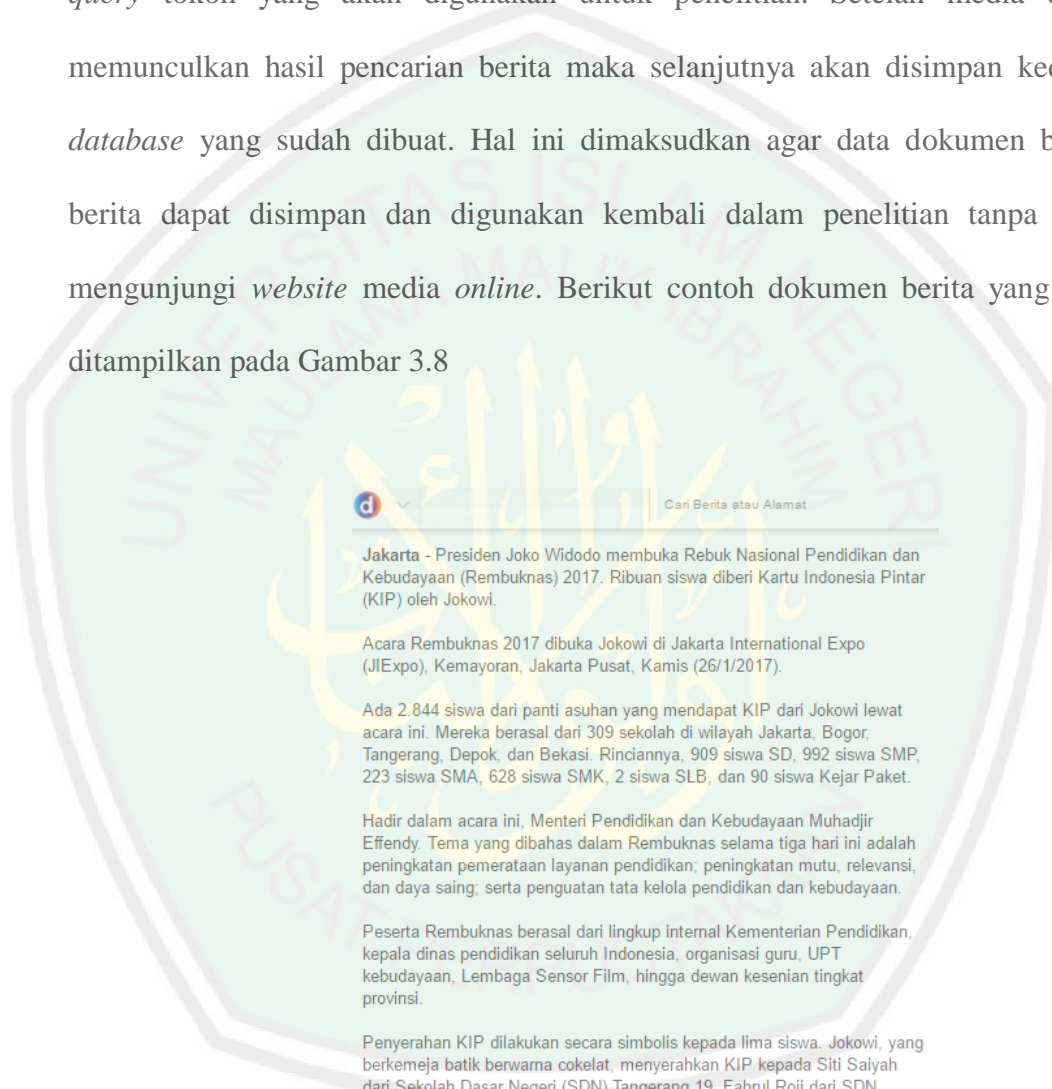
Gambar 3.7 Flowchart Evaluasi dan Validasi

3.2 Data

Ada beberapa data yang digunakan pada penelitian ini yaitu data dokumen, data *stopword*, data narasumber, dan data kata dasar. data-data tersebut digunakan sebagai sumber informasi dari beberapa proses yang ada dalam sistem.

3.2.1 Data Dokumen

Dalam penelitian ini data yang digunakan adalah berita/artikel dari media *online* yang sudah terverifikasi oleh Dewan Pers Indonesia. Berita tersebut didapatkan dengan cara mencari terlebih dahulu ke *website* media *online* dengan *query* tokoh yang akan digunakan untuk penelitian. Setelah media *online* memunculkan hasil pencarian berita maka selanjutnya akan disimpan kedalam *database* yang sudah dibuat. Hal ini dimaksudkan agar data dokumen berupa berita dapat disimpan dan digunakan kembali dalam penelitian tanpa harus mengunjungi *website* media *online*. Berikut contoh dokumen berita yang yang ditampilkan pada Gambar 3.8



Gambar 3.8 Dokumen Berita *Online*

3.2.2 Data Stopword

Data *stopword* adalah daftar kata yang tidak memiliki makna dalam penelitian dan jumlahnya biasanya banyak dalam suatu dokumen. Daftar kata tersebut akan digunakan sebagai database untuk dibandingkan dengan kata dalam dokumen yang akan diteliti. Berikut contoh daftar *stopword* pada Tabel 3.1

Tabel 3.1 Tabel Contoh *Stopword*

id_stopword	stopword
1	ada
2	adanya
3	adalah
4	adapun
5	agak
6	agar
7	akan
8	dengan

3.2.3 Data Kata Dasar

Data kata dasar yaitu kumpulan kata dasar yang bersumber dari kamus besar Bahasa Indonesia, dan sudah tersimpan dalam *database*. Rancangan tabel contoh kata dasar terdapat dalam Tabel 3.2

Tabel 3.2 Tabel Contoh Kata Dasar

id_katadasar	katadasar
1	makan
2	minum
3	lari
4	sehat
5	rapi
6	segar
7	bugar
8	siap

3.2.4 Data Kata Kunci

Data kata kunci adalah data yang diperoleh dari pemilihan kata dari *data training* yang mewakili kategori. Dalam penelitian ini terdapat lima kategori yaitu hukum, pendidikan, pembangunan, teknologi, dan sosial. Kata kunci yang disimpan dalam database tersebut bertujuan sebagai pertimbangan kemiripan dokumen yang akan akan dibandingkan. Penerapan kata kunci ditampilkan pada Tabel 3.3 sampai Tabel 3.7 sebagai bentuk penerapannya.

Tabel 3.3 Kata Kunci Kategori Pendidikan

id	term	frekuensi	kategori
103	pendidikan	171	pendidikan
104	anak	69	pendidikan
105	sekolah	68	pendidikan
106	guru	64	pendidikan
107	siswa	46	pendidikan
108	belajar	33	pendidikan
109	muhadjir	30	pendidikan
110	kebudayaan	29	pendidikan
111	jurusan	25	pendidikan
112	karakter	24	pendidikan
113	smk	22	pendidikan
114	kip	20	pendidikan
115	rektor	19	pendidikan
116	kejuruan	16	pendidikan
117	ujian	16	pendidikan
118	smp	14	pendidikan
119	didik	14	pendidikan
120	perguruan	14	pendidikan
121	sma	13	pendidikan
122	universitas	13	pendidikan
123	vokasi	13	pendidikan
124	pelatihan	12	pendidikan
125	pintar	11	pendidikan
126	mendikbud	11	pendidikan
127	mahasiswa	11	pendidikan
128	school	9	pendidikan
129	kurikulum	8	pendidikan
130	lulusan	8	pendidikan
131	pelajar	8	pendidikan

Dari Tabel 3.3 diketahui bahwa kata kunci yaitu kata yang unik sebagai perwakilan dari kategori pendidikan. Dari tabel tersebut hanya ditampilkan beberapa dari term yang mewakili kategori pendidikan. Dari kata kunci tersebut dijadikan pedoman untuk mencari dokumen yang mempunyai kemiripan dengan kategori pendidikan.

Tabel 3.4 Kata Kunci Kategori Pembangunan

id	term	freq	kategori
145	pembangunan	133	pembangunan
146	jalan	99	pembangunan
147	rumah	83	pembangunan
148	jembatan	68	pembangunan
149	perbatasan	63	pembangunan
150	tol	58	pembangunan
151	dibangun	55	pembangunan
152	perumahan	51	pembangunan
153	proyek	41	pembangunan
154	infrastruktur	40	pembangunan
155	konstruksi	38	pembangunan
156	bendungan	28	pembangunan
157	membangun	27	pembangunan
158	kereta	26	pembangunan
159	ekonomi	23	pembangunan
160	lahan	21	pembangunan
161	kota	21	pembangunan
162	bandara	18	pembangunan
163	pasar	16	pembangunan
164	pembangkit	16	pembangunan
165	subsidi	15	pembangunan
166	lokasi	14	pembangunan
167	fisik	14	pembangunan

Dari Tabel 3.4 diketahui bahwa kata kunci yaitu kata yang unik sebagai perwakilan dari kategori pembangunan. Dari tabel tersebut hanya ditampilkan beberapa dari term yang mewakili kategori pembangunan. Dari kata kunci tersebut dijadikan pedoman untuk mencari dokumen yang mirip dengan kategori pembangunan.

Tabel 3.5 Kata Kunci Kategori Sosial

id	term	freq	kategori
110	anak	22	sosial
111	kemiskinan	18	sosial
112	miskin	15	sosial
113	sosial	13	sosial
114	pengangguran	9	sosial
115	bantuan	7	sosial
116	jalan	7	sosial
117	masalah	6	sosial
118	agama	5	sosial
119	warga	5	sosial
120	bangsa	5	sosial
121	konflik	4	sosial
122	pelayanan	4	sosial
123	pemimpin	4	sosial
124	umat	4	sosial
125	keadilan	4	sosial
126	keberagaman	4	sosial
127	pengentasan	4	sosial
128	kesejahteraan	3	sosial
129	beragama	3	sosial
130	tuhan	3	sosial
131	peduli	3	sosial
132	ibadah	3	sosial

Dari Tabel 3.5 diketahui bahwa kata kunci yaitu kata yang unik sebagai perwakilan dari kategori sosial. Dari tabel tersebut hanya ditampilkan beberapa dari term yang mewakili kategori sosial. Dari kata kunci tersebut dijadikan pedoman untuk mencari dokumen yang mirip dengan kategori sosial.

Tabel 3.6 Kata Kunci Kategori Hukum

id	term	freq	kategori
163	hukum	21	hukum
164	uu	16	hukum
165	perikanan	14	hukum
166	dpr	13	hukum
167	pasal	13	hukum
168	hak	12	hukum
169	angket	11	hukum
170	kpk	9	hukum
171	ruu	8	hukum
172	peraturan	7	hukum
173	kewenangan	7	hukum
174	ilegal	7	hukum
175	aturan	6	hukum
176	konstitusi	5	hukum
177	yasonna	5	hukum
178	bertentangan	4	hukum
179	ham	4	hukum
180	pemberantasan	4	hukum
181	mahkamah	4	hukum
182	kemenkumham	4	hukum
183	illegal	4	hukum
184	adil	3	hukum
185	korupsi	3	hukum

Dari Tabel 3.6 diketahui bahwa kata kunci yaitu kata yang unik sebagai perwakilan dari kategori hukum. Dari tabel tersebut hanya ditampilkan beberapa dari term yang mewakili kategori hukum. Dari kata kunci tersebut dijadikan pedoman untuk mencari dokumen yang mirip dengan kategori hukum.

Tabel 3.7 Kata Kunci Kategori Teknologi

id	term	freq	kategori
215	mobil	34	teknologi
216	listrik	32	teknologi
217	rudiantara	20	teknologi
218	dahlan	13	teknologi
219	startup	13	teknologi
220	digital	13	teknologi
221	teknologi	12	teknologi
222	unicorn	8	teknologi
223	palapa	6	teknologi
224	ring	6	teknologi
225	transportasi	6	teknologi
226	pesawat	5	teknologi
227	tesla	5	teknologi
228	jaringan	5	teknologi
229	inovasi	4	teknologi
230	pembangkit	4	teknologi
231	komunikasi	4	teknologi
232	investor	4	teknologi
233	gojek	4	teknologi
234	kamera	4	teknologi
235	plug	3	teknologi
236	optik	3	teknologi
237	situs	2	teknologi

Dari Tabel 3.7 diketahui bahwa kata kunci yaitu kata yang unik sebagai perwakilan dari kategori teknologi. Dari tabel tersebut hanya ditampilkan beberapa dari term yang mewakili kategori teknologi. Dari kata kunci tersebut dijadikan pedoman untuk mencari dokumen yang mirip dengan kategori teknologi.

Tabel 3.8 Contoh Nilai Idf tiap Term

id	term	idf
1	pendidikan	0.316
2	anak	0.569
3	sekolah	0.502
4	guru	0.803
5	siswa	0.719
6	belajar	0.745
7	muhadjir	0.803
8	kebudayaan	0.607
9	jurusan	1.347
10	karakter	0.949
11	smk	1.046
12	kip	1.171
13	rektor	1.046
14	kejuruan	1.046
15	ujian	0.995
16	smp	0.908
17	didik	1.046
18	perguruan	1.171
19	sma	0.908
20	universitas	1.104
21	vokasi	1.250
22	pelatihan	1.104
23	pintar	0.995

Dari Tabel 3.8 diketahui bahwa setelah dilakukan proses preprocessing, pembobotan maka tiap term akan mendapatkan nilai bobot. Yang ditampilkan dalam tabel tersebut adalah kumpulan nilai idf dari tiap term yang terdapat dalam dokumen *training*.

Tabel 3.9 Contoh Nilai Bobot Dokumen

id	Doc_id	Bobot_doc
1	1	31.09
2	2	42.907
3	3	27.899
4	4	23.829
5	5	28.971
6	6	15.961
7	7	17.403
8	8	42.917
9	9	14.165
10	10	28.315
11	11	12.321
12	12	7.257
13	13	6.617
14	14	15.946
15	15	25.276
16	16	21.191
17	17	12.968
18	18	22.024
19	19	47.942
20	20	28.415
21	21	23.691
22	22	16.017
23	23	24.338
24	24	12.798

Dari Tabel 3.9 diketahui bahwa dari nilai bobot tiap term selanjutnya akan dikalkulasikan menjadi bobot dokumen yaitu dengan menjumlahkan seluruh bobor term yang ada dalam dokumen. Data bobot dokumen yang sudah didapatkan selanjutnya akan disimpan dan dijadikan acuan pada proses testing.

3.2.5 Data Narasumber

Data narasumber adalah data yang diperoleh dari narasumber. Yang memiliki fungsi sebagai pembanding dari hasil dari proses prediksi penentuan kontribusi tokoh politik. Data ini berupa hasil dari analisis yang dilakukan oleh narasumber terhadap dokumen yang diuji. Narasumber yang dipilih dalam penelitian ini adalah yang sudah mempunyai kredibilitas, terverifikasi secara

akademis dan ahli terhadap bidang bahasa Indonesia dan pendidikan. Data narasumber ini disimpan kemudian akan digunakan pada proses evaluasi dan validasi.

3.3 Analisis Sistem

Analisis bertujuan untuk menganalisa proses yang dilakukan sesuai dengan desain sistem. Hal tersebut bermaksud agar fungsi dari proses satu dengan yang lain dapat diketahui. Dalam analisis dijelaskan mengenai metode dan juga contoh dari implementasi proses yang digunakan.

3.3.1 Stemming

Stemming adalah tahapan yang dilakukan dengan sistem IR dengan mentransformasi term yang termuat dalam dokumen menjadi kata-kata akarnya (*root word*) sesuai kaidah dan logika tertentu. Seperti kata bersama, kebersamaan, menyamai, dirubah menjadi bentuk *root word*nya yaitu “sama”. Proses pada *stemming* dokumen berbahasa Indonesia berbeda prosesnya dengan menggunakan *stemming* pada teks berbahasa Inggris. Jika dalam teks yang berbahasa Inggris, proses pengerjaannya adalah dengan menghilangkan sufiks. Maka pada teks berbahasa Indonesia, selain sufiks, prefiks, dan konfiks dihilangkan. (Agusta 2009)

Stemming tahapannya dalam penelitian menggunakan algoritma yaitu *porter stemmer*. Langkah-langkah algoritma ini sebagai berikut:

1. Mengapus Partikel
2. Menghapus Possesive Pronoun.
3. Menghapus awalan pertama. Jika tidak ada dilanjutkan ke langkah 4a, jika ada dicari lalu dilanjutkan ke langkah 4b.

4. a. Menghapus awalan kedua, dilanjutkan ke langkah 5a.
 - b. Menghapus akhiran, jika tidak ada maka kata tersebut dikategorikan sebagai *root word*. Jika ada maka dilanjutkan ke langkah 5b.
5. a. Menghapus akhiran. Kemudian kata akhir dikategorikan sebagai *root word*
 - b. Menghapus awalan kedua. Kemudian kata akhir dikategorikan sebagai *root word*.

3.3.2 TF-IDF

Pada algoritma TF-IDF prosesnya yaitu melakukan perhitungan bobot pada setiap kata yang umum digunakan terhadap *information retrieval*. Proses ini menjadikan efisien, mudah dan hasilnya akurat. Proses ini melakukan perhitungan nilai *Term Frequency (TF)* dan *Inverse Document Frequency (IDF)* pada setiap kata pada tiap dokumen. Proses dari perhitungan ini dengan bobot setiap token t di dokumen dengan rumus:

$$W_{dt} = tf_{dt} * IDF_t$$

Dimana :

d : dokumen ke- d

t : kata ke- t dari kata kunci

W : bobot dokumen ke- d terhadap kata ke- t

tf : banyaknya kata yang dicari pada sebuah dokumen

IDF : *Inversed Document Frequency*

Nilai IDF didapatkan dari $IDF : \log^2 \left(\frac{D}{df} \right)$

dimana

D : total dokumen

df : banyak dokumen yang mengandung kata yang dicari

Selanjutnya pada bobot (W) tiap dokumen diketahui, akan dilakukan proses pengurutan yaitu berdasarkan semakin besarnya nilai W atau bobot, semakin besar pula tingkat similaritas dokumen tersebut terhadap kata kunci yang dicari, dan semakin kecil bobot maka semakin kecil pula tingkat similaritas dokumen tersebut terhadap kata kunci yang dicari.

3.3.3 K-Nearest Neighbor

Pada penelitian ini metode *K-Nearest Neighbor* digunakan untuk menentukan similaritas antara dokumen yang diuji dengan dokumen yang ada dalam pelatihan. Fungsi dari algoritma *K-Nearest Neighbor* yaitu untuk melakukan klasifikasi pada objek yang baru berdasarkan atribut dan *training samples*. Yaitu hasil dari sampel uji yang baru diklasifikasikan berdasarkan mayoritas kategori yang ada pada *K-Nearest Neighbor*. Algoritma *K-Nearest Neighbor* menggunakan klasifikasi kedekatan ketetanggaan dengan nilai prediksi berdasarkan sampel uji yang baru. Jarak yang dicari adalah jarak *Euclidean Distance*. (Krisandi, 2013)

$$D(a, b) = \sqrt{\sum_{k=1}^d (a_k - b_k)^2}$$

a = bobot data training

b = bobot data testing

d = jumlah atribut

D = *Euclidean Distance*

Untuk memulai proses mencari nilai jarak terdekat maka diperlukan nilai dari k terlebih dahulu. Nilai k berfungsi sebagai penentu mayoritas kemiripan dari data yang akan diklasifikasi. Untuk menentukan nilai k maka perlu dilakukan

analisis terhadap *K-Nearest Neighbor* pada penelitian sebelumnya yang sudah dilakukan.

3.3.3.1 Menentukan Nilai K

Menurut Tri Halomoan dalam penelitiannya mengenai *K-Nearest Neighbor*, metode untuk menentukan nilai k yang efektif yaitu dengan menggunakan metode *Brute Force*. Metode *Brute Force* difungsikan yaitu untuk mendapatkan nilai k yang optimal dan akurasi yang besar dari proses training atau percobaan pendahulu, maka tidak dimasukkan ke parameter nilai k secara manual dalam pengujian data (Halomoan, 2014).

Menurut Alfian Sukma Nilai k optimal untuk algoritma *k-nearest neighbor* berdasarkan data secara umumnya, dengan nilai k tinggi dapat mengurangi *noise* pada proses klasifikasi, tetapi dengan batasan pada tiap klasifikasi akan menyebabkan hasilnya kurang optimal. Untuk mendapatkan nilai k yang optimal dapat dilakukan dengan optimalisasi parameter, yaitu menggunakan *cross-validation*. Pada proses klasifikasi diprediksi berdasarkan data *training* yang memiliki kedekatan tinggi (dengan kata lain, $k = 1$) maka disebut sebagai algoritma *nearest neighbor*. Untuk mendapatkan hasil yang optimal algoritma *k-NN* dipengaruhi dengan fitur-fitur yang tidak relevan, maka bobot pada fitur tidak setara dengan relevansinya terhadap klasifikasi. Penelitian pada algoritma ini mayoritas berfokus dengan memilih dan memberikan bobot terhadap atribut atau fitur, agar hasil klasifikasi menjadi optimal (Sukma, 2014).

Menurut Nobertus Krisandi nilai k yang mayoritas digunakan adalah $k=1$, $k=3$, $k=5$ dan $k=7$. Dalam penggunaannya nilai $k=1$ dilakukan sebagai tahap utama, untuk diproses dengan *data training* dan *data testing*, lalu dilanjutkan

dengan ketentuan sebanyak 70% data *training* dan 30% untuk data *testing*. Selanjutnya menentukan *cluster* awal pada data *training* dengan berdasar pada kelas interval. kemudian *cluster* pada data *testing* dilakukan pencarian berdasarkan pengujian dengan nilai $k=3$, $k=5$ dan $k=7$. Tahapn selanjutnya adalah menghitung nilai persentase pada anggota *cluster* yang didapatkan pada proses pengujian data untuk nilai $k=3$, $k=5$ dan $k=7$ (Krisandi, 2013).

Dari beberapa penelitian diatas dapat disimpulkan bahwa untuk menentukan nilai k yaitu dengan menggunakan *Brute Force* dengan dimulai dari nilai 1,3,5,7 dan seterusnya dengan menggunakan nilai ganjil karena untuk menentukan mayoritas. Setelah dilakukan proses tersebut lalu dipilih nilai mana yang mempunyai kemiripan paling dekat atau jarak yang minimum dan mempunyai nilai akurasi terbesar.

3.3.3.2 Menghitung *Euclidean Distance*

Dalam algoritma KNN yaitu menggunakan pengelompokan ketetangaan dengan nilai prediksi dari data *testing* yang baru. Menentukan kemiripan yang digunakan yaitu dengan metode *Euclidean Distance*. Berikut penerapan untuk menghitung nilai jarak *Euclidean Distance* dalam algoritma *K-Nearest Neighbor*.

Tabel 3.3 Contoh Berita

No	Dokumen	Sumber	Kategori	Bobot	Kontribusi
1	berita 1	kompas	nasional	7,03	pendidikan
2	berita 2	kompas	nasional	6,39	hukum
3	berita 3	kompas	nasional	6,95	pendidikan
4	berita 4	kompas	nasional	3,88	pendidikan
5	berita 5	kompas	nasional	4,32	hukum
6	berita 6	kompas	nasional	5,34	pembangunan
7	berita 7	kompas	nasional	6,04	?

Dari tabel 3.12 berisi berita yang berjumlah sebanyak tujuh. Enam dari tujuh berita tersebut sudah diketahui kontribusinya. Kemudian satu berita belum diketahui kontribusinya. Maka untuk mengetahui kontribusi dari dokumen yang satu dicari dengan menggunakan jarak *Euclidean Distance* dalam algoritma *K-Nearest Neighbor* kemudian dibandingkan dengan berita yang lain untuk mengetahui berita mana yang mempunyai kesamaan terdekat. Dengan memasukkan kedalam rumus maka dapat diketahui jarak antar dokumen. Jarak antara dokumen satu dengan dokumen tujuh adalah sebagai berikut.

$$D(d_1, d_7) = \sqrt{(Wd_7 - Wd_1)^2}$$

Dari perhitungan diatas dapat diketahui bahwa jarak *Euclidean Distance* antara dokumen satu dengan dokumen tujuh. Kemudian dokumen tujuh dibandingkan dengan dokumen dua sampai dokumen enam akan menghasilkan jarak sebagai berikut.

Tabel 3.4 Hasil Perhitungan *Euclidean Distance*

No	Dokumen	Distance
1	d1,d7	0,99
2	d2,d7	0,35
3	d3,d7	0,91
4	d4,d7	2,16
5	d5,d7	1,72
6	d6,d7	0,7

Setelah diketahui jarak antara dokumen tujuh dengan dokumen yang lain maka langkah selanjutnya yaitu mengurutkan mulai dokumen yang mempunyai jarak terbesar sampai terkecil. Penerapannya sesuai dengan tabel berikut.

Tabel 3.5 *Euclidean Distance* terurut

No	Dokumen	Distance	Kontribusi
4	d4,d7	2,16	pendidikan
5	d5,d7	1,72	hukum
1	d1,d7	0,99	pendidikan
3	d3,d7	0,91	pendidikan
6	d6,d7	0,7	hukum
2	d2,d7	0,35	hukum

Setelah data dokumen terurut maka pada langkah selanjutnya yaitu mengambil dokumen yang mempunyai nilai jarak terbesar sesuai dengan nilai k . jika nilai k yaitu tiga maka akan diambil sejumlah tiga dokumen dari yang terbesar sehingga ditampilkan sebagai berikut.

Tabel 3.6 Dokumen sesuai Nilai K

No	Dokumen	Distance	Kontribusi
4	d4,d7	2,16	pendidikan
5	d5,d7	1,72	hukum
1	d1,d7	0,99	pendidikan

Dari tabel tersebut dapat diketahui bahwa dokumen tujuh mempunyai kemiripan dengan dokumen empat, lima dan juga satu. Maka sesuai dengan metode *K-Nearest Neighbor* untuk menentukan kontribusi dari dokumen tujuh yaitu dengan menentukan mayoritas dari kontribusi antara dokumen empat, lima dan satu. Maka kontribusi dari dokumen tujuh adalah pendidikan.

3.3.4 Akurasi

Akurasi adalah tingkat ketepatan antara informasi yang diminta oleh pengguna dengan jawaban yang diberikan oleh sistem.. Dalam penelitian ini untuk menentukan akurasi maka harus dilakukan pengujian terlebih dahulu. Dari data *training* dan *testing* yang sudah diuji kemudian akan dibandingkan dengan data yang diperoleh dari narasumber. Contoh penerapan seperti dalam tabel berikut.

Tabel 3.7 Contoh Hasil Pengujian

		Data Sebenarnya	
		TRUE	FALSE
Prediksi	TRUE	70	60
	FALSE	30	240

Dari data hasil pengujian maka untuk mencari akurasi dengan menggunakan rumus yang sudah ditetapkan. Contoh dari perhitungan untuk mengetahui akurasi adalah sebagai berikut.

$$\text{Akurasi} = \frac{TP+TN}{TP+PN+FP+FN}$$



BAB IV

UJI COBA DAN PEMBAHASAN

Pada tahap ini dijelaskan mengenai langkah-langkah dari proses uji coba, hasil dan pembahasannya, serta hubungan penelitian ini dengan Integrasi Islam. Hasil dari uji coba dalam penelitian ini adalah untuk mengetahui tingkat akurasi dan *error rate* dengan menggunakan metode *K-Nearest Neighbor*.

4.1 Langkah Uji Coba

Dalam tahapan ini dijelaskan langkah-langkah dari proses uji coba pada sistem penentuan kontribusi tokoh politik menggunakan *K-Nearest Neighbor*. Tahapan dari proses ini adalah sebagai berikut :

1. Pada penelitian ini menggunakan 50 dokumen berita sebagai data *training* yang terdiri dari masing-masing sepuluh dokumen tentang pendidikan, pembangunan, hukum, teknologi dan sosial. Semua dokumen tersebut dilakukan *preprocessing*, pembobotan, dan seleksi untuk dijadikan data *training*.
2. Proses seleksi kategori yaitu melakukan pemilihan kata kunci sebagai perwakilan dari kategori hukum, sosial, teknologi, pendidikan dan pembangunan. Digunakan sebagai pembanding dengan dokumen yang akan dicari jenis kategorinya.
3. Proses *testing* yaitu proses melakukan perbandingan bobot dari dokumen *testing* dengan data seleksi kategori dari dokumen pada data *training*. Untuk menghitung perbandingan yaitu dengan menggunakan rumus *Cosinus Similarity*.

4. Proses *sorting* yaitu melakukan pengurutan dari nilai *Cosinus* terbesar hingga terkecil, semakin besar nilai *Cosinus* semakin besar pula kemiripan dari dokumen tersebut. Maka dipilih nilai *Cosinus* terbesar.
5. Proses penentuan kontribusi tokoh dari hasil kemiripan terbesar dengan data pada seleksi kategori yaitu meliputi bidang hukum, sosial, teknologi, pendidikan dan pembangunan..
6. Proses *testing* dengan menggunakan dokumen sebanyak 4914 meliputi 30 tokoh politik.
7. Proses evaluasi dari hasil testing dan menentukan akurasi serta *error rate* untuk mengetahui tingkat keberhasilannya dengan membandingkan hasil dari sistem dengan hasil dari narasumber.

4.2 Hasil Uji Coba

Penelitian ini menggunakan 4914 data testing yang meliputi 30 tokoh politik di Indonesia. Dalam menentukan kontribusi tokoh politik yaitu dengan metode *K-Nearest Neighbor* untuk mengetahui kemiripan dokumen *testing* dengan data *training*.

Hasil dalam uji coba yang dilakukan adalah seperti yang ditunjukkan pada

Tabel 4.1

Tabel 4.8 Tabel Hasil Penentuan Kontribusi

No	Nama Tokoh	Hasil Sistem		Narasumber		Kesimpulan	
		1	2	1	2	1	2
1	MUHAMMAD JUSUF KALLA	Sosial	Hukum	Pembangunan	Sosial	Tidak	Tidak
2	AMIEN RAIS	Pendidikan	Sosial	Sosial	hukum	Tidak	Tidak
3	PRABOWO SUBIANTO	Hukum	Sosial	Hukum	Sosial	ya	Ya
4	ANIES BASWEDAN	Pembangunan	Hukum	Pendidikan	Pembangunan	Tidak	Tidak
5	SUSILO BAMBANG YUDHOYONO	Hukum	Sosial	Pembangunan	Sosial	Tidak	Ya
6	HATTA RADJASA	Hukum	sosial	Pembangunan	Sosial	Tidak	ya
7	BACHARUDDIN JUSUF HABIBIE	Sosial	Teknologi	Teknologi	Pembangunan	Tidak	Tidak
8	MAHFUD MD	Hukum	Pendidikan	Hukum	Pendidikan	ya	ya
9	SUSI PUDJIASTUTI	Hukum	Pendidikan	Pembangunan	Hukum	Tidak	Tidak
10	JOKO WIDODO	Hukum	Sosial	Pembangunan	Sosial	Tidak	Ya
11	DAHLAN ISKAN	Hukum	Pendidikan	Pendidikan	Teknologi	tidak	tidak
12	MUHAMMAD ZAINUL MAJDI	Sosial	Hukum	Pembangunan	Sosial	tidak	tidak
13	MEGAWATI SOEKARNOPUTRI	Pendidikan	Hukum	Pembangunan	Sosial	tidak	tidak
14	SOEKARWO	Hukum	Pendidikan	Pembangunan	Pendidikan	tidak	ya
15	WIRANTO	Hukum	Sosial	Hukum	Sosial	ya	ya
16	IGNASIUS JONAN	Pembangunan	Pembangunan	Teknologi	Pembangunan	tidak	ya
17	GANJAR PRANOWO	Sosial	Pembangunan	Pembangunan	Sosial	ya	tidak
18	RUDIANTARA	Hukum	teknologi	Teknologi	Pembangunan	tidak	tidak
19	YASONNA LAOLY	Hukum	Sosial	Hukum	Sosial	ya	tidak
20	TRI RISMA HARINI	Sosial	Hukum	Pembangunan	Sosial	tidak	tidak
21	BASUKI THAHYA PURNAMA	Sosial	Hukum	Sosial	Pembangunan	ya	tidak
22	RIDWAN KAMIL	Sosial	Pendidikan	Pembangunan	Sosial	tidak	tidak
23	BOEDIONO	Pendidikan	hukum	Pembangunan	Sosial	tidak	tidak
24	KHOFIFAH INDAR PARAWANSA	Pendidikan	sosial	Pembangunan	Sosial	tidak	tidak
25	SAIFULLAH YUSUF	Sosial	Pendidikan	Pembangunan	Sosial	tidak	tidak
26	EMIL ELISTIANTO DARDAK	Sosial	Pembangunan	Pembangunan	Sosial	tidak	tidak
27	BASUKI HADI MULJONO	Pembangunan	Hukum	Pembangunan	Teknologi	ya	tidak
28	AHMAD HERYAWAN	Sosial	Hukum	Pembangunan	Sosial	tidak	tidak
29	IMAM NAHRAWI	Pendidikan	Pembangunan	Pendidikan	Pembangunan	ya	ya
30	MUHADJIR EFFENDY	Pendidikan	hukum	Pendidikan	Teknologi	ya	tidak

Dari Tabel 4.1 diketahui hasil dari pengujian tersebut menghasilkan kesesuaian sebanyak 18 dan tidak sesuai sebanyak 42. Dari hasil tersebut dapat

diketahui

$$\text{Akurasi} = 18/60 \times 100\% = 30 \%$$

$$\text{Error Rate} = 42/60 \times 100\% = 70 \%$$

Tabel 4.2 Tabel Hasil Penentuan Kontribusi Lanjutan

No	Nama Tokoh	Hasil Sistem		Narasumber		Kesimpulan	
		1	2	1	2	1	2
1	MUHAMMAD JUSUF KALLA	Sosial	Hukum	Pembangunan	Sosial	Ya	tidak
2	AMIEN RAIS	Pendidikan	Sosial	Sosial	hukum	Tidak	ya
3	PRABOWO SUBIANTO	Hukum	Sosial	Hukum	Sosial	ya	ya
4	ANIES BASWEDAN	Pembangunan	Hukum	Pendidikan	Pembangunan	Ya	tidak
5	SUSILO BAMBANG YUDHOYONO	Hukum	Sosial	Pembangunan	Sosial	Tidak	ya
6	HATTA RADJASA	Hukum	sosial	Pembangunan	Sosial	Tidak	ya
7	BACHARUDDIN JUSUF HABIBIE	Sosial	Teknologi	Teknologi	Pembangunan	Tidak	ya
8	MAHFUD MD	Hukum	Pendidikan	Hukum	Pendidikan	Ya	ya
9	SUSI PUDJIASTUTI	Hukum	Pendidikan	Pembangunan	Hukum	Ya	tidak
10	JOKO WIDODO	Hukum	Sosial	Pembangunan	Sosial	Tidak	ya
11	DAHLAN ISKAN	Hukum	Pendidikan	Pendidikan	Teknologi	tidak	ya
12	MUHAMMAD ZAINUL MAJDI	Sosial	Hukum	Pembangunan	Sosial	Ya	tidak
13	MEGAWATI SOEKARNOPUTRI	Pendidikan	Hukum	Pembangunan	Sosial	Tidak	tidak
14	SOEKARWO	Hukum	Pendidikan	Pembangunan	Pendidikan	Tidak	ya
15	WIRANTO	Hukum	Sosial	Hukum	Sosial	Ya	ya
16	IGNASIUS JONAN	Pembangunan	Pembangunan	Teknologi	Pembangunan	Ya	ya
17	GANJAR PRANOWO	Sosial	Pembangunan	Pembangunan	Sosial	Ya	ya
18	RUDIANTARA	Hukum	teknologi	Teknologi	Pembangunan	Tidak	ya
19	YASONNA LAOLY	Hukum	Sosial	Hukum	Sosial	Ya	ya
20	TRI RISMA HARINI	Sosial	Hukum	Pembangunan	Sosial	Ya	tidak
21	BASUKI THAHYA PURNAMA	Sosial	Hukum	Sosial	Pembangunan	Ya	tidak
22	RIDWAN KAMIL	Sosial	Pendidikan	Pembangunan	Sosial	Ya	tidak
23	BOEDIONO	Pendidikan	hukum	Pembangunan	Sosial	Tidak	tidak
24	KHOFIFAH INDAR PARAWANSA	Pendidikan	sosial	Pembangunan	Sosial	Tidak	ya
25	SAIFULLAH YUSUF	Sosial	Pendidikan	Pembangunan	Sosial	Ya	tidak
26	EMIL ELISTIANTO DARDAK	Sosial	Pembangunan	Pembangunan	Sosial	Ya	ya
27	BASUKI HADI MULJONO	Pembangunan	Hukum	Pembangunan	Teknologi	Ya	tidak
28	AHMAD HERYAWAN	Sosial	Hukum	Pembangunan	Sosial	ya	tidak
29	IMAM NAHRAWI	Pendidikan	Pembangunan	Pendidikan	Pembangunan	Ya	ya
30	MUHADJIR EFFENDY	Pendidikan	hukum	Pendidikan	Teknologi	ya	tidak

Dari Tabel 4.2 diketahui hasil dari pengujian tersebut menghasilkan kesesuaian sebanyak 36 dan tidak sesuai sebanyak 24. Pada pengujian ini yang dibandingkan adalah dari tiap prioritas dengan dua prioritas yang ada yaitu dengan kemiripan silang atau sebaliknya. Dari hasil pengujian tersebut dapat diketahui

$$\text{Akurasi} = 36/60 \times 100\% = 60\%$$

$$\text{Error Rate} = 24/60 \times 100\% = 40\%$$

Dari nilai perhitungan diatas dapat diketahui bahwa sistem penentuan kontribusi tokoh politik menggunakan *K-Nearest Neighbor* yang terbaik adalah menghasilkan akurasi sebesar 60% dan juga menghasilkan *error rate* sebesar 40%.

4.3 Pembahasan

Hasil dari sistem penentuan kontribusi tokoh politik menggunakan *K-Nearest Neighbor* tersebut yaitu nilai akurasi sebesar 60% disebabkan karena dalam data testing dengan menggunakan 4914 dokumen meliputi 30 tokoh politik masing-masing tokoh mempunyai dokumen yang relevan berbeda-beda dipengaruhi oleh jumlah dokumen yang membahas tentang tokoh tersebut.

Dari dokumen relevan untuk setiap tokoh hasil dari sistem tersebut hanya dipilih dua prioritas tertinggi untuk dibandingkan. Dua prioritas tertinggi dari sistem dan dari narasumber. Untuk meningkatkan akurasi maka diperlukan dokumen tambahan sebagai sumber penentuan dan juga untuk jumlah prioritas perlu ditambahkan.

Untuk perbandingan hasil dari sistem dengan narasumber juga dipengaruhi oleh pendapat dari narasumber yang sesuai dengan pemikiran narasumber

tersebut, bukan berdasarkan dokumen yang ada dalam sistem *testing*. Menurut pendapat penulis, dengan digunakannya dua prioritas tertinggi dari sistem penentuan kontribusi tokoh politik dan menghasilkan akurasi sebesar 60% adalah sudah baik hanya saja perlu ditingkatkan tingkat akurasinya.

Dalam penentuan jenis kategori atau kontribusi dokumen, dengan dokumen 50 training sudah bisa memberikan kata kunci untuk dijadikan sebagai pembanding dengan dokumen yang akan dicari jenis kategorinya. Akan tetapi untuk lebih meningkatkan keberhasilan maka diperlukan dokumen tambahan sebagai *training* untuk menambahkan kata kunci yang ada dalam sistem.

Narasumber yang dijadikan pembanding dalam penelitian ini adalah tokoh yang berpengalaman dalam dunia pendidikan. Beliau bernama bapak Ryan Aminullah Yasin. Beliau adalah Peraih Anugerah Konstitusi Guru PKN Nasional Tahun 2016, dan peraih Guru SMK Berprestasi Nasional Tahun 2017. Saat ini beliau aktif mengajar di SMKN 1 Trenggalek dan STIT Sunan Giri Trenggalek. Dalam bidang pendidikan, konstitusi, kelembagaan, beliau berpengalaman dibuktikan dengan beberapa penghargaan dan prestasi yang telah diraih.

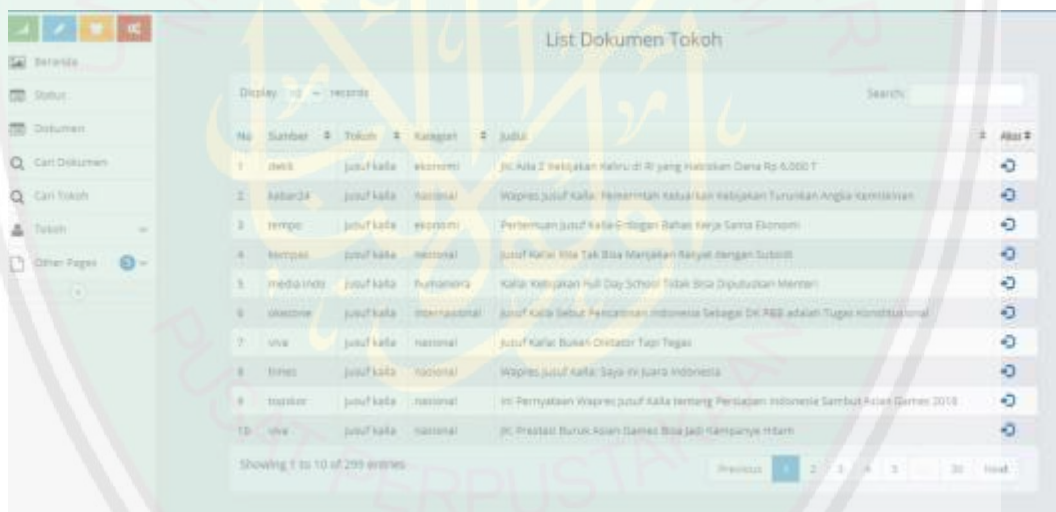
4.4 Implementasi

Dari hasil sistem yang telah dibuat, berikut tampilan dari sistem penentuan kontribusi tokoh politik menggunakan *K-Nearest Neighbor*. Mengenai tampilan dari halaman utama sistem ditunjukkan dengan Gambar 4.1



Gambar 4.1 Tampilan Halaman Utama

Dalam halaman utama sistem, dapat diketahui beberapa menu seperti menu beranda, status, dokumen, cari dokumen, cari tokoh, dan tokoh. Dari halaman ini bertujuan untuk memberikan gambaran umum mengenai sistem yang dibuat, juga untuk memudahkan user untuk menggunakan sistem ini.



Gambar 4.2 Tampilan List Dokumen

Dalam Gambar 4.2 diketahui bahwa dalam list dokumen terdapat dokumen tentang tokoh politik yang digunakan dalam sistem ini. Dari beberapa dokumen ini berguna untuk menjadi testing dari penentuan kontribusi tokoh politik yang dicari.

6	HATTA RAJARUA		hukum(1) sma(K2) pembangunan(2)
7	BACHARUDDIN Jusuf HABIBIE		sosial(2) teknologi(4) pembangunan(1)
8	MAHRUD MD		hukum(1) perdagangan(2) sma(K2)
9	SUGI PUJIASATUM		pendidikan(1) hukum(1) pembangunan(1) sma(K1)
10	JOKO WIDODO		pembangunan(6) perdagangan(6)

Gambar 4.3 Tampilan Kontribusi Tokoh

Dalam Gambar 4.3 diketahui bahwa dalam tampilan kontribusi tokoh yaitu terdapat 28 tokoh yang digunakan dalam penelitian ini. Dari tokoh tersebut dapat diketahui kontribusi dari masing masing tokoh tersebut berdasarkan dokumen ada.



Gambar 4.4 Tampilan Pencarian Dokumen

Dalam pencarian dokumen ini digunakan untuk mencari dokumen secara online dengan menggunakan google custom search dengan dikhususkan pada media *online*. Dari hasil pencarian akan disimpan kedalam *database* guna untuk menambahkan dokumen yang ada dalam sistem.



Gambar 4.5 Tampilan Pencarian Kontribusi Tokoh

Dalam pencarian kontribusi ini yaitu akan menampilkan kontribusi dari tokoh tersebut berdasarkan dokumen yang ada dalam sistem. Pencarian dilakukan dengan mengetikkan nama tokoh dalam kolom pencarian, kemudian sistem akan memunculkan hasilnya.

No	Nilai	Query	Kontribusi
1	28	derah	hakum
2	85	derah	hakum
3	120	derah	hakum
4	308	derah	perbaharuan
5	438	derah	pendidikan
6	900	derah	amal
7	794	derah	amal
8	1023	derah	pendidikan
9	1069	derah	amal
10	1229	derah	amal
11	1543	derah	pendidikan
12	1733	derah	amal

Gambar 4.6 Tampilan Hasil Similarity

Dalam proses similarity yaitu menampilkan kategori dari tiap dokumen yang mempunyai kemiripan dengan term pada tokoh yang dilakukan proses pencarian. Semakin banyak jumlah kemunculannya maka semakin besar pula tingkat kemiripannya.

4.4.1 Implementasi Sistem

Pada tahap ini dalam proses penyusunan aplikasi dibutuhkan beberapa fungsi dan komponen yang digunakan. Dalam sistem penentuan kontribusi tokoh politik menggunakan K-Nearest Neighbor terdapat beberapa aspek sebagai berikut.

Tabel 4.1. *Source Code Preprocessing*

```
<?php
include "conn.php";//sambungkan ke mysql
include "stemming.php";
//fungsi preproses menerima teks dan menerapkan beberapa tugas
awal
//fase indexing dokumen teks
function preproses($text1) {

    //1. ubah ke huruf kecil
    $text1 = strtolower(trim($text1));

    //2.hilangkan tanda baca

        $retanda = mysql_query("SELECT * FROM tbtanda ORDER BY
id");
        while($rowtanda = mysql_fetch_array($retanda)) {
            $text1 = str_replace($rowtanda['tanda'], " ", $text1);
        }
        // $text1 = str_replace(' ', ' ', $text1);

//3. hapus stoplist
    $restop = mysql_query("SELECT * FROM tbstop ORDER BY id");
    while($rowstop = mysql_fetch_array($restop)) {
        $text1 = str_replace($rowstop['stopword'], " ", $text1);
    }

//hilangkan spasi di awal & akhir teks
$text1 = trim($text1);

return $text1;
```

Dalam proses *preprocessing* yaitu melakukan perbandingan term dalam suatu dokumen dengan stopword serta tanda baca yang akan dilakukan untuk mendapatkan suatu term yang akan digunakan sebagai penentu dari testing. Tujuan dari proses ini adalah menghilangkan tanda baca, stopword dan spasi ganda dalam suatu term.

Tabel 4.2. Source Code Stemming

```

<?php
function cari1($kat){
    include "conn.php";
    $has = mysql_num_rows(mysql_query("SELECT * FROM
tb_katadasar WHERE katadasar='$kat'"));
    return $has;
}

//menghapus partikel
function happar ($kat){
if(cari($kat)!=1){
    if((substr($kat, -3) == 'kah' )||( substr($kat, -3) ==
'lah' )||( substr($kat, -3) == 'pun' )){
        $kat = substr($kat, 0, -3);
    }
}
return $kat;
}

//menghapus possessive pronoun
function hapuspp($kat){
if(cari($kat)!=1){
    if(strlen($kat) > 4){
        if((substr($kat, -2)== 'ku')||(substr($kat, -2)== 'mu')){
            $kat = substr($kat, 0, -2);
        }else if((substr($kat, -3)== 'nya')){
            $kat = substr($kat,0, -3);
        }
    }
}
return $kat;
}
}

```

Dalam proses *stemming* yaitu melakukan pengembalian suatu term agar menjadi kata dasar. Dalam proses ini dilakukan dengan menghilangkan imbuhan depan, tengah, serta belakang agar term yang digunakan sesuai dengan daftar term kata dasar. Dengan term yang sesuai kata dasar maka proses untuk mencari kemiripan dokumen menjadi lebih akurat.

Tabel 4.3. Source Code Stopword

```

function stop2($text1) {
    $stopp = mysql_query("SELECT * FROM tbstop2 ORDER BY
id");
    while($rowstopp = mysql_fetch_array($stopp)) {
        $text1 = str_replace($rowstopp['tanda'], " ", $text1);
    }
    return $text1;
}
function get_term($text1){
    // membuang tanda baca dalam text.
    $text1 = preg_replace('/[?!:.,()*"\'€\$\\n]||[-
]|\\\'/,\'', $text1);
    $text1 = str_replace('/', ' ', $text1);
    $text1 = str_replace(' ', ' ', $text1);

    // konversi ke lowercase
    $text1 = strtolower(trim($text1));

    // Tokenization
    $term = explode(" ", $text1);

    $tok = $term;
    $freq = array();

    for($i=0;$i<(count($tok));$i++){
        for($j=0;$j<(count($tok));$j++){
            if ($term[$i] == $tok[$j]){
                $freq[$term[$i]]+=1;
                array_splice($term,$i,1);
            }
        }
    }
    return $freq;
}

```

Tahapan dari proses Stopword yaitu merubah huruf kapital dalam suatu term menjadi lowercase, kemudian tahapan berikutnya yaitu dengan memecah atau explode dari dokumen menjadi term-term yang akan disimpan dalam database term beserta jumlah frekuensinya.

Tabel 4.4. *Source Code Similarity*

```

$cek_idf = mysql_fetch_array(mysql_query("select idf from
katakunci where term like '$termnya' "));
$fidfnya=$cek_idf['idf'];

$total_bobot[$b] =(($data['freq'])*($cek_idf['idf']));
//menampilkan hasil perkalian bobot
//echo
'.'. $data['freq']. 'X'. $cek_idf['idf']. '='. $total_bobot[$b]. '</br>';
//$sql11=mysql_query("select * from katakunci where term like
'$termnya' and kategori = '1'");
//satu
$cek_name1 = mysql_fetch_array(mysql_query("select * from
katakunci where term like '$termnya' and kategori = '1'"));
$jml1=mysql_fetch_array(mysql_query("select count(id) from
katakunci where term like '$termnya' and kategori = '1'"));
$total_1[$c]=(($cek_name1['idf'])*($jml1[0]));
//dua
$cek_name2 =mysql_fetch_array(mysql_query("select * from katakunci
where term like '$termnya' and kategori = '2'"));
$jml2=mysql_fetch_array(mysql_query("select count(id) from
katakunci where term like '$termnya' and kategori = '2'"));
$total_2[$d]=(($cek_name2['idf'])*($jml2[0]));
//tiga
$cek_name3 = mysql_fetch_array(mysql_query("select * from
katakunci where term like '$termnya' and kategori = '3'"));
$jml3=mysql_fetch_array(mysql_query("select count(id) from
katakunci where term like '$termnya' and kategori = '3'"));
$total_3[$e]=(($cek_name3['idf'])*($jml3[0]));
//empat
$cek_name4 =mysql_fetch_array(mysql_query("select * from katakunci
where term like '$termnya' and kategori = '4'"));
$jml4=mysql_fetch_array(mysql_query("select count(id) from
katakunci where term like '$termnya' and kategori = '4'"));
$total_4[$e]=(($cek_name4['idf'])*($jml4[0]));
//lima
$cek_name5 = mysql_fetch_array(mysql_query("select * from
katakunci where term like '$termnya' and kategori = '5'"));
$jml5=mysql_fetch_array(mysql_query("select count(id) from
katakunci where term like '$termnya' and kategori = '5'"));
$total_5[$d]=(($cek_name5['idf'])*($jml5[0]));

if($cek_name1 >0){
    $p1++;
} if($cek_name2 >0){
    $p2++;
} if($cek_name3 >0){
    $p3++;
} if($cek_name4 >0){
    $p4++;
    } if($cek_name5 >0){
        $p5++;
} else {
}
}

```

Dalam proses similarity yaitu proses penjumlahan hasil dari perkalian term yang terdapat dalam suatu dokumen dengan term yang terdapat dalam kata kunci dari tiap kategori. Semakin besar bobotnya maka semakin besar pula tingkat kemiripan dengan kategori tersebut.

4.5 Integrasi Islam

Al-Quran telah memberikan gambaran mengenai berbuat kebaikan dan memberi manfaat kepada sesama. Para nabi, dan para sahabat nabi juga sudah memberikan pelajaran bagi umat muslim bahwa salah satu ciri tanda orang beriman adalah dengan berbuat kebaikan. Sesuai dalam firman Allah Subhanahu Wa Ta'ala

وَلِكُلِّ وِجْهَةٌ هُوَ مُوَلِّيٰهَا فَاسْتَبِقُوا الْخَيْرَاتِ ۚ أَيْنَ مَا تَكُونُوا يَأْتِ بِكُمُ اللَّهُ جَمِيعًا ۚ إِنَّ اللَّهَ عَلَىٰ كُلِّ شَيْءٍ قَدِيرٌ

Terjemahan : “Dan bagi tiap-tiap umat ada kiblatnya (sendiri) yang ia menghadap kepadanya. Maka berlomba-lombalah (dalam membuat) kebaikan. Di mana saja kamu berada pasti Allah akan mengumpulkan kamu sekalian (pada hari kiamat). Sesungguhnya Allah Maha Kuasa atas segala sesuatu.” (Qur’an Surat Al-Baqarah ayat 148)

Tafsir Jalalayn : “(Dan bagi masing-masing) maksudnya masing-masing umat (ada arah dan tujuan) maksudnya kiblat (tempat ia menghadapkan wajahnya) di waktu shalatnya. Menurut suatu qiraat bukan 'muwalliihaa' tetapi 'muwallaahaa' yang berarti majikan atau yang menguasainya, (maka berlomba-lombalah berbuat kebaikan) yakni segera menaati dan menerimanya. (Di mana saja kamu berada, pastilah Allah akan mengumpulkan kamu semua) yakni di hari kiamat, lalu dibalas-Nya amal perbuatanmu. (Sesungguhnya Allah Maha Kuasa atas segala sesuatu)”.

Dari ayat tersebut Allah Subhanahu Wata'ala telah memberikan jaminan kepada umatnya yaitu barangsiapa yang melakukan kebaikan dan belomba-lomba untuk melakukan kebaikan maka nanti pada hari akhir, semua kebaikan yang telah dilakukan akan mendapatkan balasan yang baik juga.

Dijelaskan juga dalam hadist nabi mengenai pentingnya berbuat yang bermanfaat, Rasulullah Shalallahu ‘alaihi wa Salam bersabda :

خير الناس أنفعهم للناس

Terjemahan Hadist : “*Sebaik-baik manusia adalah yang paling bermanfaat bagi manusia*” (HR. Ahmad, ath-Thabrani, ad-Daruqutni. Hadits ini dihasankan oleh al-Albani di dalam *Shahihul Jami’* nomor:3289).

Dalam hadist nabi tersebut juga dijelaskan bahwa manusia yang terbaik adalah manusia yang memberikan manfaat bagi manusia lain. Bukannya manusia yang kaya, yang pintar, atau terkenal tetapi yang terbaik adalah manusia yang memberikan manfaat terhadap sesamanya.

Penelitian mengenai sistem penentuan kontribusi tokoh politik ini juga sejalan dengan hadist nabi tersebut yaitu untuk mengetahui kontribusi yang bermanfaat bagi masyarakat di Indonesia. Tokoh politik diharapkan menjadi manusia yang terbaik yaitu manusia yang bermanfaat bagi masyarakat.

BAB V

KESIMPULAN DAN SARAN

5.1 Kesimpulan

Setelah dilakukan uji coba sistem penentuan kontribusi tokoh politik menggunakan metode *K-Nearest Neighbor* dengan *testing* menggunakan 4914 dokumen menggunakan pencarian sejumlah 30 tokoh politik menghasilkan akurasi sebesar 60% dengan *error rate* sebesar 40% sehingga dapat disimpulkan bahwa sistem ini berhasil untuk mengetahui kontribusi tokoh politik yang ada di Indonesia. tetapi perlu ditingkatkan nilai akurasinya dan memperkecil nilai *error rate*.

5.2 Saran

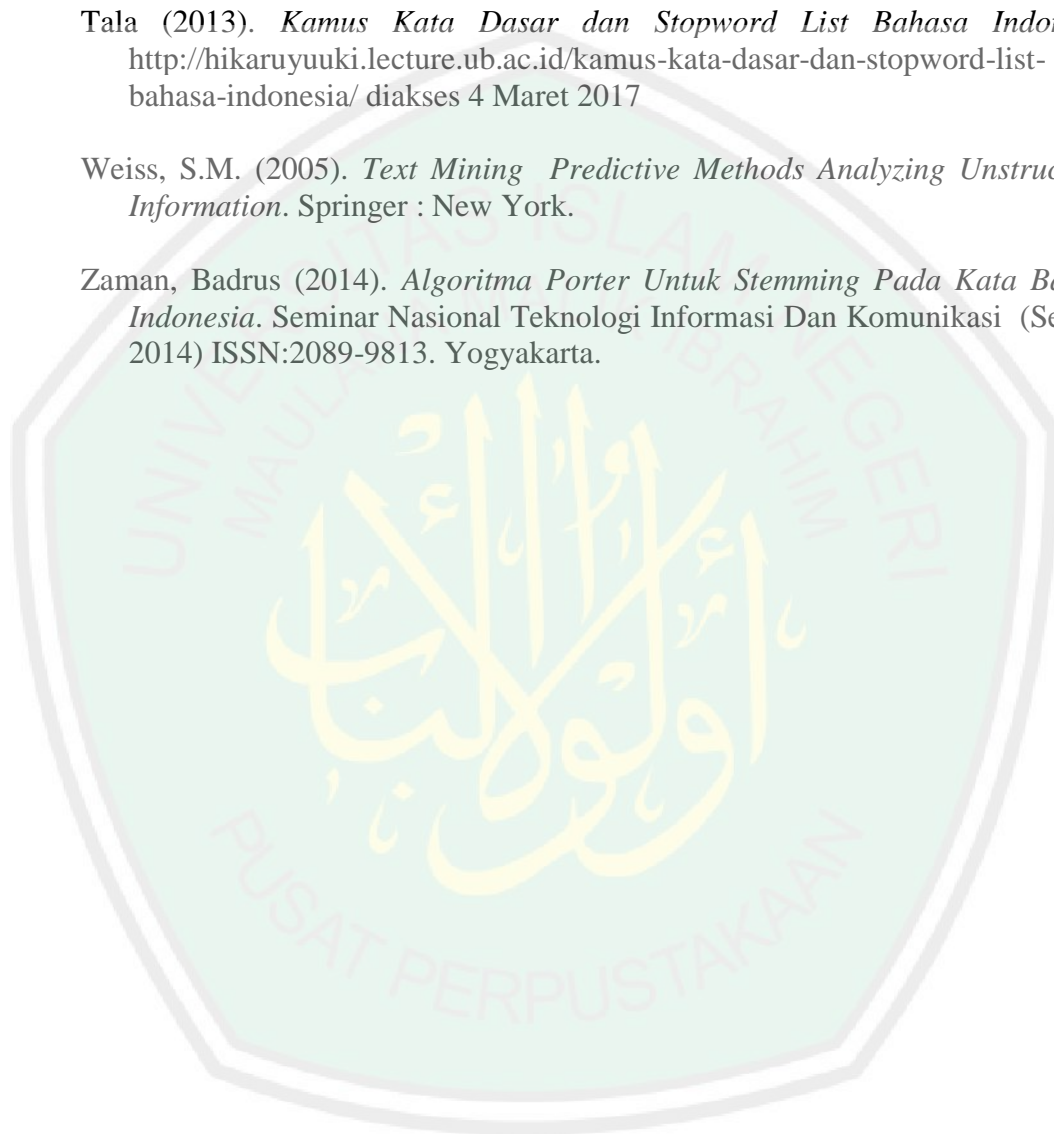
Berikut saran dari penelitian yang sudah dilakukan dan bisa dikembangkan terhadap penelitian selanjutnya, antara lain:

- a. Penelitian selanjutnya diharapkan bisa menambah jumlah prioritas menjadi lebih dari dua dan penambahan jumlah dokumen guna untuk meningkatkan nilai akurasinya.
- b. Bisa menggunakan dokumen secara *online*, dalam penelitian ini untuk pengambilan dokumen dilakukan secara manual yaitu menggunakan aplikasi *webharvy* dari media *online* untuk mendapatkan *database*.
- c. Menambahkan jumlah tokoh politik guna untuk mengetahui kontribusi tokoh politik lain yang belum terdapat dalam penelitian ini.

DAFTAR PUSTAKA

- Agusta, L. (2009). *Perbandingan Algoritma Stemming Porter Dengan Algoritma Nazief & Adriani Untuk Stemming Dokumen Teks Bahasa Indonesia*. Proceeding Konferensi Nasional Sistem Dan Informatika. Yogyakarta. Hal 196-201.
- Anggono, Riandri (2009). *Analisis Perbandingan Metode K-Nearest Neighbor dan Naive Bayes Classifier dalam Klasifikasi Teks*. Repository Universitas Atma Jaya. Yogyakarta
- Ayistiya, S (2014). *Algoritma Nearest Neighbor*. <http://ayistiya13.blogspot.co.id/2014/11/algoritma-nearest-neighbor.html> diakses 15 Maret 2017
- Bakti, Nagari (2012). *Analisis Kontribusi Pemberian Beasiswa Djarum Terhadap Peningkatan Prestasi Akademik Mahasiswa Penerima Tahun Angkatan 2010/2011 Daerah Istimewa Yogyakarta*. Repositori UNY. Yogyakarta
- Chandra, D. (2016). *Klasifikasi Berita Lokal Radar Malang Menggunakan Metode Naive Bayes Dengan Fitur N-Gram*. Journal JITIKA volume 10, No.1, Februari 2016.
- Fathan, Ahmad (2014). *Analisis Sentimen dan Klasifikasi Kategori Terhadap Tokoh Publik pada Data Twitter menggunakan Naive Bayes Classifier*. Repositori UGM. Yogyakarta
- Feldman, & Sanger (2007). *The Text Mining Handbook Advanced*. Penerbit : Cambridge University.
- Halomoan, Tri (2014). *Implementasi Modified K-Nearest Neighbor Dengan Otomatisasi Nilai K Pada Pengklasifikasian Penyakit Tanaman Kedelai*. Repositori Universitas Brawijaya. Malang
- Han, & Kamber (2006). *Data Mining Concepts and Technique*. San Francisco:Diane Cerra
- Handoko B. (2008). *Perbandingan Metode Naive Bayes Classifier (Nbc) Dengan K-Nearest Neighbor (K-Nn) Untuk Klasifikasi Dokumen*. Repositori UKDW. Yogyakarta.
- Imanuel, R (2014). *Analisis Prediksi Tingkat Pengunduran Diri Mahasiswa Dengan Metode K-Nearest Neighbor*. Repositori STIMIK AMIKOM, Yogyakarta.
- Nugraheni Y (2013). *Data Mining*. <http://yohananutgraheni.wordpress.com/2013/03/03/Data-Mining/> Diakses 9 Februari 2017.

- Romli, Asep M (2012). *Panduan Mengelola Media Online*. Nuansa Cendekia. Bandung
- Rosdiansyah D. (2014). *Analisis Sentimen Twitter Menggunakan Metode K-Nearest Neighbor Dan Pendekatan Lexicon*. Repositori UIN SUSKA. Pekanbaru
- Tala (2013). *Kamus Kata Dasar dan Stopword List Bahasa Indonesia*. <http://hikaruyuuki.lecture.ub.ac.id/kamus-kata-dasar-dan-stopword-list-bahasa-indonesia/> diakses 4 Maret 2017
- Weiss, S.M. (2005). *Text Mining Predictive Methods Analyzing Unstructured Information*. Springer : New York.
- Zaman, Badrus (2014). *Algoritma Porter Untuk Stemming Pada Kata Bahasa Indonesia*. Seminar Nasional Teknologi Informasi Dan Komunikasi (Sentika 2014) ISSN:2089-9813. Yogyakarta.



LAMPIRAN

Tabel Daftar Perusahaan Pers Indonesia Terverifikasi Administrasi dan Faktual

No	Nama Media	Jenis	Website	Provinsi	Alamat
1	Kompas TV	Siaran	www.kompas.tv	Jakarta	Jln.Palmerah Selatan No.1 Kel.Gelora Kec.Tanah Abang, Jakarta Pusat
2	Radar Pekanbaru	Cetak	www.radarpekanbaru.com	Riau	Jl. Arifin Ahmad Blok B No.8 Pekanbaru
3	telusur.co.id	Siber	www.telusur.co.id	Jakarta	Komplek Kalibata Indah Jl. Durian Blok O No.1 Rawajati, Jakarta Selatan 12750
4	Ummi	Cetak	www.ummi-online.com	Jakarta	Jl. Mede No.42a Utan Kayu Utara, Jakarta
5	Mercusuar	Cetak	www.harianmercusuar.com	Sulawesi Tengah	Jl. Yos Sudarso No. 33
6	Gatra	Cetak	www.gatra.com	Jakarta	Gedung Gatra Jl. Kalibata Timur IV/15, Jakarta Selatan. 12740
7	Bali TV	Siaran		Bali	Jl. Kebo Iwa No. 63 A
8	Rakyat Bengkulu	Cetak	www.harianrakyatbengkulu.com	Bengkulu	Jl. Natadirja No. 69 Km 6,5, Bengkulu
9	Orbit	Cetak		Sumatera Utara	Jl. Asmara Gedung Bumi Jurnalis Komplek Bumi Asri Blok C No. 61 - 62 Medan
10	LKBN ANTARA	Siaran		Jakarta	Wisma Antara Lt.19 Jl. Merdeka Selatan No. 17 Jakarta Pusat
11	Padang Ekspres	Cetak	www.padangekspres.co.id	Sumatera Barat	Jl. Adi Negoro No. 17a, Lubuk Buaya, Padang
12	RRI	Siaran		Jakarta	Jl. Medan Merdeka Barat 4-5, RT.2/RW.3 Jakarta Pusat
13	Palembang Ekspres	Cetak	www.palpres.com	Sumatera Selatan	Gedung Graha Pena Palembang, Jl. Kol H Barlian No. 773, Palembang, 30152
14	batamclick.com	Siber	www.batamclick.com	Kepulauan Riau	Kawasan Kota Mandiri Mall Botania 2 Blok E5 Kecamatan Batam Kota " Kota Batam
15	Info Bekasi	Cetak	www.infogadinggroup.com/www.infonitas.com	Jawa Barat	Komp. Gading Bukit Indah Blok K-27 Jl Bukit Gading Raya Jakarta Utara 14240
16	Metro Banjar	Cetak		Kalimantan Selatan	Gedung Djok Mentaya, Jl. AS Mustaffa No. 16, Banjarmasin 701111
17	Siwalima	Cetak	www.siwalimanews.com	Maluku	Jl. Diponegoro No.20, Ambon. 97124
18	Wawasan	Cetak		Jawa Tengah	Jl. Kawi No.20 Semarang
19	Tribun Batam	Cetak	www.batam.tribunnews.com	Kepulauan Riau	Komplek MCP Jl. Kerapu Batu Ampar, Batam Kepulauan Riau
20	Tribun Pekanbaru	Cetak	www.pekanbaru.tribunnews.com	Riau	Jl. Haji Imam Munandar/Harapan Raya No. 383, Kec. Bukit Raya Kel. Tangkerang Labui RT. 01, RW. 04 Pekanbaru 28281
21	obsessionnews.com	Siber	www.obsessionnews.com	Jakarta	Jl. Kedondong No. 161 Jagakarsa, Jaksel
22	News24x.com	Siber	www.news24x.com	Riau	Perkantoran Grand Sudirman Blok D No.19 Jl. Datuk Setia Maharaja, Kec. Tangkerang Selatan, Kel. Bukit Raya, Pekanbaru, Riau
23	Kumparan.com	Siber	kumparan.com	Jakarta	Jl Jati Murni no.1A, Jati Padang, Pasar Minggu, Jakarta Selatan
24	Media Indonesia	Cetak	www.mediaindonesia.com	Jakarta	Kompleks Delta Kedoya, Jl. Pilar Raya Kav. A-D Kedoya Selatan, Kebon Jeruk, Jakarta Barat 11520
25	BUMN Track	Cetak	www.bumntrack.com	Jakarta	Ged. Sarinah Lt. 13, Jl. MH Thamrin No. 11, Jakarta Pusat, 10350
26	Radar Banten	Cetak	www.radarbanten.co.id	Banten	Graha Pena Radar Banten Jl. Kolonel H TB. Suwandi, Lingkar Selatan, Serang, Banten
27	Radar Kepahiang	Cetak		Bengkulu	Jl. Lintas Kapahiang - Curup Desa Karang Anyar, Kec. Kapahiang, Bengkulu
28	Sniper	Cetak		Sumatera Selatan	Jl. Jenderal A. Yani 940 ULU RT 2 No.36, Palembang
29	Info Kebayoran	Cetak		Jakarta	Jl. Raya Serpong, Ruko Sutera Niaga III, Blok D, No. 10, Serpong, Tangerang Selatan
30	Radar Manado	Cetak		Sulawesi Utara	Gedung Graha Pena Manado Post Lt. 3 Jl. Babepalar No.62 Wanea, Manado
31	Tribun Jogja	Cetak	www.tribunjogja.com	Daerah Istimewa Jogjakarta	Jl. Jenderal Sudirman No. 52 RT. 03/01 Kel. Kotabaru Kec. Gondokusuman DIY 55224
32	iNEWS TV	Siaran	www.mncgroup.com	Jakarta	Jl. Kebon Sirih No. 21-27 Jakarta Pusat 10340

33	Pronews FM	Siaran		Sumatera Barat	
34	Koran Sindo Batam	Cetak		Kepulauan Riau	Jl. Tenggiri Tjg Sengkuang Batu Ampar Batam Kepri
35	Rakyat Cirebon	Cetak		Jawa Barat	Jl. Perjuangan No. 99 Cirebon
36	Tabengan	Cetak	www.hariantabengan.com	Kalimantan Tengah	Jl. Imam Bonjol 19, Palangkaraya, Kalimantan Tengah 73118
37	Bisnis Papua	Cetak	www.bisnispapua.net	Papua	Gedung Kadin Papua Lt. 1 Ruko Pasifik Permai Jayapura
38	Suara Media Nasional	Cetak	www.suaramedianasional.co.id	Jawa Timur	Jl. Durian, Ruko PG Pesantren, Kota Kediri Kantor pusat: Dsn. Temboro Desa Plaosan Kec. Wates, Kediri 64174
39	batamnews.co.id	Siber	www.batamnews.co.id	Kepulauan Riau	Jl. Timor Blok D No. 15, Bengkong Dalam, Batam
40	Pekanbaru MX	Cetak	www.pekanbaru.mx.net	Riau	Gedung Graha Pena Pekanbaru lantai 6, Jl. H.R. Soebrantas KM 10,5 Tampan
41	Men's Obsession	Cetak		Jakarta	Jl. Kedondong No. 161 Jagakarsa, Jaksel
42	Radar Pet Petulai	Cetak		Bengkulu	Jl. Jendral Sudirman No. 14 Kel. Tempel Rejo, Kec. Curup Selatan, Kab. Rejang Lebong
43	Radar Sulbar	Cetak	www.radarsulbar.com	Sulawesi Barat	Jl. Jendral Sudirman No. 50, Mamuju, Sulawesi Barat
44	Metrojambi.com	Siber	www.metrojambi.com	Jambi	Jl. A. Thalib No. 7-8 RT. 05 Pematang Sultur
45	Koran Sindo	Cetak	www.koran-sindo.com	Jakarta	Gedung Sindo, Jl. KH Wahid Hasyim No.38 Kebon Sirih, Menteng, Jakarta Pusat 10340
46	Pojoksatu.id	Siber	pojoksatu.id	Jawa Barat	Graha Pena Bogor Lantai 4 Jl KH Abdullah bin Muhammad Nuh No 30 Kota Bogor, Jawa Barat.
47	Tangerangonline.id	Siber	http://tangerangonline.id	Banten	Jl. Yapen Raya Blok QD No. 3 Sektor 14.6, Nusa Loka Bsd, Serpong Tangerang Selatan
48	Andalas	Cetak	www.harianandalas.com	Sumatera Utara	Jl. T. Amir Hamzah Ruko, Kompleks Griya Riatir Indah No.182-186 Medan
49	gatra.com	Siber	www.gatra.com	Jakarta	Gedung Gatra Jl. Kalibata Timur IV/15, Jakarta Selatan. 12740
50	manadone.ws.co.id	Siber	www.manadone.ws.co.id	Sulawesi Utara	Jl. Wakeke no. 11 Manado
51	Djaka Lodang	Cetak	www.djakalodang.co.id	Daerah Istimewa Jogjakarta	Jalan Patehan Tengah 29, Yogyakarta. 55133
52	Metro TV	Siaran		Jakarta	Jl Pilar Mas Raya Kav. A-D, Kedoya - Kebon Jeruk Jakarta 11520
53	Majalah Arung Media	Cetak	www.arungmedia.com	Sumatera Selatan	Jl. RA Abusamah/Manggis I No. 2789, Kel. Sukajaya, Kec. Sukarame, Palembang 30151
54	SBO TV	Siaran	sbo.co.id	Jawa Timur	Jl. A. Yani 88 Gedung Graha Pena Lt. 21 Ketintang, Gayungan Surabaya
55	Radar Tanggamus	Cetak		Lampung	Jl. Soekarno-Hatta No.09 Pekon Terbaya, Kec. Kota Agung Kab.Tanggamus Lampung Selatan 35684
56	Radar Semarang	Cetak	www.radarsemarang.com	Jawa Tengah	Jl.Veteran 55, Semarang 502315
57	Kaltim Post	Cetak	www.kaltimpos.co.id	Kalimantan Timur	Gedung Biru, Jl. Soekarno Hatta Km. 3,5 No. 64, Balikpapan 76126.
58	tabloidjubi.com	Siber	www.tabloidjubi.com	Papua	Jl. Sakura Gg. Jati No.A5 Perumnas II Waena, Kota Jayapura
59	Radio Suara Surabaya	Siaran		Jawa Timur	Jl. Wonokitri Besar 40 C Surabaya
60	goriau.com	Siber	www.goriau.com	Riau	Surya Kartama Agung A-4, Jl. Kartama Kel. Maharatu, Kec. Marpoyan Damai, Pekanbaru, Riau
61	tribunnews.com	Siber	www.tribunnews.com	Jakarta	Gedung Group of Regional Newspaper Kompas Gramedia, Jl Palmerah Selatan No 3, Jakarta Pusat Telp. 021-5359525-23
62	Sinar Tani	Cetak	www.tabloidsinartani.com	Jakarta	Jl. Harso RM 3 Ragunan, Jakarta 12550
63	Celebes TV	Siaran		Sulawesi Selatan	Jln. Jend.Sudirman No. 5 Makassar " Sulawesi Selatan
64	jamberita.com	Siber	www.jamberita.com	Jambi	Jl. Perum Mutiara Mayang Blok I No 4 RT 34 Kel. Mayang Mangurai Kec. Aalam Barajo
65	Suara Pembaruan	Cetak	www.suarapembaruan.com	Jakarta	Berita Satu Plaza, Lantai 11 Jl. Jend. Gatot Subroto Kav 35-36 Jakarta 12950
66	Bisnis Bali	Cetak		Bali	Gedung Pers Bali Ketut Nadha Jl. Kebo Iwa 63A, Denpasar 80116
67	Radar Bengkulu	Cetak		Bengkulu	Jl. P. Tendean No 3F Kota Bengkulu, 38224
68	Waspada	Cetak		Sumatera Utara	Jl. Brigjen Katamso No. 01

69	tempo.co	Siber	www.tempo.co	Jakarta	Jl. Palmerah Barat No.8 Kel. Grogol Utara, Kec. Kebayoran Lama, Jaksel
70	Manado Post	Cetak	www.manadopostonline.com	Sulawesi Utara	Jl. Babe Palar No. 62 Kelurahan Wanea Manado, Gedung Graha Pena
71	Tangerang Ekspres	Cetak			Graha Pena Tangerang, Moderngolf Shophouse No. 8, Modernland, Tangerang Banten
72	viva.co.id	Siber	www.viva.co.id	Jakarta	Kawasan Industri Pulogadung, Gedung tvOne Jl. Rawa Terate II No.2 Jakarta Timur 13260
73	Sriwijaya Post	Cetak	www.palembang.tribunnews.com	Sumatera Selatan	Jl. Alamsyah Ratuprawiranegara No. 120, Kel Bukit Lama Kec. Ilir Barat 1, Palembang
74	Kabar Cirebon	Cetak	www.kabar_cirebon.com	Jawa Barat	Jl. RA Kartini No. 7 Cirebon, 45123
75	Rakyat Kalbar	Cetak	www.equator-news.com	Kalimantan Barat	Ged. Graha Pena Kalbar Lt. 2&3, Jl. Arteri Supadio (Ayani 2) Km 3,5 Sungai Raya, Kalbar, 78391
76	Lintaslampung.com	Siber	www.lintaslampung.com	Lampung	Gloria Persada Blok H No.10 Rajabasa Raya, Rajbasa, Bandar Lampung
77	Solo Pos	Cetak	www.solopos.com	Jawa Tengah	Griya Solopos, Jl. Adi Sucipto No. 90, Solo 57145
78	Samarinda Pos	Cetak		Kalimantan Timur	Jl. Soekarno Hatta Km 3.5
79	Pos Metro Mandau	Cetak		Riau	Jl. KH. Ahmad Dahlan No. 14C Pekanbaru
80	Kabar Priangan	Cetak	www.kabar-priangan.com	Jawa Barat	Jl. Re Martadinata No. 215 A Kec. Indihiang, Kota Tasikmalaya
81	Pontianak Post	Cetak	www.pontianakpost.com	Kalimantan Barat	Gedung Graha Pena Pontianak Post Jl. Gajah Mada No. 2-4 Pontianak 78121
82	Faktual	Cetak		Lampung	Jl. Sisingamaraja No.44 Bandar Lampung
83	TA TV	Siaran		Jawa Tengah	Jln. Brigjend Katamso 173 Mojosongo, Solo
84	Kaltara Pos	Cetak		Kalimantan Utara	Jl. Mulawarman No.32 Rt 1 Gedung Sileur Lt. 4 Tarakan Kaltara
85	Pos Metro Indragiri	Cetak		Riau	Jl. Hr. Soebrantas km. 105, Panam, Pekanbaru
86	rmol.co	Siber	www.rmol.co	Jakarta	Graha Pena, Lt. 9, Jalan Raya Kebayoran Lama No. 12, Jakarta Selatan
87	riaumandiri.co	Siber	www.riaumandiri.co	Riau	Gedung Riau Pers Jl. Tuanku Tambusai No.7 Pekanbaru
88	IPOTNEWS.COM	Siber	http://ipotnews.com	Jakarta	Wisma GKBI 7/F Suite 718, Jl. Jend. Sudirman No. 28, Jakarta 1021
89	Kompas	Cetak	www.kompasprint.com	Jakarta	Jl. Palmerah Selatan 26-28, Jakarta 10270
90	Radar Sulteng	Cetak		Sulawesi Tengah	Gedung Graha Pena, Jl. Yos Sudarso No.9, Palu
91	Kontan	Cetak	www.kontan.co.id	Jakarta	Gedung Kontan Jl. Kebayoran Lama No. 1119, Jakarta 12210
92	balipuspanews.com	Siber	www.balipuspanews.com	Bali	Perumahan Bumi Sasih Asri Blok I No.7, Banjar Sasih, Batubulan, Sukawati - Gianyar, Bali
93	Bengkulu News	Cetak		Bengkulu	Jl. Kapten Tendean KM 6.5 Komplek PU Pengairan No. 13, Kota Bengkulu
94	LWI POS	Cetak		Sumatera Utara	jl Dr wahidin gg sawah no 28 Kel Rejo Kec Binjai Timur Kota Binjai Sumatera Utara
95	RCTI	Siaran	www.mncgroup.com	Jakarta	Jl. Kebon Sirih No. 21-27 Jakarta Pusat 10340 Indonesia
96	Singgalang	Cetak	www.hariansinggalang.co.id	Sumatera Barat	Jl. Veteran No. 17, Padang - 25116
97	Berita Satu News Channel	Siaran		Jakarta	Beritasatu Plasa Lantai 9, Jl. Jend Gatot Subroto Kav 35-36
98	News Hunter	Cetak	www.newshunter.com	Sumatera Selatan	Jl. Sapta Marga Komp. Cita Dama II Blok M No. 5 RT. 43, RW. 09 Palembang
99	Tanjungpinang Pos	Cetak	Tanjungpinangpos.id	Kepulauan Riau	Jl. Panjanitan Komp. Pinlang Mas Blok Mayang No.10-11 Tanjungpinang, Kepulauan Riau