**KLASIFIKASI BERITA HOAX MENGGUNAKAN METODE**

**APRIORI**

**LAPORAN HASIL**

OLEH

ALAM ASRORUL HAQ

150535603542



**UNIVERSITAS NEGERI MALANG**

**FAKULTAS TEKNIK**

**PROGRAM STUDI S1 TEKNIK INFORMATIKA**

**JULI 2019**

# **LEMBAR PENGESAHAN**

Hasil skripsi dengan judul Klasifikasi Berita *Hoax* Menggunakan Metode Apriori oleh Alam Asrorul Haq, telah diperiksa dan disetujui di depan pembimbing pada tanggal 1 3 Juli 2019

Pembimbing I

I Made Wirawan, S.T.,S.S.T., M.T.

NIP. 196907171998021001

Pembimbing II

Agusta Rakhmat Taufani, S.T., M.T.

NIP. 198108172014041001

**ABSTRAK**

Asrorul Haq, Alam. Klasifikasi Berita *Hoax* Menggunakan Metode Apriori, Jurusan Teknik Elektro, Fakultas Teknik, Universitas Negeri Malang. Pembimbing: (I) I Made Wirawan, S.T., S.S.T., M.T. (II) Agusta Rakhmat Taufani, S.T., M.T.

**Kata kunci :** Klasifikasi, Apriori*, fake news, fact,* Tf-Idf

Berita adalah kumpulan dari beberapa informasi yang dirangkum dan disebarkan. Berita memiliki 2 buah kategori ,berita bohong/ *fake* dan berita benar /*fact*. Berita *fake* memberikan dampak yang sangat buruk bagi semua kalangan yang menerimanya. Hal tersebut dapat menimbulkan perpecahan,pertikaian dan perubahan sikap pada seseorang. Salah satu contohnya yakni berita yang disebar di sosial media guna mengadu domba 2 buah kalangan. Oleh karena itu, pada penelitian ini akan dilakukan klasifikasi guna mengetahui apakah berita yang disajikan termasuk kategori berita *fake*  atau tidak dengan menggunakan *machine learning.*

Data yang didapat yakni dikumpulkan dari data yang terdapat di situs *turnbackhoax.id*. Penelitian ini menggunakan metode *Tf-Idf* dan Apriori untuk pembobotan kata dalam berita yang selanjutnya akan dicari nilai kata yang terindikasi berita *hoax* yang bertujuan untuk membuat library kumpulan kata *hoax*. Library kumpulan kata hoax di cari dengan menggunakan 100 buah data latih. Untuk proses pengujian, dilakukan dengan membandingkan *library* kata dan kata pada sebuah berita sehingga didapat nilai ke aslian dari berita.

Proses pengujian berita menggunakan 14 buah berita benar dan 16 buah berita salah atau *hoax.* Hasil penelitian dengan metode apriori ini yakni sebuah nilai / bobot berita terindikasi *hoax* atau tidak*.* Dari data uji yang telah digunakan, didapat nilai akurasi sebesar 83,3% untuk pengujian baik berita benar maupun berita salah. Dengan adanya penelitian ini diharapkan dapat meminimalisir penyebaran berita *hoax* dan mendeteksi berita-berita *hoax* sejak dini.

# **KATA PENGANTAR**

Puji syukur penulis panjatkan kepada Allah SWT yang telah melimpahkan rahmat dan hidayah-Nya, sehingga penulis dapat menyelesaikan laporan skripsi yang berjudul “Klasifikasi Berita *Hoax* Menggunakan Metode Apriori”. Penulis menyadari bahwa dalam penyusunan laporan skripsi ini melibatkan bantuan dan dorongan dari berbagai pihak. Oleh karena itu, penulis menyampaikan ucapan terima kasih kepada yang terhormat.

1. Dr. Marji, M.Kes. selaku Dekan Fakultas Teknik
2. Aji Prasetya Wibawa, S.T., M.M.T., Ph.D. selaku Ketua Jurusan Teknik Elektro Fakultas Teknik, Universitas Negeri Malang
3. Ilham Ari Elbaith Zaeni, S.T., M.T., Ph.D. selaku Kaprodi Teknik Informatika beserta seluruh staf Jurusan TE dan FT
4. I Made Wirawan, S.T., S.S.T., M.T., dosen pembimbing I yang telah bersedia meluangkan waktu, tenaga, dan pemikiran guna memberikan bimbingan dan saran kepada penulis, sehingga laporan skripsi ini dapat diselesaikan dengan baik.
5. Agusta Rakhmat Taufani, S.T., M.T., dosen pembimbing II yang telah bersedia meluangkan waktu, tenaga, dan pemikiran guna memberikan bimbingan dan saran kepada penulis, sehingga laporan skripsi ini dapat diselesaikan dengan baik..
6. Orang tua penulis, Bapak Misbah dan Ibu Yayuk yang telah memberikan motivasi dan doa yang tak henti-hentinya kepada penulis. Semoga kalian diberikan kesehatan oleh Allah SWT.
7. Teman-temanku Teknik Informatika offering A angkatan 2015 yang selalu memberikan dukungan, semangat, dan pengalaman. Semoga silaturahmi kita tetap terjalin.

Penulis juga mengucapkan terima kasih kepada seluruh pihak yang tidak bisa penulis sebutkan satu per satu, yang telah membantu dan berkontribusi dalam penyelesaian skripsi ini. Ucapan terima kasih memang tidak cukup untuk membalas kebaikan kalian, semoga Allah SWT membalasnya dengan kebaikan yang berlipat.

Malang, 4 Juli 2019

Alam Asrorul Haq

# **DAFTAR ISI**

Halaman

[**LEMBAR PENGESAHAN** i](#_Toc17096387)

[**ABSTRAK** ii](#_Toc17096388)

[**KATA PENGANTAR** iii](#_Toc17096389)

[**DAFTAR ISI** v](#_Toc17096390)

[**DAFTAR TABEL** vii](#_Toc17096391)

[**DAFTAR GAMBAR** viii](#_Toc17096392)

[**BAB I** 8](#_Toc17096393)

[**A.** **Latar Belakang Masalah** 2](#_Toc17096394)

[**B.** **Rumusan Masalah** 4](#_Toc17096395)

[**C.** **Batasan Masalah** 4](#_Toc17096396)

[**D.** **Tujuan Penelitian** 4](#_Toc17096397)

[**E.** **Manfaat Penelitian** 5](#_Toc17096398)

[**F.** **Definisi Operasional** 5](#_Toc17096399)

[**BAB II** 5](#_Toc17096400)

[**A.** **Dokumen Berita untuk Pengklasifikasian Data** 5](#_Toc17096401)

[**B.**  **Metode *Waterfall* untuk Pembuatan Aplikasi Klasifikasi Berita *Hoax*** 5](#_Toc17096402)

[***C.*** ***Natural Language* *Processing*** 6](#_Toc17096403)

[1. *Tokenizing* 6](#_Toc17096404)

[*2.* *Stopwords Removal* 6](#_Toc17096405)

[3. *Stemming* 6](#_Toc17096406)

[**D.** **Teknik *TF-IDF*** 7](#_Toc17096407)

[**E.**  ***Apriori* *Method*** 8](#_Toc17096408)

[**F.** **Bahasa Pemrograman *Python*** 9](#_Toc17096409)

[**G.** ***Database* MySQL** 9](#_Toc17096410)

[**H.** **Pengukuran Performa** 10](#_Toc17096411)

[**I.** **Penelitian yang Relevan** 12](#_Toc17096412)

[**J.** **Kerangka Pemikiran** 13](#_Toc17096413)

[1. Proses input data latih : 13](#_Toc17096414)

[2. Proses *input* data uji : 15](#_Toc17096415)

[**BAB III** 17](#_Toc17096416)

[**A. Desain Penelitian** 17](#_Toc17096417)

[**1.** **Identifikasi Berita *Hoax*** 18](#_Toc17096418)

[**2.** **Pengumpulan Data** 18](#_Toc17096419)

[**3.** **Pre Prossesing Menggunakan NLP (Neural Language Processing)** 18](#_Toc17096420)

[**4.** ***Processing* menggunakan Tf-Idf** 20](#_Toc17096421)

[**5.** ***Processing* menggunakan Apriori** 21](#_Toc17096422)

[**6.** **Eksperimen dan Pengujian** 22](#_Toc17096423)

[**7.**  **Evaluasi dan Validasi Hasil** 24](#_Toc17096424)

[**BAB IV** 25](#_Toc17096425)

[**A.** **Desain Aplikasi** 25](#_Toc17096426)

[**1.** **Usecase Diagram** 25](#_Toc17096427)

[**2.** ***Diagram Konteks*** 26](#_Toc17096428)

[**B.** **Penulisan Code Program** 31](#_Toc17096429)

[1. *Preprocessing* Dokumen 31](#_Toc17096430)

[2. Proses *Input* Data Latih 32](#_Toc17096431)

[3. Proses hitung Tf-Idf dan Apriori 32](#_Toc17096432)

[***4.*** Proses *Create Library Excel* 33](#_Toc17096433)

[5. Proses Input data Uji dan Pembandingan dengan Data Latih 34](#_Toc17096434)

[**C.** **Antar Muka Aplikasi** 34](#_Toc17096435)

[1. Halaman Utama 35](#_Toc17096436)

[2. Halaman Input Dataset 36](#_Toc17096437)

[**D.** **Hasil dan Pengujian *System*** 37](#_Toc17096438)

[1. Pencarian Nilai Batas 37](#_Toc17096439)

[2. Proses Uji Hasil Data Uji 37](#_Toc17096440)

[**E.**  **Pembahasan** 41](#_Toc17096441)

[**BAB V** 42](#_Toc17096442)

[**A.** **Kesimpulan** 42](#_Toc17096443)

[**B.** **Saran** 42](#_Toc17096444)

[**DAFTAR PUSTAKA** 44](#_Toc17096445)

# **DAFTAR TABEL**

Tabel Halaman

[2. 1 Precision and Recall………………………………………………………………………………..……… ……11](#_Toc13852424)  [4. 1 Usecase Deskripsi 26](#_Toc13852429)

[4. 2 Tabel Keterangan 40](#_Toc13852430)

[4. 3 Nilai dari Komponen Perhitungan dengan Batas 0.0533 41](#_Toc13852431)

[4. 4 Nilai dari Komponen Perhitungan dengan Batas 0.05 41](#_Toc13852432)

[4. 5 Nilai dari Komponen Perhitungan dengan Batas 0.0561 42](#_Toc13852433)

# **DAFTAR GAMBAR**

Gambar Halaman

[2. 1 Proses Input Data Latih 14](#_Toc13854104)

[2. 2 Proses Input Data Uji 15](#_Toc13854105)

[3. 1 Metode yang Diusulkan 17](#_Toc13854106)

[3. 2 Flowchart Stoward Removal 19](#_Toc13854107)

[3. 3 Flowchart Tf-Idf 20](#_Toc13854108)

[3. 4 Flowchart Apriori 21](#_Toc13854109)

[3. 5 Data Latih 23](#_Toc13854110)

[3. 6 Data Uji 23](#_Toc13854111)

[3. 7 Data Pencarian Limit 24](#_Toc13854112)

[4. 1 Usecase Diagram 25](#_Toc13854116)

[4. 2 Diagram Konteks 27](#_Toc13854117)

[4. 3 Aliran Data 28](#_Toc13854118)

[4. 4 Input Dataset 29](#_Toc13854119)

[4. 5 Aliran Data User 30](#_Toc13854120)

[4. 6 Penyimpanan Hasil Proses dalam Bentuk Excel 31](#_Toc13854121)

[4. 7 Source Code Preprocessing 32](#_Toc13854122)

[4. 8 Source Code Input Data Latih 33](#_Toc13854123)

[4. 9 Source Code proses hitung 34](#_Toc13854124)

[4. 10 Source Code Input Library Into Excel 35](#_Toc13854125)

[4. 11 Proses Pencarian Hasil Input Data Uji 35](#_Toc13854126)

[4. 12 Halaman Utama 36](#_Toc13854127)

[4. 13 Hasil Berita Tidak Hoax 37](#_Toc13854128)

[4. 14 Hasil Berita Terindikasi Hoax 37](#_Toc13854129)

[4. 15 Input Dataset Berita 38](#_Toc13854130)

[4. 16 Data Uji Nilai Batas 39](#_Toc13854131)

[4. 17 Data Uji Berita Hoax 39](#_Toc13854132)

[4. 18 Data Uji Berita Real 40](#_Toc13854133)

# **BAB I**

**PENDAHULUAN**

## **A. Latar Belakang Masalah**

Informasi adalah sebuah pesan yang disampaikan dari satu orang kepada orang lain baik melalui media maupun dari mulut ke mulut.Selain itu, informasi dapat diartikan dengan hasil dari pengolahan sebuah atau beberapa data (Darmawan, 2012). Tanpa adanya sebuah informasi seseorang tidak dapat saling mengetahui satu sama lain dan tidak dapat belajar dari pengalaman orang lain. Informasi juga sangat bermanfaat untuk kehidupan bersosial maupun personal. Kumpulan dari beberapa informasi baru yang tersusun dan disajikan dalam bentuk media cetak, siaran ataupun internet disebut berita. Berita juga dapat menjadi sebuah acuan informasi untuk di konsumsi khalayak umum sehingga apapun informasi yang sedang atau masih baru terjadi akan ditampilkan sehingga orang dapat mengetahui informasi tersebut dengan mudah.

Dalam berkomunikasi, informasi menjadi suatu hal yang penting untuk disampaikan. Dengan perkembangan teknologi yang begitu pesat informasi menjadi salah satu konsumsi masyarakat.Akibatnya, batas- batas ruang dan waktu menjadi semakin kabur (Ardoni, 2015). Seiring berkembangnya teknologi penyebaran informasi menjadi sangat mudah. Salah satunya yakni penyebaran informasi melalui berita baik melalui media social, situs berita *online* maupun situs-situs penyedia layanan informasi lain.Jejaring social seakan memperkuat kedudukan internet sebagai *new* *media communication*,dimana jarak seakan tidak lagi terlihat(Kesuma, 2016). Sangat tingginya pengguna internet juga mempengaruhi penyebaran berita melalui media online.Pengguna internet di Indonesia mencapai 38 juta lebih dengan penetrasi mencapai 15 persen pada tahun 2014 (Lestari, 2017). Adanya berita dapat berdampak positif maupun berdampak negative bagi pembacanya. Sehingga pembaca dituntut untuk bijak menyikapi berita yang disebar oleh penyedia layanan berita. Banyak sekali persoalan yang

timbul akibat penyalahgunaan berita sebagai sarana kampanye, politik dan lain sebagainya.

Pertukaran sebuah informasi sangatlah penting bagi kehidupan bersosial. Akan tetapi ada beberapa oknum yang menyalahgunakan informasi atau menyelewengkan informasi guna memenuhi kepentingan pribadi sehingga dapat merugikan beberapa pihak terkait khususnya dalam sebuah berita. Pembuatan berita-berita palsu oleh oknum-oknum terkait biasa disebut dengan berita *hoax* / berita yang masih diragukan fakta dan status aslinya. Berita hoax biasa didasari dari dendam, dengki atau pencarian sensasi oleh si pembuat kepada yang dituju. Seperti berita pada 5 februari 2019 tentang es krim magnum mengandung babi yang disebarkan melalui jejaring social facebook. Berita ini tidak terbukti kebenaranya sehingga disimpulkan sebagai berita hoax. Tujuan dari pembuatan berita palsu ini yakni dikarenakan adanya dendam kepada seseorang ataupun hanya untuk mencari sensasi agar mendapat view dan follower yang cukup banyak. Selain itu media sosial juga memiliki fungsi positif yakni memberikan informasi terkini tentang peristiwa yang terjadi pada masyarakat (Aulia., 2018).

Media social merupakan sarana yang efektif untuk menyampaikan suatu informasi. Dengan semakin cepatnya pertukaran informasi, semakin cepat pula penyebaran berita-beriat *hoax* yang dinilai sangat merugikan dikalangan masyarakat melalui media social (Rahadi dkk, 2017). Berita hoax juga dapat memicu perpecahan sehingga sangat berbahaya dikonsumsi oleh masyarakat umum. Dengan begitu masyarakat dituntut untuk memilah dan memilih berita serta memastikan keaslian berita terlebih dahulu sebelum mencerna mentah-mentah isi dari berita.Hoax berkembang karenda didasari beberapa faktor meliputi : Motif politik kekuasaan, Kurangnya kesadaran social masyarakat , penyalahgunaan pengaruh tokoh-tokoh guna mempengaruhi pengikutnya (Septanto, 2017). Dalam kasus berita hoax yang ada di Indonesia ini sangat beragam. Semakin banyak pula penyebaran berita hoax dan segala ujaran kebencian guna kepentingan politik yang kian masif jelang Pilpres 2019. Kementerian Sekretariat Negara menyebut penyebaran berita hoaks menjelang Pemilu merupakan satu fenomena yang timbul di tengah masyarakat. Ini berpotensi menciptakan disintegrasi dan memecah belah bangsa Indonesia.(Syafirdi, 2019). Dalam kasus ini, terbukti bahwa berita hoax sangat signifikan berperan dalam pemecah belah masyarakat. Sehingga perlu adanya solusi dan penindak lanjutan oknum-oknum pembuatan berita hoax. Dengan adanya penelitian *Klasifikasi Berita Hoax Menggunakan Metode Apriori* diharapkan dapat meminimalisir perpecahan masyarakat akibat berita hoax.

## **B. Rumusan Masalah**

Berdasarkan latar belakang yang telah dipaparkan, didapat rumusan masalah sebagai berikut :

1. Perlunya identifikasi berita *hoax* sehingga dapat menyaring antara berita real dan berita terindikasi *hoax*
2. Seberapa besar tingkat akurasi yang didapat untuk pengklasifikasian berita *hoax* dengan menggunakan metode *apriori*

## **C. Batasan Masalah**

Agar penelitian ini lebih terarah dan fokus, maka diberikan Batasan- batasan pada penelitian ini sebagai berikut :

1. Sistem yang dibuat hanya dapat menerima inputan berupa string/text

2. Program yang dibuat hanya dapat menerima *input* dan menambahkan *input* kedalam *database*.

3. Tempat penyimpanan *dictionary* / *library* kata berupa .xlsx

4. Metode Apriori yang dipakai hanya rumus pencarian nilai support

5. Masukkan berupa text

## **D. Tujuan Penelitian**

Adapun tujuan dari dibuatnya tugas akhir ini adalah untuk :

1. Identifikasi berita sehingga berita dapat digolongkan menjadi berita *hoax* dan berita *real* dengan menggunakan metode *apriori.*
2. Mencari tingkat akurasi *system* dalam pencarian berita *hoax* dan *real*.

## **E. Manfaat Penelitian**

Manfaat dari penelitian ini adalah :

1. Bagi lembaga penyedia berita dan pembaca sebagai filter berita yang dapat dipercaya dan

kurang dapat dipercaya.

2. Bagi peneliti, sebagai pembangan penggunaan algoritma *Apriori* khususnya dalam pengklasifikasian kategori berita *hoax.*

## **F. Definisi Operasional**

1.Algoritma Apriori yakni salah satu algoritma yang digunakan untuk menentukan nilai sebuah support dari kata yang ditentukan sebagai kata dalam library. Nilai ini berguna untuk pengklasifikasian berita terindikasi *hoax* ataupun tidak.

2. *Precision And Recall* adalah sebuah metode yang digunakan untuk menentukan tingkat ketepatan antara informasi yang diminta oleh pengguna dengan jawaban yang diberikan pada sistem. *Precision and recall* dapat digunakan untuk mengukur tingkat keberhasilan sistem dalam menemukan sebuah informasi.

3. *Term Frequency and Inverse Document Frequency(TF-IDF)* yakni sebuah metode menentukan nilai bobot pada sebuah kata yang nantinya digunakan untuk mengukur tingkat ke palsuan sebuah berita.

# **BAB II**

**KAJIAN PUSTAKA**

## **A. Dokumen Berita untuk Pengklasifikasian Data**

Berita sendiri memiliki arti yakni laporan tercepat dari suatu peristiwa atau kejadian yang faktual. Pada penelitian ini terdapat kumpulan dokumen berita tervalidasi berita *hoax* yangnantinya digunakan untuk pembuatan *library* kata. Berita *hoax* ini didapat dari website *turnbackhoax.id* yang memilki beberapa berita yang terbukti bukan berita benar atau tervalidasi berita *hoax.* Penggunaan klasifikasi dinilai tepat untuk identifikasi berita hoax dikarenan dalam metode klasifikasi terdapat teknik untuk mengidentifikasi pola dari kata pada berita yang terindikasi *hoax.* Data dokumen yang sudah dikumpulkan dibagi menjadi 2 buah bagian yakni data uji dan data latih. Data latih yakni berita tervalidasi hoax yang nantinya akan digunakan untuk pembuatan *library* kata. Data uji yakni data baik berita *hoax* maupun non *hoax* yang digunakan untuk mengukur performa sistem yang digunkan untuk klasifikasi berita *hoax*.

## **B. Metode *Waterfall* untuk Pembuatan Aplikasi Klasifikasi Berita *Hoax***

Metode *waterfall* yakni metode pengembangan yang digunakan untuk penelitian dalam bidang sistem informasi. Model ini melakukan pendekatan secara urut dan sistematis yang dimulai dari tahap awal kebutuhan sistem kemudian tahap analisis, selanjutnya tahap desain, tahap *coding,* tahap *testing*, tahap *verifikasi,* dan tahap *maintenance*(Pascapraharastyan dkk, 2014). Model metode ini dijuluki dengan *waterfall* dikarenakan tahapan tahapan yang dilakukan berjalan secara runtut dari tahap awal dilanjut dengan tahap selanjutnya yang dapat dikerjakan jika tahap sebelumnya telah diselesaikan.

## ***C. Natural Language* *Processing***

*Natural Languge Processing*  adalah sebuah pemrosesan bahasa alami. Digunakan untuk mengkaji bahasa manusia ke dalam bahasa computer yang mencakup pemrosesan simbolik dan static (Bird & Loper, 2016). NLP ini berguna untuk memproses dan memahami bahasa manusia ke dalam komputer sehingga maksud dari target dapat tersampaikan dengan benar ke dalam komputer. Dalam hal ini salah satu kesulitan yang dihadapi yakni soal ambiguitas dari kata yang diberikan oleh manusia sehingga komputer salah dalam memproses maksud yang diinginkan (Khurana dkk, 2017).

Penerapan teknik ini dalam dunia teknologi yakni steaming / pemotongan kata menjadi bentuk dasar, pembuatan ringkasan sebuah cerpen,pembuatan botchat untuk took-toko online dan lain sebagainya. Dalam hal ini *Natural Language processing*  sangat berperan penting guna menghubungkan antara Bahasa Manusia ke dalam komputer sehingga dapat diolah sedemikian rupa sesuai kebutuhan.

Berikut adalah tahapan-tahapan pemrosesan dari NLP :

### 1. *Tokenizing*

Yakni sebuah proses pemisahan text menjadi per kata sehingga tiap kata dapat diolah dengan mudah.

### *2. Stopwords Removal*

Stopword merupakan kata umum yang digunakan / kata yang tidak penting untuk dianalisi / diproses. Seperti contoh kata : dan, atau, akan tetapi, jika. Tujuan dari penghapusan kata ini yakni mengurangi pembengkakan jumlah index yang digunakan.

### 3. *Stemming*

Proses *stemming* yakni sebuah penghapusan kata yang memilki awalan / akhiran sehingga didapat kata dasar. Guna dari *stemming* ini yakni meminimalisir persamaan kata yang memiliki perbedaan awalan / akhiran sehingga dapat memperkecil jumlah indeks yang akan diproses (Perkin , 2014).

## **D. Teknik *TF-IDF***

Metode *TF-IDF* (*Term Frequency* *and* *Inverse Document* *Frequency)* yakni sebuah metode pembobotan sebuah kata yang digunakan untuk memberikan nilai seberapa penting kata tersebut dalam pembentukan text yang digunakan. Dalam metode ini memiliki algoritma yang digunakan untuk mengukur bobot tiap-tiap kata dalam sebuah dokumen. Semakin besar bobot sebuah kata, maka semakin penting pula kata tersebut dalam sebuah dokumen. (Gifti, 2018)

Proses dari TF-IDF yakni *Term Frequent* akan menghitung frekuensi kata yang muncul dan dibandingkan jumlah kata yang terdapat dalam sebuah dokumen. Berikut rumus persamaan dari metode *Term Frequent*:

*Tf(i)* = ………… ……..*(2.1)*

Keterangan :

*Tf (i)* = Nilai *Term Frequent* kata (i)

*freq ti* = Jumlah kemunculan kata(i) dalam sebuah dokumen

*t* = Jumlah keselurahan kata dalam sebuah dokumen

Untuk proses IDF (*Inverse Document Frequent*) yakni menghitung jumlah seluruh dokumen yang dibandingkan dengan dokumen yang memiliki kata i. Berikut rumus persamaan yang digunakan :

*Idf(i) = log ……………(2.2)*

Keterangan :

*Idf* = Nilai *Inverse Documen Frequent* yang terdapat dalam dokumen

*D* = Jumlah seluruh dokumen

= Jumlah dokumen yang mengandung kata (i)

Setelah diketahui Tf dan Idf dari 2 buah rumus persamaan yang telah dipaparkan maka langkah selanjutnya yakni mencari weight / bobot dari sebuah kata dengan mengalikan hasil dari TF dan IDF sebuah kata (Gifti, 2018).

*W = tf(i) x idf(i) ……………… (2.3)*

Keterangan :

*W* = *Weight* / Bobot dari sebuah kata

*tf(i)* = *TermFrequent* dari sebuah kata

*idf(i)* = *Inverse Document Frequent* dari sebuah kata

**E. *Apriori* *Method***

*Apriori* adalah sebuah asumsi tentang segala sesuatu sebelum bertemu dengan pengalaman sehingga dapat mengambil kesimpulan. Algoritma ini digunakan agar computer dapat mempelajari aturan asosisasi, menacri pola hubungan antara satu item dengan yang lain. Secara umum, dalam pembentukan pola asosiasi algoritma apriori memiliki 2 buah tahap yakni pencarian *frequent itemset* dan membentuk pola asosiasi (Listriani , 2016). Metode ini dapat di aplikasikan kedalam *stemming* kata dengan memanfaatkan hasil *support* dari metode *apriori.* Cara kerja algoritma ini yakni :

1. *Itemset Frequent* yakni pencarian item yang sering muncul bersamaan dalam sebuah data. Contoh : kata berita dan hoax sering muncul dalam sebuah paragraph.

2. *Knowledge* Pencarian data / informasi yang penting dalam sebuah paragraph

3. *Support* (nilai penunjang) yakni presentasi dari *record* yang mengandung kombinasi item dibanding dengan jumlah *record* contoh : jika ada kata a dan b maka *support* dari {a,b} yakni peluang sebuah kata a dan b yang muncul dalam sebuah dokumen (G. A. Saputro, 2017).

Rumus untuk menghitung nilai *support* pada suatu itemset yakni:

*Support (A) = …… (2.4)*

Keterangan :

*Support(A) =* Nilai Penunjang dari sebuah kata

## **F. Bahasa Pemrograman *Python***

Bahasa pemrograman *python* merupakan bahasa pemrograman yang termasuk kedalam *highlevel language. Hightlevel language* merupakan kategori bahasa pemrograman yang mendekati bahasa manusia. Penulisan dari bahasa python ini sangat *simple* sehingga tidak memerlukan banyak *space* untuk digunakan. Selain itu, keuntungan dari bahasa pemrograman *python* yakni tidak memakan banyak waktu dalam pembuatanya dikarenakan bahasa yang tidak terlalu rumit. Hal ini membuat bahasa *python* sangat mudah untuk dipelajari baik untuk pemula maupun bagi yang sudah menguasai (Hokya, 2013).

Bahasa pemrograman python termasuk kedalam bahasa pemrograman *open source* yakni dapat digunakan secara bebas sehingga banyak perusahaan yang memanfaatkan bahasa pemrograman *python* untuk memberikan pelayanan.Selain itu, *python* merupakan Bahasa pemrograman tingkat tinggi sehingga dapat memaksimalkan kinerja *system* yang ingindibuat. Salah satu aplikasi yang dapat digunakan untuk mengembangkan sebuah software dengan mendukungbahasa pemrograman python yakni pycharm. Python juga mendukung pengembangan pada berbagai sistem operasi seperti : Linux, Android, Mac Os, Windows, Palm. (Hendri, 2003)

## **G. *Database* MySQL**

Yakni sebuah manajemen basis data SQL yang digunakan untuk penyimpanan, pengaturan dan pengolahan data yang nantinya akan dapat diakses dengan mudah. MySQL termasuk kedalam *system database* gratis yang didistribusikan dibawah lisensi GPL (*General Public License*). Pada awalnya, MySQL ditemukan oleh Michael Monty Widenius pada tahun 1979 dari Swedia (Saputro, 2016). Kelebihan dari database MySQL ini yakni :

1. Memiliki keamanan yang cukup baik sehingga pengguna dapat mengguanakan database ini untuk keperluan pribadi maupun komersial.

2. Gratis sehingga dapat digunakan oleh siapapun

3. Stabil dalam pengoperasianya sehingga dapat digunakan dengan optimal

4. Fleksibel dalam berbagai macam program sehingga dapat di akses dalam berbagai platform dan berbagai bahasa program

5. Merupakan *Database Management System* (DBMS)

6. MySQL merupakan *database relational*

7. Menggunakan enkripsi *password* sehingga aman untuk digunakan

## **H. Pengukuran Performa**

Setelah sebuah *system* telah selesai dibuat dan menampilkan hasil yang telah diinginkan, maka langkah selanjutnya yakni akan dilakukan pengukuran performa dari *system* yang dibuat. Pengukuran performa dilakukan bertujuan menguji kinerja dan akurasi dari *system* yang sudah dibuat.Pengukuran performa ini dilakukan dengan menggunakan metode *Precision , Recall* dan *Accuracy* (Weddiningrum dkk, 2018).

*Precision dan Recall* yakni berguna untuk mengukur keefektifan pengambilan informasi (Purnama, 2012). Rumus formula dari *Precision* dan *Recall*  yakni :

*Precision = ………….. (2.5)*

*Recall = ………………... (2.6)*

**Tabel 2. 1 Precision and Recall**

|  |  |  |
| --- | --- | --- |
| Berita | Relevan | Tidak Relevan |
| Hoax  Tidak Hoax | *True positve (tp)*  *False negative (fn)* | *Flase postive (fp)*  *True negative (tn)* |

Berdasarkan Tabel 2.1 dapat dirumuskan formula perhitungan akurasi sebuah *system* yakni

*P = tp / (tp+fp) …....…………………………... (2.7)*

*R = tp / (tp+fn) …………….………………….. (2.8)*

Keterangan :

P = *Precision*

R = *Recall*

tp = Berita hoax terklasifikasi hoax (*True Positive)*

*fp =* Beritareal terklasifikasi hoax *(False Positive)*

*tn =* Beritareal terklasifikasi real *(True Negative)*

*fn =* Berita hoax terklasifikasi real *(False Negative)*

Contoh terdapat 10 buah dokumen berita yang terdiri dari 5 buah konten berisi *fake news* dan 5 buah konten berisi berita *real.* Jika dalam *system* ini mendeteksi terdapat 6 buah berita yang terindikasi *fake* yakni 4 buah berisi konten *fake* *news* dan 2 buah termasuk berita *real*  maka 4 dari berita yang telah dipilih termasuk tn (*true positive*) 2 yang dipilih termasuk tn (*false psotive*). 1 berita *fake* yang tidak terdeteksi termasuk *fn* (*false negative*). Dan 4 berita sisa yang tidak terdeteksi termasuk tn (*True Negative*)

Selain *Precision and Recall,* Perhitungan akurasi *system* juga diperlukan untuk memastikan seberapa akurat *system* dapat digunakan.maka langkah selanjutnya yakni perhitungan akurasi *system* dengan menggunakan rumus persamaan :

*ac = x 100% ………………………….. (2.9)*

Keterangan :

*ac* = Tingkat akurasi

= Berita Hoax yang terindikasi hoax

= Berita Real yang yang tidak terindikasi berita hoax

= Berita real yang terindikasi hoax

= Berita Hoax yang terdeteksi real

Jumlah deteksi benar yakni hasil dari penjumlahan *true positive* dan *True negative.*Lalu hasil tersebut akan dibagi dengan total dari seluruh data yang digunakan untuk pengujian. Setelah itu dikalikan 100% agar dapat mencari persentase tingkat akurasi dari peneltian ini (Weddiningrum , 2018).

## **I. Penelitian yang Relevan**

Dalam penelitian ini dilakukan penelusuran studi literatur pada beberapa penelitian terkait dengan klasifikasi berita hoax . Beberapa penelitian tersebut antara lain :

1. Penilitian yang dilakukan oleh Frista Gifti (2018) menjelaskan pendeteksian konten hoax berbahasa Indonesia menggunakan metode *Levenshtein Distance.* Dalam penilitan tersebut penulis menggunakan metode *Tf-Idf* untuk memberikan bobot kata yang nantinya akan dicari jarak kata asal dengan kata sumber menggunakan metode *Levenshtein Distance* pada sebuah berita. Hasil batas yang didapat yakni 0,0014 pada data 100 berita yang terindikasi berita *hoax.* Metode yang diambil dari penelitian ini yakni proses pengklasifikasian berita *hoax.*

2. Penelitian yang dilakukan oleh Marin Vukovic (2009) yang berjudul “*An Intelegent Automatic Hoax Detection Sistem*” menjelaskan tentang pengklasifikasian email hoax dengan membandingkan pola tersimpan yang sama. Kelemahan dari peneleitian ini yakni jika ada email yang memiliki pola baru, maka *system* belum dapat mengidentifikasi email tersebut. Metode yang diambil dari penelitian ini yakni proses pengklasifikasian berita *hoax*.

3. Penelitian yang dilakukan oleh Munjiah Nur Saadah, Widar Atmagi, Dyah S, Agus Zainal (2013) yang berjudul “Sistem Temu Kembali Dokumen Texs dengan pembobotan TF-Idf dan LCS” menjelaskan tentang bagaimana membangun system pengembalian sejumlah dokumen dengan metode tertentu yang memiliki relevansi tinggi sesuai dengan permintaan pengguna. Metode yang digunakan penulis yakni metode pembobotan dengan menggunakan Tf-Idf yang disesuaikan dengan menggunakan *LCS* guna mempertimbangkan kemunculan urutan kata yang sama antara *query* dengan teks dalam document. Metode yang diambil dari penelitian ini yakni metode pembobotan kata.

## **J. Kerangka Pemikiran**

Penelitian ini menggunakan parameter berupa beriita hoax yang nantinya akan di proses dengan menggunakan metode *NLP ,Apriori dan TF-Idf untuk dijadikan dataset.* Hasil dari inputan tersebut yakni berupa kata yang memiliki nilai bobot tersendiri yang nantinya digunakan untuk menilai berita baru yang di inputkan sebagai data uji. Setelah di proses, hasil dari meotde akan diuji akurasi dengan menggunakan perhitungan akurasi *system.*

1. Proses input data latih :

Dalam Prose *input* data latih, data uang di masukkan berupa data berita yang terindikasi berita hoax.Berita yang sudah dimasukkan nantinya akan diproses sehingga didapat *library* data kata-kata hoax.

Input Pre-Prosesing Proses (1)

*Input* Berita Hoax

Prosses NLP

Pencarian Bobot TF-IDF

Pencarian Nilai *Support Apriori*

Pencarian Hasil Rataan Support dan Bobot

Library Kata Didapat

Output Proses (3) Proses (2)

**Gambar 2. 1 Proses Input Data Latih**

* 1. *Input* Berita Hoax

Proses *input* berita terindikasi *hoax* guna membuat sebuah data latih.

* 1. Proses NLP (Neural Language Prosesing)

Proses untuk memisah kata dari kalimat, memilah tanda baca dan pengahapusan kata yang tidak penting.

* 1. Proses (1)

Proses pencarian nilai bobot sebuah kata dengan menggunakan rumus *Tf-Idf* yang nantinya akan disimpan menjadi sebuah data latih.

* 1. Proses (2)

Proses pencarian Nilai sebuah kata dengan menggunakan rumus dari metode *apriori* yang nantinya akan disimpan menjadi sebuah data latih.

1.5 Proses (3)

Proses untuk mencari Nilai sebuah kata dengan mencari rataan dari hasil support dan nilai bobot yang nantinya akan dijadikan acuan dari data latih.

* 1. Output

Proses penyimpanan library kata (data latih) yang nantinya digunakan sebagai pembanding data uji.

1. Proses *input* data uji :

Data *input* data uji yakni terdiri dari data berita real dan berita yang terindikasi berita *hoax.* Data ini digunakan untuk menguji tingkat keberhasilan system yang dibuat.

*Input* *Pre-Processing* Proses(1)

*Input* Berita

Prosses NLP

Pencarian kata pada data latih

Pencarian nilai hasil

Rataan Nilai Hasil

Hasil

Output Proses(3) Proses (2)

**Gambar 2. 2 Proses Input Data Uji**

* 1. *Input* Berita Hoax

Proses *input* berita terindikasi *hoax* guna membuat sebuah data latih.

* 1. Proses NLP (Neural Language Prosesing)

Proses untuk memisah kata dari kalimat, memilah tanda baca dan pengahapusan kata yang tidak penting.

* 1. Proses (1)

Proses *Searching* kata pada inputan pada data latih

*2.4* Proses (2)

Pencarian Nilai Hasil yang terdapat pada kata yang telah ditemukan

*2.5* Proses (3)

Proses mencari rata-rata dari nilai yang telah didapat dari penacrian Nilai Hasil

* 1. Hasil

Hasil yang didapat yakni apakah data yang di inputkan terindikasi data hoax atau tidak.

# **BAB III**

**METODE PENELITIAN**

## **A. Desain Penelitian**

Metode penelitian yang digunakan adalah metode eksperimen. Metode eksperimen adalah metode cara penyajian sebuah data melalui proses percobaan-percobaan untuk membuktikan sendiri suatu pernyataan atau hipotesis tertentu dalam sebuah penelitian.Penelitian eksperimen merupakan penelitian yang paling valid dikarenakan metode eksperimen dilakukan dengan pengontrolan variable yang sangat ketat sehingga didapat hasil yang maksimal(Jaedun, 2011). Keuntungan utama penelitian dengan penggunaan metode eksperimen adalah adanya kendali ditangan peneliti dan ketepatan sebuah logika yang terkandung di dalam metode tersebut. Metode ini juga sering digunakan karena dapat dibuat perencanaan yang sistematis sehingga metode eksperimen adalah metode yang paling kuat dalam sebuah penelitian. Berikut adalah tahapan metode eksperimen :

Pengumpulan Data

Pre Prosses NLP

Proses TF-IDF

Proses Apriori

Experimen dan Pengujian

Evaluasi dan Validasi Hasil

**Gambar 3. 1 Metode yang Diusulkan**

1. **Identifikasi Berita *Hoax***

Identifikasi berita *hoax* dilakukan dengan cara mencari kata – kata yang sering atau yang memungkinkan untuk digunakan sebagai kata yang terindikasi sebagai kata hoax. Seperti contoh , kata “akan”. Kata “akan” biasa digunakan untuk sebuah premis yang belum dilakukan. Sedangkan berita adalah sebuah informasi yang faktual bersifat sedang atau sudah terjadi. Sehingga kata “akan” memiliki peluang sebagai kata yang terindikasi *hoax*.Pada *system* deteksi *hoax* ini, dilakukan pencarian kata yang memungkinkan untuk pembuatan berita hoax dengan menggunakan rumus pembobotan tf-idf dan apriori sehingga didapat nilai sebuah kata berkemungkinan dijadikan untuk berita *hoax.*

### **2. Pengumpulan Data**

Data yang dikumpulkan terdiri dari 2 buah jenis data yakni data uji dan data latih. Data latih yakni data yang berisikan kumpulan berita *hoax* yang nantinya akan digunakan untuk proses pengindex an kata guna pembuatan *library*. Data uji yakni data yang berisikan berita baik *hoax* maupun real yang nantinya digunakan untuk menentukan tingkat akurasi sistem. Data yang diambil adalah data berita *hoax* dari situs turbackhoax.id dan juga data *hoax* dari broadcast whatsapp. Sedangkan berita yang nantinya dijadikan data uji berasal dari detik.com dan sebagian berasal dari turnbackhoax.id.

### **3. Pre Prossesing Menggunakan NLP (Neural Language Processing)**

Data berita yang didapat dari situs turnbackhoax.id ataupun dari detik.com adalah sebuah data mentah yang belum dapat diolah. Agar data tersebut dapat diolah, maka dilakukan pre prosessing dengan menggunakan NLP (*Neural Language Processing*). Pada tahap pre prosessing ini memiliki beberapa *step* yakni :

3.1. Tokenizing

Proses pemisahan text menjadi per kata sehingga tiap kata dapat diolah dengan mudah. Proses pemisahan kata dilakukan menggunakan fungsi dalam python.

3.2. *Stoward Removal*

Stopword merupakan kata umum yang digunakan / kata yang tidak penting untuk dianalisi / diproses. Seperti contoh kata : dan, atau, tetapi. Tujuan dari penghapusan kata ini yakni mengurangi pembengkakan jumlah index yang digunakan.

Cek pada library kata umum

Terdapat Kata Umum

Hapus Kata

*Input* Data

Yes

No

**Gambar 3. 2 Flowchart Stoward Removal**

a. Masukkan Data

b. Proses pencarian kata yang dinilai sebagai kata tidak penting pada *library*

c. Jika terdapat kata umum maka *system* akan menghapus kata tersebut

d. Jika tidak, maka *system* akan melanjutkan pencarian sampai akhir

f. *System* mengakhiri proses

3.3 *Stemming*

Sebuah penghapusan kata yang memilki awalan / akhiran sehingga didapat kata dasar. Guna dari *stemming* ini yakni meminimalisir persamaan kata yang memiliki perbedaan awalan / akhiran sehingga dapat memperkecil jumlah indeks yang akan diproses. Proses stemming ini menggunakan fungsi *stemmer factory* yang terdapat pada python dengan *import* *package* sastrawi

### **4. *Processing* menggunakan Tf-Idf**

Setelah dilakukan *pre processing* dengan menggunakan NLP, maka langkah selanjutnya yakni menacari bobot dari sebuah kata dengan menggunakan *Term Frequent dan Inverse Document Frequent.* Alur sistem dari program Tf-Idf ini digambarkan pada *flowchart* berikut :

Hitung TF

Tf=Frekuensi kemunculan kata /seluruh kata

Hitung IDF Idf=log(jml seluruh doc/jml doc mengandung kata)

Data NLP

Bobot = Tf\*Idf

Input into csv

Kata+bobot

**Gambar 3. 3 Flowchart Tf-Idf**

1. Data yang sudah di proses dengan NLP dimasukkan kedalam perhitungan Tf
2. Selanjutnya akan dihitung nilai Idf dari data
3. Pencarian bobot dengan menggabungkkan Tf dan Idf
4. Kata dan bobot di simpan dalam bentuk .csv

### **5. *Processing* menggunakan Apriori**

Tahap selanjutnya yakni prosesing dengan menggunakan metode apriori. Dalam metode ini diambil sebuah nilai support dari perbandingan antara jumlah berita yang mengandung kata dengan jumlah total berita. Yang selanjutnya akan dimasukkan kedalam data *index library* kata yang tersimpan beserta bobot. Selanjunya, data akan dicari rataan dari nilai bobot dan nilai support sehingga didapat nilai akhir. Proses alur dari sistem digambarkan pada *flowchart* berikut :

Hitung Support

Support=jumlah berita yang mengandung kata /seluruh berita

Hitung Nilai Akhir : (Support+Bobot) / 2

Data NLP

Input into csv

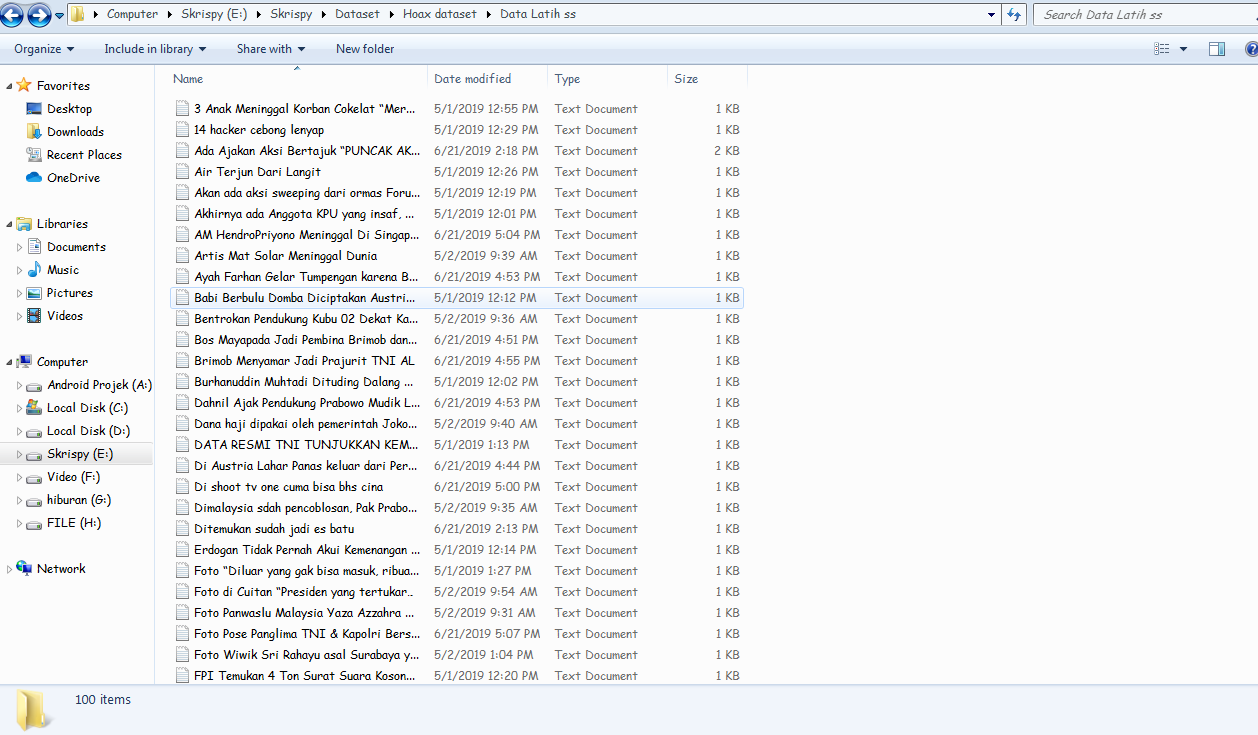
Kata+Support+Nilai Akhir

**Gambar 3. 4 Flowchart Apriori**

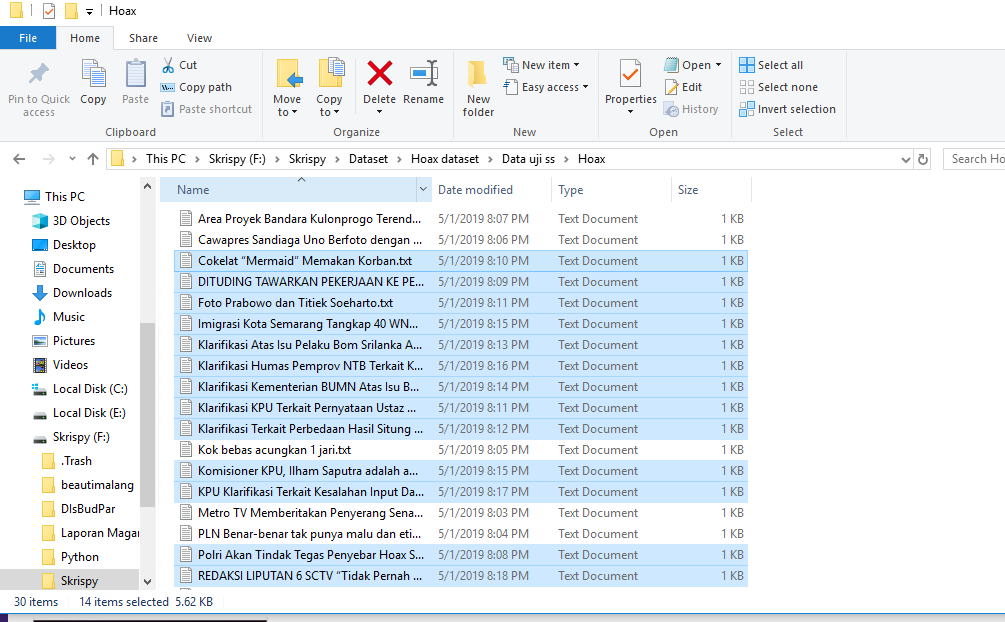
1. Masukan data yang sudah diproses NLP
2. *System* menghitung nilai *support* dengan menggunakan rumus apriori
3. Proses selanjutnya yakni pencarian nilai akhir dengan cara menari rataan dari nilai *support*  dan nilai bobot
4. Data kata dan nilai akhir di simpan dalam bentuk .csv

### **6. Eksperimen dan Pengujian**

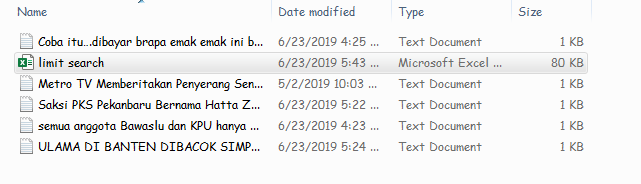
Penelitian ini akan menggunakan data latih sebanyak 100 buah berita *hoax* yang dapat di lihat pada Lampiran 1. Dan 30 buah data uji dengan rincian 16 buah berita hoax dan 14 buah berita real yang dapat dilihat pada Lampiran 2. Dengan data uji tersebut dilakukan pengujian sistem untuk mencari tingkat akurasi dari *sytem*.Untuk pencarian nilai limit akan digunakan 5 buah berita *hoax* yang nantinya akan diproses sehingga didapat batas yang sesuai untuk dijadikan acuan.Data Uji dan data latih dapat dilihat pada Gambar 3.5 dan 3.6. Pada Gambar 3.5 terdapat 100 buah data latih berita *hoax* yang nantinya akan diproses oleh *system* sehingga dapat menghasilkan *library* kata *hoax.* Pada Gambar 3.6 yakni terdapat 30 buah berita dengan rincian : 16 buah berita hoax dan 14 buah berita real yang nantinya akan digunakan untuk pengujian akurasi *system.*Dan pada Gambar 3.7proses pencarian nilai limit. Terdapat 5 buah berita terindikasi *hoax* yang nantinya akan digunakan sebagai *sample* data guna mencari nilai limit dari system yang digunakan sebagai acuan pengklasifikasian berita.

****

**Gambar 3. 5 Data Latih**

****

**Gambar 3. 6 Data Uji**



**Gambar 3. 7 Data Pencarian Limit**

### **7. Evaluasi dan Validasi Hasil**

Evaluasi dilakukan dengan memasukkan hasil dari data uji yang telah diuji pada tahap sebelumnya. Digunakan rumus *precision* *and* *recall* dan *accuracy* untuk menentukan tingkat keakuratan sistem.

# **BAB IV**

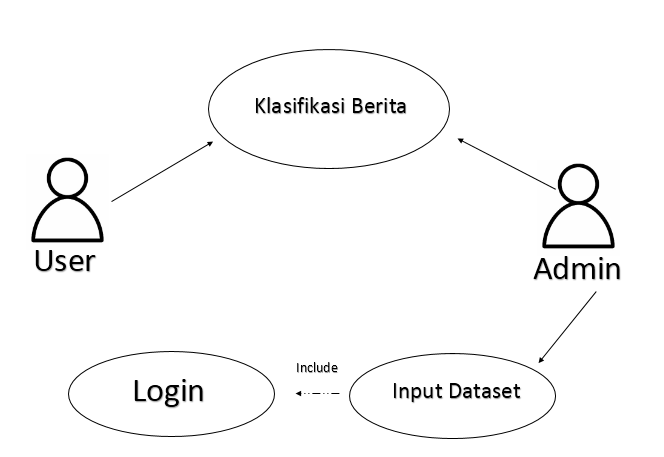
**HASIL**

Pada tahap ini dilakukan analisi secara menyeluruh dari *system* yang telah dibuat. Akan dijelaskan rincian perancangan system dan implementasi dalam sebuah bahasa pemrograman. Selanjutnya akan dilakukan ujicoba/ *testing* guna mencari kekurangan dan kelebihan pada *system* yang dibuat sehingga dapat dijadikan bahan pertimbangan guna melakukan pengembangan penelitian selanjutnya.

## **A. Desain Aplikasi**

Pada tahap ini terdapat beberapa desain perancangan *system* meliputi *Usecase Diagram,Diagram Konteks,* Alurdata *user* , Alur data *admin*, Penyimpanan data*,*

### **1. Usecase Diagram**

****

**Gambar 4. 1 Usecase Diagram**

Pada gambar 4.1 di atas menjelaskan tentang alur *system*  yang dibuat. Dalam diagram di atas dapat dilihat bahwa user hanya bias memberikan inputa untuk klasifikasi berita sedangkan admin dapat menambah dataset ke dalam database.

**Tabel 4. 1 Usecase Deskripsi**

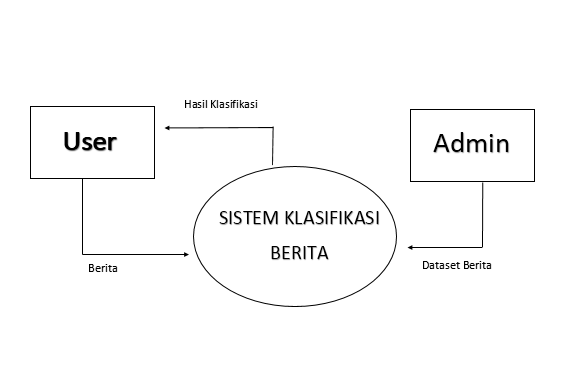
|  |
| --- |
| Nama *Usecase* Alur Sistem |
| Aktor *User, Admin*  *Normal Course* 1. User Memasukkan berita yang ingin diklasifikasi  2. Admin login untuk menginputkan *database*  *Pre-Condition* 1. User memiliki berita yang ingin diklasifikasi  2. Admin terdaftar dalam system  *Post-Condition* 1. User mendapatkan hasil klasifikasi  2. Admin Berhasil mengiputkan data |

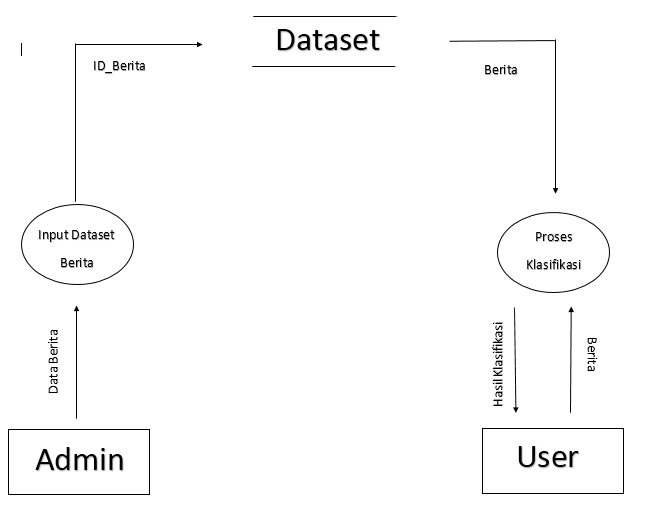
### **2. *Diagram Konteks***

Dalam pembuatan system diperlukan *diagram konteks* yang digunakan untuk mempermudah memahami aliran data dari aplikasi yang dibuat Pada Gambar 4.2 diketahui bahwa user memberikan *input* kedalam system klasifikasi berita dan admin memberikan *input* berupa dataset berita sehingga *system* dapat memberikan *output* hasil identifikasi berita kepada user.

**2.1 Aliran Data**

Pada tahap ini, alur diagram yang sudah dibuat pada diagram konteks akan diuraikan dengan rinci sehingga didapat aliran yang detail dan dapat dilihat dengan jelas.Berikut adalah Aliran data pada *system* klasifikasi berita *hoax* yangd apat dilihat pada Gambar 4.3.

 **Gambar 4. 2 Diagram Konteks**

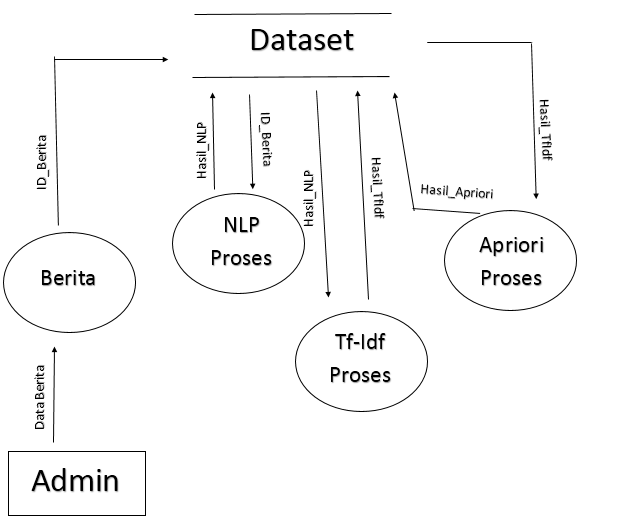
**

**Gambar 4. 3 Aliran Data**

Pada gambar 4.3 Admin dapat memberikan *input* berupa *input dataset* berita yang nantinya akan dijadikan *dataset* untuk proses klasifikasi sedangkan user dapat memberikan *input* berupa berita yang nantinya akan diklasifikasi dengan menggunakan dataset input dari admin.

**2.2 Aliran Data *Admin***

Dalam tahap ini, aliran data yang telah dibuat pada aliran data akan dijabarkan lebih rinci sehingga dapat dipahami dengan jelas. Berikut aliran data admin :

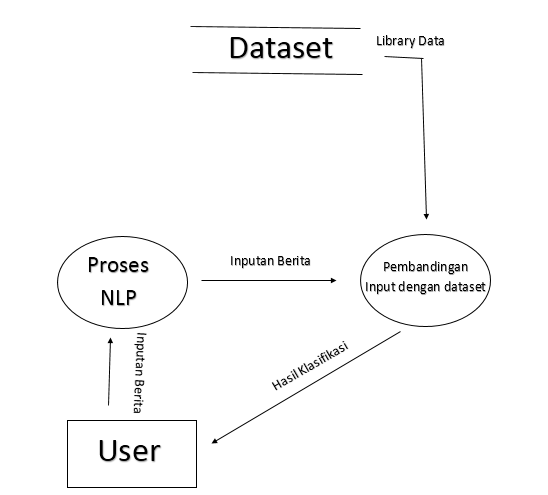


**Gambar 4. 4 Input Dataset**

Pada gambar 4.4 dapat diamati bahwa *input* dari admin akan di lanjutkan kedalam *database* lalu di proses dengan NLP proses pencarian kata umum dari sebuah kata sebelum diolah ,Tf-Idf yakni proses penacrian bobot kata dan Apriori *methode* untuk mencari nilai support *.*Sehingga didapat dataset yang bias digunakan untuk proses klasifikasi pada *input* user.

**2.3 Aliran Data *User***

Setelah dilakukan *input dataset,* Selanjutnya adalah proses *input* dari user. Berikut adalah Aliran data *User* pada gambar4.5 *:*

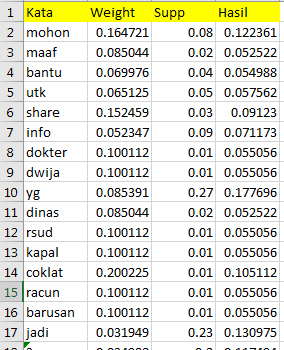


**Gambar 4. 5 Aliran Data User**

Pada tahap ini dilakukan proses NLP(*Natural Language Processing*) yang digunakan untuk pencarian kata natural.Selanjutnya, dilakukan proses pencarian kata pada dataset yang telah dibuat.Setelah proses pencarian maka akan mucul hasil dari klasifikasi berita.

**2.4 Penyimpanan *Data***

Untuk penyimpanan dataset , aplikasi yang dibuat memiliki 2 buah tempat penyimpanan yakni berupa excel. Proses pertama yakni penyimpanan data yang di *input* oleh admin kedalam excel. Langkah selanjutnya yakni proses pengolahan data sehingga didapat nilai suatu kata yang disimpan dalam bentuk .xlsx. Dilakukan penyimpanan dalam bentuk excel dikarenakan data yang disimpan sangat banyak sehingga untuk memaksimalkan kinerja proses maka penyimpanan dilakukan didalam excel.

******

**Gambar 4. 6 Penyimpanan Hasil Proses dalam Bentuk Excel**

Gambar 4.6 yakni penyimpanan excel yang digunakan untuk menyimpan data yang telah di proses.

## **B. Penulisan Code Program**

Pada bab ini akan dijelaskan secara detail penulisan code program yang digunakan untuk proses klasifikasi berita *hoax* menggunakan apriori.

1. *Preprocessing* Dokumen

Pada tahap ini yakni pengolahan dokumen yang masih mentah menjadi dokumen yang dapat diproses oleh system. Data latih yang digunakan yakni 100 buah data berita yang terindikasi berita *hoax.* Alur proses pada *preprocessing* ini yakni *tokenizing* (pemisahan kalimat) ,*Case Folding* (penyetaraan huruf menjadi huruf kecil), *Stopwords removal* (Penghilangan kata tidak penting) dan Stemming (Perubahan menjadi kata dasar). Berikut adalah source code yang digunakan untuk *preprocessing :*

|  |
| --- |
| #Case Folding /Huruf Kecil  text =text.casefold()  #tokenizin  text=text.split()    #Stopword Removal / Menghilangkan kata tidak penting  factory = StopWordRemoverFactory()  stopword = factory.create\_stop\_word\_remover()  text = stopword.remove(text)  #Steeming kata dasar  factory = StemmerFactory()  stemmer = factory.create\_stemmer()  text=stemmer.stem(text) |

**Gambar 4. 7 Source Code Preprocessing**

1. Proses *Input* Data Latih

Setelah melalui *preprocessing* maka selanjutnya data akan di masukkan ke dalam *database* . *Source code input* data kedalam *database* dapat dilihat pada gambar 4.8*.*

|  |
| --- |
| def database(x,b,c):  import pymysql  # Open database connection  db = pymysql.connect("localhost","xiinlaw","Arcasevenvold04","skripsi" )  # prepare a cursor object using cursor() method  cursor = db.cursor()    sql = """INSERT INTO beritax (Judul\_Berita, #ganti  Berita,JumlahKata)  VALUES (%s,%s,%s)"""  val=(x,b,c)  try:  # Execute the SQL command  cursor.execute(sql,val)  # Commit your changes in the database  db.commit()  except:  # Rollback in case there is any error  db.rollback()  # disconnect from server  db.close() |

**Gambar 4. 8 Source Code Input Data Latih**

1. Proses hitung Tf-Idf dan Apriori

Setelah melalui proses input db, maka langkah selanjutnya yakni proses perhitungan untuk membuat *library* kata yang nantinya digunakan untuk pembanding data uji. Perhitungan dilakukan dengan menggunakan metode Tf-Idf dan Apriori. S*ource code* dari proses hitung dapat dilihat pada gambar 4.9.

|  |
| --- |
| # idf dan apriori  #jumlah dokumen yang mengandung kata  sqli = "SELECT ID FROM beritax WHERE Berita LIKE %s"  b = berita[r]  a = ('%' + b + '%')  val = (a)  cursor.execute(sqli, val)  a = cursor.fetchall()  rr = len(a)  rr=float(rr)  Kata.append(berita[r])  print('Kata ke : ',r)  print(berita[r],' Jmlh :',rr)  Idf = math.log(100 / rr)  apriori = rr/100  TfIdf=Tf\*Idf |

**Gambar 4. 9 Source Code proses hitung**

1. Proses *Create Library Excel*

Dalam tahap ini yakni tahap penyimpanan kata yang sudah diproses dan dihitung dengan menggunakan apriori dan tf-idf. Hasil perhitungan tersebut akan disimpan dalam bentuk .xlsx. S*ource code* yang digunakan untuk proses *create library excel* dapat dilihat pada gambar 4.10*.*

|  |
| --- |
| # input kata on excel  kata = sh.cell(column=1, row=roww)  kata.value = berita[r]  # Weight  weight = sh.cell(column=2, row=roww)  weight.value = TfIdf  # Support  supp = sh.cell(column=3, row=roww)  supp.value = apriori  # Hasil  hasil = sh.cell(column=4, row=roww)  hasil.value = (apriori+TfIdf)/2  print('Jumlah Kata Yang tersimpan : ', roww-1)  r += 1  roww += 1  wk.save('uji.xlsx') #ganti |

**Gambar 4. 10 Source Code Input Library Into Excel**

1. Proses Input data Uji dan Pembandingan dengan Data Latih

Pada tahap ini user akan memasukkan data berita yang nantinya data akan diproses oleh system untuk dibandingkan dengan data latih sehingga didapat hasil *output*  yang diinginkan oleh user. S*ource code* yang digunakan dapat dilihat pada gambar 4.11.

|  |
| --- |
| def hitung(text):  #pembandingan dengan librari  nlpcount=0  text=text.split()  jmlkata=len(text)  sigmakatasama=0  #print(jmlkata)  for g in text:  wk = openpyxl.load\_workbook("kata.xlsx") #ganti  sh = wk.active  rows = sh.max\_row  for i in range(1, rows + 1):  c = sh.cell(i,1)  h=sh.cell(i,4)  #print(c.value)  if(text[nlpcount]==c.value):  sigmakatasama+=h.value  # print(c.value,h.value,Hasilnlp[nlpcount])  nlpcount+=1  print(sigmakatasama,jmlkata)  ratahasil=sigmakatasama/jmlkata  return ratahasil |

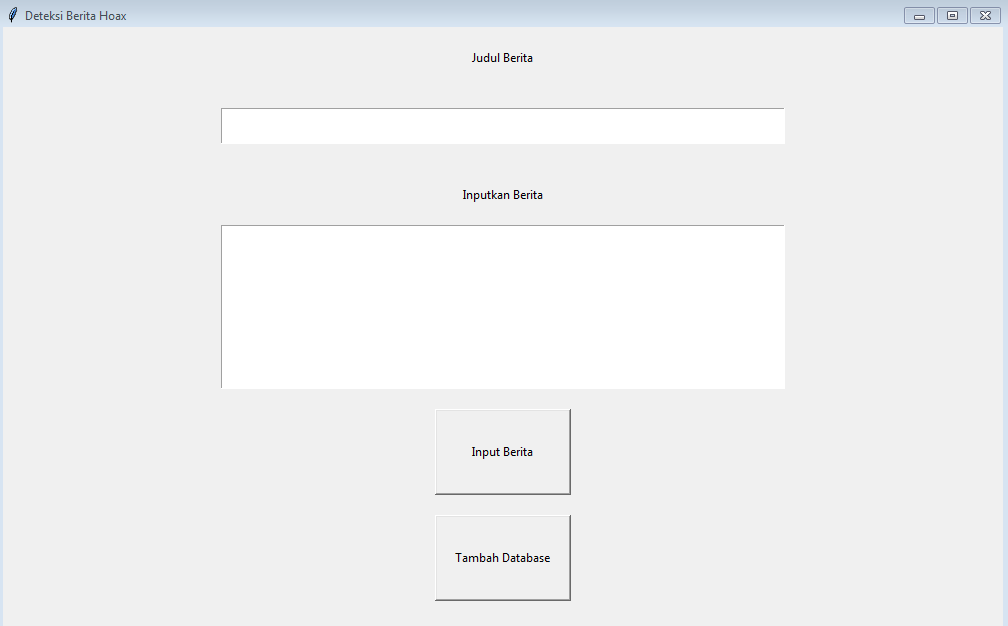
**Gambar 4. 11 Proses Pencarian Hasil Input Data Uji**

## **C. Antar Muka Aplikasi**

Pada tahap ini yakni menjelaskan antar muka *system* yang telah dibuat guna mempermudah pengguna untuk mengoprasikan *system* yang telah dibuat. Terdapat 2 buah antar muka yakni halaman utama (halaman *input* data uji) dan halaman input dataset.

1. Halaman Utama

Pada halaman utama ini terdapat satu buah text kolom dan 2 buah button yang digunakan untuk memasukkan data dan untuk menuju ke halaman input dataset. User dapat memasukkan berita yang ingin diidentifikasi ke dalam text kolom. Selanjutnya user dapat menekan tombol input untuk melihat hasil. Hasil akan di tampilkan berbetuk pop up window yang berisi Nilai hoax suatu berita dan terindikasi hoax atau tidak.



**Gambar 4. 12 Halaman Utama**

****

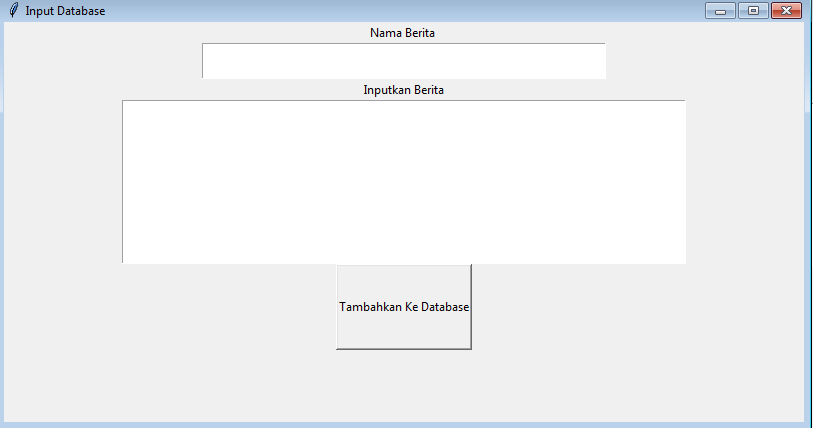
**Gambar 4. 13 Hasil Berita Tidak Hoax**

****

**Gambar 4. 14 Hasil Berita Terindikasi Hoax**

1. Halaman Input Dataset

Pada halaman ini terdapat 2 buah text kolom meliputi judul berita dan isi berita. Setelah admin memasukkan data maka secara otomatis data yang telah di masukkan akan masuk kedalam penyimpanan.

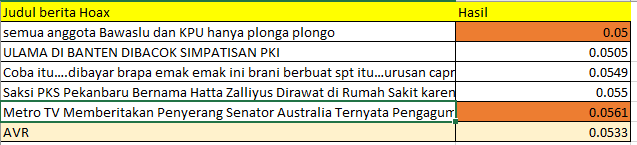


**Gambar 4. 15 Input Dataset Berita**

## **D. Hasil dan Pengujian *System***

Pada tahap ini adalah tahap pengujian *system* dengan menggunakan data uji yang telah di siapkan berupa 16 data terindikasi berita *hoax* dan 14 buah data berita berupa berita real. Untuk menentukan nilai batas dilakukan uji coba menggunakan 5 buah data terindikasi *hoax*  yang nantinya dicari rataan nilai sehingga didapat hasil nilai batas. Selanjutnya, *system* akandiuji menggunakan batas yang telah ditentukan dan dicari hasil dari *precision and recall*  serta tingkat akurasi menggunakan rumus yang telah dijabarkan pada bab II.

1. Pencarian Nilai Batas

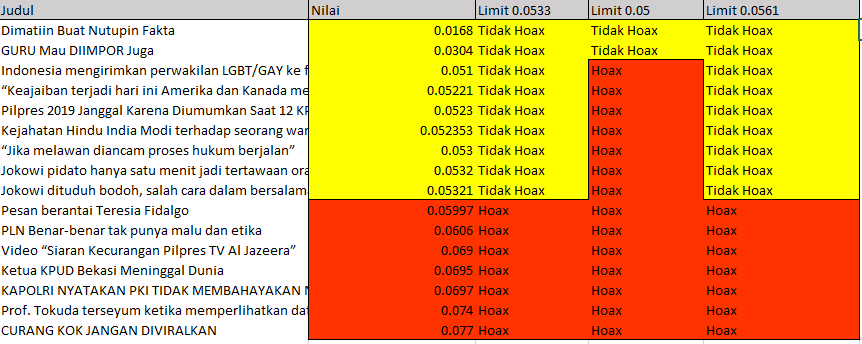


**Gambar 4. 16 Data Uji Nilai Batas**

Untuk mencari nilai batas dilakukan uji coba dengan menggunakan 5 buah data uji berita terindikasi *hoax*.Hasil rata-rata, batas atas dan batas bawah dari nilai data uji akan digunakan untuk nilai batas dari pengujian data uji selanjutnya.

1. Proses Uji Hasil Data Uji

Pada tahap ini dilakukan 2 buah pengujian yakni pengujian data berupa berita hoax dan pengujian data berupa data real.Proses pengujian dilakukan dengan menggunakan 3 buah nilai batas yang telah ditentukan yakni 0.0533, 0.0561 dan 0.05.



**Gambar 4. 17 Data Uji Berita Hoax**



**Gambar 4. 18 Data Uji Berita Real**

**Tabel 4. 2 Tabel Keterangan**

|  |  |
| --- | --- |
| **Keterangan** | **Posisi** |
| *True Positive* (Berita *Hoax* Terklasifikasi hoax) | Gambar 4.17 bagian bawah |
| *False Negative* (Berita *Hoax* Terklasifikasi Real) | Gambar 4.17 bagian atas |
| *True Negative* (Berita Real Terklasifikasi Real) | Gambar 4.18 bagian atas |
| *False Positive* (Berita Real Terklasifikasi *Hoax*) | Gambar 4.18 bagian bawah |

2.1 Skenario 1

Skenario 1 proses uji dengan menggunakan nilai batas rata-rata dari pencarian nilai limit yakni 0,0533.

**Tabel 4. 3 Nilai dari Komponen Perhitungan dengan Batas 0.0533**

|  |
| --- |
| Berita Jumlah |
| *True Positive* (tp) 7  *False Positive* (fp) 3  *False Negative* (fn) 9  *True Negative* (tn) 11 |

Berdasarkan nilai dari setiap komponen, maka dapat ditentukan nilai *Precision, Recall* dan *Accuracy* sebagai berikut :

*Precision :* tp/(tp+fp) *Recall :* tp/(tp+fn)

= 7/10 = 7/16

= 0,7 = 0,437

*Accuracy = x100*

*=* (7 + 11) / 30 x100

= 60 %

2.2 Skenario 2

Skenario 2 proses uji dengan menggunakan nilai batas bawah dari pencarian nilai limit yakni 0,05.

**Tabel 4. 4 Nilai dari Komponen Perhitungan dengan Batas 0.05**

|  |
| --- |
| Berita Jumlah |
| *True Positive* (tp) 14  *False Positive* (fp) 3  *False Negative* (fn) 2  *True Negative* (tn) 11 |

Berdasarkan nilai dari setiap komponen, maka dapat ditentukan nilai *Precision, Recall* dan *Accuracy* sebagai berikut :

*Precision :* tp/(tp+fp) *Recall :* tp/(tp+fn)

= 14/17 = 14/16

= 0,82 = 0,875

*Accuracy = x100*

*=* (14 + 11 ) / 30 x100

= 83,33 %

2.3 Skenario 3

Skenario 3 proses uji dengan menggunakan nilai batas atas dari pencarian nilai limit yakni 0,0561.

**Tabel 4. 5 Nilai dari Komponen Perhitungan dengan Batas 0.0561**

|  |
| --- |
| Berita Jumlah |
| *True Positive* (tp) 7  *False Positive* (fp) 2  *False Negative* (fn) 9  *True Negative* (tn) 12 |

Berdasarkan nilai dari setiap komponen, maka dapat ditentukan nilai *Precision, Recall* dan *Accuracy* sebagai berikut :

*Precision :* tp/(tp+fp) *Recall :* tp/(tp+fn)

= 7/9 = 7/16

= 0,77 = 0,437

*Accuracy = x100*

*=* (7 + 12) / 30 x100

= 63,33 %

## **E. Pembahasan**

Berdasarkan hasil pengujian *system* nilai batas 0.0533 menghasilkan akurasi sebesar 60% , untuk nilai batas 0,561 mendapatkan akurasi sebesar 63,33 % dan untuk nilai batas 0.5 mendapatkan nilai akurasi sebesar 83,33%.Dengan data dari ke 3 buah skenario dapat disimpulkan bahwa nilai batas yang paling efektif digunakan untuk proses klasifikasi berita *hoax* pada system yakni 0.5 dengan tingkat akurasi 83,33 % dengan nilai *precission and recall* 82% dan 87% . Dapat dijelaskan bahwa pengubahan nilai limit sangatlah berpengaruh pada tingkat akurasi system.Dikarenakan nilai dari data uji yang berbeda-beda sehingga untuk dapat mencari kefektifan sebuah system digunakan ujicoba beberapa skenario.

# **BAB V**

**PENUTUP**

## **A. Kesimpulan**

Berdasarkan dari hasil penelitian yang telah dilakukan, maka dapat disimpulkan bahwa :

1. Metode Apriori dapat digunakan untuk mengklasifikasi berita terindikasi hoax.
2. Pada penelitian yang dilakukan metode apriori dapat menghasilkan tingkat akurasi terbesar 83,33 %
3. Perpaduan antara *apriori* dan Tf-Idf terbukti mampu membedakan antara berita *hoax* dan *real.*
4. Kelebihan dari algoritma ini yakni dapat digunakan untuk mendeteksi berita hoax murni (bukan berita yang diubah).
5. Kekurangan dari algoritma ini yakni sangat bergantung pada seberapa banyak library yang dipunyai oleh system dan proses pembuatan library yang memakan cukup waktu.
6. Dalam skenario 2 digunakan nilai batas 0.5 untuk mendeteksi 16 buah berita *hoax* dan 14 buah berita *real* memiliki nilai *precission* 82% dan nilai *recall* 87% sehingga semakin banyak data latih yang dimasukkan semakin akurat pula *system* melakukan proses klasifikasi

## **B. Saran**

Berdasarkan kesimpulan yang telah dijabarkan, berikut adalah saran untuk penelitian yang akan datang :

1. Berita yang dipakai sebagai acuan diharapkan memiliki rentan waktu tertentu sehingga berita tidak kadaluarsa guna meningkatkan akurasi pengecek an berita
2. Untuk meningkatkan akurasi, beberapa hal yang diprediksi dapat meningkatkan akurasi *system* yakni memperbarui *library* yang dibuat sehingga banyak kata yang tersimpan untuk penilaian berita.
3. Mengelompokkan berita sesuai dengan topik dan memberikan pemilihan pendeteksian berita berdasarkan topik dapat meningkatkan efektifitas proses klasifikasi.Dengan begitu, diharapkan penghitungan kata akan lebih spesifik dan dapat meningkatkan akurasi dari *system* yang akan dibuat.
4. Diharapkan pada penelitian selanjutnya dapat menerapkan metode apriori dengan metode lain sehingga bisa mendapatkan kombinasi metode yang lebih akurat untuk proses klasifikasi berita *hoax*

# **DAFTAR PUSTAKA**

Ardoni. (2015). Teknologi Informasi: Kesiapan Pustakawan Memanfaatkannya Ardoni. *Jurnal Aplikasi Iptek Untuk Masyarakat*, *4*(1), 59–62. Retrieved from http://jurnal.unpad.ac.id/dharmakarya/article/download/9042/4064

Bim, I. M. (2012). *Menerapkan Metode Probabilistik Binary Metode Probabilistik Binary Independence Model ( Bim ) Periode Wisuda : 28 Februari 2013 Fakultas Sains Dan Teknologi*.

Bird, S., & Loper, E. (2016). The natural language toolkit NLTK: The Natural Language Toolkit. *Proceedings of the ACL-02 Workshop on Effective Tools and Methodologies for Teaching Natural Language Processing and Computational Linguistics*, (March), 63–70. https://doi.org/10.3115/1225403.1225421

Cardiologia, S. B. de, Francisco, P. M. S. B., Segri, N. J., Borim, F. S. A., Malta, D. C., Fontbonne, A., … Brasil, N. (2018). https://doi.org/10.1590/s1809-98232013000400007

Darmawan, D. (2012). *Mengenal Teknologi Informasi*. 92.

Hendri. (2003). *Cepat Mahir python*. 1–92.

Jaedun, A. (2011). Oleh : Amat Jaedun. *Metodologi Penelitian Eksperimen*, 0–12.

Kesuma, H. W. A. (2016). *Penerapan Metode TF-IDF dan Cosine Similarity dalam Aplikasi Kitab Undang-Undang Hukum Dagang*.

*2014-Python 3 Text Processing with NLTK 3 Cookbook*. (n.d.). https://doi.org/10.1017/CBO9781107415324.004\

Khurana, D., Koli, A., Khatter, K., & Singh, S. (2017). *Natural Language Processing: State of The Art, Current Trends and Challenges*. (Figure 1). Retrieved from http://arxiv.org/abs/1708.05148

KPM. (2013). *Buku Panduan Pemrograman Python. Buku*, *84*, 487–492. Retrieved from http://ir.obihiro.ac.jp/dspace/handle/10322/3933

Lestari, R. D. (2017). *Quality News dan Popular News Sebagai Trend Pemberitaan Media Online (Studi Deskriptif Kualitatif Trend Pemberitaan Quality News dan Popular News pada Media Online Nasional di Indonesia Periode 2016)*. *Channel*, *5*(1), 83–94.

Listriani, D., Setyaningrum, A. H., & M.A, F. E. (2016). *Penerapan Metode Asosiasi Menggunakan Algoritma Apriori Pada Aplikasi Pola Belanja Konsumen ( Studi Kasus Toko Buku Gramedia Bintaro )*. *Jurnal Teknik Informatika Vol 9 No. 2, Universitas Islam Negeri Jakarta*, *9*(2), 120–127. Retrieved from http://journal.uinjkt.ac.id/index.php/ti/article/view/5602/3619

Pascapraharastyan, R. A., Supriyanto, A., & Sudarmaningtyas, P. (2014). *Rancang Bangun Sistem Informasi Manajemen Arsip Rumah Sakit Bedah Surabaya Berbasis Web*. *Sistem Informasi*, *3*(1), 72–77.

Saputro, G. A. (2017). *Perapan Algoritma Apriori Untuk Mencari Pola Penjualan di Cafe*.

Saputro, H. (2016). Pembelajaran Basis Data (Mysql). *Modul Pembelajaran Praktek Basis Data (MySQL)*, 1–34. Retrieved from http://dinus.ac.id/repository/docs/ajar/materi\_1.pdf

Septanto, H. (2017). *Jurnal Sosbud 3*. 157–162.

Vibriza, J., Rahadi, D. R., Marwan, M. R., & Ahyad. (2017). Perilaku pengguna dan informasi. *Jurnal Gunadarma*, *4*(1), 192–208. https://doi.org/https://doi.org/10.22146/jps.v4i2.28586

Weddiningrum, F. G., Studi, P., Informasi, S., Sains, F., Teknologi, D. A. N., Islam, U., & Ampel, N. (2018). *Deteksi Konten Hoax Berbahasa Indonesia Pada Media Sosial Menggunakan Metode Levenshtein Distance*.

Adi Syafitrah, Februaru 2019, “*[SALAH] Es Krim Magnum Mengandung Babi*”, Online (Turnbackhoax.id), Diakses Maret 2019 ).