KLASIFIKASI ANJURAN, LARANGAN, DAN INFORMASI PADA HADITS MENGGUNAKAN *NAIVE BAYES CLASSIFIER*

Imam Fahrur Rozi[1], M. Hasyim Ratsanjani[2], Riki Ahmad Afandi[3]

[1,2,3] Program Studi Teknik Informatika, Jurusan Teknologi Informasi, Politeknik Negeri Malang
[1]imam.rozi@polinema.ac.id, [2]hsy@polinema.ac.id, [3]rikiahmadafandi@gmail.com

Abstrak— Pertumbuhan informasi khususnya dalam bentuk data teks saat ini sangat pesat sehingga mendorong berbagai penelitian baru khususnya dalam klasifikasi teks, namun dalam penelitian untuk kasus kitab agama masih jarang dilakukan, adapun digitalisasi hanya mencakup bab yang dikandungnya belum menyentuh kategori anjuran, larangan atau sekedar hadits informasi sehingga membuat umat muslim kesulitan dalam memaknainya. Permasalahan mengenai kategori hadits dapat diselesaikan dengan klasifikasi menggunakan metode *naïve bayes classifier*, hal ini dikarenakan *naïve bayes classifier* dapat mengklasifikasikan jumlah data yang sedikit maupun banyak dan lebih dari dua kategori. Klasifikasi pada hadits yang akan terklasifikasikan ada tiga ketegori yaitu kategori anjuran, larangan dan informasi. Proses pengklasifikasian hadits akan dimulai dari pengumpulan data hadits yang sudah terlabeli kemudian akan diproses dalam tahap *preprocessing*, pembobotan, TF-IDF, pengkalsifikasian menggunakan *naïve bayes classifier* dan akan dihitung hasil kalsifikasinya menggunakan metode *f-measure*. Berdasarkan penelitian yang telah dilakukan, untuk mengetahui hasil pengujian terbaik dalam klasifikasi yaitu dengan melihat akurasi tertinggi pada proses perhitungan menggunakan *f-measure*. Pada pengujian yang telah dilakukan dengan beberapa percobaan menggunakan data seimbang dan pengujian menggunakan data tidak simbang, hasil tertinggi yang diperoleh yaitu dengan menggunakan data seimbang dengan *stoplist* dengan akurasi mencapai 79,14%. Hal itu dikarenakan untuk membuktikan bahwa pengumpulan kata pada kamus *stoplist* yang mengandung perawi dan sanat hadits mempunyai pengaruh besar dalam proses pengklasifikasian.

Kata kunci—*Klasifikasi teks, Hadits, Naïve Bayes Classifier*.

I. PENDAHULUAN

Perkembangan hadits saat ini sudah banyak mengalami kemajuan seiring dengan berkembanganya teknologi dan ilmu pengetahuan. Dulunya hadits dikemas dalam bentuk kumpulan buku sehingga jika ingin mencari informasi tentang suatu hadits harus membuka buku dari bebearapa kumpulan buku hadits. Hal ini yang dirasa masih kurang efisien. Dan sekarang hadits sudah dikemas dalam bentuk digital yang dilengkapi dengan menu pencarian sehingga semakin memudahkan dalam pencarian informasi ataupun mempelajari hadits. Selain itu banyak media sosial yang memposting penggalan hadits dari berbagai sumber.

Pada hadits digital yang ada saat ini, kebanyakan hanya berupa terjemahanya tanpa diikuti keteranganya tentang faedah apakah hadits tersebut merupakan larangan, anjuran ataupun hadits informasi. Dan kendalanya, bagi orang yang awam pasti akan kesulitan dalam memaknai hadits dikarenakan ada ribuan hadits yang terdapat dalam kitab-kitab induk. Dengan sedikitnya pengetahuan akan memaknai hadits tersebut tentu ditakutkan menjadi salah pemahaman dan tidak sesuai apa yang diharapkan atau diajarkan oleh Rosullullah. Namun salah satu dari buku hadits yang menjadi perhatian khusus bagi para ulama adalah Targhip wa Tarhib karya Al-Hafizh Ibnu Hajar Al-Asqalani (773 H - 852 H) dimana hadits yang terkandung di dalamnya merupakan kumpulan hadits shohih dari berbagai kitab hadits lainya yang sudah terlabeli antara hadits anjuran dan larangan sehingga memudahkan para pembaca untuk memahaminya dan dapat digunakan sebagai acuan hadits anjuran an larangan.

Maka dari itu dibutuhkan sebuah sistem baru yang tidak hanya mudah untuk digunakan, tetapi juga mampu memberikan informasi dari pertanyaan pengguna secara langsung agar memudahkan pengguna dalam memaknai hadits yang mereka pelajari tetapi masih belum paham akan makna hukum hadits tersebut. Sistem ini dikenal dengan nama *text classification system* atau biasa dikenal dengan sistem klasifikasi text.

Dari beberapa penelitian tentang *text classification system* yang telah dilakukan sebelumnya (Erfian Junianto, 2015) pada Penerapan PSO Untuk Seleksi Fitur Pada Klasifikasi Dokumen Berita Menggunakan NBC, yang menghasilkan akurasi diatas 98% dengan data testing 20% dari data training. Untuk menguji akurasi *Naive Bayes Classificatio*n akan digunakan dalam pengujian Sistem Klasifikasi hadits berdasarkan anjuran, larangan dan informasi untuk mendapatkan jawaban klasifikasi hadits yang sesuai dan lebih akurat, dan titik fokus pada sistem ini untuk implementasi metode naive bayes sebagai metode untuk mendapatkan jawaban.

II. KAJIAN PUSTAKA

A. *Text Mining*

Text mining merupakan proses menambang data yang berupa teks dimana sumber data biasanya didapatkan dari dokumen dan tujuannya adalah mencari kata-kata yang dapat mewakili isi dari dokumen sehingga dapat dilakukan analisa keterhubungan antar dokumen. Tujuan dari *text mining*

adalah mengekstrak informasi yang berguna dari sumber data. Jadi, sumber data yang digunakan pada *text mining* adalah sekumpulan dokumen yang memiliki format yang tidak terstruktur melalui identifikasi dan eksplorasi pola yang menarik. Adapan tugas khusus *text mining* antara lain, pengkategorisasian teks (*text categorization*) dan pengelompokan teks (*text clustering*) [1]. Prosedur utama dalam metode ini terkait dengan menemukan kata-kata yang dapat mewakili isi dari dokumen untuk selanjutnya dilakukan analisis keterhubungan antar dokumen dengan menggunakan metode statistik tertentu seperti analisis kelompok, klasifikasi dan asosiasi. Tahapan dalam text mining secara umum adalah *tokenizing, case folding dan filtering.*

B. *Preprocessing*

Preprocessing merupakan tahapan sebelum proses pengklasifikasian yang diperlukan untuk membersihkan, menghilangkan, mengubah sumber data, baik itu berupa karakter non alfabet maupun kata-kata yang tidak diperlukan. Hal ini bertujuan agar data yang digunakan lebih optimal ketika digunakan pada proses pengklasifikasiannya [2]. Tahapan *preprocessing* setiap kasus dapat berbeda- beda. Berikut merupakan tahapan preprocessing yang digunakan dalam penelitian ini:

1. *Tokenizing*

   Tokenizing adalah proses memecah dokumen menjadi kumpulan kata. *Tokenization* dapat dilakukan dengan menghilangkan tanda baca dan memisahkannya perspasi.

2. *Case Folding*

   Tidak semua dokumen teks konsisten dalam penggunaan huruf kapital. Oleh karena itu, peran *Case Folding* dibutuhkan dalam mengonversi keseluruhan teks dalam dokumen menjadi suatu bentuk standar (biasanya huruf kecil atau *lowercase*).

3. *Filtering*

   filtering adalah tahap mengambil kata-kata penting dari hasil token. Dalam tahap ini bisa digunakakan algoritma stoplist (membuang kata yang kurang penting) atau *wordlist* (menyimpan kata penting).

C. Tf-idf

Term frequncy inverse document frequency (TF•IDF) adalah metode pembobotan yang menggabungkan metode TF dan IDF. Metode ini diusulkan oleh Salton sebagai sebuah kombinasi metode yang dapat

memberikan performansi yang lebih baik, khususnya dalam memperbaiki nilai *recall* dan *precision* (Tokunaga & Iwayama, 1994). Berikut ini merupakan perhitungannya:

$$TF.IDF = TF * \log(N/DF) \quad (1)$$

Keterangan:

TF : Jumlah term tersebut
N : Total dokumen
DF : Jumlah dokumen yang mengandung suatu term

D. *Naïve Bayes Classifier*

Naive bayes classifier merupakan metode probabilistik yang mengasumsikan bahwa keberadaan atribut sebuah kelas tertentu tidak memiliki keterkaitan dengan atribut-atribut lain. Asumsi tersebut dikenal dengan *class conditional independence* yang dibuat untuk memudahkan perhitungan-perhitungan, sehingga perhitungan tersebut dianggap "*naive*", dalam implementasinya maksud naïve tersebut adalah bahwa kemunculan suatu term kata dalam suatu kalimat tidak dipengaruhi kemungkinan kata – kata yang lain dalam kalimat padahal dalam kenyataanya bahwa kemungkinan kata dalam kalimat sangat dipengaruhi kemungkinan keberadaan kata-kata yang dalam kalimat [3].

Salah satu kelebihan penggunaan *naive bayes classifier* adalah metode ini hanya membutuhkan jumlah data latih yang sedikit untuk menentukan estimasi parameter yang diperlukan dalam proses klasifikasi serta menghasilkan akurasi yang tinggi. Sehingga membuat klasifikasi ini prosesnya menjadi cepat dan sederhana. *Naive Bayes* menggunakan konsep dasar *teorema bayes*, yaitu melakukan klasifikasi dengan melakukan nilai probabilitas, berikut adalah persamaan dari *Naïve Bayes*:

$$p(wkj|ci)=\frac{f(wkj,ci)+1}{f(ci)+|w|} \quad (2)$$

$f\ (wkj, ci)$ adalah nilai kemunculan kata w_kj pada kategori ci
$f(ci)$ adalah jumlah keseluruhan kata pada kategori ci
$|w|$ adalah jumlah keseluruhan fitur/kata yang digunakan
Dan

$$p(ci) = \frac{fd\,(ci)}{|D|} \quad (3)$$

$fd(ci)$adalah jumlah opini yang memiliki kategori ci
$|D|$adalah jumlah seluruh training opini
Dibentuk sebuah model probabilistic

$$p(wkj|ci) = \frac{f(wkj,ci)+1}{f(ci)+|w|} \quad (4)$$

Berdasarkan persamaan (3) di atas merupakan model dari *teorema Naive Bayes* yang selanjutnya akan digunakan dalam proses klasifikasi.

E. Hadits

Hadits menurut bahasa yaitu sesuatu yang baru, menunjukkan sesuatu yang dekat atau waktu yang singkat. Hadits juga berarti berita yaitu sesuatu yang diberitakan, diperbincangkan, dan dipindahkan dari seorang kepada orang lain. Hadits menurut istilah *syara'* ialah hal-hal yang datang dari Rasulullah SAW, baik itu ucapan, perbuatan, atau pengakuan (*taqrir*). Berikut ini adalah penjelasan mengenai ucapan, perbuatan, dan perkataan [4].

Hadits *Qauliyah* (ucapan) yaitu hadits hadits Rasulullah SAW, yang diucapkannya dalam berbagai tujuan dan persuaian (situasi). *Hadits Fi'liyah* yaitu perbuatan-perbuatan Nabi Muhammad SAW, seperti pekerjaan melakukan shalat lima waktu dengan tatacaranya dan rukun-rukunnya, pekerjaan menunaikan ibadah hajinya dan pekerjaannya mengadili dengan satu saksi dan sumpah dari pihak penuduh. Hadits *Taqririyah* yaitu perbuatan sebagian para sahabat Nabi yang telah diikrarkan oleh Nabi SAW,

baik perbuatan itu berbentuk ucapan atau perbuatan, sedangkan ikrar itu adakalanya dengan cara mendiamkannya, dan atau melahirkan anggapan baik terhadap perbuatan itu, sehingga dengan adanya ikrar dan persetujuan itu. Bila seseorang melakukan suatu perbuatan atau mengemukakan suatu ucapan dihadapan Nabi atau pada masa Nabi, Nabi mengetahui apa yang dilakukan orang itu dan mampu menyanggahnya, namun Nabi diam dan tidak menyanggahnya, maka hal itu merupakan pengakuan dari Nabi [4]. Keadaan diamnya Nabi itu dapat dilakukan pada dua bentuk:

Pertama, Nabi mengetahui bahwa perbuatan itu pernah dibenci dan dilarang oleh Nabi. Dalam hal ini kadang-kadang Nabi mengetahui bahwa siapa pelaku berketerusan melakukan perbuatan yag pernah dibenci dan dilarang itu. Diamnya Nabi dalam bentuk ini tidaklah menunjukkan bahwa perbuatan tersebut boleh dilakukannya. Dalam bentuk lain, Nabi tidak mengetahui berketerusannya si pelaku itu melakukan perbuatan yang di benci dan dilarang itu. Diamnya Nabi dalam bentuk ini menunjukkan pencabutan larangan sebelumnya [4].

Kedua, Nabi belum pernah melarang perbuatan itu sebelumnya dan tidak diketahui pula haramnya. Diamnya Nabi dalam hal ini menunjukkan hukumnya adalah meniadakan keberatan untuk diperbuat. Karena seandainya perbuatan itu dilarang, tetapi Nabi mendiamkannya padahal ia mampu untuk mencegahnya, berarti Nabi berbuat kesaahan; sedangkan Nabi terhindar bersifat terhindar dari kesalahan [4].

F. *Unified Modeling Language*

Unified Modeling Language adalah bahasa spesifikasi standar yang dipergunakan untuk mendokumentasikan, menspesifikasikan dan membanngun perangkat lunak. UML merupakan metodologi dalam mengembangkan sistem berorientasi objek dan juga merupakan alat untuk mendukung pengembangan sistem [5].

G. *F-measure*

Kinerja dari sebuah sistem klasifikasi dapat diukur dengan menggunakan metode berupa *confusion matrix*. *Confusion matrix* mengandung informasi yang membandingkan hasil klasifikasi yang dilakukan oleh sistem dengan hasil klasifikasi yang seharusnya. Tempat istilah representasi hasil proses klasifikasi pada pengukuran kerja menggunakan *confusion matrix*, yaitu: *True Positives* (TP), *True Negatives* (TN), *False Positives* (FP), dan *False Negatives* (FN) (Silvin, 2019). *True Positives* (TP) merupakan jumlah data positif yang terdeteksi benar, sedangkan *False Negatives* (FN) merupakan jumlah data positif yang terdeteksi sebagai data negatif. *True Negatives* (TN) merupakan jumlah data *negative* yang terdeteksi secara benar, sedangkan *False Positives* merupakan data negatif yang terdeteksi sebagai data positif

Berdasarkan nilai *True Positives* (TP), *True Negatives* (TN), *False Positives* (FP), dan *False Negatives* (FN) dapat diperoleh nilai *precision*, *recall* dan *f-measure*. *Precision*, yang disebut juga *Positive Predictive Value* (PPV), merupakan jumlah data positif yang diklasifikasikan dengan benar dibagi dengan jumlah kelas bernilai positif yang diprediksi. *Precision* membantu jika nilai *false positives* tergolong tinggi. Seperti rumus yang dapat dilihat pada Persamaan (5), *precision* merupakan jumlah dari *True Positives* dibagi dengan jumlah *True Positives* (TP) dan *False Positives* (FP) [6].

$$Precision = \frac{TruePrositives}{TruPrositives + FalsePositives} \quad (5)$$

Recall, yang disebut juga *Sensitivity* atau *True Positive Rate*, menunjukkan persentase data kategori positif yang diprediksi dengan benar oleh sistem. *Recall* merupakan jumlah prediksi positif yang dibagi dengan jumlah kelas yang bernilai positif dalam data *testing*. *Recall* sangat membantu ketika nilai *false negatives* tergolong tinggi. Persamaan (6) menunjukkan persamaan untuk mendapatkan nilai *recall*.

$$Recall = \frac{TruePrositives}{TruPrositives + FalseNegatives} \quad (6)$$

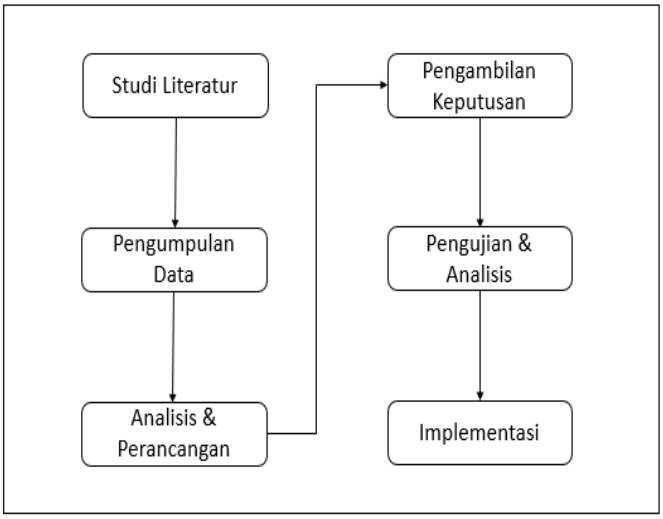
F-measure, yang juga disebut F1 score atau *F-score*, merupakan sebuah ukuran akurasi dari testing yang digunakan untuk menghasilkan rata-rata harmonik berbobot dari *precision* dan *recall*. *F-measure* mencapai nilai terbaiknya, yang berarti *precision* dan *recall* yang sempurna, pada nilai 1. Sedangkan *F-measure* dengan *precision* dan *recall* terendah akan menjadi nilai 0. Persamaan *F-measure* dapat dilihat pada Persamaan 7.

$$F\ Measure = 2 * \frac{1}{\frac{1}{precision} + \frac{1}{recall}} \quad (7)$$

III. METODOLOGI PENELITIAN

A. Tahapan Penelitian

Dalam metodologi penelitian, menjelaskan langkah-langkah yang dilakukan untuk merancang sistem. Metodologi penelitian yang dilakukan dalam penelitian ini melalui beberapa tahap, yaitu studi literature, pengumpulan data, analisa dan perancangan, implementasi sistem, uji coba sistem, dan kesimpulan. Tahapan dalam penelitian ini terdapat pada bagan berikut.



Gambar 1 Tahapan Penelitian

B. Metode Pengumpulan Data

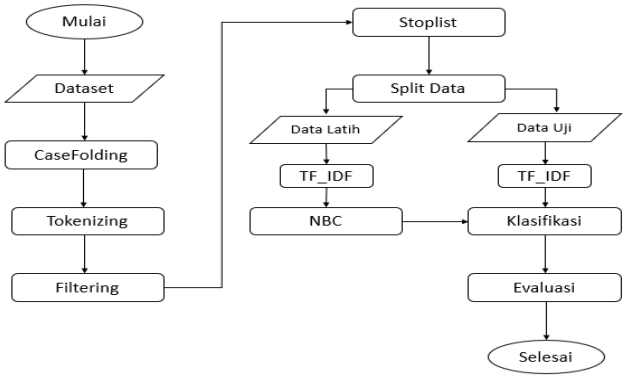
Dalam pengambilan data, Penulis langsung menginputkan data secara manual yaitu berisikan bab tentang hadits, matan, kategori dan asal dari kitab yang

menghimpunnya melalui dua kitab hadits yang menjadi rujukan dan sumber penelitian, dalam hal ini penulis menggunakn kitab At Targhib wa At Targhib dan Bulughul Maram karya Al-Hafizh Ibnu Hajar Asqoloni, karena pada kitab ini sudah terhimpun pelabelan sesuai dengan kelasnya alasan lain peneliti menggunakan kitab tersebut adalah karena di dalam kitab At Targhib wa At Targhib dan Bulughul Maram terdapat hadits dari beberapa sumber utama kitab hadits seperti kitab Hadits Al-Bukhari, Ibnu Majah, Muslim dll. Untuk dataset dalam penelitian ini terdapat tiga kategori yaitu kategori anjuran sebanyak 240 dihimpun dari kitab At Targhib wa At Targhib, larangan 240 hadits dari kitab hadits At Targhib wa At Targhib dan informasi 240 hadits dari kitab Hadits Bulughul Maram sebanyak sihingga terkumpul sebanyak 720 Hadits.

IV. PERANCANGAN

A. Gambaran Umum Sistem

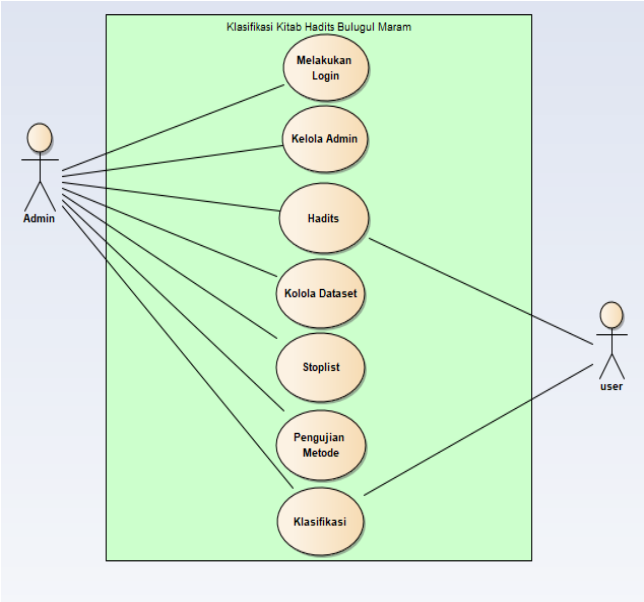
Sistem yang dibangun dalam penelitian ini merupakan sistem untuk mengklasifikasikan teks hadits berdasakan nilai anjuran, larangan dan informasi, klasifikasi dilakukan dengan melalui beberapa tahap seperti pada Gambar 3.1 berikut.



Gambar 2 Gambaran Umum Sistem

B. Analisa Kebutuhan

Use case diagram menggambarkan alursistem yang dirancang untuk memenuhi kebutuhan *fungsional*. Berdasarkan analisis diatas terdapat dua jenis pengguna dan terdapat 6 (enam) kegiatan di dalam sistem, hubungan antara pengguna melakukan suatu kegiatan tersebut diilustrasikan pada Gambar 3



Gambar 3 *Use Case*

Pada *use case* di atas dapat di definisikan bahwa pengguna dapat melakukan proses klasifikasi yang terdiri dari beberapa fitur yaitu:

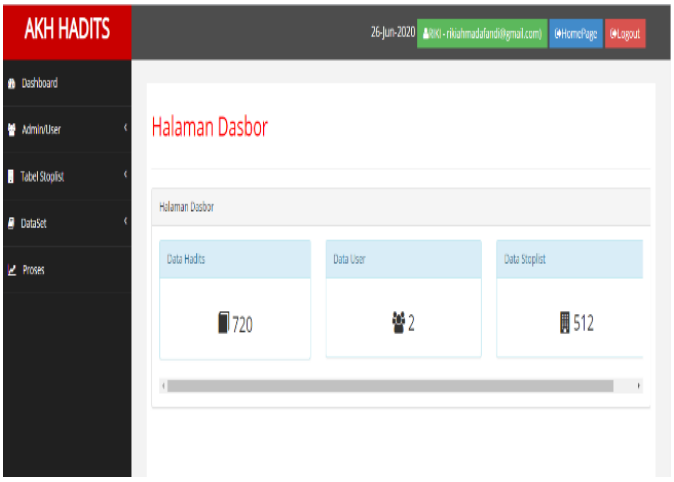
1. Pengguna dapat melakukan akses pada halaman sistem dengan melakukan *login.*
2. Pengguna dapat mengelola data admin untuk proses login.
3. Pengguna dapat mengelola data stoplist untuk proses filtering.
4. Pengguna dapat mengelola dataset untuk dijadikan dataset dalam proses klasifikasi.
5. Pengguna dapat melakukan pengujian dan pembelaaran pada metode untuk membagi data latih dan data uji.
6. Pengguna dapat melakukan klasifikasi hadits untuk mengetahu akurasi pada klasifikasi.
7. Pengguna dapat melakukan klasifikasi hadits untuk mengetahu akurasi pada klasifikasi.

V. IMPLEMENTASI DAN PENGUJIAN

A. Implementasi Tampilan Sistem

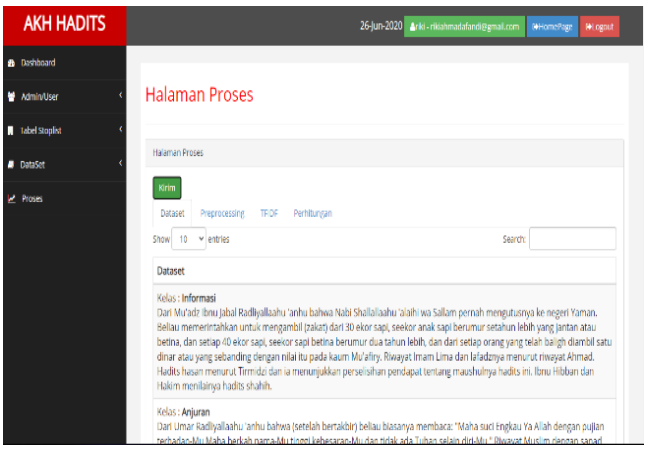
Pada implementasi ini merupakan hasil antarmuka sistem yang sudah terrealisasikan berikut adalah implementasi antar muka.

Tampilan pada menu utama adalah terdapat tabel dengan jumlah data yang ada pada kamus filtering, dataset dan data admin, dan menu pada tampilan ini terdapat 5 menu pada navbar, seperti pada Gambar 4



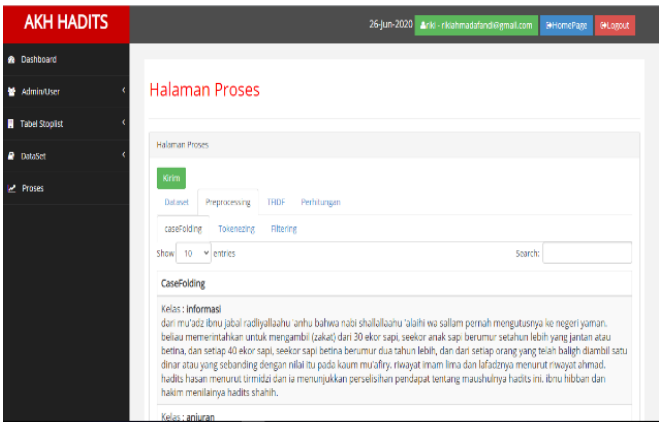
Gambar 4 Halaman Utama

Pada halaman *get dataset* merupakan proses untuk memaggil *dataset* yang terdapat pada menu atau database *dataset* untuk dijadikat data klasifikasi dimana pada get *dataset* dilakukan secara acak untuk proses pengujian.



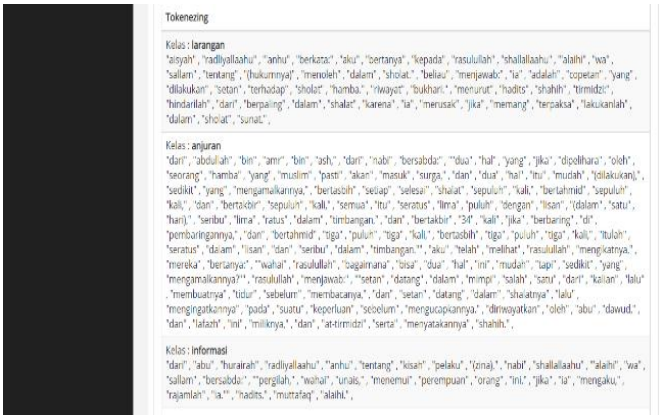
Gambar 5 Halaman *Get Dataset*

Pada menu *case folding* merupakan proses perubahan huruf kapital menjadi kecil atau *lowers* untuk mempermudah dahan proses pengklasifikasian untuk hasil dari tampilan proses case folding terdapat pada Gambar 6



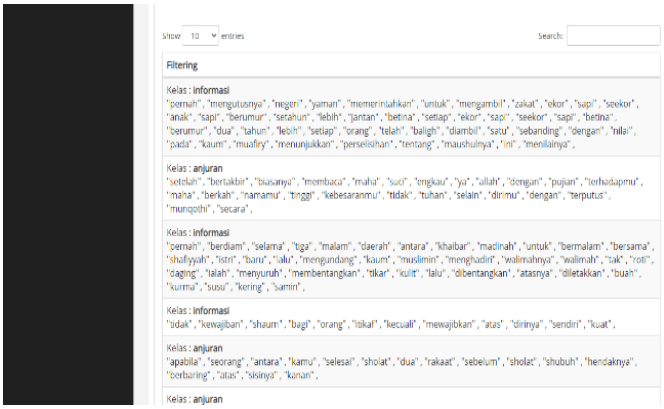
Gambar 6 Halaman *Preprocessing*

Pada proses atau tampilan halaman *tokenizing* merupakan pemisahan dari rangkaian menjadi rangkaian perkata sebelum penghilangan fitur kata yang tidak memiliki bobot pada proses *filtering*, seperti pada Gambar 7 yang tertera di bawah ini.



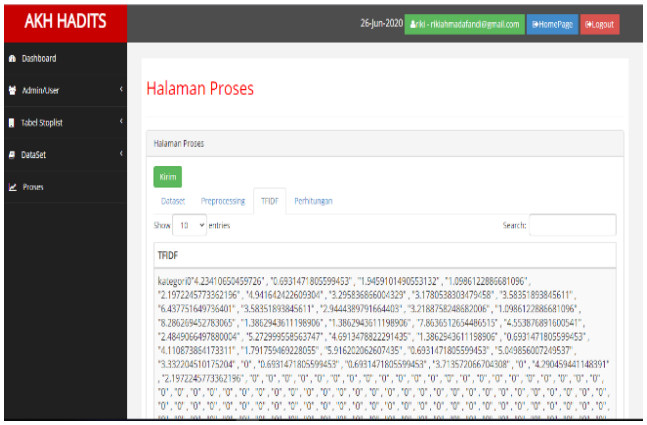
Gambar 7 Halaman *Tokenizing*

Pada halaman *filtering* merupakan proses penghilangan kata yang dianggap tidak memiliki bobot yang terkonfigurasi pada kamus data stoplist untuk mengetahui bobot pada setiap kata sehingga akan terbentuk pola hadits atau *dataset* berdasarkan kategori yang dikandungnya.



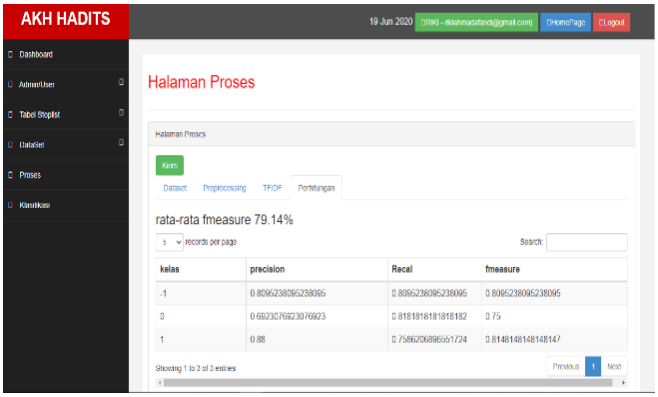
Gambar 8 Halaman *Filtering*

Pada Halaman tf-idf merupakan pembobotan untuk mencari metrik sebelum data diolah menggunakan metode *naïve bayes classifier*. Sehingga pada tiap kata mengandung bobot yang di bawanya untuk proses pengklasifikasian.



Gambar 9 Halaman TF-IDF

Pada Halaman pengujian atau perhitungan merupakan hasil akhir yang diperoleh dalam proses pengklasifikasian untuk mengetahui akurasi yang diperoleh pada tahap pengujian metode untuk halaman hasil perhitungan sendiri memiliki tabel *f-measure* sebagai metode pengkur akurasi dalam proses klasifikasi menggunakan *naïve bayes classifier* pada penelitian ini. Sehingga hasil akurasi akan ditampilkan berdasarkan rata-rata nilai akurasi perkelas menggunakan *preccision*, *recall* dan ditentukan hasil akhir berupa nilai rata-rata *f-measure.*



Gambar 10 Halaman Perhitungan

B. Pengujian Akurasi Sistem

Pada proses ini dilakukan pengujian dengan porsi 90:10 untuk data latih dan data uji. Pengujian klasifikasi anjuran, larangan dan informasi dilakukan sebanyak dua kali pada masing – masing jenis data, yaitu data dengan porsi kelas yang seimbang, dan data dengan porsi kelas yang tidak seimbang. Untuk pengujian *dataset* seimbang digunakan data sebanyak 720 dengan pembagian 240 untuk kelas anjuran, 240 kelas larangan, 240 untuk kelas informasi. Untuk pengujian *dataset* tidak seimbang digunakan data sebanyak 720 dengan pembagian anjuran 250, larangan 275, dan informasi 195. Pengujian dilakukan dengan data yang melewati proses *filtering* dan tanpa *filtering*. Pengujian dilakukan dengan klasifikasi *naïve bayes* dengan menggunakan *f-measure*.

1) Hasil Pengujian Data Seimbang Dengan *Stoplist*

Pada tabel dibawah menjelaskan tentang perhitungan yang diperoleh setelah melakukan pengujian klasifikasi menggunakan *naïve bayes classifier* menggunakan *dataset* seimbang dengan *stoplis* sebanyak 720 data dengan proporsi data split 90:10 atau 640 data training dan 72 data *testing* sehingga diperoleh hasil pengujian seperti pada Tabel 1.

TABEL 1 HASIL PENGUJIAN DATA SEIMBANG DENGAN *STOPLIST*

Kategori	*Precision*	*Recall*	*F-Measure*
Anjuran	80.95%	80.95%	80.95%
Informasi	69.23%	81.81%	75%
Larangan	88%	75.86%	81.48%

2) Hasil Pengujian Data Seimbang Tanpa *Stoplist*

Pada tabel ini menjelaskan tentang perhitungan yang diperoleh setelah melakukan pengujian klasifikasi menggunakan *naïve bayes classifier* menggunakan data seimabang tanpa mengunakan kamus *stoplis* dengan jumlah data yang sama ketika melakukan pengujian dengan *dataset* seimbang dengan *stoplist* yaitu 720 data dengan proporsi *split* data 90:10 yaitu 648 data untuk data training dan 72 data untuk data *testing* sehingga diperoleh hasil sebagai berikut.

TABEL 2 HASIL PENGUJIAN DATA SEIMBANG TANPA *STOPLIST*

Kategori	Precision	Recall	F-Measure
Anjuran	81,81%	75,00	78,26
Informasi	68,18%	83,33	74,99
Larangan	82,14%	76,66	79,31

3) Hasil Pengujian Data Tidak Seimbang Dengan *Stoplist*

Pada tabel ini menjelaskan tentang perhitungan yang diperoleh setelah melakukan pengujian klasifikasi menggunakan *naïve bayes classifier* dengan jumlah data 720 data, dengan proporsi split data 90:10 yaitu 648 data untuk data *training* dan 72 data untuk data testing sehingga diperoleh hasil sebagai berikut.

TABEL 3 HASIL PENGUJIAN DATA TIDAK SEIMBANG DENGAN *STOPLIST*

Kategori	Precision	Recall	F-Measure
Anjuran	74,07%	74,07%	74,07%
Informasi	61,11%	78,57%	68,75%
Larangan	81,48%	70,96%	7586%

4) Hasil Pengujian Data Tidak Seimbang Tanpa *Stoplist*

Pada tabel ini menjelaskan tentang perhitungan yang diperoleh setelah melakukan pengujian klasifikasi menggunakan *naïve bayes classifier* dengan jumlah data 720 data, dengan proporsi split data 90:10 yaitu 648 data untuk data training dan 72 data untuk data testing sehingga diperoleh nilai precission, recall dan fmeasure seperti pada Tabel 5.3.

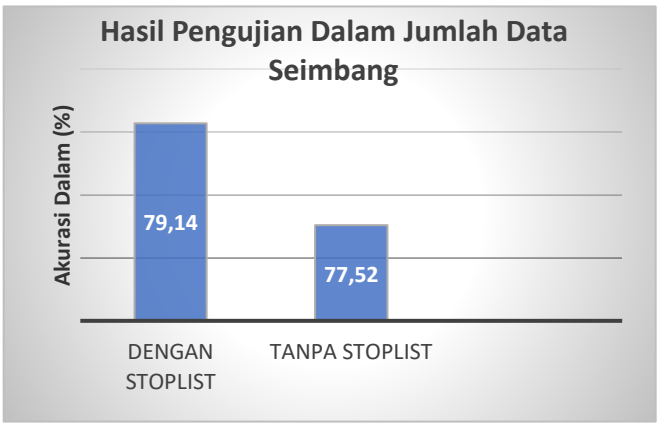
TABEL HASIL PENGUJIAN DATA TIDAK SEIMBANG DENGAN *STOPLIST*

Kategori	Precision	Recall	F-Measure
Anjuran	81,81%	62,06%	70,58%
Informasi	52,17%	92,30%	66,66%
Kategori	*Precision*	*Recall*	*F-Measure*
Larangan	77,77%	70.00%	73,68%

VI. HASIL DAN PEMBAHASAN

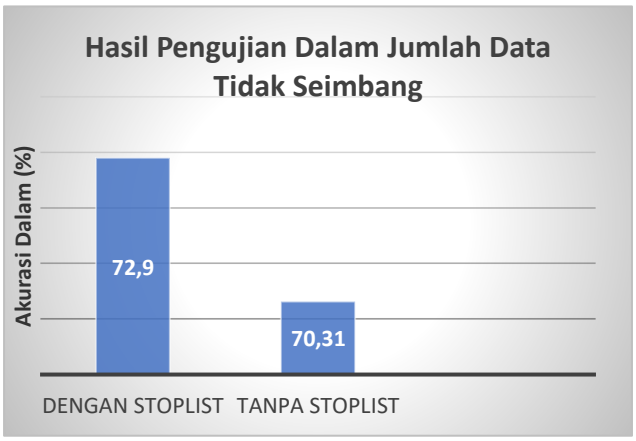
Hasil dari penelitian ini diperoleh untuk mengukur keefektifan klasifikasi hadits berdasarkan kategori yang sudah ditentukan, dari beberapa percobaan menggunakan data seimbang diperoleh hasil yang paling tinggi dibandingkan dengan pengujian dengan menggunakan data tidak seimbang, pengujian dengan data seimbang sendiri terdapat dua kali pengujian yaitu dengan menggunakan *stoplist* dan tanpa menggunakan stoplis dengan proporsi data

720 data maka dari hasil pengujia dibawah dapat disimpulkan bahwa pengujian dengan menggunakan data seimbang dengan *stoplist* akurasinya lebih tinggi yaitu mencapai 79,14 % seperti dapat dilihat pada Gambar 11



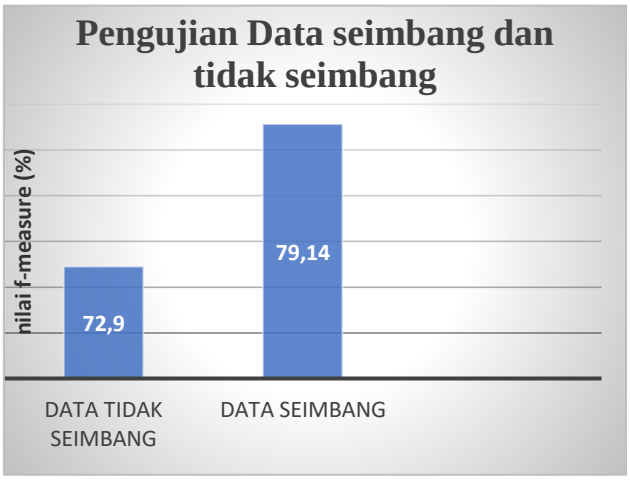
Gambar 11 Grafik Hasil Pengujian Data Seimbang

Selanjutnya parameter yang juga diuji adalah menggunakan komposisi data tidak seimbang masing-masing kelas yaitu dengan 720 data dibagi 250 untuk kategori anjuran, 270 untuk kategori larangan dan 195 untuk kategori informasi pengujian juga dilakukan dengan menggunakan data stoplist dan pengujian tanpa menggunakan data *stoplist*, pada pengujian menggunakan data seimbang perolehan hasil pengujian terlihat grafik yang menurun yaitu hanya mencapai hasil tertinggi 72,90% pada pengujian data tidak simbang menggunakan *stoplist*, sehingga hasil akurasi dapat dilihat sebagai berikut.



Gambar 12 Grafik Hasil Pengujian Data Tidak Seimbang

Selanjutnya parameter yang juga diuji adalah menggunakan komposisi data tidak seimbang dan menggunakan data tidak seimbang, dengan data seimbang pada kategori anjuran 240, kategori larangan 240 dan kategori informasi 240, dan pada kelas data seimbang dengan jumlah data yang sama pada kelas seimbang yaitu 720 tetapi dengan pembagian pada tiap kategori tidak merata yaitu 275 untuk kelas larangan, 250 untuk kelas anjuran dan 195 untuk kelas informasi maka dihasilkan nilai tertinggi yaitu pengujian dengan menggunakan data seimang yang mencapai 79,14%, seperti pada Gambar 6.3.



Gambar 12 Grafik Hasil Pengujian Data Seimbang dan Data Tidak Seimbang

VII. KESIMPULAN DAN SARAN

A. Kesimpulan

Adapun Kesimpulan Berdasarkan hasil pengujian dan pembahasan dari klasifikasi anjuran, larangan dan informasi hadits menggunakan metode *naïve bayes classifier*, maka diperoleh kesimpulan sebagai berikut:

1. Data Hadits dapat terdigitalisasikan dari kumpulan kitab hadits dan berdasarkan per kategori yang dikandungnya sehingga mumudahkan dalam proses klasifikasi.
2. Proses klasifikasi hadits dilakukan dengan lebih dari satu percobaan dengan proporsi data seimbang dan menggunakan proporsi data seimbang untuk mengetahui hasil yang paling tinggi pada proses klasifikasi.
3. Penerapan pengujian klasifikasi menggunakan data *stoplist* mempengaruhi hasil performa klasifikasi karena pada kata perawinya dan sanad pada setiap hadits baik anjuran, larangan dan informasi yang memungkinkan sama baik di dalam hadits anjuran, larangan maupun informasi.
4. Penggujian dengan menggunakan data seimbang 720 data hadits mengggunakan stoplist menghasilkan akurasi yang paling tinggi mencapai 79,14% dibanding dengan pengujian yang lain, pengujian menggunakan data seimbang tanpa *stoplist* mencapai 77,52%, pengujian menggunakan data tidak seimbang dengan *stoplist* mencapai 72,90% dan pengujian menggunakan data tidak seimbang tanpa stoplist cenderung memiliki hasil pengujian paling rendah yaitu 70,31% dengan jumlah data seimbang mengunakan proporsi 720 data dan pembagian data split 90:10 648 untuk data training 72 untuk data testing, sedangkan pada pengujian data tidak seimbang menggunakan proporsi data yang sama 720 dengan pembangian 250 untuk kategori anjuran, 275 untuk kategori larangan dan 195 untuk kategori informasi pada skala data split 90:10, 648 untuk data training dan 72 untuk data testing.

B. Saran

Penelitian tentang klasifikasi anjuran, larangan dan informasi hadits menggunakan metode *naïve bayes classifier* dengan beberapa saran sebagai berikut:

1. Pada penelitian ini sebaiknya menggunakan dataset yang lebih banyak untuk melihat atau memperoleh akurasi yang lebih tinggi karena jumlah dataset yang lebih banyak akan mempengaruhi akurasi permormasi dari klasifikasi nmenggunakan naïve bayes.
2. Diharapkan penelitian yang akan datang menggunakan berbagai macam percobaan lagi dengan menambahkan *k-fold croos validation* untuk mengetahui bobot data training yang baik dan sesuai sehingga akan diharapkan memilihi akurasi yang lebih tinggi dari penelitian yang sudah ada.

DAFTAR PUSTAKA

[1] Ervita , P. K., & Tedy , S. (2014). PENERAPAN TEXT MINING PADA SISTEM KLASIFIKASI. *Jurnal Sarjana Teknik Informatika Volume 2 Nomor 3*, 11.

[2] Firdaus , M. A., & Adam , B. M. (n.d.). IMPLEMENTASI TEKS MINING PADA APLIKASI PENGAWASAN. *Jurnal Ilmiah Komputer dan Informatika (KOMPUTA)*, 8.,

[3] Syair , S. A., Said , A., & Danang , T. M. (2017). KLASIFIKASI ANJURAN, LARANGAN, DAN INFORMASI PADA HADITS SHAHIH BUKHARI MENGGUNAKAN NAÏVE BAYES CLASSIFIER. *e-Proceeding of Engineering*, 9.

[4] JAMARIL, S. (2017). PENGERTIAN, KEDUDUKAN DAN FUNGSI HADITS. *https://sumbar.kemenag.go.id/v2/post/1952/pengertian-kedudukan-dan-fungsi-hadits.html*.

[5] Elissa, "Contoh Ade , H. (2016). PEMODELAN UML SISTEM INFORMASI MONITORING PENJUALAN DAN STOK. *JURNAL KHATULISTIWA INFORMATIKA, VOL. IV, NO. 2*, 10.

[6] Silvin. (2019). *Analisis Sentimen Media Twitter Menggunakan Long Short-Term Memory Recurrent Neural Network (Studi Kasus: Prediksi Polling Kandidat Presiden Indonesia 2019).* Tangerang: Universitas Multimedia Nusantara.