

Sistem Informasi *Retrieval* Antara Hadis Arbain dan Alquran dengan *Vector Space Model*

Imam Amirulloh^{1*}, Taufik Wibisono², Yanti Apriyani³, Melisa Winda Pertiwi⁴, Mira Kusmira⁵

^{1,2,3} Program Studi Sistem Informasi, Universitas Bina Sarana Informatika Kampus Tasikmalaya, Tasikmalaya, Indonesia

^{4,5} Program Studi Sistem Informasi, Universitas Nusa Mandiri, Jakarta, Indonesia

**e-mail Corresponding Author: imam.iau@bsi.ac.id*

Abstract

The Alquran and Hadith have complementary relationships and explain each other. On this basis, in explaining the contents of the Hadith, it is necessary to look into the contents of the Alquran and vice versa. However, the large amount of Alquran and Al-Hadith data impacts the execution time of data searching from the two sources, especially searching Alquran verses based on 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.

Kata kunci: *Alquran; Hadith; System Information Retrieval; Vector Space Model*

Abstrak

Alquran dan Hadis mempunyai keterkaitan saling melengkapi dan menjelaskan satu sama lain. Atas dasar tersebut dalam menjelaskan isi Hadis perlu di lihat juga dari pandangan isi Alquran begitu juga sebaliknya. Namun dengan jumlah data Alquran dan Hadis yang sangat besar berdampak pada lamanya eksekusi pencarian data dari kedua kitab tersebut terutama pencarian dari Ayat Alquran ke Hadis 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: *Alquran; Hadis; Sistem Informasi Retrieval; Vector Space Model*

1. Pendahuluan

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 Alquran dan Hadis kedua kitab ini sudah menjadi rujukan bagi orang yang menganut agama Islam.

Alquran dan Hadis mempunyai keterkaitan, keduanya saling melengkapi dan menjelaskan satu sama lain. Atas dasar tersebut dalam menjelaskan isi Hadis perlu di lihat juga dari pandangan isi Alquran begitu juga sebaliknya. Namun dengan jumlah data Alquran dan Hadis yang sangat besar berdampak pada lamanya eksekusi pencarian data dari kedua kitab tersebut terutama pencarian dari Ayat Alquran ke Hadis yang membutuhkan waktu yang lama dan ketepatan dalam pencarian dikarenakan volume data Alquran lebih banyak dibanding volume data Hadis. Teknologi yang semakin berkembang, permasalahan tersebut sudah bukan

menjadi suatu yang sulit terutama dalam penyimpanan data Alquran dan Hadis yang sangat banyak bisa di simpan dalam bentuk elektronik dan digital. Sudah banyak pengembangan perangkat lunak di dunia telah mencoba mendigitalkan Alquran dan Hadis, akan tetapi belum adanya yang mendigitalisasikan langsung keterkaitan Alquran terhadap Hadis 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.

Maka pada penelitian ini akan menghasilkan sebuah sistem informasi yang mengklasifikasikan Ayat Alquran berdasarkan Hadis Arbain yang bertujuan agar masyarakat khususnya penganut agama Islam dapat mengetahui penjelasan Hadis tersebut berdasarkan Alquran ataupun sebaliknya dengan tingkat ketepatan pencarian dan penjelasan yang baik dan akurat.

2. 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 meneliti tentang sistem information retrieval pencarian kesamaan ayat terjemahan Alquran berbahasa Indonesia dengan *query expansion* dari tafsirnya [1]. Hasil dari penelitian tersebut stemming pada tahapan *pre-processing* menggunakan algoritme Nazief Adriani menghasilkan akurasi 95% untuk teks terjemahan Alquran 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, dan Dian Nursantika, meneliti tentang perbandingan metode cosine similarity dengan metode *jaccard similarity* pada aplikasi pencarian terjemah Alquran dalam bahasa Indonesia [2]. Pada penelitian ini 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%.

Julham Makatita meneliti tentang pembuatan aplikasi kumpulan doa harian bersumber Alquran dan Hadis berbasis android [3]. Pada penelitiannya hanya memaparkan pembuatan sebuah aplikasi android untuk doa harian yang berdasarkan pada Alquran dan Hadis tanpa melakukan text mining.

Geehan Sabah Hassan, meneliti dalam mengkategorikan Tafsir Alquran menggunakan Algoritme K-NN. Text Categorization menjadi langkah dalam pelabelan pada bahasa text alami dengan menggunakan KNN Algoritme [4]. hasil yang didapat dari penelitian tersebut Algoritme K-NN menunjukkan hasil yang baik.

Suhaib Kh. Hamed and Mohd Juzaidin Ab Aziz, meneliti tentang sistem penjawab pertanyaan pada terjemahan Alquran berdasarkan Teknik Ekspansi *Query* dan klasifikasi neural network [5]. 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 Faidi, meneliti tentang membandingkan Arabaci NLP Tools untuk klasifikasi Hadis [6]. 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, meneliti mengenai Implementasi Algoritme *String Matching* Dalam Pencarian Surah Dan Ayat Dalam Alquran berbasis Web [7]. dari hasil penelitiannya tersebut berkesimpulan implementasi algoritme *string matching* berbasis web membuat orang-orang lebih mudah dalam mencari kata atau *statement* pada terjemahan Alquran.

Tutik Khotimah, meneliti mengenai pengelompokan surat dalam al qur'an menggunakan algoritme *K-Means* [8]. 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, meneliti mengenai 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 [9]. Semua hasil ini didasarkan pada frekuensi awal yang dihitung dengan menggunakan metode Frekuensi *Frekuensi-Inverse Document Frequency* (TF-IDF).

Gagak Susanto dan Hari Lugis Purwanto meneliti mengenai Information Retrieval Menggunakan *Latent Semantic Indexing* Pada Ebook [10], dari hasil penelitiannya menyatakan bahwa sistem information retrieval menggunakan LSI dapat berjalan sesuai dengan yang diharapkan peneliti, dan bisa melakukan pencarian kata sesuai dengan kata kunci dan menampilkan tingkat kemiripan dokumen ebook yang tersimpan di database.

Abdul Karim dan Moh. Ainol Yaqin meneliti mengenai Implementasi *Vector Space Model* untuk Meningkatkan Kualitas Pencarian dan Penentuan Derajat Hadis pada Kitab-Kitab Hadis [11]. Hasilnya analisis keshahihan Hadis dengan pendekatan information retrieval didapatkan tingkat akurasi keabsahan perolehan dokumen (analisis Hadis dan *Vector Space Model*) rata-rata adalah 98%

Amalia Khasanah meneliti mengenai Perancangan Aplikasi Alquran Menggunakan Voice Recognition Sebagai Media Pencarian Terjemahan Al Qur'an Berbasis Android [12]. Hasil dari penelitian tersebut terciptanya sebuah aplikasi android dengan fitur voice recognition dalam proses pencarian terjemahan Alquran yang mengacu pada kitab Al Quran terbitan Departemen Agama RI.

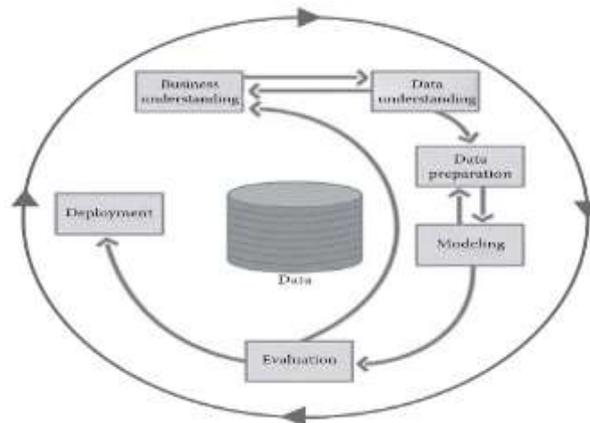
Meisya Fitri meneliti mengenai Perancangan Sistem Temu Balik Informasi Dengan Metode Pembobotan Kombinasi Tf-Idf Untuk Pencarian Dokumen Berbahasa Indonesia [13], dari penelitian tersebut didapatkan hasil Sistem mampu mengumpulkan dokumen berita melalui proses crawling website dan memberikan bobot dengan mengimplementasikan metode pembobotan kata dengan metode kombinasi Tf-Idf secara lengkap sehingga lebih banyak data relevan yang dapat diperhitungkan dalam pencarian.

Miftachur Robani dan Achmad Widodo meneliti mengenai Algoritme *K-Means Clustering* Untuk Pengelompokan Ayat Alquran Pada Terjemahan Bahasa Indonesia [14]. Dari hasil penelitiannya tersebut proses clustering dengan *K-Means* memiliki hasil pengujian silhouette pada Surat Al-Fatihah bernilai positif sebesar 0,336 yang artinya data pada kelompok yang tepat. Frekuensi tiap jumlah data sebesar 53% yang artinya kelompok yang dihasilkan memiliki kemiripan yang sedang. Sedangkan untuk hasil perhitungan rata-rata precision sebesar 53% dan perhitungan rata-rata *recall* sebesar 100%.

Chaerul Hadi dan Muhammad Rifqi Ma'arif meneliti Implementasi *Cosine Similarity* Dalam Aplikasi Pencarian Ayat Alquran Berbasis Android [15]. berdasarkan hasil pengujian, 70% responden menyatakan bahwa aplikasi ini dapat berguna sebagai pembantu pengguna perangkat bergerak android untuk mencari dan melihat ayat-ayat Alquran secara tematik.

Dari beberapa penelitian terdahulu yang ditinjau memiliki kesamaan melakukan pencarian terhadap surat dan ayat Alquran, namun tidak ada yang mencoba untuk mengklasifikasi berdasarkan Hadis terutama objek Hadis yang lebih khusus. Sehingga letak perbedaan pada penelitian ini ada pada objek klasifikasinya yaitu Hadis, dengan Hadis yang dipakai lebih khusus yaitu Hadis arbain. Objek tersebut dilakukan preprocessing dengan Alquran terjemahan indonesia, dan *State of Art* pada penelitian ini hasil dari preprocessing akan dilakukan tingkat kemiripan dengan *Vector Space Model* dari kedua object tersebut. Perbedaan yang lainnya pada penelitian ini adalah terbuatnya aplikasi sistem informasi dari hasil proses sistem informasi retrieval tersebut.

3. Metodologi



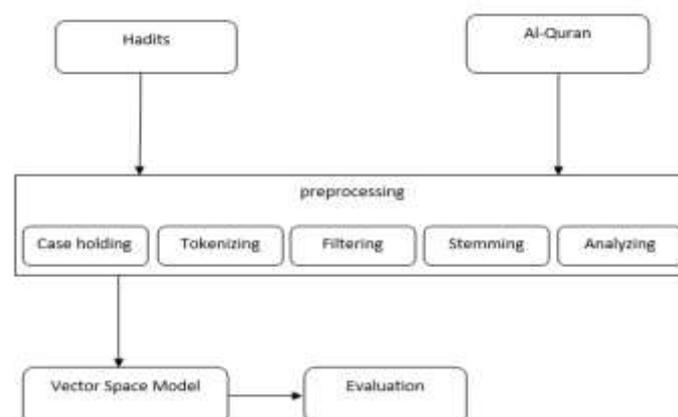
Gambar 1. Metode CRISP-DM

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 Hadis dan quran beserta kandungan ayat didalamnya, yang satu sama lain akan dikaitkan mana saja ayat Alquran yang ada hubungannya dengan salah satu Hadis yang dimaksud.
- *Data Understanding*
Tahapan kedua ini mengumpulkan Hadis-Hadis yang ada dalam kita Hadis arbain yang berjumlah 42 beserta ayat-ayat Alquran yang berjumlah kurang lebih 6236 ayat.
- *Data Preparation*
Tahapan ini akan dilakukan proses proses pre-processing terhadap dokumen Alquran dan beberapa tema Hadis
- *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 Simalrity*, *Dice* dan *Jaccard* sehingga dapat diketahui ayat Alquran mana saja yang paling mendekati atau ada kaitannya dengan Hadis yang dimaksud dan metode mana yang mempunyai performa terbaik
- *Deployment*
Pada tahapan ini akan dibuat aplikasinya dalam berbasis web.

3.2. Kerangka Kerja Penelitian



Gambar 2. Kerangka Kerja Sistem

1. *Query / Text Document*

Pada bagian ini menentukan *Query* dan *Text Document*, *Query* pada penelitian ini adalah Hadis, 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 dokumen 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 Alquran dan Hadis 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 Alquran dan Hadis 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 t 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 Alquran dan Hadis.

4. Hasil dan Pembahasan

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 Alquran berdasarkan kategori Tema Hadis Arb'in dengan teknik sistem informasi retrieval.
- Mengapa memilih Hadis Arbain sebagai kategori pengelompok ayat Alquran? Hadis Arba'in Hadis merupakan kitab Hadis yang pertama kali banyak dikaji oleh orang – orang yang baru mempelajari hadist.
- Seberapa tepatkah metode sistem informasi retrieval yang akan dipakai untuk mengelompok ayat Alquran? 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 Alquran dan Hadis, berikut pemaparan data-data yang akan dijadikan objek:

- Alquran, data-data quran di ambil dari website www.qurandatabase.org dengan format file database SQL, Alquran terdiri dari 114 surat dan 6236 ayat
- Hadis, Dari sekian banyak kitab-kitab hadis yang ada, pada penelitian ini memilih salah satu Hadis yaitu hadis arbain yang mempunyai jumlah Hadis sebanyak 42 Hadis dengan tema yang berbeda.

3. Data Preparation

proses pengolahan terhadap dokumen dan *query* dengan melakukan *pre-processing* 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 Alquran dan Hadis dilakukan tokenizing
- *Filtering tanda baca*
Dilakukan proses menghilangkan tanda baca seperti:

Tabel 1. Tanda Baca

| No. | 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 algoritme *stoplist* dan *wordlist*, pada proses peneliti menggunakan *stopword* dari website <https://www.kaggle.com/>, pada website tersebut terdapat 758 *stopword*

- *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

| id | term | dokid | tf |
|--------|---------|-------|----|
| 215286 | nama | 1 - 1 | 1 |
| 215287 | aloh | 1 - 1 | 1 |
| 215288 | maha | 1 - 1 | 2 |
| 215289 | murah | 1 - 1 | 1 |
| 215290 | sayang | 1 - 1 | 1 |
| 215291 | puji | 1 - 2 | 1 |
| 215292 | alfah | 1 - 2 | 1 |
| 215293 | tuhan | 1 - 2 | 1 |
| 215294 | semesta | 1 - 2 | 1 |
| 215295 | alam | 1 - 2 | 1 |

Gambar 3. Tabel tf dokumen Alquran

| term | dokid | tf |
|---------|-------|----|
| nal | 1 | 1 |
| ikhlas | 1 | 1 |
| hijrah | 1 | 1 |
| fitriah | 1 | 1 |
| dunia | 1 | 1 |
| iman | 2 | 1 |
| islam | 2 | 2 |
| irisan | 2 | 1 |
| ilmu | 2 | 1 |

Gambar 4. Tabel tf dokumen Hadis

- 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:

$$IDF_j = \text{Log} \left(\frac{D}{df_j} \right) \dots\dots\dots(1)$$

Contoh sebagian hasil dari perhitungan idf pada gambar berikut:

| 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 | 0,82 |
| 215290 | sayang | 1 - 1 | 1 | 5,53 |
| 215291 | puji | 1 - 2 | 1 | 0,0 |
| 215292 | allah | 1 - 2 | 1 | 1,55 |
| 215293 | tuhan | 1 - 2 | 1 | 2,01 |
| 215294 | semesta | 1 - 2 | 1 | 7,14 |
| 215295 | alam | 1 - 2 | 1 | 6,96 |

Gambar 5. Tabel *idf* Dokumen Alquran

| 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 |
| ihlan | 2 | 1 | 0 |
| ilmu | 2 | 1 | 6,96 |

Gambar 6. Tabel *idf* Dokumen Hadis

- 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:

| 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 | 0,82 | 0,82 |
| 215290 | sayang | 1 - 1 | 1 | 5,53 | 5,53 |
| 215291 | puji | 1 - 2 | 1 | 0,0 | 0,0 |
| 215292 | allah | 1 - 2 | 1 | 1,55 | 1,55 |
| 215293 | tuhan | 1 - 2 | 1 | 2,01 | 2,01 |
| 215294 | semesta | 1 - 2 | 1 | 7,14 | 7,14 |
| 215295 | alam | 1 - 2 | 1 | 6,96 | 6,96 |

Gambar 7. Tabel *tf-idf* dokumen Alquran

| 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 |
| fitriah | 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 |
| ihlan | 2 | 1 | 0 | 0 |
| ilmu | 2 | 1 | 6,96 | 6,96 |

Gambar 8. Tabel *tf-idf* dokumen Hadis

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

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

Sebagian hasil perhitungan panjang *vector*, sebagai berikut:

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 Hadis arbain dan Alquran:

| ID Hadis | Ayat Al-Quran | MAJMA' Ayat |
|----------|---|-------------|
| 1 | 86- 5,7- 28,8- 72,3- 195,2- 218,4- 27,8- 87,6- 70,4- 156,3- 345, | 10 |
| 2 | 5- 91,17- 7,18- 77,18- 30,5- 3,3- 39,40- 88,71- 14,4- 125,2- 112, 8- 81,5- 5,2- 285,9- 123,3- 79,21- 7,18- 43,7- 188,6- 90,27- 89,2- 3,31- 94,22- 7,7- 187, | 24 |
| 3 | 2- 185,2- 183,22- 27,2- 286,3- 67,75- 20,19- 95,9- 71,20- 112,10- 33, 2- 3,08- 38,2- 256, | 13 |
| 4 | 4- 18,2- 132,64- 11,57- 22,28- 14,23- 5,21- 14, | 7 |
| 5 | 17- 36,57- 27,3- 3, | 3 |
| 6 | 5- 87,28- 113,7- 275,224- 1,48- 12,5- 40,22- 46,08- 206,25- 89, | 9 |
| 7 | 87- 9,31- 95,42- 33,3- 138,3- 104, | 5 |
| 8 | 42- 13, | 1 |
| 9 | 40- 5,84- 38,8- 46,55- 7, | 4 |
| 10 | 32- 16,19- 5,5- 88,28- 77, | 4 |
| 11 | 2- 2,48- 15,14- 10, | 3 |
| 12 | 31- 6,83- 9,2- 148,109- 3,103- 2,103- 1, | 8 |
| 13 | 3- 103,49- 10,49- 11, | 3 |
| 14 | 2- 179,4- 87, | 2 |
| 15 | 51- 77,21- 29,21- 25,21- 24,4- 36,50- 18,16- 97, | 7 |
| 16 | 3- 136,3- 159, | 2 |
| 17 | 2- 195,29- 77, | 2 |
| 18 | 88- 4,2- 177, | 2 |
| 19 | 10- 49,7- 288,9- 51,64- 11,112- 2,1- 5,48- 75,25- 74,4- 9, | 9 |
| 20 | 33- 55, | 1 |
| 21 | 13- 99,72- 16,48- 33,43- 30,7- 106,7- 216,7- 217,2- 912,2- 349, | 9 |
| 22 | 7- 157,88- 1,8- 29,66- 11,3- 135,59- 18, | 6 |
| 23 | 17- 7,25- 93,7- 264,4- 82,18- 28,8- 45,2- 212, 9- 106,18- 26,57- 18,3- 281, | 11 |
| 24 | 7- 160,47- 88,27- 7,7- 158,28- 15,18- 17,10- 13,7- 44, | 8 |
| 25 | 7- 157,3- 114,2- 148,3- 2,2- 177, | 3 |
| 26 | 4- 8,18- 12,17- 23,7- 85,7- 96,97- 7,107- 6,107- 5,107- 4,107- 3,107- 2,107- 1,3- 2, | 11 |
| 27 | 3- 125,17- 36,18- 28,8- 16, | 4 |
| 28 | 2- 180,4- 94,17- 27,48- 7, | 4 |
| 29 | 9- 19,61- 11,17- 78,9- 185,2- 185, | 3 |
| 30 | 23- 7, | 1 |
| 31 | 113- 5,103- 6,202- 4,102- 3,102- 2,102- 1,20- 64,18- 46,18- 49, | 9 |
| 32 | 2- 195, | 1 |
| 33 | 24- 23,24- 4,34- 71,4- 68, | 4 |
| 34 | 8- 2,5- 23,3- 159,2- 270,7- 105,5- 78, | 6 |
| 35 | 22- 30,5- 32,48- 11,9- 71,48- 16, | 3 |
| 36 | 33- 36,47- 24,75- 4,70- 24,107- 7,107- 6, 107- 5, 107- 4,107- 3,107- 2,107- 1,70- 36,95- | 14 |
| 37 | 1,49- 11, | 2 |
| 38 | 18- 50,22- 40,21- 31,21- 83,10- 84,10- 61, | 8 |
| 39 | 84- 16,2- 295,22- 78, | 3 |
| 40 | 84- 7,103- 1,103- 2,103- 1,10- 24,2- 185, | 8 |
| 41 | 10- 16, | 1 |
| 42 | 3- 120,3- 74,32- 53,7- 56,6- 132,22- 118, | 8 |

Gambar 15. Tabel Koleksi Data Relevan

Pada Gambar 15 terkumpul koleksi data relevan sebanyak 246 ayat Alquran. Berdasarkan koleksi data relevan dan hasil klasifikasi ayat terhadap Hadis, akan diketahui nilai *recall*, *precision*, dan *accuracy*, sehingga dapat dilihat perbandingan dari ketiga metode tersebut.

| Hadis 1 | Jumlah Data | Retrieved | | Net Retrieved Relevan | Recall % | Precision % | Akurasi % |
|-------------|-------------|-----------|-------------|-----------------------|----------|-------------|-----------|
| | | Relevan | Not Relevan | | | | |
| 1 | 200 | 8 | 192 | 3 | 80 | 4 | 96,89 |
| 2 | 694 | 11 | 683 | 12 | 43,8 | 1,8 | 39,857 |
| 3 | 77 | 9 | 68 | 4 | 89,5 | 11,3 | 99,187 |
| 4 | 478 | 2 | 456 | 7 | 28,8 | 0,2 | 97,811 |
| 5 | 79 | 1 | 78 | 2 | 23,2 | 1,3 | 98,718 |
| 6 | 103 | 6 | 99 | 3 | 86,7 | 2,2 | 91,454 |
| 7 | 14 | 2 | 12 | 3 | 40 | 14,3 | 99,76 |
| 8 | 160 | 1 | 159 | 0 | 100 | 0,2 | 91,762 |
| 9 | 142 | 2 | 140 | 2 | 50 | 0,8 | 94,489 |
| 10 | 246 | 3 | 243 | 1 | 75 | 1,2 | 96,088 |
| 11 | 31 | 3 | 28 | 0 | 100 | 5,2 | 99,514 |
| 12 | 202 | 0 | 202 | 6 | 0 | 0 | 96,668 |
| 13 | 220 | 3 | 217 | 0 | 100 | 1,1 | 96,31 |
| 14 | 226 | 0 | 226 | 3 | 0 | 0 | 96,347 |
| 15 | 707 | 3 | 704 | 4 | 43,2 | 0,4 | 88,614 |
| 16 | 34 | 0 | 33 | 3 | 0 | 0 | 99,439 |
| 17 | 15 | 0 | 15 | 2 | 0 | 0 | 99,487 |
| 18 | 418 | 1 | 417 | 1 | 50 | 0,2 | 95,471 |
| 19 | 754 | 4 | 750 | 2 | 44,8 | 0,5 | 87,903 |
| 20 | 44 | 1 | 43 | 0 | 100 | 2,3 | 89,31 |
| 21 | 603 | 0 | 603 | 9 | 0 | 0 | 90,1 |
| 22 | 562 | 2 | 557 | 1 | 83,3 | 0,9 | 91,671 |
| 23 | 869 | 3 | 858 | 6 | 47,3 | 0,8 | 86,574 |
| 24 | 978 | 4 | 952 | 4 | 50 | 0,4 | 84,873 |
| 25 | 107 | 2 | 105 | 3 | 80 | 1 | 96,666 |
| 26 | 489 | 3 | 486 | 10 | 23,1 | 0,3 | 90,418 |
| 27 | 106 | 3 | 103 | 1 | 75 | 0,8 | 91,839 |
| 28 | 611 | 4 | 605 | 0 | 100 | 6,7 | 80,214 |
| 29 | 796 | 4 | 792 | 3 | 87 | 0,8 | 88,719 |
| 30 | 403 | 0 | 403 | 1 | 0 | 0 | 91,48 |
| 31 | 84 | 1 | 83 | 8 | 11,1 | 1,1 | 95,182 |
| 32 | 69 | 0 | 69 | 1 | 0 | 0 | 98,131 |
| 33 | 81 | 1 | 81 | 3 | 21 | 1,2 | 95,674 |
| 34 | 623 | 4 | 619 | 2 | 86,7 | 0,8 | 90,042 |
| 35 | 727 | 3 | 719 | 3 | 80 | 0,4 | 88,447 |
| 36 | 146 | 1 | 145 | 11 | 2,1 | 0,7 | 97,477 |
| 37 | 7 | 1 | 6 | 1 | 50 | 14,3 | 69,188 |
| 38 | 113 | 3 | 110 | 3 | 50 | 3,0 | 97,884 |
| 39 | 83 | 0 | 83 | 3 | 0 | 0 | 96,461 |
| 40 | 212 | 2 | 210 | 4 | 33,5 | 0,9 | 86,522 |
| 41 | 883 | 0 | 883 | 1 | 0 | 0 | 83,827 |
| 42 | 138 | 2 | 136 | 4 | 81,9 | 0,8 | 96,154 |
| RATA - RATA | | | | 44,27 | 1,862 | 84,288 | |

Gambar 16. Tabel Perhitungan Recall, precision, dan accuracy (Cosine Similarity)

| Hadis 1 | Jumlah Data | Retrieved | | Net Retrieved Relevan | Recall % | Precision % | Akurasi % |
|-------------|-------------|-----------|-------------|-----------------------|----------|-------------|-----------|
| | | Relevan | Not Relevan | | | | |
| 1 | 200 | 8 | 192 | 3 | 80 | 4 | 96,89 |
| 2 | 694 | 11 | 683 | 13 | 43,8 | 1,6 | 86,862 |
| 3 | 77 | 9 | 68 | 4 | 69,2 | 15,3 | 99,167 |
| 4 | 478 | 2 | 456 | 5 | 28,8 | 0,4 | 92,613 |
| 5 | 79 | 1 | 78 | 2 | 33,3 | 1,3 | 98,718 |
| 6 | 103 | 6 | 99 | 3 | 66,7 | 1,2 | 91,854 |
| 7 | 14 | 2 | 12 | 3 | 40 | 14,3 | 99,76 |
| 8 | 160 | 1 | 159 | 0 | 100 | 0,3 | 93,762 |
| 9 | 142 | 2 | 140 | 2 | 50 | 0,6 | 94,469 |
| 10 | 246 | 3 | 243 | 1 | 75 | 1,2 | 96,088 |
| 11 | 31 | 3 | 28 | 0 | 100 | 5,3 | 99,134 |
| 12 | 202 | 0 | 202 | 6 | 0 | 0 | 96,668 |
| 13 | 220 | 3 | 217 | 0 | 100 | 1,4 | 96,32 |
| 14 | 226 | 0 | 226 | 3 | 0 | 0 | 96,345 |
| 15 | 707 | 3 | 704 | 4 | 43,2 | 0,4 | 88,614 |
| 16 | 34 | 0 | 33 | 3 | 0 | 0 | 99,439 |
| 17 | 15 | 0 | 15 | 2 | 0 | 0 | 99,487 |
| 18 | 418 | 1 | 417 | 1 | 50 | 0,2 | 92,977 |
| 19 | 754 | 4 | 750 | 5 | 44,4 | 0,5 | 87,603 |
| 20 | 44 | 1 | 43 | 0 | 100 | 2,3 | 90,31 |
| 21 | 603 | 0 | 603 | 9 | 0 | 0 | 90,2 |
| 22 | 562 | 2 | 557 | 3 | 83,3 | 0,9 | 91,653 |
| 23 | 869 | 3 | 858 | 6 | 45,5 | 0,6 | 86,574 |
| 24 | 978 | 4 | 952 | 4 | 50 | 0,4 | 84,679 |
| 25 | 107 | 2 | 105 | 3 | 40 | 1 | 96,666 |
| 26 | 489 | 3 | 486 | 10 | 23,1 | 0,5 | 90,418 |
| 27 | 106 | 3 | 103 | 1 | 75 | 0,6 | 91,919 |
| 28 | 611 | 4 | 605 | 0 | 100 | 0,7 | 90,214 |
| 29 | 796 | 4 | 792 | 3 | 80 | 0,6 | 88,729 |
| 30 | 403 | 0 | 403 | 1 | 0 | 0 | 91,48 |
| 31 | 84 | 1 | 83 | 8 | 11,1 | 1,1 | 95,182 |
| 32 | 69 | 0 | 69 | 1 | 0 | 0 | 98,131 |
| 33 | 81 | 1 | 81 | 3 | 21 | 1,2 | 95,654 |
| 34 | 623 | 4 | 619 | 2 | 69,7 | 0,6 | 90,043 |
| 35 | 727 | 3 | 719 | 2 | 60 | 0,4 | 88,442 |
| 36 | 146 | 1 | 145 | 13 | 2,1 | 0,7 | 97,472 |
| 37 | 7 | 1 | 6 | 1 | 50 | 14,3 | 99,889 |
| 38 | 113 | 3 | 110 | 3 | 50 | 3,0 | 97,884 |
| 39 | 83 | 0 | 83 | 3 | 0 | 0 | 96,461 |
| 40 | 212 | 2 | 210 | 4 | 33,3 | 0,9 | 96,522 |
| 41 | 883 | 0 | 883 | 1 | 0 | 0 | 83,827 |
| 42 | 138 | 2 | 136 | 4 | 33,3 | 0,8 | 96,154 |
| RATA - RATA | | | | 44,27 | 1,862 | 84,288 | |

Gambar 17. Tabel 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 | 43,8 | 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 | 96,32 |
| 14 | 226 | 0 | 226 | 2 | 0 | 0 | 96,345 |
| 15 | 707 | 3 | 704 | 4 | 42,9 | 0,3 | 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 | 87,803 |
| 20 | 44 | 1 | 43 | 0 | 100 | 2,3 | 99,31 |
| 21 | 603 | 0 | 603 | 9 | 0 | 0 | 90,2 |
| 22 | 563 | 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 | 306 | 3 | 303 | 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 | 1 | 25 | 1,2 | 98,654 |
| 34 | 623 | 4 | 619 | 2 | 66,7 | 0,6 | 90,043 |
| 35 | 722 | 3 | 719 | 3 | 60 | 0,4 | 88,442 |
| 36 | 146 | 1 | 145 | 13 | 7,1 | 0,7 | 97,472 |
| 37 | 7 | 1 | 6 | 1 | 50 | 14,3 | 99,889 |
| 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 | | | | | 44,17 | 1,862 | 94,288 |

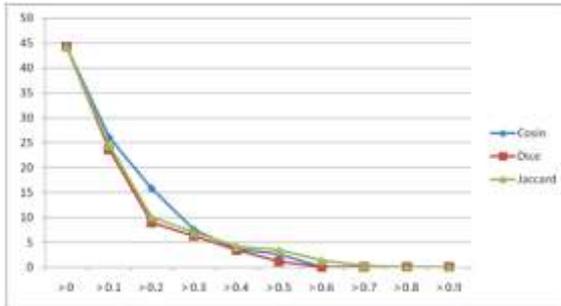
Gambar 18. Tabel Perhitungan *Recall*, *precision*, dan *accuracy* (Jaccard)

Gambar 16, 17, dan 18 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 menunjukkan perbandingan ketiga metode tersebut bernilai sama, sehingga tidak terlalu terlihat perbandingan yang signifikan.

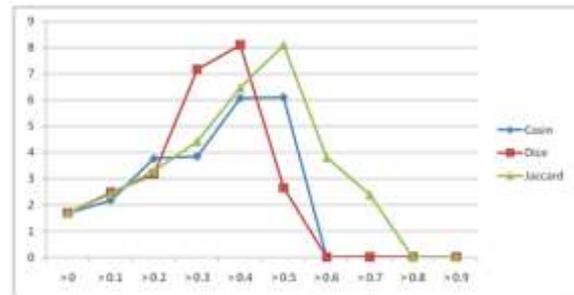
| Metode | Tingkat | Nilai Persamaan | | | | | | | | | |
|---------|---------|-----------------|--------|--------|--------|--------|--------|--------|--------|--------|--------|
| | | >0 | >0,1 | >0,2 | >0,3 | >0,4 | >0,5 | >0,6 | >0,7 | >0,8 | >0,9 |
| Cosin | Recall | 44,271 | 26,027 | 15,82 | 7,443 | 3,726 | 2,513 | 0 | 0 | 0 | 0 |
| | Precisi | 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 |
| | Precisi | 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 | 0,205 | 0 | 0 |
| | Precisi | 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 |

Gambar 19. Tabel Perbandingan *recall*, *precision*, dan *accuracy*

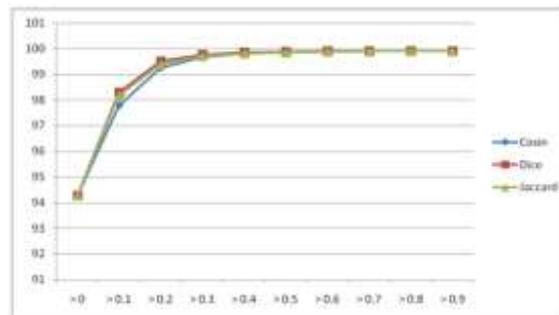
Gambar 19 menyajikan perbandingan ketiga metode tersebut berdasarkan nilai persamaan dari hasil perhitungan *recall*, *precision*, dan *accuracy*. percobaan nilai persamaan yang dilakukan dengan interval 0,1 yang dimulai dari 0 sampai 1.



Gambar 20. Grafik Hasil Percobaan Perbandingan nilai *recall*



Gambar 21. Grafik Hasil Percobaan Perbandingan nilai *precision*

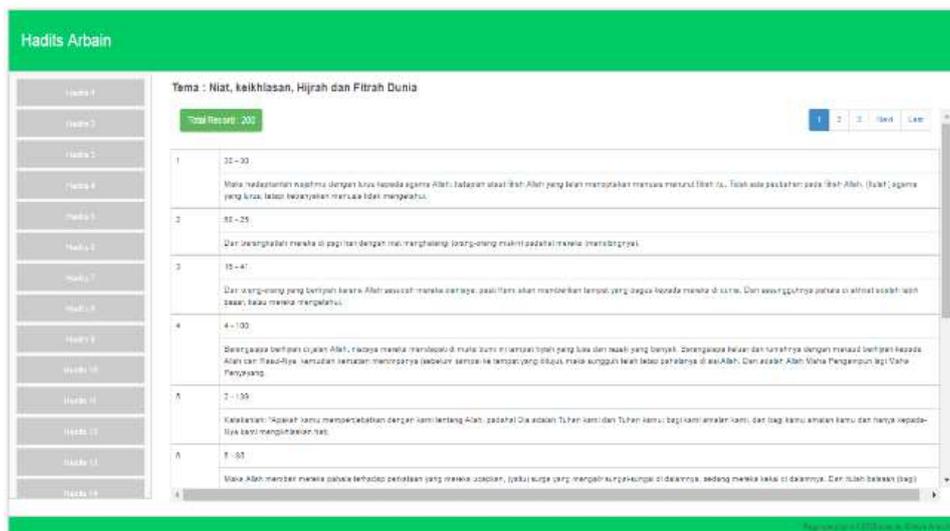


Gambar 22. Grafik Hasil Percobaan Perbandingan nilai *accuracy*

Dari hasil perbandingan pada Grafik Gambar 20-22, 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 dilakukan pembangunan aplikasi berbasis web klasifikasi ayat Alquran berdasarkan Hadis arbain dengan menggunakan metode *jaccard* menggunakan pemrograman PHP.



Gambar 23. Aplikasi Web Hadis Arbain

Gambar 23 merupakan aplikasi web Hadis Arbain yang akan menampilkan data klasifikasi ayat Alquran berdasarkan tema-tema pada Hadis arbain yang mana klasifikasi

tersebut merupakan hasil dari proses data Preparation, modelling dan *evaluation* hasil perbandingan *recall*, *precision*, dan *accuracy* dengan metode Jaccard

5. Simpulan

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. 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*.

Dari tingkat *accuracy*, metode *dice* menunjukkan hasil yang lebih baik dibandingkan dengan metode *Jaccard* dan *Cosine*. Hasil *recall*, *precision* dan *accuracy* cenderung bernilai kecil, dikarenakan jumlah data yang ditemukan dengan koleksi data relevan berbanding 1:61

Berdasarkan koleksi data relevan sebanyak 246 ayat ditemukan data relevan yang ditemukan sebanyak 108

Daftar Referensi

- [1] T.P.B. Poernomo & Gunawan. "Sistem information retrieval pencarian kesamaan ayat terjemahan al quran berbahasa indonesia dengan query expansion dari tafsirnya". Seminar Nasional "Inovasi dalam Desain dan Teknologi" -IDeaTech, ISSN 2089-1121, 2015
- [2] O. Nurdiana, Jumadi & D. Nursantika. "Perbandingan metode cosine similarity dengan metode jaccard similarity pada aplikasi pencarian terjemah Alquran dalam bahasa indonesia", *Jurnal Online Informatika*, Vol 1, No.1, pp.59-63, 2016
- [3] J. Makatita, L. Marlinda & R Yuniarto. "Aplikasi Kumpulan Doa Harian Bersumber Alquran Dan Hadis Berbasis Android", Seminar Nasional Ilmu Pengetahuan dan Teknologi Komputer (SINPTEK), Vol. 4, No. 1, pp.262-268, 2016
- [4] G.S. Hassan & S.K. Mohammad. "Categorization of 'Holy Quran-Tafseer' using K-NearestNeighbor Algorithm", *International Journal of Computer Applications*, Vol 129, No.12, pp.1-6, 2015
- [5] S.K. Hamed, M.J.A. Aziz. "A Question Answering System on Holy Quran Translation Based on Question Expansion Technique and Neural Network Classification". *Journal of Computer Sciences*. Vol 12, No.3, pp.169-177, 2016
- [6] K. Faidi, R. Ayed, I. Bounhas, & B. Elayeb. "Comparing Arabic NLP tools for Hadith Classification". *International Journal on Islamic Applications in Computer Science and Technology*, Vol 3, no. 3, pp. 1-12, 2015
- [7] M. Syarif. "Implementasi Algoritme String Matching Dalam Pencarian Surah Dan Ayat Dalam Alquran Berbasis Web". *Indonesian Journal on Networking and Security*, Vol 6, No.2, pp.70-76, 2017
- [8] T. Khotimah. "Pengelompokan surat dalam al qur'an menggunakan algoritme k-means". *Jurnal SIMETRIS*, Vol. 5, No.1, pp.83-88, 2014
- [9] M. Alhawarat, M Hegazi & Anwer Hilal. "Processing the Text of the Holy Quran: A Text Mining Study", (IJACSA) *International Journal of Advanced Computer Science and Applications*, Vol 6, No.2, pp.262-267, 2015
- [10] G. Susanto & H. L. Purwanto, "Information retrieval menggunakan latent semantic indexing pada ebook," *SMATIKA Jurnal STIKI Informatika Jurnal*, vol. 8, no. 2, pp. 74-79, 2018.
- [11] A. Karim & M.A. Yaqin. "Implementasi Vector Space Model untuk Meningkatkan Kualitas Pencarian dan Penentuan Derajat Hadis pada Kitab-Kitab Hadis". *IncomTech: Jurnal Telekomunikasi dan Komputer*, Vol.10, No.1, pp.1-7, 2020
- [12] A. Khasanah. "Perancangan Aplikasi Al Qur'an Menggunakan Voice Recognition Sebagai Media Pencarian Terjemahan Al Qur'an Berbasis Android." Yogyakarta: Amikom Yogyakarta, 2013.
- [13] M. Fitri. "Perancangan Sistem Temu Balik Informasi Dengan Metode Pembobotan Kombinasi Tf-Idf Untuk Pencarian Dokumen Berbahasa Indonesia". *Jurnal Sistem dan Teknologi dan Informasi*, Vol 1, No.1, pp.1-6, 2013
- [14] M. Robani & A. Widodo. "Algoritme K-Means Clustering Untuk Pengelompokan Ayat Al Quran Pada Terjemahan Bahasa Indonesia". *Jurnal Sistem Informasi Bisnis*, Vol 6, No.2, pp. 164-176, 2016

- [15] C. Hadi & M.R. Ma'arif. "Implementasi Cosine Similarity Dalam Aplikasi Pencarian Ayat Alquran Berbasis Android", Jurnal Teknologi Informasi dan Komunikasi, Vol.6, No. 2, pp.70-79, 2017