

SISTEM INFORMASI RETRIEVAL KETERKAITAN ANTARA HADITS ARBAIN DAN ALQURAN (TERJEMAHAN INDONESIA) DENGAN VECTOR SPACE MODEL

¹⁾ Imam Amirulloh ²⁾ Yanti Apriyani ³⁾ Taufik Wibisono

⁴⁾ Melisa Winda Pertiwi ⁵⁾ Mira Kusmira

^{1,2,3)} Program Studi Sistem Informasi, Universitas Bina Sarana Informatika Kampus Tasikmalaya
Jl. Tanuwijaya No.4, Empangsari, Kec. Tawang, Kab. Tasikmalaya, Jawa Barat 46113

^{4,5)} Program Studi Sistem Informasi, Teknologi Informasi, Universitas Nusa Mandiri,
Jl. Jatiwaringin Raya No.2, Jakarta Timur 13620, Indonesia

e-mail: imam.iau@bsi.ac.id¹⁾ yanti.ynp@bsi.ac.id²⁾ taufik.tik@bsi.ac.id³⁾
melisa.mwp@bsi.ac.id⁴⁾ mira.mik@bsi.ac.id⁵⁾

ABSTRAK

Al-Quran dan Al-Hadits mempunyai keterkaitan saling melengkapi dan menjelaskan satu sama lain. Atas dasar tersebut dalam menjelaskan isi Al-hadits perlu di lihat juga dari pandangan isi Al-Quran begitu juga sebaliknya. Namun dengan jumlah data Al-Quran dan Al-Hadits yang sangat besar berdampak pada lamanya eksekusi pencarian data dari kedua kitab tersebut terutama pencarian dari Ayat Al-Quran ke Hadits yang membutuhkan waktu yang lama dan ketepatan dalam pencarian. Untuk memecahkan permasalahan ini, ilmu penelitian sistem informasi retrieval bisa menjadi solusi, salah satu bagian dari ilmu sistem informasi retrieval adalah Pre-Processing dengan Model Vector Space Model (VSM) dan untuk pengklasifikasian menggunakan metode Cosine Similarity, Dice dan Jaccard sehingga dapat membantu dalam menemukan hasil pencarian dengan tingkat ketepatan yang baik. Berdasarkan hasil penelitian tersebut Cosine, Dice Dan Jaccard mempunyai hasil yang sama berdasarkan rata-rata recall, precision dan accuracy secara keseluruhan yaitu dengan nilai recall 44,27 %, precision 1,8 %, dan accuracy 94 %.

Kata Kunci: Al-Quran, Al-Hadits , Sistem Informasi Retrieval, Vector Space Model.

ABSTRACT

The Al-Quran and Al-Hadith have complementary relationships and explain each other. On this basis, in explaining the contents of the Al-Hadith, it is necessary to look into the contents of the Al-Quran and vice versa. However, the large amount of Al-Quran and Al-Hadith data impacts the execution time of data searching from the two sources, especially searching Al-Quran verses based on Al-Hadith which requires much time and accuracy. Regarding this problem, we propose information retrieval system research as a solution. One part of the information retrieval system is Pre-Processing using Vector Space Model (VSM) and classification using Cosine Similarity, Dice, and Jaccard methods. That method can assist in finding search results with a great degree of accuracy. Based on the results of this study Cosine, Dice and Jaccard have the same results based on the average recall, precision, and overall accuracy, with a recall value of 44.27%, 1.8% precision, and 94% accuracy

Keywords: Al-Quran, Al-Hadith , System Information Retrieval, Vector Space Model.

I. PENDAHULUAN

P Kehidupan beragama menjadi bagian terpenting bagi manusia dalam menjalankan kehidupannya, adanya aturan-aturan dan kewajiban untuk manusia yang sudah tercantum dalam kitab agamanya masing-masing yang harus dijalankan dalam kehidupan sehari-hari. Islam merupakan salah satu agama yang ada dan diakui oleh dunia, dengan penganut agama yang paling banyak dibanding dengan agama yang lain terutama di Indonesia. Agama islam mempunyai panduan bagi penganutnya dalam menjalankan hidupnya yaitu Al-Quran dan Al-hadits kedua kitab ini sudah menjadi rujukan bagi orang yang menganut agama islam.

Al-Quran dan Al-Hadits mempunyai keterkaitan, keduanya saling melengkapi dan menjelaskan satu sama lain. Atas dasar tersebut dalam menjelaskan isi Al-hadits perlu di lihat juga dari pandangan isi Al-Quran begitu juga sebaliknya. Namun dengan jumlah data Al-Quran dan Al-Hadits yang sangat besar berdampak pada lamanya eksekusi pencarian data dari kedua kitab tersebut terutama pencarian dari Ayat Al-Quran ke Hadits yang membutuhkan waktu yang lama dan ketepatan dalam pencarian dikarenakan volume data Al-Quran lebih banyak dibanding volume data Al-Hadits.

Teknologi yang semakin berkembang, permasalahan tersebut sudah bukan menjadi suatu yang sulit terutama dalam penyimpanan data Al-Quran dan Al-Hadits yang sangat banyak bisa di simpan dalam bentuk elektronik dan digital. Sudah banyak pengembangan perangkat lunak di dunia telah mencoba mendigitalkan Al-

Qur'an dan Al-Hadits, akan tetapi belum adanya yang mendigitalkan langsung keterkaitan Al-Quran terhadap Hadits dengan memperhatikan tingkat ketepatan dalam eksekusi hasil pencarian keterkaitannya.

Untuk memecahkan permasalahan ini, ilmu penelitian sistem informasi retrieval bisa menjadi solusi, salah satu bagian dari ilmu sistem informasi retrieval adalah Pre-Processing, Model Vector Space Model (VSM) dan untuk pengklasifikasian menggunakan metode Cosine Similarity, Dice dan Jaccard, dari ketiga metode tersebut akan dibandingkan dan dicari metode terbaik, sehingga dapat membantu dalam menemukan hasil pencarian dengan tingkat ketepatan yang baik.

II. TINJAUAN PUSTAKA

Ada beberapa penelitian yang dilakukan sebelumnya yang ada keterkaitan serta dijadikan landasan dari penelitian ini, berikut literatur yang sudah dikumpulkan beserta hasil dari penelitiannya tersebut diantaranya:

Broto Poernomo, 2015. Sistem information retrieval pencarian kesamaan ayat terjemahan al quran berbahasa indonesia dengan query expansion dari tafsirnya. Hasil dari penelitian tersebut stemming pada tahapan pre-processing menggunakan algoritma Nazief Adriani menghasilkan akurasi 95% untuk teks terjemahan Al Quran dan tafsir berbahasa Indonesia. Kegagalan pembentukan kata dasar pada tahap ini dikarenakan masih adanya kata dalam bahasa Arab yang tidak diterjemahkan ke bahasa Indonesia.

Ogie Nurdiana, Jumadi, Dian Nursantika, 2016. Perbandingan metode cosine similarity dengan metode jaccard similarity pada aplikasi pencarian terjemah al-qur'an dalam bahasa indonesia. Metode cosine, jaccard dan k-nearest neighbor (K-NN) yang digunakan pada proses klasifikasi dokumen teks dengan hasil akhir dari percobaan 33 kali dengan key yang berbeda dan total 6326 dokumen di dapat metode cosine yang nilai kemiripannya tertinggi yaitu 41% dari metode jaccard 19% dan k-nearest neighbor (K-NN) 40%.

Evana Ainul, Faturahman, Zainal Abidin. Rancang bangun search engine terjemahan tafsir ayat-ayat al-quran pada dokumen teks berbahasa indonesia menggunakan metode dice similarity. 30 dokumen yang telah diuji coba, ada 80% atau 24 dokumen yang ditemukan padanan ayatnya dan sebanyak 20% atau 6 dokumen tidak ditemukan padanan ayatnya.

Julham Makatita, 2016. Aplikasi kumpulan doa harian bersumber al-qur'an dan hadits berbasis android. Pada paper hanya memaparkan pembuatan sebuah aplikasi android untuk doa harian yang berdasarkan pada Al-Quran dan Al-Hadits tanpa melakukan text mining.

Geehan Sabah Hassan, 2015. Mengkategorikan Tafsir Alquran menggunakan Algoritma K-NN. Text Categorization menjadi langkah dalam pelabelan pada bahasa text alami dengan menggunakan KNN Algoritma, hasil yang didapat Algoritma K-NN menunjukan hasil yang baik.

Suhaib Kh. Hamed and Mohd Juzaidin Ab Aziz, 2016. Sistem penjawab pertanyaan pada terjemahan alquran berdasarkan Teknik Ekspansi Query dan klasifikasi neural network, dengan hasilnya Klasifikasi dengan menggunakan NN telah menunjukkan tingkatan sekitar 90% dan evaluasi dari pendekatan yang diusulkan dari penelitian ini berdasarkan keseluruhan QAS telah menunjukkan tingkat sekitar 87%. Ini menunjukkan bahwa QAS berhasil memberikan hasil yang menjanjikan di bidang kritis ini.

Kaouther Faïdi, 2015. Membandingkan Arabaci NLP Tools untuk klasifikasi hadits, dari penelitian ini didapatkan hasil Khoja's stemmer mengungguli alat lainnya dan pengelompokan SVM itu mencapai keakuratan tertinggi diikuti oleh pembimbing Naïve Bayes, dan masing-masing pembuat pohon pengambilan keputusan.

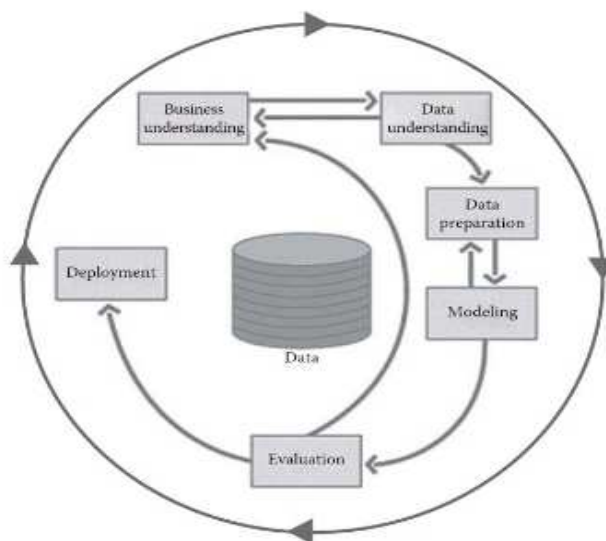
Muhammad Syarif, 2017. Implementasi Algoritma String Matching Dalam Pencarian Surah Dan Ayat Dalam Al-Quran Berbasis Web, dari hasil penelitiannya tersebut berkesimpulan implementasi algoritma string matching berbasis web membuat orang – orang lebih mudah dalam mencari kata atau statement pada terjemahan alquran.

Tutik Khotimah, 2014. Pengelompokan surat dalam al qur'an menggunakan algoritma k-means. Dari percobaan yang dilakukan, nilai optimal diperoleh pada iterasi ke-14 dengan nilai k sebanyak 4 cluster. Cluster_0 terdiri dari 18 anggota dengan nilai centroid pada variabel ayat sebesar 106,389, variabel ruku' sebesar 9,778, dan variabel lama sebesar 1333,5. Cluster_1 terdiri dari 26 anggota dengan nilai centroid pada variabel ayat sebesar 62,038, variabel ruku' sebesar 4,885, dan variabel lama sebesar 662,5. Cluster_2 memiliki 7 anggota dengan nilai centroid pada variabel ayat sebesar 183,143, variabel ruku' sebesar 22,857 dan variabel lama sebesar 3310,571. Sedangkan cluster_3 memiliki 63 anggota dengan nilai centroid pada variabel ayat sebesar 22,635, variabel ruku' sebesar 1,508, dan variabel lama sebesar 145,635.

Mohammad Alhawarat, 2015. pemrosesan text pada alquran studi text mining, Hasilnya menunjukkan beragam karakteristik Alquran seperti kata-kata yang paling penting, kata-katanya dan bab-babnya dengan frekuensi tinggi. Semua hasil ini didasarkan pada frekuensi awal yang dihitung dengan menggunakan metode Frekuensi Frekuensi-Inverse Document Frequency (TF-IDF)

III. METODOLOGI PENELITIAN

CRISP-DM dimulai pada tahun 1996 dan merupakan pola umum yang berlaku untuk proyek data mining manapun. Peta CRISP-DM dari proses CRISP-DM umum menjadi proses aplikasi tertentu. Intinya, model proses menggambarkan siklus hidup proses data mining yang terdiri dari enam langkah dasar (Vivake Kale, 2016).



Gambar 1. Metode CRISP-DM

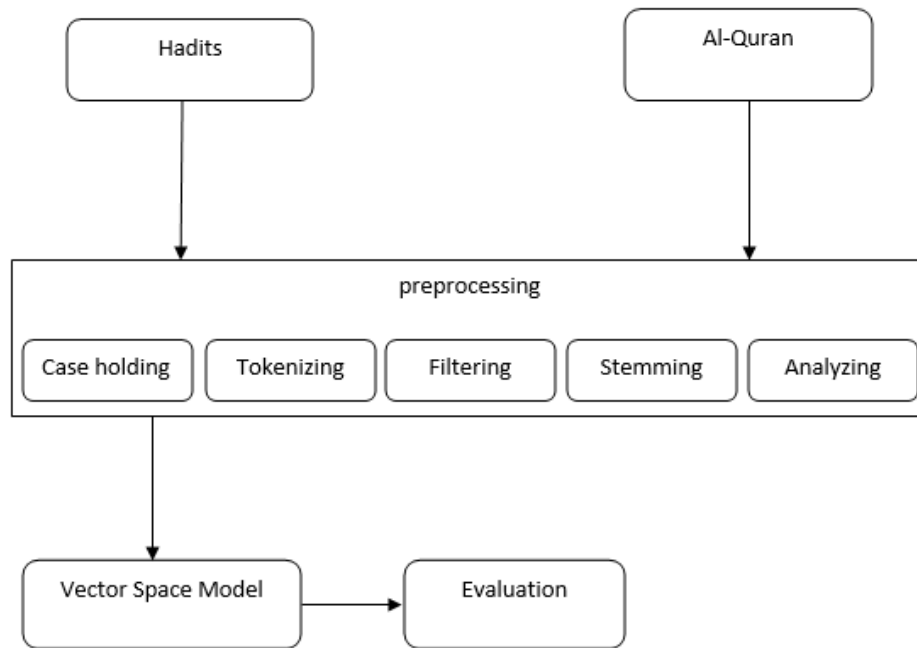
3.1. Metodologi

Berdasarkan metodologi yang dipakai yaitu CRISP-DM ada beberapa tahapan antara lain:

- Business Understanding
Pada tahapan pertama ini dilakukannya proses mencari pengetahuan mengenai hadits dan quran beserta kandungan ayat didalamnya, yang satu sama lain akan dikaitkan mana saja ayat alquran yang ada hubungannya dengan salah satu hadits yang dimaksud.
 - Data Understanding
Tahapan kedua ini mengumpulkan hadits-hadits yang ada dalam kita hadits arbain yang berjumlah 42 beserta ayat-ayat alquran yang berjumlah kurang lebih 6236 ayat.
 - Data Preparation
Tahapan ini akan dilakukan proses pre-processing terhadap dokumen alquran dan beberapa tema hadits
 - Modelling
Dari hasil data preparation tersebut akan dilakukan proses perhitungan vector space model sehingga dapat diketahui nilai qosin dari masing – masing dokumen.
 - Evaluation
Pada tahapan ini dilakukannya proses perbandingan dan ranking dari nilai hasil metode Cosine Simlarity, Dice dan Jaccard sehingga dapat diketahui ayat alquran mana saja yang paling mendekati atau ada kaitannya dengan hadits yang dimaksud dan metode mana yang mempunyai performa terbaik
 - Deployment
Pada tahapan ini akan dibuat aplikasinya dalam berbasis web.
- Kerangka Kerja Penelitian

3.2. Kerangka Kerja Penelitian

Ada dua pekerjaan utama yang dikerjakan oleh Information Retrieval, yaitu melakukan pre-processing terhadap database dan kemudian menerapkan metode tertentu untuk menghitung kedekatan (relevansi atau similarity) antara dokumen di dalam database yang telah di-preprocess dengan query pengguna, berikut kerangka kerja sistemnya:



Gambar 2. Kerangka Kerja Sistem

1. Query / Text Document

Pada bagian ini menentukan Query dan Text Document, Query pada penelitian ini adalah Hadits, sedangkan untuk text document adalah Alquran, kedua data tersebut akan dilakukan proses pre processing

2. Pre Processing

Pada tahapan akan ada beberapa tahapan yang dilakukan diantaranya:

a. Case Holding

Merupakan tahapan dimana setiap kata pada document akan disamakan besar kecilnya sebuah term

b. Tokenizing

proses pemotongan string masukan berdasarkan tiap kata yang menyusunnya. prinsipnya proses ini adalah memisahkan setiap kata yang menyusun suatu dokumen, pada tiap satu ayat al-quran dan hadits ini akan dilakukan pemisahan menjadi setiap kata.

c. Filtering

proses dimana diambil sebagian sinyal dari frekwensi tertentu, dan membuang sinyal pada frekwensi yang lain. Filtering pada citra juga menggunakan prinsip yang sama, yaitu mengambil fungsi citra pada frekwensi-frekwensi tertentu dan membuang fungsi citra pada frekwensi-frekwensi tertentu

d. Stemming

proses untuk menemukan kata dasar dari sebuah kata. Dengan menghilangkan semua imbuhan (*affixes*) baik yang terdiri dari awalan (*prefixes*), sisipan (*infixes*), akhiran (*suffixes*) dan *confixes* (kombinasi dari awalan dan akhiran) pada kata turunan. Stemming digunakan untuk mengganti bentuk dari suatu kata menjadi kata dasar dari kata tersebut yang sesuai dengan struktur morfologi Bahasa yang baik dan benar. Pada penelitian ini setiap kata arab dalam al-quran dan al-hadits akan dilakukan penghapusan imbuhan, awalan dan akhiran contohnya didalam kata arab terdapat huruf arab alif lam, iya, nun, dll akan di hilangkan sehingga terbentuk menjadi kata dasar.

3. Vector Space Model

Vector space model adalah suatu model yang digunakan untuk mengukur kemiripan antara suatu dokumen dengan suatu query. *Query* dan dokumen dianggap sebagai vektor-vektor pada ruang n-dimensi, dimana n adalah jumlah dari seluruh term yang ada dalam leksikon. Leksikon adalah daftar semua term yang ada dalam index

4. Evaluation

Pada tahapan ini akan dievaluasi dengan melihat *ranking* hasil dari perhitungan cosim dari setiap query Al-Quran dan Al-Hadits

IV. HASIL DAN PEMBAHASAN

Berdasarkan metodologi yang dipakai pada penelitian ini dan juga telah dibahas pada bab III, kemudian pada bab ini akan dipaparkan tentang implementasi metodologi CRISP-DM tersebut ke dalam penelitian ini, berikut pemaparannya:

1. Business Understanding

Untuk mendapatkan pengetahuan tersebut maka teknisnya dibuat ke dalam bentuk pertanyaan – pertanyaan sebagai berikut:

- Apa tujuan dari penelitian ini? Jawabannya penelitian ini dilakukan untuk mengetahui klasifikasi ayat Al-Quran berdasarkan kategori Tema Hadits Arb'in dengan teknik sistem informasi retrieval.
- Mengapa memilih hadits Arba'in sebagai kategori pengelompok ayat Al-Quran? Hadits Arba'in hadits merupakan kitab hadits yang pertama kali banyak dikaji oleh orang – orang yang baru mempelajari hadist.
- Seberapa tepatkah metode sistem informasi retrieval yang akan dipakai untuk pengelompokan ayat A-Quran? Untuk mendapatkan jawaban perlu dilakukan penelitian ini dengan pendekatan metode Vector Space Model

2. Data Understanding

Sesuai dengan judul penelitian, yang menjadi objek data adalah Al-Quran dan Al-hadits, berikut pemaparan data-data yang akan dijadikan objek:

- Al-Quran, data – data quran di ambil dari website www.qurandatabase.org dengan format file database SQL, alquran terdiri dari 114 surat dan 6236 ayat
- Al-Hadits, Dari sekian banyak kitab – kitab hadits yang ada, pada penelitian ini memilih salah satu hadits yaitu hadis arba'in yang mempunyai jumlah hadits sebanyak 42 hadits dengan tema yang berbeda.

3. Data Preparation

proses pengolahan terhadap dokumen dan queri dengan melakukan preprocessing terhadap kedua objek data tersebut yang disimpan didalam database MySQL, untuk prosesnya menggunakan pemrograman PHP. berikut beberapa proses kegiatan yang dilakukan secara bertahap:

- Case holding
Dilakukan proses perubahan teks menjadi huruf kecil semua
- Tokenizing
Dilakukan proses pemotongan kalimat menjadi kata-kata terhadap kedua objek data Al-Quran dan Al-Hadis dilakukan tokenizing
- Filtering tanda baca
Dilakukan proses menghilangkan tanda baca seperti:

Tabel 1. Tanda Baca

1	Tanda Titik (.)
2	Tanda Koma (,)
3	Tanda Titik Koma (;)
4	Tanda Titik Dua (:)
5	Tanda Hubung (-)
6	Tanda Pisah (—, —)
7	Tanda Elipsis (...)
8	Tanda Tanya (?)
9	Tanda Seru (!)
10	Tanda Kurung (...)
11	Tanda Kurung Siku ([...])
12	Tanda Petik ("...")
13	Tanda Petik Tunggal ('...')
14	Tanda Garis Miring (/)
15	Tanda Penyingkat (Apostrof)(')

- Filtering Word

Tahap mengambil kata-kata penting dari hasil tokenizing, dengan cara membuang kata – kata menggunakan algoritma stoplist dan wordlist, pada proses peneliti menggunakan stopword dari website <https://www.kaggle.com/>, pada website tersebut terdapat 758 stopwords

- Stemming

Tahap merubah kata-kata pada dokument menjadi kata dasarnya, hasil dari stemming pada kedua objek data tersebut adalah terhadap Alquran menghasilkan jumlah term 42161 dan stemming terhadap hadis menghasilkan jumlah term 242.

- Pembobotan menggunakan tf-idf
Hasil Perhitungan term frekuensi

Tabel 2. tf document Al-quran

id	term	dokid	tf
215286	nama	1 - 1	1
215287	allah	1 - 1	1
215288	maha	1 - 1	2
215289	murah	1 - 1	1
215290	sayang	1 - 1	1
215291	puji	1 - 2	1
215292	allah	1 - 2	1
215293	tuhan	1 - 2	1
215294	semesta	1 - 2	1
215295	alam	1 - 2	1
215296	maha	1 - 3	2
215297	murah	1 - 3	1
215298	sayang	1 - 3	1
215299	kuasa	1 - 4	1
215300	balas	1 - 4	1
215301	engkau	1 - 5	2
215302	sembah	1 - 5	1
215303	tolong	1 - 5	1
215304	jalan	1 - 6	1
215305	lurus	1 - 6	1

Tabel 3. tf document Al-Hadits

term	dokid	tf
niat	1	1
ikhlas	1	1
hijrah	1	1
fitriah	1	1
dunia	1	1
iman	2	1
islam	2	2
ihsan	2	1
ilmu	2	1
gaib	2	1
ajar	2	2
syahadat	3	1
sholat	3	1
zakat	3	1
haji	3	1
puasa	3	1

- Menghitung inverse document frequency (idf)
Proses perhitungan dengan cara melihat ketersediaan term dalam seluruh dokumen, Sedangkan untuk Inverse Document Frequency (IDF) dihitung dengan menggunakan formula sebagai berikut:

$$IDF_j = \log \left(\frac{D}{df_j} \right)$$

Persamaan 4.1. Rumus *Idf*

Contoh sebagian hasil dari perhitungan idf pada gambar berikut:

Tabel 4. *idf* Dokumen Al-Quran

id	term	dokid	tf	idf
215286	nama	1 - 1	1	6.37
215287	allah	1 - 1	1	1.55
215288	maha	1 - 1	2	3.26
215289	murah	1 - 1	1	6.82
215290	sayang	1 - 1	1	5.53
215291	puji	1 - 2	1	6.6
215292	allah	1 - 2	1	1.55
215293	tuhan	1 - 2	1	2.61
215294	semesta	1 - 2	1	7.14
215295	alam	1 - 2	1	6.96
215296	maha	1 - 3	2	3.26
215297	murah	1 - 3	1	6.82
215298	sayang	1 - 3	1	5.53
215299	kuasa	1 - 4	1	4.79
215300	balas	1 - 4	1	5.11
215301	engkau	1 - 5	2	5.81
215302	sembah	1 - 5	1	5.34
215303	tolong	1 - 5	1	5.17
215304	jalan	1 - 6	1	4.19
215305	lurus	1 - 6	1	6.45
215306	jalan	1 - 7	3	4.19

Tabel 5. *idf query* hadis

term	dokid	tf	idf
niat	1	1	11.6
ikhlas	1	1	9.28
hijrah	1	1	8.43
fitriah	1	1	12.6
dunia	1	1	5.19
iman	2	1	3.37
islam	2	2	7.96
ihsan	2	1	0
ilmu	2	1	6.96
gaib	2	1	11.6
ajar	2	2	7.35
syahadat	3	1	0
sholat	3	1	11.02
zakat	3	1	7.56
haji	3	1	9.02
puasa	3	1	9.02
orb	4	1	12.6
ibu	4	1	7.02

- Menghitung *tf-idf*

Untuk menghitung *tf-idf* setiap term pada document dan query dengan cara mengalikan nilai *tf* dan *idf* pada term yang didapatnya.

Contoh sebagian hasil perhitungan *tf-idf* sebagai berikut:

Tabel 6. *tfidf* dokumen Al-quran

id	term	dokid	tf	idf	tfidf
215286	nama	1 - 1	1	6.37	6.37
215287	allah	1 - 1	1	1.55	1.55
215288	maha	1 - 1	2	3.26	6.52
215289	murah	1 - 1	1	6.82	6.82
215290	sayang	1 - 1	1	5.53	5.53
215291	puji	1 - 2	1	6.6	6.6
215292	allah	1 - 2	1	1.55	1.55
215293	tuhan	1 - 2	1	2.61	2.61
215294	semesta	1 - 2	1	7.14	7.14
215295	alam	1 - 2	1	6.96	6.96
215296	maha	1 - 3	2	3.26	6.52
215297	murah	1 - 3	1	6.82	6.82
215298	sayang	1 - 3	1	5.53	5.53
215299	kuasa	1 - 4	1	4.79	4.79
215300	balas	1 - 4	1	5.11	5.11
215301	engkau	1 - 5	2	5.81	11.62
215302	sembah	1 - 5	1	5.34	5.34
215303	tolong	1 - 5	1	5.17	5.17
215304	jalan	1 - 6	1	4.19	4.19
215305	lurus	1 - 6	1	6.45	6.45
215306	jalan	1 - 7	3	4.19	12.57

Tabel 7. *tf-idf query* Hadis

term	dokid	tf	idf	tfidf
niat	1	1	11.6	11.6
ikhlas	1	1	9.28	9.28
hijrah	1	1	8.43	8.43
fitrah	1	1	12.6	12.6
dunia	1	1	5.19	5.19
iman	2	1	3.37	3.37
islam	2	2	7.96	15.92
ihsan	2	1	0	0
ilmu	2	1	6.96	6.96
gaib	2	1	11.6	11.6
ajar	2	2	7.35	14.7
syahadat	3	1	0	0
sholat	3	1	11.02	11.02
zakat	3	1	7.56	7.56
haji	3	1	9.02	9.02
puasa	3	1	9.02	9.02

- Menghitung panjang *vector*

Untuk menghitung panjang *vector* setiap document dan query dengan cara menghitung akar dari penjumlahan kuadrat pada setiap term.

$$|D| = \sqrt{\sum (tfidf^2)}$$

Persamaan 4.2. Rumus Panjang Vektor

Sebagian hasil perhitungan panjang *vector*, sebagai berikut:

Tabel 8. Panjang Vektor Dokumen Alquran

id	chapter	verse	meaning	vektor
2	1	1	Dengan menyebut nama Allah Yang Maha Pemurah lagi...	12.64
3	1	2	Segala puji bagi Allah, Tuhan semesta alam.	11.95
4	1	3	Maha Pemurah lagi Maha Penyayang.	10.92
5	1	4	Yang menguasai di Hari Pembalasan.	7
6	1	5	Hanya Engkaulah yang kami sembah, dan hanya kepada...	7.42
7	1	6	Tunjukkan kami jalan yang lurus.	7.69
8	1	7	(yaitu) Jalan orang-orang yang telah Engkau beri n...	16.65
9	2	1	Alif laam mim.	15.24
10	2	2	Kitab (Al Quran) ini tidak ada keraguan padanya; p...	11.54
11	2	3	(yaitu) mereka yang beriman kepada yang ghaib, ya...	18.28
12	2	4	dan mereka yang beriman kepada Kitab (Al Quran) ya...	19.13
13	2	5	Mereka itulah yang tetap mendapat petunjuk dari Tu...	6.67
14	2	6	Sesungguhnya orang-orang kafir, sama saja bagi mer...	5.1
15	2	7	Allah telah mengunci-mati hati dan pendengaran mer...	16.6
16	2	8	Di antara manusia ada yang mengatakan: "Kami berim...	7.8
17	2	9	Mereka hendak menipu Allah dan orang-orang yang be...	15.98
18	2	10	Dalam hati mereka ada penyakit, lalu ditambah Alla...	17.61
19	2	11	Dan bila dikatakan kepada mereka: "Janganlah kamu ...	14.43
20	2	12	Ingatlah, sesungguhnya mereka itulah orang-orang y...	10.62
21	2	13	Apabila dikatakan kepada mereka: "Berimanlah kamu ...	23.44
22	2	14	Dan bila mereka berjumpa dengan orang-orang yang b...	17.44
23	2	15	Allah akan (membalas) olok-olokan mereka dan membi...	16.31
24	2	16	Mereka itulah orang yang membeli kesesatan dengan ...	15.64
25	2	17	Perumpamaan mereka adalah seperti orang yang meny...	25.99
26	2	18	Mereka tuli, bisu dan buta, maka tidaklah mereka p...	16.28
27	2	19	atau seperti (orang-orang yang ditimpa) hujan leba...	34.42
28	2	20	Hampir-hampir kilat itu menyambar penglihatan mere...	36.53
29	2	21	Hai manusia, sembahlah Tuhanmu yang telah mencipta...	14.72

Tabel 9. Panjang *Vector Query* Hadis

id	hadis	vektor
1	Niat, keikhlasan, Hijrah dan Fitrah Dunia	21.85
2	Iman, Islam, Ihsan, Hari Akhir, Ilmu Gaib, Belajar...	25.75
3	Syahadat, sholat, zakat, haji, dan puasa	16.45
4	Pengorbanan ibu yang mengandung, teori reproduksi ...	19.95
5	Kesempurnaan islam, bida' dan taklid	10.49
6	penetapan halal, haram, dan syubhat, kedudukan hat...	12.51
7	Da'wah dan menegah kemunkaran, dan pentingnya sal...	18.39
8	Aqidah dan syariat harus ditegakkan, perlindungan n...	21.26
9	Patuh kepada rosululloh, bertakwa sebatas kemampua...	19.18
10	mempersembahkan yang terbaik untuk Allah, mengkons...	11.41
11	Meninggalkan keraguraguan	6.79
12	Optimalisasi waktu dan potensi, meninggalkan waktu...	9.88
13	menyakitkan saudara dan diri sendiri, persatuan umat	11.14
14	nyawa seorang muslim dilindungi, dan hukuman dalam...	22.76
15	pengaruh iman dalam keseharian, menjaga perkataan...	25.81
16	meninggalkan sifat marah	12.24
17	profesionalisme, berbuat baik pada seluruh makhluk	7.47
18	Takwa, bekal disetiap tempat dan waktu, Akhirat Mul...	20.37
19	Menyapkan generasi beriman, Allah tempat bergantung...	21.14
20	Menumbuhkan rasa malu	12.27
21	bertanya untuk mendapatkan kebaikan, iman dan isti...	3.36
22	Munasabah atau evaluasi diri, rindu surga, memperh...	12.24
23	keutamaan bersuci, kekuatan dzikir, sedekah, berko...	19.8
24	besarnya bahaya zalim, Allah sumber hidayah dan re...	21.73
25	Ini terhadap kebaikan orang lain, pintu-pintu keba...	23.96
26	menolong sesama manusia, menjaga kepentingan bera...	9.9
27	kebenaran mengalahkan ketenangan hati, hati - hati ...	28.49
28	Anjuran berwasiat menjelang kematian, berpegang te...	24.37
29	Hakikat keselamatan, masuk surga, dan terhindar da...	30.38
30	Sesungguhnya Allah ta'ala telah menetapkan kewajib...	24.39

5. Evaluasi

Dari sekian tahapan proses diatas sampai dengan ditemukan nilai cosine pada setiap dokumen dan queri yang saling keterkaitan, langkah selanjutnya kita mengklasifikasikan dan mengurutkan dokumen dari yang mempunyai tingkat kemiripan paling mendekati terhadap queri, sehingga dapat diketahui jumlah ayat yang mempunyai keterkaitan dengan hadits tertentu, berikut hasilnya:

Tabel 13. Hasil Klasifikasi Ayat

a. Cosine Similarity			b. Dice Distance			c. Jaccard		
Hadist	Jumlah Ayat	%	Hadist	Jumlah Ayat	%	Hadist	Jumlah Ayat	%
H - 1	200	3,21	H - 1	200	3,21	H - 1	200	3,21
H - 2	694	11,13	H - 2	694	11,13	H - 2	694	11,13
H - 3	57	0,91	H - 3	57	0,91	H - 3	57	0,91
H - 4	458	7,34	H - 4	458	7,34	H - 4	458	7,34
H - 5	79	1,27	H - 5	79	1,27	H - 5	79	1,27
H - 6	505	8,1	H - 6	505	8,1	H - 6	505	8,1
H - 7	14	0,22	H - 7	14	0,22	H - 7	14	0,22
H - 8	390	6,25	H - 8	390	6,25	H - 8	390	6,25
H - 9	345	5,53	H - 9	345	5,53	H - 9	345	5,53
H - 10	246	3,94	H - 10	246	3,94	H - 10	246	3,94
H - 11	57	0,91	H - 11	57	0,91	H - 11	57	0,91
H - 12	202	3,24	H - 12	202	3,24	H - 12	202	3,24
H - 13	220	3,53	H - 13	220	3,53	H - 13	220	3,53
H - 14	226	3,62	H - 14	226	3,62	H - 14	226	3,62
H - 15	707	11,34	H - 15	707	11,34	H - 15	707	11,34
H - 16	33	0,53	H - 16	33	0,53	H - 16	33	0,53
H - 17	35	0,56	H - 17	35	0,56	H - 17	35	0,56
H - 18	438	7,02	H - 18	438	7,02	H - 18	438	7,02
H - 19	754	12,09	H - 19	754	12,09	H - 19	754	12,09
H - 20	44	0,71	H - 20	44	0,71	H - 20	44	0,71
H - 21	603	9,67	H - 21	603	9,67	H - 21	603	9,67
H - 22	562	9,01	H - 22	562	9,01	H - 22	562	9,01
H - 23	863	13,84	H - 23	863	13,84	H - 23	863	13,84
H - 24	956	15,33	H - 24	956	15,33	H - 24	956	15,33
H - 25	207	3,32	H - 25	207	3,32	H - 25	207	3,32
H - 26	589	9,45	H - 26	589	9,45	H - 26	589	9,45
H - 27	506	8,11	H - 27	506	8,11	H - 27	506	8,11
H - 28	613	9,83	H - 28	613	9,83	H - 28	613	9,83
H - 29	706	11,32	H - 29	706	11,32	H - 29	706	11,32
H - 30	405	6,49	H - 30	405	6,49	H - 30	405	6,49
H - 31	94	1,51	H - 31	94	1,51	H - 31	94	1,51
H - 32	69	1,11	H - 32	69	1,11	H - 32	69	1,11
H - 33	82	1,31	H - 33	82	1,31	H - 33	82	1,31
H - 34	623	9,99	H - 34	623	9,99	H - 34	623	9,99
H - 35	722	11,58	H - 35	722	11,58	H - 35	722	11,58
H - 36	146	2,34	H - 36	146	2,34	H - 36	146	2,34
H - 37	7	0,11	H - 37	7	0,11	H - 37	7	0,11
H - 38	132	2,12	H - 38	132	2,12	H - 38	132	2,12
H - 39	93	1,49	H - 39	93	1,49	H - 39	93	1,49
H - 40	215	3,45	H - 40	215	3,45	H - 40	215	3,45
H - 41	883	14,16	H - 41	883	14,16	H - 41	883	14,16
H - 42	238	3,82	H - 42	238	3,82	H - 42	238	3,82

Dari keterangan pada tabel 13. hasil klasifikasi tiap hadits dengan menggunakan ketiga metode tersebut dilihat dari banyak ayat yang terkait adalah sama.

Untuk membandingkan kualitas ketiga metode tersebut jumlah ayat dari hasil klasifikasi data tersebut akan dilakukan pencocokan data dengan koleksi data relevan yang ada, berikut koleksi data relevan keterkaitan hadits arbain dan alquran (Mu'inudinillah Bashri, Maerwandi Tarmizi, 2010):

Tabel 14. Koleksi Data Relevan

ID Hadist	Ayat Al-Quran	Jumlah Ayat
1	98 - 5,7 - 29,8 - 72,3 - 195,2 - 218,4 - 97,8 - 67,6 - 70,4 - 134,3 - 145,	10
2	5 - 93,17 - 7,28 - 77,18 - 30,5 - 3,3 - 19,40 - 66,72 - 14,4 - 125,2 - 112,6 - 82,5 - 5,2 - 285,9 - 122,3 - 79,21 - 7,16 - 43,7 - 188,6 - 50,27 - 65,2 - 3,31 - 34,22 - 7,7 - 187,	24
3	2 - 185,2 - 183,22 - 27,2 - 196,3 - 97,73 - 20,19 - 55,9 - 71,20 - 132,19 - 31,2 - 3,16 - 36,2 - 256,	13
4	4 - 18,2 - 132,64 - 11,57 - 22,23 - 14,22 - 5,31 - 14,	7
5	17 - 36,57 - 27,5 - 3,	3
6	5 - 87,16 - 115,2 - 275,114 - 2,49 - 12,5 - 40,22 - 46,16 - 106,26 - 89,	9
7	87 - 9,51 - 55,41 - 33,3 - 110,3 - 104,	5
8	42 - 13,	1
9	40 - 5,64 - 16,8 - 46,59 - 7,	4
10	32 - 16,19 - 3,5 - 88,28 - 77,	4
11	2 - 2,49 - 15,14 - 10,	3
12	31 - 6,63 - 9,2 - 148,103 - 3,103 - 2,103 - 1,	6
13	3 - 103,49 - 10,49 - 12,	3
14	2 - 179,4 - 93,	2
15	51 - 27,51 - 26,51 - 25,51 - 24,4 - 36,50 - 18,16 - 97,	7
16	3 - 134,3 - 159,	2
17	2 - 195,28 - 77,	2
18	68 - 4,2 - 197,	2
19	10 - 49,7 - 188,9 - 51,64 - 11,112 - 2,1 - 5,46 - 15,25 - 74,4 - 9,	9
20	33 - 53,	1
21	15 - 99,72 - 16,46 - 13,41 - 30,2 - 200,2 - 219,2 - 217,2 - 512,2 - 149,	9
22	7 - 157,66 - 1,9 - 29,66 - 11,3 - 133,59 - 18,	6
23	17 - 7,25 - 30,7 - 204,4 - 82,13 - 28,8 - 45,2 - 222,9 - 108,33 - 35,57 - 18,2 - 261,	11
24	7 - 160,47 - 38,17 - 7,7 - 156,39 - 53,18 - 17,10 - 13,7 - 44,	8
25	7 - 157,3 - 114,2 - 148,5 - 2,2 - 177,	5
26	4 - 9,33 - 32,17 - 23,7 - 85,7 - 56,107 - 7,107 - 6,107 - 5,107 - 4,107 - 3,107 - 2,107 - 1,5 - 2,	13
27	3 - 135,17 - 36,13 - 28,8 - 10,	4
28	2 - 180,4 - 59,57 - 27,59 - 7,	4
29	9 - 19,61 - 11,17 - 79,3 - 185,2 - 185,	5
30	23 - 7,	1
31	113 - 5,102 - 5,102 - 4,102 - 3,102 - 2,102 - 1,29 - 64,18 - 46,18 - 45,	9
32	2 - 195,	1
33	24 - 23,24 - 4,24 - 51,4 - 65,	4
34	8 - 2,5 - 23,3 - 139,2 - 278,7 - 165,5 - 78,	6
35	22 - 30,5 - 32,49 - 13,9 - 71,49 - 10,	5
36	33 - 36,47 - 24,73 - 4,70 - 24,107 - 7,107 - 6,107 - 5,107 - 4,107 - 3,107 - 2,107 - 1,170 - 36,96 - 1,49 - 11,	14
37	23 - 61,2 - 148,	2
38	18 - 39,22 - 40,35 - 32,10 - 63,10 - 64,10 - 62,	6
39	64 - 16,2 - 196,22 - 78,	3
40	94 - 7,103 - 3,103 - 2,103 - 1,10 - 24,3 - 185,	6
41	20 - 16,	1
42	3 - 135,5 - 74,39 - 53,7 - 56,6 - 133,23 - 118,	6

Pada Tabel 14 terkumpul koleksi data relevan sebanyak 246 ayat Al-Quran. Berdasarkan koleksi data relevan dan hasil klasifikasi ayat terhadap hadits, akan diketahui nilai recall, precision, dan accuracy, sehingga dapat dilihat perbandingan dari ketiga metode tersebut.

Tabel 15. Perhitungan recall, precision,dan accuracy (Cosine Similarity)

Hadits 1	Jumlah Data	Retrieved		Not Retrieved Relevant	Recall %	Precision %	Akurasi %
		Relevant	Not Relevant				
1	200	8	192	2	80	4	96,89
2	694	11	683	13	45,8	1,6	88,862
3	57	9	48	4	69,2	15,8	99,167
4	458	2	456	5	28,6	0,4	92,613
5	79	1	78	2	33,3	1,3	98,718
6	505	6	499	3	66,7	1,2	91,954
7	14	2	12	3	40	14,3	99,76
8	390	1	389	0	100	0,3	93,762
9	345	2	343	2	50	0,6	94,469
10	246	3	243	1	75	1,2	96,088
11	57	3	54	0	100	5,3	99,134
12	202	0	202	6	0	0	96,668
13	220	3	217	0	100	1,4	96,52
14	226	0	226	2	0	0	96,345
15	707	3	704	4	42,9	0,4	88,654
16	33	0	33	2	0	0	99,439
17	35	0	35	2	0	0	99,407
18	438	1	437	1	50	0,2	92,977
19	754	4	750	5	44,4	0,5	87,903
20	44	1	43	0	100	2,3	99,31
21	603	0	603	9	0	0	90,2
22	562	5	557	1	83,3	0,9	91,053
23	863	5	858	6	45,5	0,6	86,158
24	956	4	952	4	50	0,4	84,679
25	207	2	205	3	40	1	96,666
26	589	3	586	10	23,1	0,5	90,458
27	506	3	503	1	75	0,6	91,919
28	613	4	609	0	100	0,7	90,234
29	706	4	702	1	80	0,6	88,729
30	405	0	405	1	0	0	93,49
31	94	1	93	8	11,1	1,1	98,382
32	69	0	69	1	0	0	98,878
33	82	1	81	3	25	1,2	98,654
34	623	4	619	2	66,7	0,6	90,045
35	722	3	719	2	60	0,4	88,442
36	146	1	145	13	7,1	0,7	97,472
37	7	1	6	1	50	14,3	99,888
38	132	3	129	3	50	2,3	97,884
39	93	0	93	3	0	0	98,461
40	215	2	213	4	33,3	0,9	96,522
41	883	0	883	1	0	0	85,827
42	238	2	236	4	33,3	0,8	96,154
RATA - RATA					44,27	1,862	94,258

Tabel 16. Perhitungan recall, precision,dan accuracy (Dice Distance)

Hadits 1	Jumlah Data	Retrieved		Not Retrieved Relevant	Recall %	Precision %	Akurasi %
		Relevant	Not Relevant				
1	200	8	192	2	80	4	96,89
2	694	11	683	13	45,8	1,6	88,862
3	57	9	48	4	69,2	15,8	99,167
4	458	2	456	5	28,6	0,4	92,613
5	79	1	78	2	33,3	1,3	98,718
6	505	6	499	3	66,7	1,2	91,954
7	14	2	12	3	40	14,3	99,76
8	390	1	389	0	100	0,3	93,762
9	345	2	343	2	50	0,6	94,469
10	246	3	243	1	75	1,2	96,088
11	57	3	54	0	100	5,3	99,134
12	202	0	202	6	0	0	96,668
13	220	3	217	0	100	1,4	96,52
14	226	0	226	2	0	0	96,345
15	707	3	704	4	42,9	0,4	88,654
16	33	0	33	2	0	0	99,439
17	35	0	35	2	0	0	99,407
18	438	1	437	1	50	0,2	92,977
19	754	4	750	5	44,4	0,5	87,903
20	44	1	43	0	100	2,3	99,31
21	603	0	603	9	0	0	90,2
22	562	5	557	1	83,3	0,9	91,053
23	863	5	858	6	45,5	0,6	86,158
24	956	4	952	4	50	0,4	84,679
25	207	2	205	3	40	1	96,666
26	589	3	586	10	23,1	0,5	90,458
27	506	3	503	1	75	0,6	91,919
28	613	4	609	0	100	0,7	90,234
29	706	4	702	1	80	0,6	88,729
30	405	0	405	1	0	0	93,49
31	94	1	93	8	11,1	1,1	98,382
32	69	0	69	1	0	0	98,878
33	82	1	81	3	25	1,2	98,654
34	623	4	619	2	66,7	0,6	90,045
35	722	3	719	2	60	0,4	88,442
36	146	1	145	13	7,1	0,7	97,472
37	7	1	6	1	50	14,3	99,888
38	132	3	129	3	50	2,3	97,884
39	93	0	93	3	0	0	98,461
40	215	2	213	4	33,3	0,9	96,522
41	883	0	883	1	0	0	85,827
42	238	2	236	4	33,3	0,8	96,154
RATA - RATA					44,27	1,862	94,258

Tabel 17. Perhitungan recall, precision,dan accuracy (Jaccard)

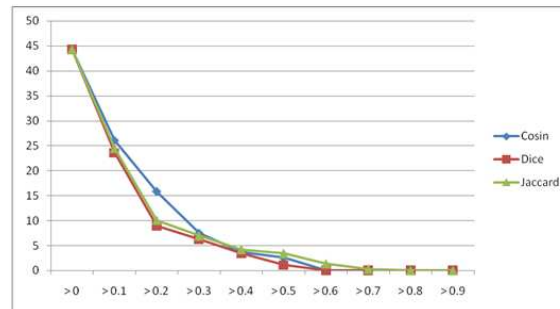
Hadits 1	Jumlah Data	Retrieved		Not Retrieved Relevant	Recall %	Precision %	Akurasi %
		Relevant	Not Relevant				
1	200	8	192	2	80	4	96,89
2	694	11	683	13	45,8	1,6	88,862
3	57	9	48	4	69,2	15,8	99,167
4	458	2	456	5	28,6	0,4	92,613
5	79	1	78	2	33,3	1,3	98,718
6	505	6	499	3	66,7	1,2	91,954
7	14	2	12	3	40	14,3	99,76
8	390	1	389	0	100	0,3	93,762
9	345	2	343	2	50	0,6	94,469
10	246	3	243	1	75	1,2	96,088
11	57	3	54	0	100	5,3	99,134
12	202	0	202	6	0	0	96,668
13	220	3	217	0	100	1,4	96,52
14	226	0	226	2	0	0	96,345
15	707	3	704	4	42,9	0,4	88,654
16	33	0	33	2	0	0	99,439
17	35	0	35	2	0	0	99,407
18	438	1	437	1	50	0,2	92,977
19	754	4	750	5	44,4	0,5	87,903
20	44	1	43	0	100	2,3	99,31
21	603	0	603	9	0	0	90,2
22	562	5	557	1	83,3	0,9	91,053
23	863	5	858	6	45,5	0,6	86,158
24	956	4	952	4	50	0,4	84,679
25	207	2	205	3	40	1	96,666
26	589	3	586	10	23,1	0,5	90,458
27	506	3	503	1	75	0,6	91,919
28	613	4	609	0	100	0,7	90,234
29	706	4	702	1	80	0,6	88,729
30	405	0	405	1	0	0	93,49
31	94	1	93	8	11,1	1,1	98,382
32	69	0	69	1	0	0	98,878
33	82	1	81	3	25	1,2	98,654
34	623	4	619	2	66,7	0,6	90,045
35	722	3	719	2	60	0,4	88,442
36	146	1	145	13	7,1	0,7	97,472
37	7	1	6	1	50	14,3	99,888
38	132	3	129	3	50	2,3	97,884
39	93	0	93	3	0	0	98,461
40	215	2	213	4	33,3	0,9	96,522
41	883	0	883	1	0	0	85,827
42	238	2	236	4	33,3	0,8	96,154
RATA - RATA					44,27	1,862	94,258

Pada Tabel 15, 16, dan 17 diatas memperlihatkan rata – rata nilai recall dibawah 50% yaitu 44,2% disebabkan karena jumlah data relevan yang ditemukan lebih sedikit dari pada jumlah data yang tidak ditemukan, sedangkan untuk precision bernilai 1,8 % hal ini disebabkan nilai karena jumlah data yang ditemukan sangat lebih banyak dari pada jumlah koleksi data relevan. Melihat nilai rata – rata recall, precision dan accuracy dari setiap metode menunjukan perbandingan ketiga metode tersebut bernilai sama, sehingga tidak terlalu terlihat perbandingan yang signifikan.

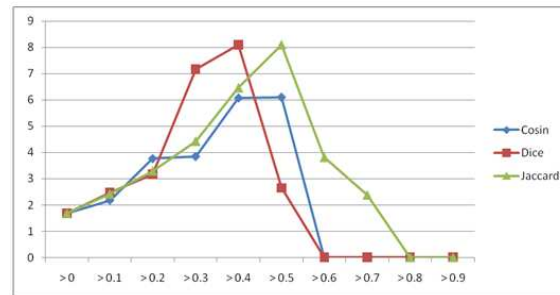
Tabel 17. Perbandingan recall, precision, dan accuracy

Metode	Tingkat	Nilai Persamaan									
		> 0	> 0.1	> 0.2	> 0.3	> 0.4	> 0.5	> 0.6	> 0.7	> 0.8	> 0.9
Cosin	Recall	44,271	26,027	15,82	7,443	3,726	2,513	0	0	0	0
	Presisi	1,682	2,177	3,761	3,846	6,071	6,101	0	0	0	0
	Akurasi	94,258	97,791	99,256	99,67	99,81	99,868	99,887	99,899	99,903	99,904
Dice	Recall	44,271	23,547	8,91	6,253	3,386	1,058	0	0	0	0
	Presisi	1,682	2,465	3,175	7,177	8,107	2,646	0	0	0	0
	Akurasi	94,258	98,285	99,498	99,773	99,854	99,883	99,894	99,902	99,903	99,905
Jaccard	Recall	44,271	24,44	10,055	6,914	4,203	3,386	1,323	0,265	0	0
	Presisi	1,682	2,417	3,296	4,431	6,468	8,107	3,81	2,381	0	0
	Akurasi	94,258	98,165	99,384	99,707	99,81	99,854	99,874	99,885	99,892	99,896

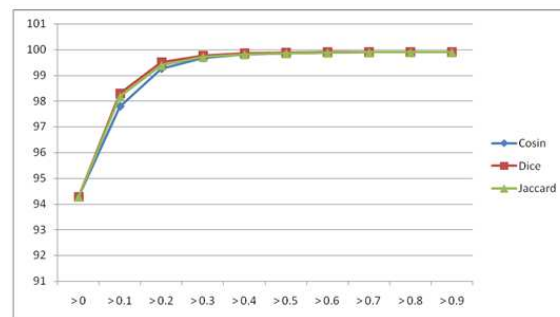
Pada tabel 17 memaparkan tentang perbandingan ketiga metode tersebut berdasarkan nilai persamaan dari hasil perhitungan recall, precision, dan accuracy. percobaan nilai persamaan yang dilakukan dengan interval 0,1 yang dimulai dari 0 sampai 1.



Gambar 3. Grafik Hasil Percobaan Perbandingan nilai recall



Gambar 4. Grafik Hasil Percobaan Perbandingan nilai precision

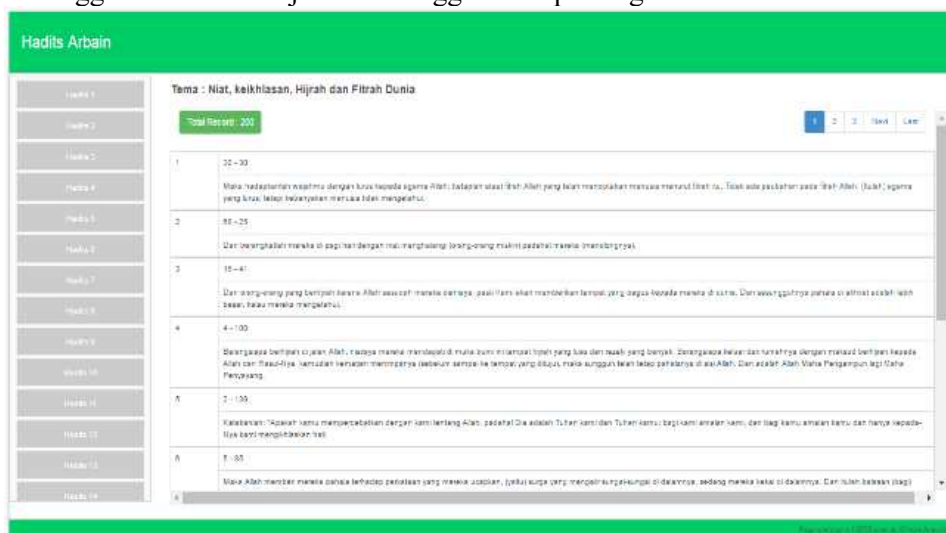


Gambar 5. Grafik Hasil Percobaan Perbandingan nilai accuracy

Dari hasil perbandingan diatas, pada penelitian ini jaccard mempunyai performa yang lebih baik dibanding kedua metode lainnya yaitu cosine dan dice distance dikarenakan mempunyai nilai rata – rata recall dan precision yang lebih baik walaupun tingkat akurasi lebih baik dari dice distance.

6. Deployment

pada tahapan ini akan dilakukan pembangunan aplikasi berbasis web klasifikasi ayat al-quran berdasarkan hadits arbain dengan menggunakan metode jaccard menggunakan pemrograman PHP.



Gambar 6. Aplikasi Web Hadis Arbain

V. KESIMPULAN DAN SARAN

5.1. Kesimpulan

Berdasarkan hasil penelitian dapat ditarik beberapa kesimpulan sebagai berikut:

1. Cosine, Dice Dan Jaccard mempunyai hasil yang sama berdasarkan rata-rata recall, precision dan accuracy secara keseluruhan yaitu dengan nilai recall 44,27 %, precision 1,8 %, dan accuracy 94 %, sehingga sulit diputuskan metode mana yang terbaik.
2. Akan tetapi dengan percobaan pada batasan terhadap hasil nilai persamaan, Metode Jaccard lebih baik dibanding dengan metode Cosine dan Dice berdasarkan perbandingan rata – rata recall dan precision
3. Dari tingkat Accuracy metode dice menunjukkan hasil yang baik dibandingkan dengan metode Jaccard dan Cosine.
4. Hasil recall, precision dan Accuracy cenderung bernilai kecil, dikarenakan jumlah data yang ditemukan dengan koleksi data relevan berbanding 1:61
5. Berdasarkan koleksi data relevan sebanyak 246 ayat ditemukan data relevan yang ditemukan sebanyak 108

5.2. Saran

Berdasarkan hasil penelitian dapat ditarik beberapa kesimpulan sebagai berikut:

1. Cosine, Dice Dan Jaccard mempunyai hasil yang sama berdasarkan rata-rata recall, precision dan accuracy secara keseluruhan yaitu dengan nilai recall 44,27 %, precision 1,8 %, dan accuracy 94 %, sehingga sulit diputuskan metode mana yang terbaik.
2. Akan tetapi dengan percobaan pada batasan terhadap hasil nilai persamaan, Metode Jaccard lebih baik dibanding dengan metode Cosine dan Dice berdasarkan perbandingan rata – rata recall dan precision
3. Dari tingkat Accuracy metode dice menunjukkan hasil yang baik dibandingkan dengan metode Jaccard dan Cosine.
4. Hasil recall, precision dan Accuracy cenderung bernilai kecil, dikarenakan jumlah data yang ditemukan dengan koleksi data relevan berbanding 1:61
5. Berdasarkan koleksi data relevan sebanyak 246 ayat ditemukan data relevan yang ditemukan sebanyak 108

DAFTAR PUSTAKA

- [1] Budiharto, Widodo. (2016). Knowledge dan Information Retrieval. Yogyakarta:DeePublish
- [2] Broto Poernomo T.P, Ir. Gunawan. (2015). Sistem information retrieval pencarian kesamaan ayat terjemahan al quran berbahasa indonesia dengan query expansion dari tafsirnya. Teknik Informtika Sekolah Tinggi Manajemen Informatika dan Komputer Asia, Teknik Informatika Sekolah Tinggi Teknik Surabaya.
- [3] Geehan Sabah Hassan, Siti Khaotijah Mohammad.(2015). Categorization of ‘Holy Quran-Tafseer’ using K-Nearest Neighbor Algorithm, International Journal of Computer Applications (0975 - 8887).
- [4] Kaouther Faidi, Raja Ayed, Ibrahim Bounhas, Bilel Elayeb. (2015). International Journal on Islamic Applications in Computer Science And Technology Vol 3.
- [5] Mohammad Alhawarat, Mohamed Hegazi, Anwer Hilal. (2015) (IJACSA) International Journal of Advanced Computer Science and Applications Vol 6 No 2.
- [6] Muhamad Syarif. (2017). Implementasi Algoritma String Matching Dalam Pencarian Surah Dan Ayat Dalam Al-Quran Berbasis Web. Indonesian Journal on Networking and Security - Volume 6 No 2.
- [7] North, Matthew (2012). Data Mining for The Masses. A Global Text Project Book.
- [8] Ogie Nurdiana, Jumadi, Dian Nursantika (2016). Perbandingan metode cosine similarity dengan metode jaccard similarity pada aplikasi pencarian terjemah al-qur'an dalam bahasa indonesia. JOIN Volume I No. ISSN 2527-9165.
- [9] Suhaib Kh. Hamed, Mohd Juzaidin Ab Aziz.(2016). A Question Answering System on Holy Quran Translation Based on Question Expansion Technique and Neural Network Classification. Center for Artificial Intelligence Technology (CAIT).
- [10] Tutik Khotimah.(2014). Pengelompokan surat dalam al qur'an menggunakan algoritma k-means. Jurnal SIMETRIS, Vol 5 No 1 ISSN: 2252-4983.
- [11] Vivek Kale (2017). Enterprise Performance Intelligence and Desicion Patterns.