

SKRIPSI

KLASIFIKASI DOKUMEN TERJEMAHAN MENGGUNAKAN ALGORITMA NAÏVE BAYES PADA AL-QUR'AN VERSI BERBAHASA INDONESIA (JUZ 1 SAMPAI JUZ 4)



Skripsi

Sebagai Salah Satu Syarat Untuk Memperoleh Gelar Sarjana Komputer

Fakultas Sains dan Teknologi

Universitas Islam Negeri Syarif Hidayatullah Jakarta

Disusun Oleh:

DIMAS GALUH FITRIANI

1113093000095

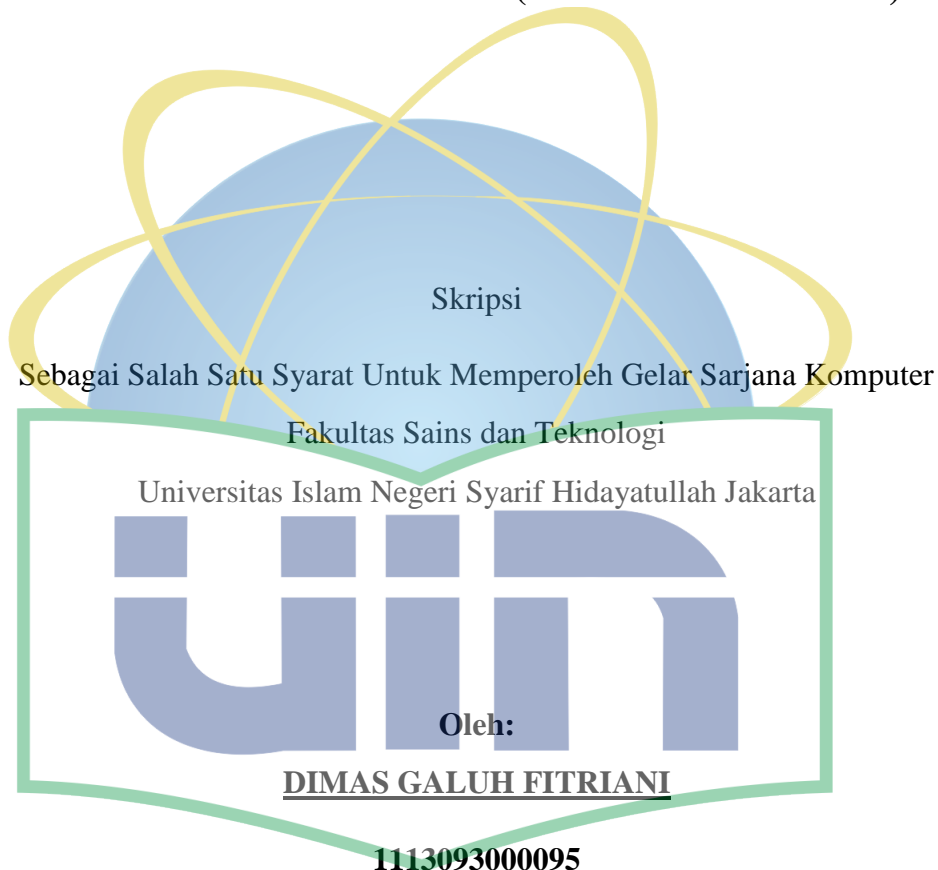
PROGRAM STUDI SISTEM INFORMASI

FAKULTAS SAINS DAN TEKNOLOGI

UNIVERSITAS ISLAM NEGERI SYARIF HIDAYATULLAH JAKARTA

2018 M/1440 H

SKRIPSI
KLASIFIKASI DOKUMEN TERJEMAHAN MENGGUNAKAN
ALGORITMA NAÏVE BAYES PADA AL-QUR'AN VERSI
BERBAHASA INDONESIA (JUZ 1 SAMPAI JUZ 4)



Universitas Islam Negeri
SYARIF HIDAYATULLAH JAKARTA

PROGRAM STUDI SISTEM INFORMASI
FAKULTAS SAINS DAN TEKNOLOGI
UNIVERSITAS ISLAM NEGERI SYARIF HIDAYATULLAH JAKARTA
2018 M/1440 H

**KLASIFIKASI DOKUMEN TERJEMAHAN MENGGUNAKAN
ALGORITMA *NAÏVE BAYES* PADA AL-QUR'AN VERSI BERBAHASA
INDONESIA (JUZ 1 SAMPAI JUZ 4)**

Skripsi

Diajukan sebagai salah satu syarat untuk memperoleh gelar Sarjana Komputer

Fakultas Sains dan Teknologi

Universitas Islam Negeri Syarif Hidayatullah Jakarta

Oleh:

DIMAS GALUH FITRIANI

1113093000095

Menyetujui,

Pembimbing I

Dr. Syopiansyah Jaya Putra, M.Sis

NIP. 19680117-2001121 001

Pembimbing II

Yuni Sugiarti, S.T., M.Kom

NIDN:0206067620

Mengetahui,

Ketua Program Studi Sistem Informasi Fakultas Sains Dan Teknologi

Universitas Islam Negeri Syarif Hidayatullah Jakarta

Universitas Islam Negeri
SYARIF HIDAYATULLAH JAKARTA
Nia Kurniadewi, MMSI
NIP. 19750412 200710 2 002

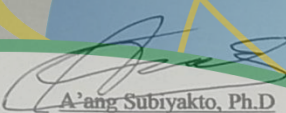
PENGESAHAN UJIAN


Skripsi berjudul “Klasifikasi Dokumen Terjemahan Menggunakan Algoritma *Naïve Bayes* Pada Al-Qur’an Versi Berbahasa Indonesia (Juz 1 Sampai Juz 4)” yang ditulis oleh Dimas Galuh Fitriani, NIM 1113093000095 telah diuji dan dinyatakan LULUS dalam Sidang *Munagosa* Fakultas Sains dan Teknologi Universitas Islam Negeri Syarif Hidayatullah Jakarta, pada hari Kamis, 20 September 2018. Skripsi ini telah diterima sebagai salah satu syarat untuk memperoleh gelar Sarjana Komputer.

Menyetujui,

Penguji I

Penguji II


A'ang Subiyakto, Ph.D



Nuryasin, M.Kom

NIP. 19760219 200710 1 002

NIP. 19760715 201101 1 005

Pembimbing I

Pembimbing II


Dr. Syopiansyah Jaya Putra, M.Sis


Yuni Sugiarti, S.T., M.Kom

NIP. 19680117 2001121 001

NIDN. 0206067620

Mengetahui,

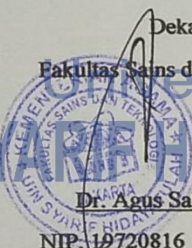
Dekan

Ketua

Fakultas Sains dan Teknologi

Program Studi Sistem Informasi


UNIVERSITAS ISLAM NEGERI
SYARIF HIDAYATULLAH JAKARTA


Dr. Agus Salim, M.Si


Nia Kumaladewi, MMSI

NIP. 19720816 199903 1 003

NIP. 19750412 200710 2002

PERNYATAAN

DENGAN INI SAYA MENYATAKAN BAHWA SKRIPSI INI BENAR-BENAR
HASIL KARYA SENDIRI DAN BELUM PERNAH DIAJUKAN SEBAGAI
SKRIPSI ATAU KARYA ILMIAH PADA PERGURUAN TINGGI ATAU
LEMBAGA MANAPUN.



Universitas Islam Negeri
SYARIF HIDAYATULLAH JAKARTA

ABSTRAK

Dimas Galuh Fitriani – 1113093000095, Klasifikasi Dokumen Terjemahan Menggunakan Algoritma Naïve Bayes Pada Al-Qur'an Versi Berbahasa Indonesia (Juz 1 Sampai Juz 4), di bawah bimbingan bapak **Syopiansyah Jaya Putra** dan ibu **Yuni Sugiarti**

Perkembangan yang pesat dalam informasi digital telah menyebabkan semakin meningkat pula volume informasi yang berbentuk teks. Klasifikasi dokumen merupakan bidang penelitian dalam pencarian informasi dan pengelompokan dokumen berdasarkan pembahasan tertentu. Permasalahan yang terjadi seperti hasil kategorisasi memberikan informasi yang sangat luas sehingga kurang relevan, terjemahan ada yang belum memiliki label kelas dan ada juga yang memiliki lebih dari satu label (*multi-label*). Penelitian ini bertujuan untuk mengklasifikasikan dokumen terjemahan Al-Qur'an berbahasa Indonesia menjadi 5 kategori yaitu, akhlak, akidah, ibadah, kisah, muamalah. Dalam penelitian ini dokumen yang digunakan yaitu, data set terjemahan Al-Qur'an berbahasa Indonesia juz 1 sampai juz 4 yang terdapat dalam www.tanzil.net. Klasifikasi ini dilakukan dengan metode *Naïve Bayes* yang merupakan metode probabilistik memiliki beberapa keunggulan kesederhanaan dalam komputasi. Hasil dari penelitian ini yaitu 5 kategori dengan tingkat keberhasilan sebagai berikut akhlak 80%, akidah 80%, ibadah 100%, kisah 93 dan muamalah 90%. Penelitian ini menunjukkan bahwa algoritma *Naïve Bayes* memiliki ketepatan dengan tingkat akurasi yang besar dalam proses klasifikasi.

Kata Kunci : *Al – Qur'an, Klasifikasi Dokumen, Naïve Bayes*

V Bab + 175 halaman + xv halaman + 8 gambar + 8 tabel + 5 lampiran

22 Pustaka (2000-2017)

Universitas Islam Negeri
SYARIF HIDAYATULLAH JAKARTA

KATA PENGANTAR

Bismillahirrahmanirrahim

Puji serta rasa syukur peneliti panjatkan kehadiran Allah SWT, yang telah melimpahkan segala nikmat dan karunia-Nya kepada peneliti sehingga peneliti dapat menyelesaikan skripsi ini dengan sebaik-baiknya. Shalawat serta salam semoga tercurah kepada junjungan Nabi Muhammad SAW beserta keluarga, sahabat dan kerabat serta kaum muslimin dan muslimat, semoga kita semua mendapatkan syafa'at dari beliau di akhirat kelak. Aamiin.

Adapun judul skripsi ini adalah **“Klasifikasi Dokumen Terjemahan Menggunakan Algoritma Naïve Bayes Pada Al-Qur'an Versi Berbahasa Indonesia (Juz 1 sampai Juz 4)”** yang disusun untuk memenuhi salah satu syarat dalam menyelesaikan program Strata Satu (S1) pada Program Studi Sistem Informasi, Fakultas Sains dan Teknologi, Universitas Islam Negeri Syarif Hidayatullah Jakarta. Semoga skripsi ini dapat memberikan manfaat kepada semua pihak dan memberikan wawasan kepada pembaca.

Dengan selesainya skripsi ini, peneliti ingin mengucapkan rasa terima kasih yang sebesar-besarnya kepada:

1. Bapak Dr. Agus Salim, M.Si, selaku Dekan Fakultas Sains dan Teknologi,
Universitas Islam Negeri Syarif Hidayatullah Jakarta

2. Ibu Nia Kumaladewi, MMSI, sebagai Ketua Program Studi dan Ibu Meinarini Catur Utami, MT selaku Sekretaris Program Studi Sistem Informasi Fakultas Sains dan teknologi UIN Syarif Hidayatullah Jakarta.
3. Bapak Dr. Syopiansyah Jaya Putra, M.Sis, sebagai Dosen Pembimbing pertama dan Ibu Yuni Sugiarti, M.kom, sebagai pembimbing kedua yang tidak pernah bosan untuk membimbing dan memberikan arahan baik secara moral maupun teknis dalam menyelesaikan skripsi ini.
4. Bapak Muhammad Nur Gunawan, MBA, sebagai dosen yang telah meluangkan waktu dan membantu dalam eksperimen peneliti dan juga memberikan arahan dalam penulisan skripsi ini.
5. Seluruh Dosen dan Staff, khususnya Program Studi Sistem Informasi yang telah memberikan ilmu serta bantuannya kepada peneliti yang tidak mungkin peneliti sebutkan satu persatu.
6. Kedua Orang Tua peneliti yang sangat peneliti sayangi dan hormati yang telah mendidik peneliti dengan kesabaran dan kasih sayang yang tiada henti serta selalu memberikan doa, semangat dan motivasi yang terus menerus sehingga segala proses dalam pengerjaan skripsi ini berjalan dengan lancar. Terimakasih atas doa yang selalu diberikan kepada peneliti. Semoga Allah SWT selalu memberikan kesehatan dan senantiasa melimpahkan rahmat dan karunianya kepada mereka. Aamiin.
7. Terimakasih untuk sahabat-sahabat peneliti, Anisa Citra Mutia, Afiati Azhami, Yuliani, Lisa Galuh, Nurul Fadhylah, Amanda, Syarah, Tini yang telah memberikan peneliti motivasi, semangat serta bantuannya, dari

perjuangan awal kuliah sampai pada tahap akhir kuliah yang peneliti belum dapat balas jasanya.

8. Terimakasih untuk sahabat peneliti Fitri Dewi Haryanti, Frida Silviani, dan Syifa Safira yang telah meluangkan waktunya untuk menemani dan membantu peneliti dalam mengerjakan skripsi ini.
9. Keluarga Besar Sistem Informasi 2013 terimakasih atas dukungan dan semua perjuangan yang sudah dilalui bersama selama duduk di bangku perkuliahan.
10. Semua pihak yang telah membantu terselesaikannya skripsi ini, maaf jika peneliti tidak bisa menyebutkan satu per satu.

Pada kesempatan ini peneliti juga memohon maaf yang sebesar-besarnya dan menyadari bahwa karya ini banyak kekurangan. Akhir kata peneliti berharap skripsi ini dapat bermanfaat bagi semua pihak. Aamiin.

Universitas Islam Negeri
SYARIF HIDAYATULLAH JAKARTA

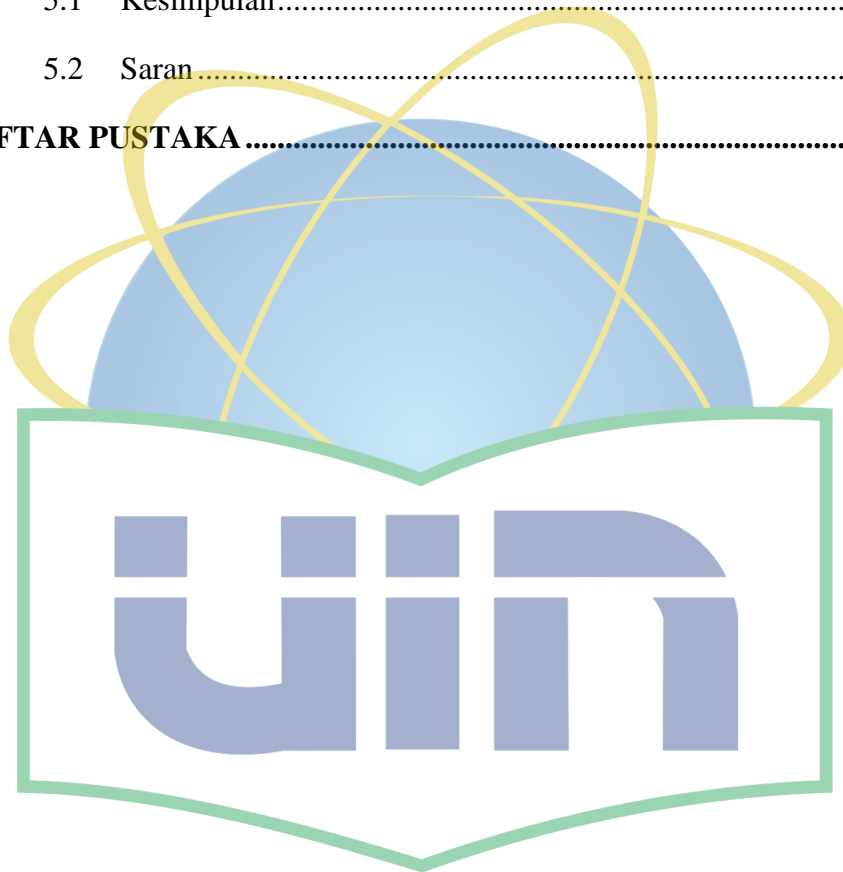
DAFTAR ISI

HALAMAN JUDUL	i
LEMBAR PERSETUJUAN	ii
LEMBAR PENGESAHAN	iii
PERNYATAAN.....	iv
ABSTRAK	v
KATA PENGANTAR.....	vi
DAFTAR ISI.....	ix
DAFTAR GAMBAR.....	xiii
DAFTAR TABEL	xiv
DAFTAR LAMPIRAN.....	xv
BAB 1 PENDAHULUAN	1
1.1 Latar Belakang.....	1
1.2 Identifikasi Masalah.....	5
1.3 Rumusan Masalah	5
1.4 Batasan Masalah.....	5
1.5 Tujuan Penelitian.....	6
1.6 Manfaat Penelitian.....	7
1.7 Metodologi Penelitian	7
1.8 Sistematika Penulisan.....	9
BAB 2 LANDASAN TEORI	11
2.1 Kategorisasi	11
2.2 Text Mining.....	12

2.2.1	Information Extraction.....	13
2.2.2	Information Retrieval.....	14
2.2.3	Natural Language Processing	14
2.2.4	Data Mining	15
2.3	Text Pre-Processing.....	17
2.3.1	Tokenisasi	17
2.3.2	Stopword Elimination.....	19
2.3.3	Stemming	19
2.4	Text Transformation	20
2.4.1	Model Ruang Vektor	22
2.4.2	Feature Selection	23
2.4.3	Feature Extraction.....	24
2.5	Pattern Discovery	25
2.6	Clustering	25
2.7	Classification	26
2.7.1	Proses Klasifikasi.....	27
2.7.2	Memilih Skema Klasifikasi	29
2.8	Perbedaan Clustering dan Classification	30
2.9	Naïve Bayes Clasifier	31
2.10	Classification Tools	34
2.10.1	Mahout.....	34
2.11	Confusion Matrix.....	37
2.12	Al-Qur'an	38
2.12.1	Akhlak	39
2.12.2	Akidah	41

2.12.3 Ibadah	43
2.12.4 Kisah	45
2.12.5 Muamalah	47
2.13 Literature Sejenis	53
BAB 3 METODOLOGI PENELITIAN	57
3.1 Metodologi Pengumpulan Data	57
3.1.1 Observasi	57
3.1.2 Studi Pustaka	58
3.1.3 Studi Literatur Sejenis	58
3.2 Material	59
3.3 Kerangka Berfikir	59
3.4 Metodologi Kategorisasi	61
3.4.1 Extraction Data	62
3.4.2 Text Pre-Processing	62
3.4.3 Feature Selection	62
3.4.4 Feature Extraction	63
3.4.5 Classification Stage	63
3.5 Performance Metrics	63
BAB 4 HASIL DAN PEMBAHASAN	66
4.1 Hasil Penelitian	66
4.1.1 Use case Diagram	66
4.1.2 Extracing Data	68
4.1.3 Text Preprocessing	69
4.1.4 Feature Selection	70
4.1.5 Feature Extraction	70

4.1.6 Classification Stage	71
4.1.7 Performance Metrics.....	74
4.2 Pembahasan	76
BAB 5 KESIMPULAN DAN SARAN	79
5.1 Kesimpulan.....	79
5.2 Saran.....	80
DAFTAR PUSTAKA	81



Universitas Islam Negeri
SYARIF HIDAYATULLAH JAKARTA

DAFTAR GAMBAR

Gambar 2. 1 Venn Diagram Of Text Mining Interaction.....	13
Gambar 2. 2 Contoh matriks term-document.....	22
Gambar 2. 3 Contoh Model ruang Vektor VSM.....	23
Gambar 3. 1 Kerangka Berfikir.....	60
Gambar 3. 2 Metode Klasifikasi	65
Gambar 4. 1 <i>Use Case Diagram</i>	68
Gambar 4. 2 Ekstraksi Data	69
Gambar 4. 3 Hasil Klasifikasi	72



Universitas Islam Negeri
SYARIF HIDAYATULLAH JAKARTA

DAFTAR TABEL

Tabel 2. 1 Perbedaan Clustering dan Classification	31
Tabel 2. 2 Confusion Matrix	37
Tabel 2. 3 Studi Literature Sejenis	53
Tabel 4. 1 Identifikasi Aktor.....	66
Tabel 4. 2 Identifikasi Use case	67
Tabel 4. 3 Hasil klasifikasi.....	73
Tabel 4. 4 Perhitungan precision.....	75
Tabel 4. 5 Perhitungan Nilai akurasi.....	76



Universitas Islam Negeri
SYARIF HIDAYATULLAH JAKARTA

DAFTAR LAMPIRAN

LAMPIRAN 1 INTERFACE MAHOUT.....	85
LAMPIRAN 2 KLASIFIKASI DALAM AL-QUR'AN VERSI THE WISDOM	92
LAMPIRAN 3 HASIL KLASIFIKASI SURAT DAN AYAT	106
LAMPIRAN 4 DATASET TERJEMAHAN AL-QUR'AN	108
LAMPIRAN 5 SURAT-SURAT PENDUKUNG PENELITIAN.....	160



Universitas Islam Negeri
SYARIF HIDAYATULLAH JAKARTA

BAB 1

PENDAHULUAN

Bab ini menjelaskan mengenai gambaran umum pelaksanaan penelitian yang berisi tentang, latar belakang penelitian, identifikasi masalah, rumusan masalah, batasan masalah, tujuan penelitian, manfaat penelitian, metodologi penelitian dan sistematika penulisan.

1.1 Latar Belakang

Perkembangan teknologi informasi pada jaman modern ini telah sampai kepada era elektronik, ditandai dengan semakin digunakannya teknologi berupa komputer dan jaringan *internet* sebagai sarana utama penyampaian informasi. Dokumen yang beredar di dunia maya terus tumbuh dan mungkin menjadi kurang efektif dalam pencarian dan penyajian informasi (Efendi & Malik, 2012). Seiring dengan kemajuan ilmu pengetahuan dan teknologi mayoritas orang lebih memanfaatkan teknologi sebagai media untuk belajar dan mencari informasi. Seperti halnya dalam mencari makna dari terjemahan Al-Qur'an.

Dalam penyajian Al-Qur'an telah di klasifikasikan berdasarkan pembahasan dari ayat tersebut, baik Al-Qur'an dalam bentuk fisik ataupun digital. Pembahasan yang tidak disatukan memungkinkan umat islam sedikit kesulitan dalam pencarian ayat berdasarkan pembahasaanya, banyaknya ayat Al-Qur'an mendorong umat islam untuk mencari cara mendapatkan ayat yang tepat dalam waktu yang singkat. Apabila ayat yang akan di cari berada pada kumpulan ayat yang berjumlah sedikit, pencarian

dilakukan belum terkomputerisasi. Namun, apabila jumlah ayat yang tersedia sangat besar, proses pencarian secara manual akan menghabiskan waktu dan tenaga, oleh karena itu diperlukan suatu cara untuk memperoleh ayat secara cepat dan tepat (Setiawati, Taufik, Jumadi, & Budiawan, 2016)

Data yang sangat banyak sekali menjadi bermasalah ketika data tersebut tidak terstruktur. Untuk itu, *text mining* dapat mengubah data yang tidak terstruktur menjadi data yang lebih terstruktur. Pengelompokan dokumen-dokumen berbasis teks ke dalam kategori-kategori tertentu berdasarkan isi kontennya, seperti mengelompokkan berita, dokumen dan *customer feedback*. Selain itu ia berguna untuk manajemen dokumen agar lebih baik, sehingga berdampak pada proses *retrieval document* yang menjadi lebih mudah. (Ingersol, Grant, S; Morton, Thomas, S; Farris, 2012)

Teknik klasifikasi suatu proses untuk menemukan model atau fungsi yang menjelaskan atau membedakan konsep atau kelas data, dengan tujuan untuk dapat memperkirakan kelas dari suatu objek (Natalius, 2011). Ada

beberapa teknik klasifikasi dokumen diantaranya adalah *Naïve Bayes*

Classifier, *Decision Tree* dan *Support Vector Machines* (Lestari et al. 2013; Natalius, 2011)

Metoda *naïve bayes* pada klasifikasi dokumen telah dilakukan oleh (Natalius, 2011). Metoda *Naïve Bayes* sangat luas dipakai dalam berbagai bidang, khususnya dalam proses klasifikasi dokumen. Metode *naïve bayes* merupakan *Supervised learning* yang membutuhkan data pelatihan terlebih dahulu (Efendi & Malik, 2012; Lestari et al. 2013). Metode *Naïve Bayes* bisa

memprediksi probabilitas bersyarat berdasarkan data pelatihan yang telah dibuat, kemudian kelebihan dari metode ini adalah tingkat akurasi yang tinggi juga waktu komputasi yang lebih cepat (Lestari et al. 2013; Setiawan & Nursantika, 2017).

Klasifikasi terjemahan Al-Qur'an tentang ilmu *sains* dengan algoritma *decision tree* berbasis *mobile* (Setiawati et al., 2016). Algoritma *decision tree* dapat dipakai untuk mengklasifikasikan terjemahan ayat Al-Qur'an dalam kategori *sains* yaitu biologi, fisika, dan kimia pada aplikasi *android* dan pada aplikasi yang dibuat dapat menampilkan informasi klasifikasi terjemahan ayat.

Klasifikasi dokumen berbahasa Indonesia menggunakan *naïve bayes classifier* (Efendi & Malik, 2012; Natalius, 2011). *Naïve Bayes Classifier* digunakan dalam proses klasifikasi terhadap dokumen berita dalam bahasa Indonesia dan penyaringan spam (*spam filtering*). Klasifikasi terhadap dokumen bertujuan untuk mengklasifikasikan dokumen tidak terstruktur ke dalam kelompok yang menggambarkan isi dari dokumen (Efendi & Malik, 2012).

Personality types classification (Lestari et al.s, 2013) mengkategorikan dan mengklasifikasikan data tidak terstruktur. Proses pengklasifikasian data antara lain *text preprocessing*, *tokenizing*, *filtering*, dan *stemming* (Efendi & Malik, 2012; Lestari et al., 2013). Proses klasifikasi dimulai dari *input* kosakata yang belum diketahui kategorisasinya, kosakatanya terdiri dari

himpunan semua kata unik yang terdapat dalam data pelatihan (Efendi & Malik, 2012; Lestari et al., 2013)

Klasifikasi *text mining* untuk terjemahan ayat-ayat Al-Qur'an menggunakan metode klasifikasi *naïve bayes* (Hilwah, Kudus, & Sunendiari, n.d.). Dalam Al-Qur'an dan Terjemahan yang diterbitkan oleh penerbit *cordova*, melampirkan bahwa Al-Qur'an terdapat 15 klasifikasi ayat-ayat Al-Qur'an. Namun tidak semua ayat memiliki label kelas, ada beberapa ayat yang belum memiliki kelas. Sehingga ayat-ayat yang belum memiliki label kelas tersebut dapat di prediksi dengan menggali informasi dari ayat yang sudah memiliki label kelasnya.

Berdasarkan laporan penelitian-penelitian sebelumnya algoritma *Naïve Bayes* memberikan hasil yang bagus ketika diimplementasikan pada dokumen *text* Bahasa Inggris (Arifin & Purnama, 2012), Chinese (Wei, n.d.), Arab (Syiam, Fayed, & Habib, 2006). Hasil algoritma *Naïve Bayes* juga bagus pada Bahasa Indonesia (Natalius, 2011). Sehingga dari latar belakang dan beberapa penelitian yang disebutkan untuk itu peneliti tertarik untuk melakukan penelitian mengenai metode *naïve bayes* untuk klasifikasi terjemahan ayat Al-Qur'an dengan 5 label kelas atau 5 klasifikasi yaitu, akhlaq, akidah, ibadah, kisah dan muamalah dengan *dataset* Al-Qur'an terjemahan berbahasa Indonesia yang terdapat di *www.tanzil.net*. Maka peneliti mengambil judul dalam kegiatan penelitian ini adalah “Klasifikasi Dokumen Terjemahan Menggunakan Algoritma *Naïve Bayes* Pada Al-Qur'an Versi Berbahasa Indonesia (Juz 1 Sampai Juz 4)”.

1.2 Identifikasi Masalah

Berdasarkan latar belakang yang sudah dijelaskan sebelumnya dan dapat disimpulkan bahwa permasalahan yang ada yaitu :

1. Hasil kategorisasi teks terkadang memberikan hasil yang sangat luas, sehingga informasi yang diterima pengguna kurang relevan.
2. Terjemahan ayat ada yang belum memiliki kategori atau label kelas. Sehingga ayat yang belum memiliki label kelas diprediksi dengan mencari informasi dari ayat yang sudah memiliki label kelas.
3. Terjemahan ayat juga seringkali memiliki lebih dari satu kategori (*Multi-label*). Sehingga diperlukannya pemodelan ulang dengan menggunakan Algoritma *Naïve Bayes*.

1.3 Rumusan Masalah

Berdasarkan identifikasi masalah tersebut, maka permasalahan yang akan dibahas dalam penelitian ini adalah “Bagaimana mengklasifikasikan terjemahan Al-Qur'an berbahasa Indonesia dengan menggunakan algoritma *naïve bayes*?”

1.4 Batasan Masalah

Agar penyusunan tugas akhir ini tidak keluar dari pokok permasalahan yang dirumuskan, maka ruang lingkup pembahasan dibatasi pada:

1. Data yang digunakan terdiri dari terjemahan Al-Qur'an bahasa Indonesia yang terdapat dalam tanzil Juz 1 sampai Juz 4 khusus surat Al-Baqarah, Ali 'Imran, dan An- Nisa. (www.tanzil.net)

2. Terdapat 5 topik yang di klasifikasikan berdasarkan Al-Qur'an terjemahan versi *The Wisdom* yaitu, Akhlaq, Akidah, Ibadah, Muamalah, dan kisah. (*Agama, Al- Qur'an The wisdom*)
3. Algoritma yang digunakan dalam pengklasifikasian ini adalah *Naïve Bayes Classifier* dan tidak membandingkannya dengan algoritma lain.
4. *Tools* yang digunakan yaitu menggunakan *Apache Mahout* dengan *localhost Hadoop* , tampilan hasil berupa *Confusion Matrix* dan tidak terdapat fitur pencarian hanya pengelompokan terjemahan Al-Qur'an.

1.5 Tujuan Penelitian

Berdasarkan penjelasan di dalam latar belakang, penelitian ini memiliki 2 tujuan umum dan tujuan khusus, yaitu:

1. Tujuan Umum

Adapun tujuan umum dari penelitian ini adalah untuk mengklasifikasi terjemahan Al-Qur'an berbahasa Indonesia menjadi

lima topik pada surat Al-Baqarah, Ali 'Imran dan An-Nisa. (*Agama, Al-Qur'an The wisdom*)

2. Tujuan Khusus

Adapun tujuan khusus dari penelitian ini adalah

1. Menghasilkan kategorisasi teks yang dapat memberikan informasi yang relevan.

2. Memprediksi terjemahan ayat Al-Qur'an yang belum memiliki label kelas.
3. Mengklasifikasikan terjemahan berdasarkan satu label kelas sesuai kategori yang ditentukan.

1.6 Manfaat Penelitian

Manfaat penelitian ini adalah sebagai berikut:

1. Membantu pengguna untuk menemukan informasi yang relevan dari data terjemahan Al-Qur'an berbahasa Indonesia yang telah di klasifikasi.
2. Memberikan referensi mengenai implementasi algoritma *naïve bayes* terhadap klasifikasi terjemahan Al-Qur'an Bahasa Indonesia.

1.7 Metodologi Penelitian

Metodologi penelitian yang digunakan pada penelitian ini menggunakan ini adalah :

1. Metodologi Pengumpulan data

Dalam penelitian ini untuk mengumpulkan data, peneliti melakukan tiga

cara yaitu :

1. Observasi

Observasi peneliti lakukan di Pusat Laboratorium Terpadu UIN

Jakarta kurang lebih selama satu tahun.

2. Studi Pustaka

Pengumpulan data dengan cara mengambil dari sumber-sumber media cetak maupun elektronik, jurnal, skripsi, *e-book* dan yang dapat dijadikan acuan pemecahan masalah.

3. Studi *Literature* Sejenis

Studi *literature* dilakukan untuk bahan acuan yang dijadikan sebagai landasan kegiatan penelitian.

2. Metodologi Kategorisasi Teks

Metodologi kategorisasi teks menggunakan Algoritma *Naïve Bayes*, memiliki tahapan-tahapan yaitu:

1. *Extraction Data*

Extraction data yaitu membaca dan menerjemahkan input dataset agar menjadi objek yang dimengerti sistem.

2. *Text Pre-Processing*

Fase ini merupakan proses awal untuk mempersiapkan teks menjadi data yang akan diolah lebih lanjut.

3. *Feature Selection*

Feature Selection bertujuan untuk mengurangi dimensi dari suatu kumpulan teks, atau dengan kata lain menghapus kata-kata yang dianggap tidak penting.

4. *Feature Extraction*

Fase ini, kata-kata yang sering muncul diekstrak berdasarkan data *tf-idf* (*term frequency-inverse document frequency*) yang telah

dibuat pada fase sebelumnya.

5. *Classification Stage*

Classification stage merupakan fase terakhir yaitu menghasilkan kategorisasi berdasarkan data test yang dibuat.

1.8 Sistematika Penulisan

Untuk memberikan gambaran secara menyeluruh terhadap masalah yang sedang di analisis, maka sistematika penulisan dapat dibagi dalam beberapa bab sebagai berikut :

BAB 1 PENDAHULUAN

Bab ini membahas mengenai latar belakang masalah, identifikasi masalah, rumusan masalah, batasan masalah, tujuan penelitian, manfaat penelitian, metodologi penelitian dan sistematika penulisan.

BAB 2 LANDASAN TEORI

Bab ini membahas mengenai definisi dan teori-teori yang digunakan sebagai acuan atau dasar dalam penelitian seperti pengertian

klasifikasi, kategorisasi, *text mining*, *information extraction*, *information retrieval*, *natural language processing*, *data mining*, *text-preprocessing*, tokenisasi, *stopword elimination*, *stemming*, *text transformation*, model ruang vector, *feature selection*, *feature extraction*, *patern discovery*, *clustering*, *classification*, perbedaan *clustering* dan *classification*, *naïve bayes clasifier*, *mahout*, *confusion matrix*.

BAB 3 METODOLOGI PENELITIAN

Bab ini membahas mengenai metodologi yang digunakan dalam penelitian yang mencakup metode klasifikasi teks menggunakan algoritma *Naïve Bayes*, yaitu metodologi pengumpulan data, material, *pre-processing*, *Feature Selection*, *Feature extraction*, *Clasification Stage*.

BAB 4 HASIL DAN PEMBAHASAN

Bab ini berisi penjelasan hasil dan pembahasan mengenai klasifikasi pada terjemahan Al-Qur'an. Hasil klasifikasi yang didapatkan menggunakan algoritma *naïve bayes* berdasarkan 4 tahap proses klasifikasi *pre-processing*, *Feature Selection*, *Feature extraction*, *Clasification Stage*. Selanjutnya pembahasan dilakukan dengan melihat pada teori sebelumnya dengan memperhatikan dan menimbang berdasarkan eksperimen yang dilakukan.

BAB 5 PENUTUP

Bab ini merupakan bab terakhir yang berisi kesimpulan dari apa yang telah diuraikan pada bab sebelumnya serta saran untuk perkembangan penelitian yang akan datang.

BAB 2

LANDASAN TEORI

Bab ini menjelaskan tentang teori-teori terkait landasan pelaksanaan penelitian, meliputi klasifikasi, kategorisasi, *text mining*, *information extraction*, *information retrieval*, *natural language processing*, *data mining*, *text-preprocessing*, tokenisasi, *stopword elimination*, *stemming*, *text transformation*, model ruang vector, *feature selection*, *feature extraction*, *patern discovery*, *clustering*, *classification*, perbedaan *clustering* dan *classification*, *naïve bayes clasifier*, *mahout*, *confusion matrix*. Teori-teori tersebut akan dijelaskan dalam bab ini secara berurutan.

2.1 Kategorisasi

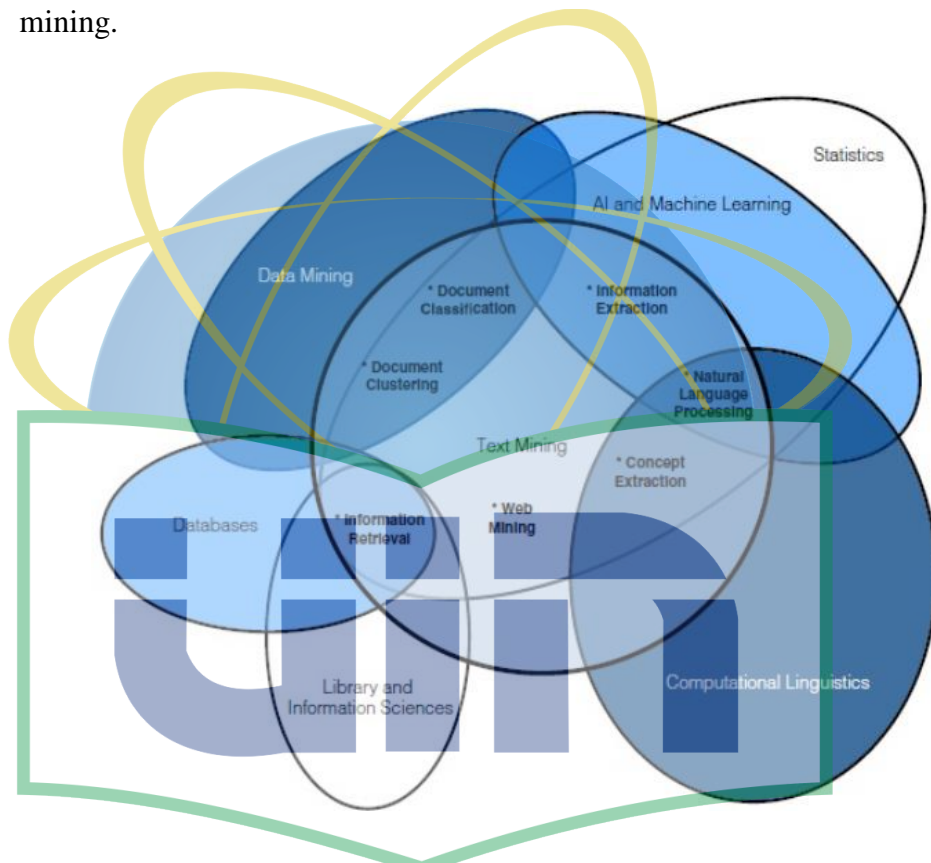
Kategorisasi adalah spesialisasi klasifikasi. Ini berhubungan dengan menugaskan kategori ke objek. Algoritma klasifikasi lainnya mungkin hanya membuat keputusan ya / tidak berdasarkan masukan, seperti *detektor* penipuan yang mengindikasikan transaksi pada kartu kredit itu palsu atau tidak palsu. Algoritma kategorisasi menempatkan suatu objek di salah satu sekumpulan kategori kecil, seperti mengelompokkan mobil sebagai *coupe*, *sedan*, *SUV*, atau *van*. Banyak konsep yang kita bahas dalam bab ini berkaitan dengan klasifikasi secara keseluruhan, sedangkan yang lain lebih terkait erat dengan kategorisasi. Anda akan melihat beberapa istilah yang digunakan secara bergantian (Ingersol, Grant, S; Morton, Thomas, S; Farris, 2012).

2.2 Text Mining

Teks adalah artian paling umum untuk bertukar informasi. Data yang dikumpulkan dalam komputer bisa menjadi bentuk terstruktur, semi terstruktur dan tidak terstruktur. Data yang disimpan dalam basis data sebagai contoh untuk sekumpulan data terstruktur. Contoh data semi terstruktur dan kumpulan data tidak terstruktur termasuk *email*, dokumen *full-text* dan file *HTML* (Putra, et.al. 2018)

Sekarang banyak jumlah data yang disimpan dalam basis data teks dan tidak dalam basis data terstruktur. Penelusuran teks didefinisikan sebagai proses menemukan bentuk yang tersembunyi, bermanfaat dan menarik dari teks dokumen yang tidak terstruktur. *Text mining* juga diketahui sebagai analisis teks yang pintar atau penemuan pengetahuan dalam teks atau teks penelusuran data. Kira-kira 80% dari data *corporate* adalah dalam format yang tidak terstruktur. Penelusuran informasi dari teks yang tidak terstruktur sangat kompleks sebagaimana berisi informasi massal yang membutuhkan metode pemrosesan khusus dan algoritma untuk mengekstrak bentuk yang berguna. Sebagai bentuk yang paling mungkin dalam penyimpanan informasi adalah teks, penelusuran teks dipertimbangkan untuk memiliki nilai yang tinggi dari itu dalam penelusuran data. Penelusuran teks adalah bidang studi yang melingkupi penelusuran data, penelusuran *web*, penelusuran informasi, ekstraksi informasi, bahasa komputasi, dan pemrosesan bahasa murni (Sumathy & Chidambaram, 2013).

Teks mining adalah cabang bidang ilmu yang menggabungkan data mining, web mining, information retrieval, information extraction, computational linguistics dan natural language processing (Sumathy & Chidambaram, 2013). Gambar 2.1 menggambarkan diagram venn dari text mining.



Gambar 2. 1 Venn Diagram Of Text Mining Interaction With Other Field (Liao, Chu, & Hsiao, 2012)

2.2.1 Information Extraction

Information extraction adalah proses mengekstrak struktur informasi dokumen teks secara otomatis dari tidak terstruktur dan/atau semi terstruktur. Sebuah sistem *Information extraction* melibatkan pengidentifikasian entitas seperti nama orang, perusahaan dan lokasi, atribut dan relasi antar entitas. *Information extraction* melakukannya berdasarkan

aturan yang diakui. *Information extraction* adalah proses pencarian untuk rangkaian yang belum terdefinisi dari teks dalam dokumen teks. Sejak *Information extraction* menunjukan masalah perubahan sebuah kesatuan dokumen tekstual menjadi database yang lebih terstruktur, database terkonstruksi oleh sebuah modul *Information extraction* dapat dilengkapi ke modul KDD untuk penggalian lebih lanjut tentang pengetahuan (Sumathy & Chidambaram, 2013).

2.2.2 *Information Retrieval*

Information Retrieval didefinisikan sebagai metode yang digunakan untuk representasi, penyimpanan dan akses informasi di mana informasi ditangani sebagian besar dalam bentuk dokumen *tekstual*, koran iklan buku yang diambil dari *database* berdasarkan pada permintaan ataupun *query* pengguna. *Information Retrieval* dianggap sebagai perluasan pengambilan dokumen di mana dokumen diproses untuk di singkat ataupun ekstrak sebagian informasi yang diminta pengguna. Sebuah sistem IR bisa mempercepat analisa secara signifikan dengan mengurangi jumlah dokumen untuk analisa (Sumathy & Chidambaram, 2013).

2.2.3 *Natural Language Processing*

Natural Language Processing adalah masalah yang paling menantang dibidang kecerdasan buatan. *Natural Language Processing* mempelajari bahasa manusia sehingga komputer bisa mengerti bahasa alami yang sama dengan manusia. *Natural Language Processing* berfokus pada *Natural*

Language Generation (NLG) dan *Natural Language Understanding* (NLU). NLG memastikan teks yang diturunkan benar secara tata bahasa. Kebanyakan sistem NLG termasuk *syntactic realizer* untuk memastikan aturan tata bahasa seperti kata kerja persetujuan subjek yang ditaati dan *text planner* untuk menentukan bagaimana mengurutkan kalimat, paragraf, dan bagian lainnya secara koheren. Aplikasi NLG yang terkenal adalah mesin translasi. NLU terdiri setidaknya satu dari komponen berikut: penanda, penganalisa *lexical*, penganalisa *sintaks*, dan penganalisa *semantic* (Sumathy & Chidambaram, 2013).

2.2.4 Data Mining

Data *mining* mengarah kepada penemuan informasi relevan atau penemuan pengetahuan dari jumlah data yang banyak. Data *mining* mencoba untuk menemukan aturan statistik dan pola secara otomatis. *Tools* data *mining* dapat memprediksi kebiasaan dan tren ke depan yang bermanfaat bagi perusahaan untuk membuat pengetahuan positif. Tujuan keseluruhan dari data *mining* adalah untuk mengekstrak informasi dari kumpulan data dan mengubahnya menjadi struktur yang dapat dimengerti untuk analisa selanjutnya (Sumathy & Chidambaram, 2013).

Data *mining* merupakan suatu kegiatan yang meliputi pengumpulan, pemakaian data historis untuk menentukan keteraturan, pola atau hubungan dalam set data berukuran besar. Salah satu tugas utama dari data *mining* adalah pengelompokan *clustering* dimana data yang dikelompokkan belum mempunyai contoh kelompok.

Secara garis besar data *mining* dapat dikelompokkan menjadi 2 kategori utama :

- a. *Descriptive mining*, yaitu proses untuk menemukan karakteristik penting dari data dalam suatu basis data. Teknik data *mining* yang termasuk dalam *descriptive mining* adalah *clustering*, *association*, dan *sequential mining*.
- b. *Predictive mining*, yaitu proses untuk menemukan pola dari data dengan menggunakan beberapa variabel lain di masa depan. Salah satu teknik yang terdapat dalam *predictive mining* adalah klasifikasi.

Pengelompokan data *mining* berdasarkan tugas bisa dikategorikan sebagai berikut:

a. Deskripsi

Deskripsi adalah menggambarkan pola dan kecenderungan yang terdapat dalam data yang memungkinkan memberikan penjelasan dari suatu pola atau kecenderungan tersebut.

b. Estimasi

Estimasi hampir sama dengan klasifikasi, akan tetapi variabel target estimasi lebih ke arah *numerik* daripada ke arah kategori.

c. Prediksi

Prediksi hampir sama dengan klasifikasi dan estimasi, akan tetapi dalam prediksi nilai dari hasil akan terwujud di masa yang akan datang.

d. Klasifikasi

Klasifikasi adalah proses untuk menemukan model atau fungsi yang menggambarkan dan membedakan kelas data atau konsep dengan tujuan memprediksikan kelas untuk data yang tidak diketahui kelasnya (Han and Kamber, 2006).

e. *Clustering*

Clustering atau analisis *cluster* adalah proses pengelompokan satu set benda- benda fisik atau abstrak ke dalam kelas objek yang sama (Han & Kamber, 2006).

f. Asosiasi

Asosiasi dalam data *mining* adalah menemukan atribut yang muncul dalam satu waktu.

Adapun langkah-langkah yang umumnya dilakukan dalam data *mining* adalah meliputi *text pre-processing*, *tokenization*, *stopword elimination*, *stemming*, *text transformation*, *feature selection*, dan *pattern discovery*.

2.3 Text Pre-Processing

Text pre-processing dilakukan untuk membuang data yang tidak konsisten dan *noise*, duplikasi data, memperbaiki kesalahan data. *Text preprocessing* bertujuan untuk mempersiapkan teks menjadi data yang diproses pada tahapan berikutnya.

2.3.1 Tokenisasi

Langkah pertama setelah mengekstraksi konten dari sebuah berkas adalah

memilih teks berukuran kecil yang bisa digunakan kembali, yang biasa disebut dengan *tokens*. *Tokens* sering kali mempresentasikan satu kata. Cara pertama paling umum dalam melakukan *tokenization* adalah dengan memilah *string* berdasarkan karakter spasi yang ditemukan (Putra, et.al. 2018)

Jika *tokenization* berbasis karakter spasi diaplikasikan pada contoh terjemahan surat Al-Fatihah ayat pertama “Dengan menyebut nama Allah Yang Maha Pemurah lagi Maha Penyayang”.

Adapun teknik-teknik umum yang biasa diaplikasikan pada *token level* adalah termasuk berikut ini:

- Pengubahan format kata - Pengubahan kata menjadi huruf kecil semua dapat membantu dalam pencarian.
- Menghilangkan *stopword* - Menghilangkan kata-kata umum seperti “dan”, “sebuah”, dan “itu”.
- *Expansion* - Menambahkan sinonim, akronim, dan abreviasi pada *token stream* bisa membuat aplikasi untuk menangani berbagai macam *input* dari *user*.
- *Part of Speech Tagging* - Menentukan jenis kata terhadap *token*, misalnya kata benda, kerja, keterangan dan lain-lain.
- *Stemming* - Mengubah kata menjadi kata dasar, misalnya “menulis” menjadi “tulis”.

2.3.2 Stopword Elimination

Stopwords adalah kata-kata yang terlalu sering diantara *file* teks yang tidak memiliki arti yang khusus dalam *Information Retrieval*. Penghilangan *stopwords* memiliki keunggulan yang penting. Mengurangi ukuran struktur pengindeksan. Untuk setiap kata pada teks jika dia adalah *stopword* maka kata tersebut termasuk kata yang sia-sia (Al-taani & Al-gharaibeh, 2010).

2.3.3 Stemming

Stemming adalah proses mengubah kata menjadi kata dasar atau kata yang lebih mudah. *Stemming* sering digunakan dalam aplikasi pencarian karena *user* biasanya berharap untuk mendapatkan hasil mengenai perbankan ketika menggunakan kata “*bank*” sebagai kata kuncinya. Ada banyak berbagai macam pendekatan untuk melakukan *stemming*, masing-masing memiliki desain dengan tujuan yang berbeda (Ingersol, Grant, S; Morton, Thomas, S; Farris, 2012).

Stemming merupakan suatu proses atau cara dalam menemukan kata dasar dari suatu kata. *Stemming* sendiri berfungsi untuk menghilangkan variasi-variasi morfologi yang melekat pada sebuah kata dengan cara menghilangkan imbuhan-imbuhan pada kata tersebut, sehingga nantinya di dapat suatu kata yang benar sesuai struktur morfologi bahasa Indonesia yang benar.

Ada tiga jenis metode *stemming*, antara lain :

1. **Successor Variety (SV)** : lebih mengutamakan penyusunan huruf dalam kata dibandingkan dengan pertimbangan atas fonem. Contoh untuk kata-kata : *corpus, able, axle, accident, ape, about* menghasilkan SV untuk kata *apple*:

- Karena huruf pertama dari kata "*apple*" adalah "a", maka kumpulan kata yang ada substring "a" diikuti "b", "x", "c", "p" disebut SV dari "a" sehingga "a" memiliki 4 SV.
- Karena dua huruf pertama dari kata "*apple*" adalah "ap", maka kumpulan kata yang ada substring "ap" hanya diikuti "e" disebut SV dari "ap" sehingga "ap" memiliki 1 SV.

2. **N-Gram Conflation** : ide dasarnya adalah pengelompokan kata-kata secara bersama berdasarkan karakter-karakter (*substring*) yang teridentifikasi sepanjang N karakter.

3. **Affix Removal** : membuang *suffix* dan *prefix* dari term menjadi suatu *stemming*. Yang paling sering digunakan adalah algoritma *Porter Stemmer* karena modelnya sederhana dan efisien.

2.4 Text Transformation

Text transformation dilakukan penyaringan (*filtration*) dengan menentukan *term* mana yang digunakan untuk merepresentasikan dokumen dengan cara menghilangkan *stop-word*, yaitu *term* yang tidak berhubungan (*irrelevant*) dengan subjek utama dari *database* meskipun kata tersebut

sering kali hadir di dalam dokumen.

Dokumen teks direpresentasikan dengan isi kata-kata dan kejadiannya.

Dua pendekatan yang digunakan dalam representasi *document*, yaitu:

1. *Bag of words*

Digunakan untuk menunjukkan representasi sederhana dari teks yang digunakan dalam model pengambilan dan klasifikasi. Dalam representasi ini, sebuah dokumen dipertimbangkan sebagai koleksi kata yang tidak punya hubungan, tidak *syntactic* ataupun *statistical*, atau diantaranya. Model *bag of words* mengasumsikan tidak adanya hubungan antar kata, jadi kita melihat bagaimana istilah ketergantungan dapat diambil dan digunakan dalam model *linear* berbasis fitur. Aspek pertama struktur *document* diabaikan dalam representasi *bag of words*, tapi dilihat dalam pencarian web. Aspek kedua dilihat dari representasi terstruktur yang digunakan dalam sistem *database* dalam sebuah mesin pencarian. Aspek ketiga dari representasi kita lihat bagaimana struktur *query* digunakan untuk menjawab pertanyaan. Akhirnya, *bag of words* didasarkan pada kata dan banyak aplikasi, seperti pencarian gambar ataupun pencarian musik dimana fitur digunakan untuk merepresentasikan objek yang diambil bukanlah kata.

2. *Vector spaces model*.

Dalam mengukur kemiripan, dokumen didefinisikan berdasarkan representasi *bag-of-words* yang dikonversi ke suatu model ruang vektor (*vector space model*, VSM). Setiap dokumen di dalam *database*

dan *query* pengguna direpresentasikan oleh suatu vektor multi-dimensi. Dimensi sesuai dengan jumlah *term* dalam dokumen yang terlibat. Contoh dari model ruang vektor tiga dimensi untuk dua dokumen D1 dan D2, satu *query* pengguna Q1, dan tiga *term* T1, T2 dan T3 diperlihatkan pada gambar berikut.

2.4.1 Model Ruang Vektor

Dalam model ruang vektor, koleksi dokumen direpresentasikan oleh matriks *term-document* (atau matriks *term-frequency*). Setiap sel dalam matriks bersesuaian dengan bobot yang diberikan dari suatu *term* dalam dokumen yang ditentukan. Nilai nol berarti bahwa *term* tersebut tidak hadir di dalam dokumen.

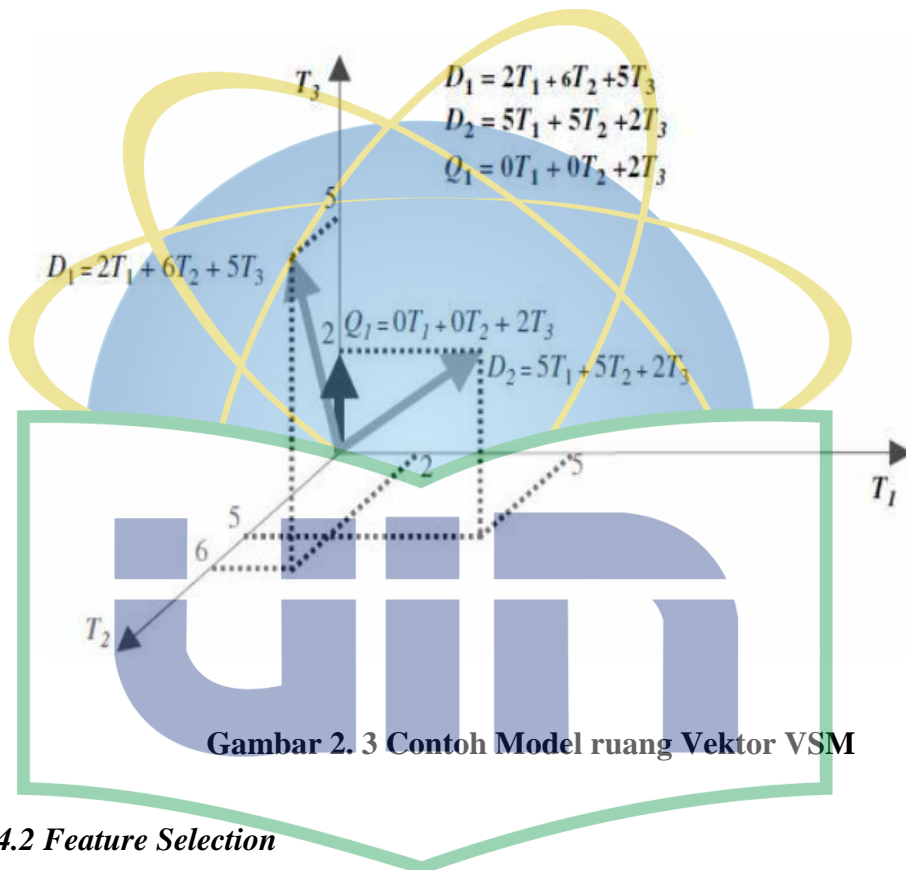
Pada Gambar 2.2 menampilkan contoh matriks *term-document* untuk database dengan n dokumen dan t *term*.

	T_1	T_2	\dots	T_t
D_1	w_{11}	w_{21}	\dots	w_{t1}
D_2	w_{12}	w_{22}	\dots	w_{t2}
\dots	\dots	\dots	\dots	\dots
D_n	w_{1n}	w_{2n}	\dots	w_{tn}

Gambar 2. 2 Contoh matriks *term-document*

Keberhasilan dari model VSM ini ditentukan oleh skema pembobotan terhadap suatu *term* baik untuk cakupan lokal maupun global, dan faktor

normalisasi. Pembobotan lokal hanya berpedoman pada frekuensi munculnya *term* dalam suatu dokumen dan tidak melihat frekuensi kemunculan term tersebut di dokumen lainnya. Contoh model ruang vector (VSM) dapat dilihat pada Gambar 2.3.



2.4.2 Feature Selection

Feature selection adalah sebuah proses pemilihan subset fitur dari fitur asli. Tahap seleksi fitur (*feature selection*) bertujuan untuk mengurangi dimensi dari suatu kumpulan teks, atau dengan kata lain menghapus kata-kata yang dianggap tidak penting atau tidak menggambarkan isi dokumen sehingga proses pengklasifikasian lebih efektif dan akurat (Berry et al, 2010). Berdasarkan ada atau tidaknya informasi label, *feature selection* dapat dibedakan menjadi 2 jenis yaitu *supervised feature selection* dan

unsupervised feature selection. Pada *supervised feature selection* dibutuhkan label kelas sedangkan pada *unsupervised feature selection* tidak membutuhkan label kelas.

2.4.3 Feature Extraction

Ekstraksi Fitur adalah proses reduksi atribut. Tidak seperti pemilihan fitur, yang memilih dan mempertahankan atribut yang paling signifikan, Ekstraksi Fitur benar-benar mengubah atribut. Atribut atau fitur yang ditransformasikan adalah kombinasi linier dari atribut aslinya.

Proses ekstraksi fitur menghasilkan seperangkat atribut yang jauh lebih kecil dan lebih kaya. Jumlah maksimum fitur dapat ditentukan pengguna atau ditentukan oleh algoritma. Secara *default*, algoritma menentukannya. Model yang dibangun dengan fitur yang diekstraksi bisa lebih berkualitas, karena atribut yang lebih sedikit dan lebih bermakna menggambarkan data.

Fitur Ekstraksi memproyeksikan kumpulan data dengan dimensi yang lebih tinggi ke sejumlah dimensi yang lebih kecil. Karena itu, ini berguna untuk visualisasi data, karena kumpulan data yang kompleks dapat divisualisasikan secara efektif saat dikurangi menjadi dua atau tiga dimensi.

Beberapa aplikasi ekstraksi fitur adalah analisis semantik laten, kompresi data, dekomposisi data dan proyeksi, dan pengenalan pola. Ekstraksi Fitur juga dapat digunakan untuk meningkatkan kecepatan dan efektivitas pembelajaran yang diawasi.

Ekstraksi Fitur dapat digunakan untuk mengekstrak tema koleksi dokumen, di mana dokumen ditunjukkan oleh serangkaian kata kunci dan frekuensinya. Setiap tema (fitur) diwakili oleh kombinasi kata kunci. Dokumen dalam koleksi kemudian dapat diungkapkan berdasarkan tema yang ditemukan.

2.5 *Pattern Discovery*

Pattern discovery adalah penemuan pola yang terdapat pada dataset. Pola adalah kumpulan item yang seringkali muncul pada dataset dan berkorelasi secara kuat. Pola biasanya bersifat *intrinsic* dan properti yang penting dari suatu data. Oleh karena itu perlu ditemukan dan dalam hal ini *text mining* menjadi data mining. Metode data mining seperti *clustering*, *classification*, *information retrieval*, dan lain-lain dapat digunakan untuk *pattern discovery*.

2.6 *Clustering*

Clustering adalah pengelompokan data tanpa berdasarkan kelas kata tertentu, sehingga dapat pula dipakai untuk memberikan label pada kelas kata yang belum diketahui. Oleh karena itu, *clustering* digolongkan sebagai metode *unsupervised learning*. Prinsip dari *clustering* adalah memaksimalkan kesamaan antar anggota satu kelas dan meminimalkan kesamaan antar kelas atau *cluster*. Salah satu contoh dari aplikasi yang menggunakan *clustering* adalah *Google News* dan *online retail industry* (Choiruddin, 2005).

Clustering dapat diimplementasikan pada berbagai macam jenis aspek teks, termasuk kata-kata pada dokumen, dokumen itu sendiri atau hasil dari pencarian. Sebagai catatan, *clustering* juga bias digunakan untuk banyak hal.

Selain teks, misalnya untuk mengelompokkan *user*, mengidentifikasi daerah padat dan lain-lain.

2.7 Classification

Classification/Klasifikasi dalam pengertian komputasi, berusaha memberi label pada data. Dengan satu set fitur untuk sebuah objek, *classifier* mencoba menetapkan label pada objek itu. *Classifier* melakukan ini dengan menggambar berdasarkan pengetahuan yang berasal dari contoh bagaimana benda lain diberi label. Contoh-contoh ini, disebut sebagai data pelatihan, berfungsi sebagai sumber pengetahuan sebelumnya yang digunakan oleh pengklasifikasi untuk membuat keputusan tentang objek yang sebelumnya tak terlihat (Ingersol, Grant, S; Morton, Thomas, S; Farris, 2012).

Classification/klasifikasi dalam bidang komputasi adalah pencarian untuk menentukan label-label terhadap data. *Classifier* melakukan ini melalui *knowledge* yang di dapat dari label-label yang terdapat pada data contoh. Contoh-contoh tersebut disebut dengan *training data*, disajikan sebagai sumber pengetahuan yang digunakan oleh *classifier* untuk membuat keputusan terhadap objek-objek yang belum terlihat sebelumnya.

Pengklasifikasian teks memberikan satu atau lebih kelas-kelas *document* sesuai dengan isinya. Kelas dipilih dari taksonomi yang telah dibuat sebelumnya (hirarki dari kategori atau kelas). API pengklasifikasian teks bertanggung jawab pada semua *preprocessing tasks* (mengeksrak teks, *tokenisasi*, *stopword removal*, dan *lemmatization*) yang dibutuhkan untuk pengklasifikasian otomatis.

API ini mendukung jenis dari skenario pengklasifikasian teks, seperti :

1. Pengklasifikasian biner, seperti penyaringan spam.
2. Pengklasifikasian banyak kelas seperti pemilihan satu kategori dari beberapa alternatif : pengklasifikasian jenis film (horror, komedi, romansa, dan lain-lain)
3. Pengkategorian banyak label : menetapkan semua kategori yang diaplikasikan menjadi sebuah dokumen.
4. Pengkategorian taksonomi kompleks : menetapkan penyusunan kategori dalam sebuah taksonomi multi level.

Algoritma menggabungkan klasifikasi dokumen statistik dengan aturan dasar penyaringan, yang mengizinkan untuk memperoleh presisi tingkat atas dalam sebuah batasan yang luas pada suatu lingkungan. Pengklasifikasi statistik menyediakan arti-arti untuk menggunakan contoh-contoh dokumen untuk mendefinisikan tiap kategori. Sebaliknya, aturan dasar pengklasifikasian memungkinkan bantuan untuk penyetelan yang tepat sebuah klasifikasi dan memeriksa hasil dari penyaring statistik.

2.7.1 Proses Klasifikasi

Mengembangkan klasifikasi otomatis mengikuti proses umum yang sama terlepas dari algoritma yang digunakan. Proses ini terdiri dari beberapa tahap: persiapan, pelatihan, pengujian, dan produksi. Seringkali

melakukan iterasi melalui proses ini, mengulang setiap fase beberapa kali, baik secara otomatis maupun manual, untuk memudahkan penyesuaian perilaku penggolong dan menghasilkan hasil terbaik. Umpan balik dari setiap fase membantu menentukan cara fase persiapan dan pelatihan agar dimodifikasi untuk menghasilkan hasil yang lebih baik. *Training* mewakili pengambilan hasil dari tahap pengujian dan menggunakannya untuk memperbaiki proses pelatihan. Setelah pengklasifikasi telah dimasukkan ke dalam produksi, seringkali perlu diperluas untuk mencakup kasus-kasus tambahan yang tidak tercakup dalam data pelatihan (Ingersoll, Morton, & Farris, 2013).

Tahap persiapan melibatkan data siap untuk proses pelatihan. Di sini Anda memilih seperangkat label yang akan dilatih untuk mengidentifikasi, cara fitur yang Anda gunakan untuk pelatihan akan diidentifikasi, dan item dalam kumpulan data yang akan ditahan untuk pengujian (Ingersol, Grant, S; Morton, Thomas, S; Farris, 2012).

Pada tahap pengujian, algoritma klasifikasi dievaluasi dengan menggunakan data contoh tambahan yang dikenal sebagai data uji. Setelah pengklasifikasi dimasukkan ke dalam produksi, seringkali nantinya perlu dilatih ulang untuk memperluas pengetahuan tentang domain yang dipakainya. Hal ini biasanya terjadi ketika kosakata baru muncul dari waktu ke waktu yang memainkan peran kunci dalam membedakan antara satu kelas dengan kelas lainnya (Ingersol, Grant, S; Morton, Thomas, S; Farris, 2012).

2.7.2 Memilih Skema Klasifikasi

Algoritma klasifikasi dipelajari dengan contoh menggunakan data yang telah diorganisasikan ke dalam kelas secara manual atau melalui beberapa proses otomatis lainnya. Kelas atau kategori yang Anda tetapkan ke objek memiliki nama dan makna yang terpisah dari objek yang diklasifikasikan. Setiap kelas ada dalam konteks kelas lain, sebuah sistem yang dikenal sebagai skema klasifikasi atau kategorisasi. Beberapa skema klasifikasi mungkin kaku, memungkinkan setiap objek jatuh ke dalam satu kelas saja dalam skema ini. Skema klasifikasi lainnya lebih fleksibel dan mengakui bahwa di dunia nyata banyak hal seringkali memiliki aspek atau aspek yang berbeda. Beberapa skema klasifikasi, seperti taksonomi *Linnaean* untuk klasifikasi organisme biologis, mungkin memiliki hierarki makna yang ketat. Orang lain mungkin tidak memiliki struktur tertentu selain yang dikonotasikan oleh hubungan linguistik, seperti *tag* kata kunci sederhana yang ditemukan di *Flickr* atau *Technorati*. Skema klasifikasi bisa sangat bervariasi dalam lingkungannya. Mereka mungkin mencakup area subjek yang luas, seperti Sistem Desimal *Dewey* yang digunakan di perpustakaan, atau skema spesifik domain yang sempit seperti skema yang digunakan untuk menjelaskan bantuan teknis bagi orang-orang penyandang cacat. Dalam banyak konteks, skema klasifikasi bersifat organik dan berkembang. Skema klasifikasi muncul dan berubah dari waktu ke waktu dari bawah ke atas, berlawanan dengan metode *top-down* untuk klasifikasi di mana seperangkat subjek yang telah ditentukan

sebelumnya, mungkin hierarkis tersedia bagi pengguna untuk dipilih (Ingersol, Grant, S; Morton, Thomas, S; Farris, 2012).

2.8 Perbedaan Clustering dan Classification

Secara umum, pada *classification*, kelas-kelas data telah ditentukan, kemudian kelas-kelas data tersebut digunakan untuk menentukan kelas suatu objek baru, sedangkan pada *clustering*, ia mencoba untuk mengelompokkan sekumpulan objek dan menemukan apakah ada suatu hubungan antar objeknya.

Adapun dalam konteks *machine learning*, *classification* masuk dalam kategori *supervised learning*, sedangkan *clustering* masuk dalam kategori *unsupervised learning*. Sebagai contoh adalah pada kasus *face recognition*. Pada kasus *face recognition* menggunakan *classification*, kita harus melatih algoritma terlebih dahulu dengan memasukkan data-data yang telah dilabeli terlebih dahulu, apakah suatu gambar wajah atau tidak, kemudian algoritmanya bisa menentukan apakah suatu foto baru yang di-*input* merupakan suatu gambar wajah atau tidak, sedangkan contoh kasus pada *clustering*, ia tidak bisa menentukan bagaimana gambar suatu wajah karena data yang digunakan tidak berlabel, tetapi ia bisa mencoba untuk mengelompokkan mana foto-foto yang memiliki gambar wajah, gambar pemandangan, gambar kuda, dan lain-lain. Dalam Tabel 2.2 terdapat penjelasan mengenai perbedaan *clustering* dan *classification*.

Tabel 2. 1 Perbedaan Clustering dan Classification

Classification	Clustering
<i>Supervised</i>	<i>Unsupervised</i>
Mengklasifikasikan sampel baru ke dalam kelas yang dikenal.	Mengelompokkan berdasarkan pola dalam data
Algoritma <i>Decision Trees, Bayesian classifiers, K-NN</i> .	Algoritma <i>K-means, Expectation Maximization, Lingo</i>
Data sampel diketahui dari label kelas.	Sampel tidak berlabel

Clustering merupakan bagian dari *text mining unsupervised* yang berarti tidak diperlukannya *training* data untuk memberi sampel pada label, pengelompokan datanya berdasarkan pola tidak dibuat kelas terlebih dahulu. Kemudian Algoritma dalam clustering yaitu *K-means, Expectation Maximization* dan *Lingo*.

Classification merupakan bagian dari *text mining supervised* yang berarti diperlukannya *training* data untuk membuat sampel data agar diketahui dari label kelasnya dan mengklasifikasikan sampel baru kedalam kelas yang dikenal. Kemudian algoritma yang digunakan dalam klasifikasi yaitu Algoritma *Decision Tree, Bayesian Classifier* dan *K-NN*

2.9 Naïve Bayes Classifier

Algoritma *naive bayes classifier* merupakan algoritma yang digunakan untuk mencari nilai probabilitas tertinggi untuk mengklasifikasi data uji pada kategori yang paling tepat. Klasifikasi *Bayesian* didasarkan pada *teorema Bayes*. Studi yang membandingkan algoritma-algoritma klasifikasi telah menemukan sebuah klasifikasi *Bayes* yang sederhana yang dikenal sebagai klasifikasi *Naive Bayes* yang dapat dibandingkan *performance*-nya dengan

klasifikasi keputusan dan jaringan syarat tiruan. Klasifikasi *Bayes* juga telah memperlihatkan keakurasian yang tinggi dan kecepatan yang baik ketika dijalankan pada *database* yang besar.

Ada dua tahap pada klasifikasi dokumen. Tahap pertama adalah pelatihan terhadap dokumen yang sudah diketahui kategorinya. Sedangkan tahap kedua adalah proses klasifikasi dokumen yang belum diketahui kategorinya.

Dalam algoritma *naïve bayes classifier* setiap dokumen direpresentasikan dengan pasangan atribut " $x_1, x_2, x_3, \dots, x_n$ " dimana x_1 adalah kata pertama, x_2 adalah kata kedua dan seterusnya. Sedangkan V adalah himpunan kategori terjemahan. Pada saat klasifikasi algoritma akan mencari probabilitas tertinggi dari semua kategori dokumen yang diujikan (V_{MAP}), dimana persamaannya adalah sebagai berikut:

$$V_{MAP} = \operatorname{argmax} \frac{P(x_1, x_2, x_3, \dots, x_n | V_j) P(V_j)}{P(x_1, x_2, x_3, \dots, x_n)} \quad (2.1)$$

Untuk $P(x_1, x_2, x_3, \dots, x_n)$ nilainya konstan untuk semua kategori (V_j) sehingga persamaan dapat ditulis sebagai berikut :

$$V_{MAP} = \operatorname{argmax} P(x_1, x_2, x_3, \dots, x_n | V_j) P(V_j) \quad (2.2)$$

Persamaan diatas dapat disederhanakan menjadi sebagai berikut:

$$V_{MAP} = \operatorname{argmax} \pi_{i=1}^n P(x_i | V_j) P(V_j) \quad (2.3)$$

Keterangan :

V_j = Kategori terjemahan $j = 1, 2, 3, 4, 5, \dots, n$. Dimana penelitian ini kategori terjemahan, $j_1 =$ akhlak, $j_2 =$ akidah, $j_3 =$ ibadah, $j_4 =$ kisah, $j_5 =$ muamalah.

$P(x_1|V_j)$ = Probalitas x_1 pada kategori V_j

$P(V_j)$ = Probalitas dari V_j

Untuk $P(V_j)$ dan $P(x_1|V_j)$ dihitung pada saat pelatihan dimana persamaannya adalah sebagai berikut :

$$P(V_j) = \frac{|docs\ j|}{|contoh|} \quad (2.4)$$

$$P(x_i|V_j) = \frac{n_k + 1}{n + |kosakata|} \quad (2.5)$$

Keterangan :

$|docs\ j|$ = jumlah dokumen setiap kategori

$|contoh|$ = jumlah dokumen dari semua kategori

n_k = jumlah frekuensi kemunculan setiap kata

n = jumlah frekuensi kemunculan kata dari setiap kategori

$|kosakata|$ = jumlah semua kata dari semua kategori

Sebagai contoh dari kategori akhlak pada surah Al-Baqarah ayat 9 yang berbunyi “*Mereka hendak menipu Allah dan orang-orang yang beriman, padahal mereka hanya menipu dirinya sendiri sedang mereka tidak sadar.*” didapat kata dari hasil preprocessing yaitu: “menipu Allah orang iman sadar, kemudian dihitung nilai probabilitasnya menggunakan rumus di atas.

$$P(\text{menipu} | \text{Akhlak}) = (1+1) / (5+215) = 0.009$$

$$P(\text{Allah} | \text{Akhlak}) = (1+1) / (5+215) = 0.009$$

$$P(\text{Orang} | \text{Akhlak}) = (1+1) / (5+215) = 0.009$$

$$P(\text{Iman} | \text{Akhlak}) = (1+1) / (5+215) = 0.009$$

$$P(\text{Sadar} | \text{Akhlak}) = (1+1) / (5+215) = 0.009$$

2.10 Classification Tools

Dalam pengolahan data untuk klasifikasi ada beberapa tools yang digunakan salah satunya adalah mahout, untuk penjelasannya dapat dilihat di bawah ini.

2.10.1. Mahout

Mahout memulai kehidupan di tahun 2008 sebagai subproyek proyek *Lucene Apache* yang menyediakannya. Mesin pencari *open source* yang terkenal dengan nama yang sama. *Lucene* menyediakan implementasi lanjutan teknik pencarian, penambangan teks, dan pencarian informasi. Di dunia ilmu komputer, konsep-konsep ini bersebelahan dengan mesin teknik belajar seperti *clustering* dan, sampai batas tertentu, klasifikasi. Akibatnya, beberapa dari karya *Lucene committers* yang jatuh lebih ke dalam mesin yang dipecah menjadi subproyek sendiri. Segera setelah itu, *Mahout* menyerap *Taste* proyek penyaringan kolaboratif *open source* (Owen, Anil, Dunning, & Friedman, 2011)

Mahout menjadi proyek *Apache* tingkat atas dengan haknya sendiri, dan mendapat logo pengendara gajah baru untuk *boot*. Sebagian besar karya *Mahout* tidak hanya menerapkan algoritma ini secara konvensional, dengan cara yang efisien dan terukur, tapi juga mengubah beberapa algoritma ini

untuk bekerja pada skala di atas *Hadoop*. Maskot *Hadoop* adalah seekor gajah, *Mahout* menginkubasi sejumlah teknik dan algoritma, banyak yang masih dalam pengembangan atau dalam fase percobaan. Mahout dapat mengimplementasikan tiga tema inti yaitu: *collaborative filtering*, *clustering* dan *classification* (Owen et al., 2011). Di dalam mahout terdapat tahap-tahap yang harus disiapkan yaitu :

1. Mengkategorikan Teks dengan Klasifikasi Naïve Bayes

Algoritma *Naïve Bayes* adalah algoritma klasifikasi probabilistik. Ini membuat keputusan tentang kelas mana yang akan diberikan ke dokumen masukan dengan menggunakan probabilitas yang berasal dari data pelatihan (Ingersol, Grant, S; Morton, Thomas, S; Farris, 2012). Proses pelatihan menganalisis hubungan antara kata-kata dalam dokumen dan kategori pelatihan, dan hubungan antara kategori dan keseluruhan rangkaian pelatihan. Fakta yang ada dikumpulkan dengan menggunakan perhitungan berdasarkan *Teorema Bayes* untuk menghasilkan probabilitas bahwa kumpulan kata-kata (dokumen) termasuk dalam kelas tertentu.

2. Menyiapkan Data Pelatihan

Pengelompokan hanya tampil sebaik masukannya. Jumlah data pelatihan, cara pengorganisasiannya, dan fitur yang dipilih sebagai masukan untuk proses pelatihan semuanya memainkan peran penting dalam kemampuan pengklasifikasi untuk secara akurat mengkategorikan

dokumen baru. Bagian ini menjelaskan bagaimana data pelatihan harus disiapkan untuk digunakan dengan klasifikasi *Mahout Bayes*.

3. Memegang Data Uji

Sekarang Anda perlu mencadangkan beberapa data pelatihan yang telah Anda hasilkan untuk pengujian. Setelah Anda melatih pengklasifikasi, Anda akan menggunakan model ini untuk mengklasifikasikan data uji dan memastikan bahwa kategori yang dihasilkan oleh pengklasifikasi identik dengan dokumen yang sudah dikenal.

4. Latih Pengklasifikasi

Setelah data pelatihan disiapkan dengan *SplitBayesInput*, saatnya untuk melatih data yang pertama dimasukan ke data pelatihan. Jika Anda menjalankan *cluster Hadoop*, salin data pelatihan dan uji coba ke *filesystem Hadoop* dan jalankan perintah berikut untuk membuat model *classifier*.

5. Menguji Pengklasifikasi

Setelah *classifier* dilatih, Anda bisa mengevaluasi kinerjanya dengan menggunakan data uji yang Anda tahan tadi. Perintah berikut akan memuat model yang dihasilkan oleh fase pelatihan ke dalam memori dan mengklasifikasikan masing-masing dokumen dalam rangkaian tes. Saat proses pengujian selesai, Anda akan diberi dua alat bantu evaluasi: persentase akurasi klasifikasi dan *Confusion Matrix*.

2.11 Confusion Matrix

Confusion matrix merupakan sebuah matriks kebingungan menggambarkan keakuratan solusi masalah untuk masalah klasifikasi. Sebuah matriks kebingungan berisi informasi tentang *actual* dan prediksi klasifikasi. Kinerja sistem seperti ini biasanya dievaluasi menggunakan data dalam matriks. Pada Tabel 2.2 menjelaskan mengenai *confusion matrix* dimana terdapat kelas prediksi dan kelas sebenarnya.

Tabel 2. 2 Confusion Matrix

		Kelas Prediksi	
		1	0
Kelas Sebenarnya	0	TP	FN
	1	FP	TN

1. *True Positive* (TP), yaitu jumlah dokumen dari kelas 1 yang benar dan diklasifikasikan sebagai kelas 1.
2. *True Negative* (TN), yaitu jumlah dokumen dari kelas 0 yang benar diklasifikasikan sebagai kelas 0.
3. *False Positive* (FP), yaitu jumlah dokumen dari kelas 0 yang salah diklasifikasikan sebagai kelas 1.
4. *False Negative* (FN) yaitu jumlah dokumen dari kelas 1 yang salah diklasifikasikan sebagai kelas 0.

Di dalam hasil *confusion matrix* terdapat perhitungan akurasi, *Presisi* dan *recall*. *Presisi* adalah fraksi dari contoh yang diambil yang relevan,

sementara *Recall* adalah fraksi dari kasus yang relevan yang diambil baik presisi maupun *recall*. Oleh karena itu, berdasarkan pengertian dan ukuran relevansi *presisi* bisa dilihat sebagai ukuran ketepatan atau kualitas. Sedangkan *recall* adalah ukuran kelengkapan atau kuantitas. Recall bukan apa-apa kecuali kenyataan tingkat positif untuk kelas. (Sumathy, K.L.; Chidambaram, 2013) (Al-taani & Al-gharaibeh, 2010)

2.12 Al-Qur'an

Al-Qur'an adalah kitab suci yang merupakan pedoman hidup dan dasar setiap langkah hidup. Al-Qur'an bukan sekedar mengatur hubungan manusia dengan Rabbnya, tetapi juga mengatur hubungan manusia dengan manusia dan alam sekitarnya. Klasifikasi ayat Al-Qur'an perlu kita ketahui, selain memudahkan kita untuk memahami semua ayat-ayat Al-Qur'an, juga agar kita lebih cepat untuk menemukan arti dari masing-masing ayat Al-Qur'an tersebut. Al-Qur'an dapat diklasifikasikan kedalam berbagai klasifikasi seperti klasifikasi Al-Qur'an berdasarkan susunan surat, berdasarkan turunnya ayat Al-Qur'an atau berdasarkan pembahasannya dan lain-lain (Setiawati et al., 2016). Dalam Al-Qur'an terjemahan versi *The Wisdom* terdapat 6 klasifikasi yaitu akhlak, akidah, ibadah, kisah, ilmu dan muamalah. Namun dalam penjelasan dibawah ini akan di jelaskan 5 klasifikasi saja yaitu akhlak, akidah, ibadah, kisah dan muamalah.

2.12.1 Akhlak

Kata Akhlak berasal dari bahasa Arab *Al-akhlaq* yang merupakan bentuk jamak dari kata *khuluk* yang berarti budi pekerti, perangai, tingkah laku atau tabiat. Sinonim dari kata akhlak ini adalah etika dan moral. Akhlak secara Bahasa adalah keadaan gerak jiwa yang mendorong ke arah melakukan perbuatan dengan tidak menghajatkan pikiran (Marzuki, 2019)

Dari pengertian di atas jelaslah bahwa kajian akhlak adalah tingkah laku manusia, atau tepatnya nilai dari tingkah lakunya, yang bisa bernilai baik (mulia) atau sebaliknya bernilai buruk (tercela). Yang dinilai di sini adalah tingkah laku manusia dalam berhubungan dengan Tuhan, yakni dalam melakukan ibadah, dalam berhubungan dengan sesamanya, yakni dalam bermuamalah atau dalam melakukan hubungan sosial antar manusia, dalam berhubungan dengan makhluk hidup yang lain seperti binatang dan tumbuhan, serta dalam berhubungan dengan lingkungan atau benda-benda mati yang juga merupakan makhluk Tuhan. Secara singkat hubungan akhlak ini terbagi menjadi dua, yaitu akhlak kepada *Khaliq* (Allah Sang Pencipta) dan akhlak kepada *makhluk* (ciptaan-Nya).

Akhlak merupakan konsep kajian terhadap *ihsan*. *Ihsan* merupakan ajaran tentang penghayatan akan hadirnya Tuhan dalam hidup, melalui penghayatan diri yang sedang menghadap dan berada di depan Tuhan ketika beribadah. *Ihsan* juga merupakan suatu pendidikan atau latihan untuk mencapai kesempurnaan Islam dalam arti sepenuhnya (*kaffah*), sehingga *ihsan* merupakan puncak tertinggi dari keislaman seseorang. *Ihsan* ini baru

tercapai kalau sudah dilalui dua tahapan sebelumnya, yaitu *iman* dan *islam*. Orang yang mencapai predikat *ihsan* ini disebut *muhsin*. Dalam kehidupan sehari-hari *ihsan* tercermin dalam bentuk akhlak yang mulia (*al-akhlak al-karimah*). Inilah yang menjadi misi utama diutusny Nabi Saw., ke dunia, seperti yang ditegaskannya dalam sebuah hadisnya: "Sesungguhnya aku diutus hanyalah untuk menyempurnakan akhlak mulia". Indeks akhlak dapat dilihat pada lampiran 2

Tugas yang amat berat dan sangat mulia itu dapat dilaksanakan dengan baik oleh Nabi berkat bimbingan langsung dari Allah Swt. dan juga didukung oleh kepribadian beliau yang sangat agung. Terkait dengan ini

Allah Swt. berfirman (Marzuki, 2009) :



Artinya: "Dan sesungguhnya kamu benar-benar berbudi pekerti yang agung." (QS.Al-Qalam (68):4)

Untuk memudahkan umat Islam dalam bersikap dan berperilaku sehari-hari, di samping memberikan aturan yang jelas dalam Al-Qur'an, Allah juga menunjuk Nabi Muhammad Saw., sebagai teladan baik dalam bersikap, berperilaku, dan bertutur kata. Dengan dua sumber inilah setiap Muslim dapat membangun kepribadiannya. Keteladanan Nabi untuk setiap Muslim ini tegaskan oleh Allah Swt. dalam firman-Nya (Marzuki, 2009) :

لَقَدْ كَانَ لَكُمْ فِي رَسُولِ اللَّهِ أُسْوَةٌ حَسَنَةٌ لِّمَن كَانَ يَرْجُوا اللَّهَ وَالْيَوْمَ
الْآخِرَ وَذَكَرَ اللَّهَ كَثِيرًا ﴿٢١﴾

Artinya: “Sesungguhnya telah ada pada (diri) Rasulullah itu suritелadan yang baik bagimu (yaitu) bagi orang yang mengharap (rahmat) Allah dan (kedatangan) hari kiamat dan dia banyak menyebut Allah.” (QS. Al-Ahzab (21): 33).

2.12.2 Akidah

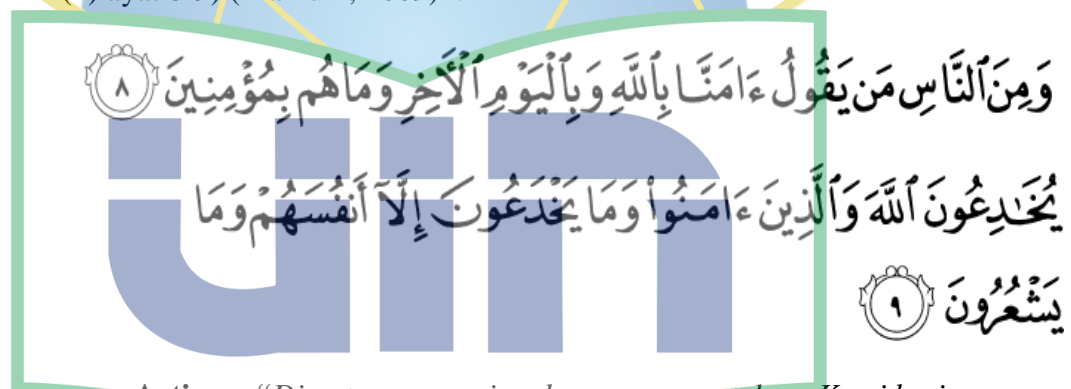
Akidah secara teknis berarti keyakinan atau iman. Dengan demikian, akidah merupakan asas tempat mendirikan seluruh bangunan (ajaran)

Islam menjadi sangkutan semua ajaran dalam Islam. Akidah juga merupakan sistem keyakinan Islam yang mendasari seluruh aktivitas umat Islam dalam kehidupannya. Akidah atau sistem keyakinan Islam dibangun atas dasar enam keyakinan atau yang disebut dengan rukun Iman yang enam (Marzuki, 2009)

Adapun kata Iman, secara etimologis, berarti percaya atau membenarkan dengan hati. Sedang menurut istilah syara', Iman berarti membenarkan dengan hati, mengucapkan dengan lisan, dan melakukan dengan anggota badan. Dengan pengertian ini, berarti iman tidak hanya terkait dengan membenaran dengan hati atau sekedar meyakini adanya Allah Swt. saja, misalnya. Iman kepada Allah berarti meyakini bahwa Allah itu ada, membuktikannya dengan ikrar syahadat atau mengucapkan kalimat-kalimat *dzikir* kepada Allah; dan mengamalkan semua perintah Allah dan

menjauhi semua larangan-Nya. Inilah makna iman yang sebenarnya, sehingga orang yang beriman berarti orang yang hatinya mengakui adanya Allah (*dzikir* hati), lisannya selalu melafalkan kalimat-kalimat Allah (*dzikir* lisan), dan anggota badannya selalu melakukan perintah-perintah Allah dan menjauhi semua larangannya. Indeks akidah dapat dilihat pada lampiran 2

Dari uraian di atas dapat juga dipahami bahwa iman tidak hanya tertumpu pada ucapan lidah semata. Kalau iman hanya didasarkan pada ucapan lidah semata, berarti iman yang setengah-setengah atau imannya orang munafik, seperti yang ditegaskan Al-Qur'an dalam surat al-Baqarah (2) ayat 8-9) (Marzuki, 2009) :



Artinya: *"Di antara manusia ada yang mengatakan: Kami beriman kepada Allah dan hari kemudian, padahal mereka itu sesungguhnya bukan orang-orang yang beriman. Mereka hendak menipu Allah dan orang-orang yang beriman, padahal mereka hanya menipu dirinya sendiri, sedang mereka tidak sadar" (QS. Al-Baqarah (2): 8-9).*

Dan iman juga tidak dapat ditunjukkan dalam bentuk amal (perbuatan) semata. Kalau hal itu saja yang ditonjolkan, maka tidak ubahnya seperti perbuatan orang munafik, sebagaimana yang disebutkan dalam Al-Qur'an surat An-Nisa (4) ayat 142 (Marzuki, 2009):

إِنَّ الْمُنَافِقِينَ يُخَادِعُونَ اللَّهَ وَهُوَ خَادِعُهُمْ وَإِذَا قَامُوا إِلَى الصَّلَاةِ
قَامُوا كُسَالَى يُرَاءُونَ النَّاسَ وَلَا يَذْكُرُونَ اللَّهَ إِلَّا قَلِيلًا ﴿١٤٢﴾

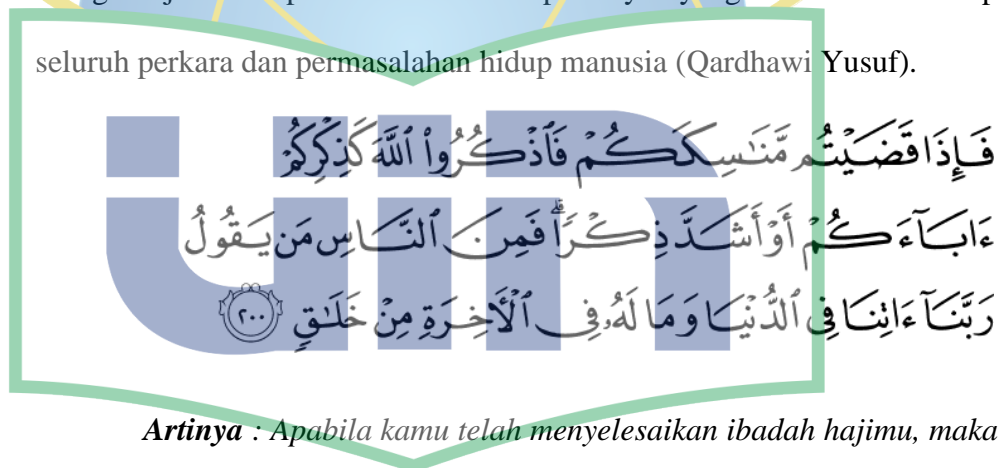
Artinya: “Sesungguhnya orang-orang munafik itu menipu Allah, dan Allah akan membalas tipuan mereka. Dan apabila mereka berdiri untuk shalat mereka berdiri dengan malas. Mereka bermaksud riya’ atau pamer dengan (shalat) di hadapan manusia. Dan tidaklah mereka menyebut Allah kecuali sedikit sekali.” (QS. An-Nisa (4): 142)

2.12.3 Ibadah

Secara bahasa Ibadah artinya: taat. Secara Istilah: menurut Syaikhul Islam Ibnu Taimiyyah Ibadah adalah kata yang mencakup semua yang disukai dan diridhai oleh Allah SWT baik berupa perkataan maupun perbuatan yang tersembunyi dan yang terang-terangan, seperti Shalat, Zakat, Puasa, dan Haji, juga berbicara benar, menunaikan amanah, berbakti kepada kedua orang tua dan menyambung silaturahmi. Juga menepati janji, amar ma'ruf nahi munkar, jihad melawan orang kafir dan munafik, berbuat baik kepada tetangga, anak yatim, orang miskin, *ibnu sabil*, dan budak baik manusia atau binatang, juga berdo'a, *dzikir* dan membaca Al-Qur'an. Itu merupakan contoh Ibadah, selain itu juga mencintai Allah SWT dan Rasul-Nya, takut kepada Allah SWT dan kembali kepada-Nya, memurnikan Ibadah karena-Nya dan sabar akan hukumNya. Begitu juga bersyukur akan nikmat-Nya dan ridha kepada

ketentuan-Nya. Juga berserah diri pada-Nya, mengharap rahmat-Nya, dan takut kepada azab-Nya. Kesemua itu bagian dari Ibadah kepada Allah SWT (Qardhawi Yusuf).

Dengan ini kita mengetahui bahwa Ibadah sangat luas cakupannya. Sebagian manusia tidak memahami makna kata 'Ibadah' kecuali shalat, puasa, sedekah, haji dan umrah, atau doa dan *dzikir*, dan mereka menyangka bahwa Ibadah tidak ada kaitannya dengan akhlak dan adab, apalagi sistem hukum dan undang-undang, atau adat dan kebiasaan. Indeks Ibadah\ dapat dilihat pada lampiran 2. Padahal hakikat ibadah yang Allah SWT jadikan sebagai tujuan hidup manusia mencakup wilayah yang luas dan mencakup seluruh perkara dan permasalahan hidup manusia (Qardhawi Yusuf).



Artinya : Apabila kamu telah menyelesaikan ibadah hajimu, maka berdzikirlah dengan menyebut Allah, sebagaimana kamu menyebut-nyebut (membangga-banggakan) nenek moyangmu, atau (bahkan) berdzikirlah lebih banyak dari itu. Maka di antara manusia ada orang yang bendoa: "Ya Tuhan kami, berilah kami (kebaikan) di dunia", dan tiadalah baginya bahagian (yang menyenangkan) di akhirat. (Al-Baqarah 2:200)

2.12.4 Kisah

Kisah berasal dari kata *al-qashshu* yang berarti mencari atau mengikuti jejak. Dikatakan, “*qashashtu atsarahu*” artinya. “saya mengikuti atau mencari jejaknya.” Kata *al-qashash* adalah bentuk *masdar*. Kalimat *Al-qashshu* atau *Al-qashshatu* searti dengan *tatabbu'ul* *atsar* yaitu pengulangan kembali masa lalu, 3 Kata *Al-qashah* adalah bentuk *masdar*, seperti tersebut dalam Al-Qur'an, QS. Al-Kahfi ayat 64 (Qardhawi Yusuf):

قَالَ ذَلِكَ مَا كُنَّا نَبِغُ فَأَرْتَدَّ إِلَىٰ آثَارِهِمَا قَصَصًا ۖ

Artinya : Musa berkata: "Itulah (tempat) yang kita cari". Lalu keduanya kembali, mengikuti jejak mereka semula. (Al-Kahf 18:64)

Macam-macam Kisah dalam Al-Qur'an Diambil dari sebuah buku yang membahas Ulumul Qur'an, dijelaskan bahwa kisah-kisah dalam Al-Qur'an itu terbagi menjadi tiga bagian, 9 penjelasannya adalah sebagai berikut (Qardhawi Yusuf):

Universitas Islam Negeri SYARIF HIDAYATULLAH JAKARTA

1. Kisah-kisah para nabi dan rasul terdahulu

Tentunya kita semua tahu bahwa tidak semua nabi dan rasul itu disebutkan kisahnya di dalam Al-Qur'an, nabi dan rasul yang disebutkan dalam Al-Qur'an hanyalah 25 orang, dimulai dari Nabi Adam AS sampai dengan Nabi Muhammad SAW. Kemudian dari 25 orang ini, secara garis besar dilihat dari sisi panjang atau singkat kisahnya, dapat dijadikan

kepada tiga kelompok yaitu, Kisah yang disebutkan dengan panjang lebar, kisah yang masuk dalam kategori ini adalah kisah dari Nabi Adam, surat Al-Baqarah : 31-37, Surat Ali 'Imran : 33 dan 59, surat Al-Maidah :27, surat Al-A'raf : 11, 19, 26, 27, 31, 35 dan 172, surat al-Isra' : 61-70, surat al-Kahfi : 50, surat Maryam : 58 dan surat Thaha : 115-121. Kisah Nabi Idris dalam surat Maryam : 56 dan seterusnya.

2. Kisah ummat, tokoh, atau pribadi (bukan nabi), dan peristiwa-peristiwa masa lalu.

Tokoh yang pertama kali kisahnya diceritakan dalam Al-Qur'an adalah dua orang putra Nabi Adam sendiri yaitu Qabil dan Habil, Al-Qur'an menceritakan kisah ketika Qabil membunuh saudaranya sendiri karena akibat dari sifat dengki. Inilah pembunuhan pertama yang terjadi dalam sejarah umat Islam. Dan masih banyak lagi kisah-kisah seorang tokoh yang diceritakan dalam Al-Qur'an, sebagian dari kisah ini antara lain:

1. Kisah Qarun yang hidup pada zaman Nabi Musa As. yang sombong dan kufur setelah kaya raya yang terdapat dalam surat Al-Qashash : 76- 79, surat Al-Ankabut : 39 dan surat Ghafir : 24. Kisah peperangan antara Jalut dan Thalut, Kisah tentang Ashabul Kahfi, Kisah Raja Dzul Qarnain, Kisah kaum Ashabul Ukhdud. Kisah Maryam yang diasuh oleh Nabi Zakaria, ibu Nabi Isa AS yang terdapat dalam surat Ali 'Imran : 36-45, An-Nisa' : 156, 171, Al-

Maidah: 17, 110, Maryam : 16, 27, Al-Mukminun : 50 dan surat At-Tahrim : 12.

3. Kisah-kisah yang terjadi pada zaman Nabi Muhammad SAW

Beberapa kisah yang terjadi pada masa Nabi Muhammad juga disebutkan dalam Al-Qur'an, salah satunya yaitu ketika sebelum Nabi lahir, Tentara Bergajah melakukan penyerbuan ke Makkah yang bertujuan untuk menghancurkan Ka'bah, yang dipimpin oleh Raja Abrahah. Diceritakan pula kisah Nabi Muhammad waktu kecil dengan statusnya sebagai anak yatim yang miskin dan belum mendapat bimbingan wahyu, dengan bahasa yang singkat dan puitis. Dan juga peristiwa setelah beliau diangkat menjadi Rasul, yaitu peristiwa Isra' dan Mi'raj, hijrah, perang Badar, perang Uhud, perang Ahzab atau perang Khandaq, dan perang Hunain, juga kisah-kisah seputar Fathul Makkah dan peristiwa lainnya yang juga tidak bisa disebutkan oleh penulis secara lengkap. Untuk indeks kisah dapat dilihat pada lampiran

2.

Universitas Islam Negeri SYARIF HIDAYATULLAH JAKARTA

2.12.5 Muamalah

Manusia dijadikan Allah SWT sebagai makhluk sosial yang saling membutuhkan antara satu dengan yang lain. Untuk memenuhi kebutuhan hidupnya, manusia harus berusaha mencari karunia Allah yang ada dimuka bumi ini sebagai sumber ekonomi. Allah SWT berfirman :

وَابْتَغِ فِيمَا آتَاكَ اللَّهُ الدَّارَ الْآخِرَةَ وَلَا تَنْسَ نَصِيبَكَ
 مِنَ الدُّنْيَا وَأَحْسِنَ كَمَا أَحْسَنَ اللَّهُ إِلَيْكَ وَلَا تَبْغِ الْفُسَادَ فِي
 الْأَرْضِ إِنَّ اللَّهَ لَا يُحِبُّ الْمُفْسِدِينَ ﴿٧٧﴾

Artinya : “Dan Carilah pada apa yang telah dianugerahkan Allah kepadamu (kebahagiaan) negeri akhirat, dan janganlah kamu melupakan bagianmu dari (kenikmatan) duniawi dan berbuat baiklah (kepada orang lain) sebagai mana Allah telah berbuat baik kepadamu, dan janganlah kamu berbuat kerusakan di (muka) bumi. Sesungguhnya Allah tidak menyukai orang-orang yang berbuat kerusakan.” (QS Al-Qoshosh : 77)

Muamalah dalam ilmu ekonomi Islam memiliki makna hukum yang bertalian dengan harta, hak milik, perjanjian, jual beli, utang piutang, sewa menyewa, pinjam-meminjam dan sebagainya. Juga hukum yang mengatur keuangan serta segala hal yang merupakan hubungan manusia dengan sesamanya, baik secara individu maupun masyarakat. Tujuannya adalah agar tercapai suatu kehidupan yang tentram, damai, bahagia dan sejahtera.

Indeks muamalah dapat dilihat pada lampiran 2. Adapun transaksi-transaksi ekonomi dalam Islam tersebut antara lain :

1. JUAL BELI

Jual beli dalam bahasa arab terdiri dari dua kata yang mengandung makna berlawanan yaitu *Al-bai'* yang artinya jual dan *Asy-syira'a* yang artinya beli. Menurut istilah hukum syara, jual beli ialah menukar suatu barang/uang dengan barang yang lain dengan cara akad (ijab/qobul). Di

zaman yang modern seperti sekarang ini transaksi jual beli dapat dilakukan dengan berbagai cara seperti lewat internet, telpon dan lain sebagainya. Demikian juga sistem pembayarannya bisa lewat cek, surat berharga dan semacamnya. Allah swt berfirman:

يَا أَيُّهَا الَّذِينَ ءَامَنُوا لَا تَأْكُلُوا أَمْوَالَكُمْ بَيْنَكُمْ
بِالْبَاطِلِ إِلَّا أَنْ تَكُونَ تِجَارَةً عَنْ تَرَاضٍ مِّنْكُمْ وَلَا تَقْتُلُوا
أَنْفُسَكُمْ إِنَّ اللَّهَ كَانَ بِكُمْ رَحِيمًا ﴿٢٩﴾

Artinya: “Hai orang-orang yang beriman, janganlah kamu saling memakan harta sesamamu dengan jalan yang batil, kecuali dengan jalan perniagaan yang berlaku dengan suka sama-suka di antara kamu. dan janganlah kamu membunuh dirimu. Sesungguhnya Allah adalah Maha Penyayang kepadamu. (An-Nisa :29)

Rasulullah saw bersabda :

أَفْضَلُ الْكَسْبِ عَمَلُ الرَّجُلِ بِيَدِهِ وَكُلُّ بَيْعٍ مَبْرُورٌ (رواه احمد)

Artinya : ” Perolehan yang paling afdhal adalah hasil karya tangan seseorang dan jual beli yang mabrur”. (HR. Ahmad)

1. Jual Beli Yang Dilarang Agama

- a) Membeli barang dengan harga yang lebih mahal dari harga pasar sedang ia tidak ingin kepada barang itu, tetapi semata-mata supaya orang lain tidak dapat membeli barang tersebut.

- b) Membeli barang untuk di tahan agar dapat dijual dengan harga yang lebih mahal, sedang masyarakat umum sangat membutuhkan barang tersebut.
- c) Menjual suatu barang untuk menjadi alat maksiat.
- d) Jual beli yang dapat menimbulkan kericuhan baik dari pihak pembeli dan penjual-nya. Seperti barang yang jelek ditutupi dengan barang yang baik.
- e) Membeli barang yang sudah dibeli orang lain yang masih dalam keadaan khiyar.

2. Manfaat Jual Beli

- a) Agar manusia saling tolong menolong antara satu dengan lainnya untuk memenuhi kebutuhan hidupnya.
- b) Manusia dituntut untuk selalu berhubungan dengan yang lain karena tak ada seorangpun yang dapat memenuhi kebutuhan hidupnya sendiri.
- c) Untuk memperluas hubungan antar desa, kota bahkan antar negara sehingga dapat diperoleh pemerataan ekonomi.
- d) Untuk menumbuhkan kreatifitas manusia agar dapat menghasilkan dan memproduksi barang-barang yang dapat dipergunakan untuk kemaslahatan manusia.

2. RIBA

Riba menurut bahasa, riba memiliki beberapa pengertian, yaitu:

1. Bertambah, karena salah satu perbuatan riba adalah meminta tambahan dari sesuatu yang dihutangkan.
2. Berkembang, berbunga, karena salah satu perbuatan riba adalah membungakan harta uang atau yang lainnya yang dipinjamkan kepada orang lain.

Sedangkan menurut istilah, yang dimaksud dengan riba menurut Al Mali ialah: “Akad yang terjadi atas penukaran barang tertentu yang tidak diketahui pertimbangannya menurut ukuran *syara*’, ketika berakad atau dengan mengakhirkan tukaran kedua belah pihak salah satu keduanya”.

Menurut Muhammad Abduh, yang dimaksud dengan riba ialah penambahan-penambahan diisyaratkan oleh orang yang memiliki harta kepada orang yang meminjam hartanya (uangnya), karena pengunduran janji pembayaran oleh peminjam dari waktu yang telah ditentukan.

Menurut Abdurrahman Al-Jaziri, yang dimaksud dengan riba ialah akad yang terjadi dengan penukaran tertentu, tidak diketahui sama atau tidak menurut aturan *syara*’ atau terlambat salah satunya.

Sedangkan menurut terminologi *syara*’, riba berarti: “Akad untuk satu ganti khusus tanpa diketahui perbandingannya dalam penilaian syariat ketika berakad atau bersama dengan mengakhirkan kedua ganti atau salah satunya.”

Dengan demikian, riba menurut istilah ahli fiqih adalah penambahan pada salah satu dari dua ganti yang sejenis tanpa ada ganti dari tambahan ini. Tidak semua tambahan dianggap riba, karena tambahan terkadang dihasilkan dalam sebuah perdagangan dan tidak ada riba didalamnya hanya saja tambahan yang diistilahkan dengan nama “riba” dan Al-Qur’an datang menerangkan pengharamannya adalah tambahan tempo.



Universitas Islam Negeri
SYARIF HIDAYATULLAH JAKARTA

2.13 Literature Sejenis

Berikut tabel 2.4 merupakan daftar penelitian sejenis yang telah dilakukan sebelumnya yang digunakan penulis sebagai referensi untuk mengerjakan penelitian ini:

Tabel 2. 3 Studi Literature Sejenis

Peneliti	Lestari, dkk (2013)	Efendi, dkk (2012)	Hilwah, dkk (2017)	Natalius (2011)	Setiawan, dkk (2017)
Judul	<i>Personality Types Classification for Indonesian Text in Partners Searching Website Using Naïve Bayes Methods.</i>	Klasifikasi Dokumen Berbahasa Indonesia Menggunakan <i>Naïve Bayes Classifier</i>	Klasifikasi <i>Text Mining</i> untuk Terjemahan Ayat-Ayat Al-Qur'an menggunakan Metode Klasifikasi <i>Naïve Bayes Text</i>	Metoda <i>Naïve Bayes Classifier</i> dan Penggunaannya pada Klasifikasi Dokumen	Klasifikasi Artikel Berita Menggunakan Metode <i>Text Mining</i> Dan <i>Naïve Bayes Classifier</i>
Model	<i>Naïve Bayes</i>	<i>Naïve Bayes</i>	<i>Naïve Bayes</i>	<i>Naïve Bayes</i>	<i>Naïve Bayes</i>
Tujuan Penelitian	Menggunakan algoritma <i>naïve bayes</i> untuk mengklasifikasikan tipe kepribadian pengguna menggunakan jenisnya.	Untuk mengklasifikasikan dokumen yang tidak terstruktur ke dalam kategori yang menggambarkan isi dokumen.	Mengklasifikasikan terjemahan Al-Qur'an menjadi 15 kategori dengan metode <i>naïve bayes</i>	Menjelaskan mengenai metoda <i>naïve bayes</i> untuk klasifikasi dokumen	Mengklasifikasikan artikel berita berbahasa Indonesia ke dalam 9 kanal yaitu, <i>news, finance, sport, otomotif, entertaint, healty, food, travel</i> dan <i>teknologi</i> .

Peneliti	Lestari, dkk (2013)	Efendi, dkk (2012)	Hilwah, dkk (2017)	Natalius (2011)	Setiawan, dkk (2017)
Hasil Penelitian	Penelitian ini berhasil memperoleh jenis kepribadian dan menemukan pasangan berdasarkan tipe kepribadian dengan menggunakan penambahan teks dengan metode <i>Naïve Bayes</i> untuk klasifikasi kepribadian.	Menghasilkan 6 klasifikasi berita ekonomi, kesehatan, olahraga, teknologi, politik dan pendidikan. Dengan tingkat akurasi 86,67%.	1. Terdapat 198 <i>wordlist</i> dari hasil <i>preprocessing</i> 2. Hasil prediksi dari dokumen baru menggunakan 15 kategori dan terdapat 707 ayat yang masuk ke dalam kategori tersebut.	Metoda <i>Naïve Bayes classifier</i> ini tidaklah 100% sempurna. Namun dalam <i>spam filtering</i> pengklasifikasian spam otomatis ini dapat disesuaikan oleh masing-masing orang sehingga tidak ada aksi salah pengklasifikasian secara personal	Menghasilkan 9 kategori berita yaitu, <i>news, finance, sport, otomotif, entertaint, healty, food, travel</i> dan <i>teknologi</i> .
Kelebihan	Klasifikasi kepribadian proses dilakukan dengan penentuan nilai VMap terbesar dari masing-masing kategori. Untuk hasil beberapa pasangan yang cocok, program menggunakan teori Kepekaan kepribadian, dimana pasangan yang cocok adalah pasangan yang memiliki kepribadian yang berlawanan.	Tingkat akurasi perangkat lunak ini mencapai 86,67% dengan menggunakan 60 dokumen yang terdiri dari 30 dokumen latih dan 30 dokumen uji.	Sudah memiliki tingkat akurasi yang tinggi yaitu pada kategori pertanian dan perdagangan dengan tingkat akurasi 97,7381%.	Penggunaan Algoritma <i>Naïve Bayes classifier</i> dalam klasifikasi dokumen dapat ditinjau dari prosesnya yang mengambil aksi berdasarkan data-data yang telah ada sebelumnya.	Sistem klasifikasi artikel berita yang dilakukan dengan menggunakan algoritma <i>naive bayes classifier</i> menerima data masukkan berupa data <i>text</i> artikel yang diproses dengan <i>text mining</i> yaitu proses <i>casefolding, tokenizing</i> dan <i>filtering</i> .

Peneliti	Lestari, dkk (2013)	Efendi, dkk (2012)	Hilwah, dkk (2017)	Natalius (2011)	Setiawan, dkk (2017)
Kekurangan	Tingkat keberhasilan klasifikasi tergantung pada jumlahnya dokumen pembelajaran yang digunakan. Klasifikasi kepribadian proses dilakukan dengan penentuan yang terbesar.	Dokumen pelatihan yang dibuat sedikit, oleh karena itu kurangnya kata - kata yang penting yang mencirikan suatu dokumen dan juga terdapat kata - kata yang dominan ke kategori lain yang bukan kategorinya sehingga dapat menimbulkan kesalahan dalam pengklasifikasian dokumen.	Dari hasil klasifikasi menggunakan model <i>naïve bayes</i> masih terdapat ayat- ayat yang tidak termasuk kedalam 15 kategori yang sudah ditentukan.	Banyaknya celah untuk mengurangi keefektifan metode ini dan akibatnya meloloskan dokumen ke dalam kelas tertentu padahal jelas-jelas dokumen tersebut tidak layak berada di kelas tersebut. Dalam kasus <i>spam filtering</i> , kelemahan ini banyak digunakan oleh <i>spammers</i> berpengalaman untuk meloloskan <i>spam</i> ke dalam kelas bukan <i>spam</i> (menganggap surat elektronik bukan <i>spam</i> padahal sebenarnya adalah <i>spam</i>).	Dokumen pelatihan yang dibuat harus banyak karena ketepatan kategori dipengaruhi oleh data pelatihan setiap kategori. Data latih ini berisi kata yang sering muncul pada masing-masing kategori atau kata-kata yang dapat mewakili kategori tertentu.

Dari tabel perbandingan diatas penulis membuat sebuah eksperimen mengenai klasifikasi dokumen terjemahan Al-Qur'an berbahasa Indonesia dengan menggunakan algoritma *naïve bayes*. Kelebihan klasifikasi yang dibangun oleh penulis dibandingkan beberapa penelitian sebelumnya, yaitu :

1. Peneliti dapat membuat klasifikasi terjemahan Al-Qur'an dengan 5 klasifikasi yaitu, akhlak, akidah, ibadah, kisah, muamalah.
2. Hasil akurasi keseluruhan dari klasifikasi yang dilakukan peneliti sebesar 82,8996 dengan hanya memasukan 292 dokumen yang menjadi data *training* dari 695 *document input*.
3. Peneliti membuat model algoritma *Naïve Bayes* di dalam *apache mahout*.
4. Sistem klasifikasi yang di bangun melalui *Feature Selection* untuk menghapus kata-kata yang dianggap tidak penting atau tidak menggambarkan isi dokumen sehingga proses pengklasifikasian lebih efektif dan akurat.
5. Peneliti sudah membuat *label indeks* terlebih dahulu di dokumen klasifikasi.

Untuk kekurangannya dari sistem klasifikasi yang dibangun oleh penulis dibandingkan beberapa penelitian sebelumnya antara lain:

1. Hanya berfokus pada algoritma *naïve bayes* tidak membuat klasifikasi dengan algoritma lain sebagai pembanding.
2. Belum diterapkannya sistem real berupa *user interface* secara umum untuk menampilkan hasil klasifikasi terjemahan Al-Qur'an.

BAB 3

METODOLOGI PENELITIAN

Bab ini memaparkan secara metodis proses pelaksanaan penelitian, mencakup penjelasan-penjelasan tentang studi pustaka, material eksperimen, metodologi klasifikasi yang menjelaskan mengenai proses algoritma *Naïve Bayes*. Tujuannya secara tidak langsung memberikan gambaran ruang lingkup dan batasan penelitian kepada para pembaca.

3.1 Metodologi Pengumpulan Data

Metode pengumpulan data dilakukan untuk memperoleh informasi yang dibutuhkan peneliti untuk mencapai tujuan penelitian. Dalam penelitian ini, metode pengumpulan data yang digunakan oleh penulis ada tiga cara, yaitu: Observasi, Studi Pustaka dan Studi Literatur Sejenis.

3.1.1 Observasi

Observasi yang penulis lakukan, dengan cara melakukan pengamatan langsung terhadap objek penelitian, yang dilakukan pada :

Tanggal : 1 Agustus 2017 – 18 Mei 2018

Nama Instansi : Pusat Laboratorium Terpadu UIN Jakarta

Alamat : Jl. Ir. H. Juanda No 95, Ciputat

3.1.2 Studi Pustaka

Kepustakaan dilakukan dengan mempelajari teori-teori terkait dari hasil penelitian sebelumnya yang mendukung pemecahan masalah. Pengumpulan data dengan cara mengambil dari sumber-sumber media cetak maupun elektronik, jurnal, skripsi, *e-book* dan *browsing internet* yang dapat dijadikan acuan pemecahan masalah. Adapun data-data buku dan pencarian melalui media elektronik seperti *internet* yang digunakan dalam penulisan terdapat didalam daftar pustaka.

3.1.3 Studi Literatur Sejenis

Studi literatur sejenis dilakukan dengan mempelajari penelitian sebelumnya yang memiliki judul dan bahasan sama dengan penelitian yang peneliti lakukan. Hasil dari analisis dan perbandingan dengan skripsi yang penulis buat. Berikut ini beberapa literatur yang peneliti gunakan yaitu :

No	Peneliti	Penelitian Sejenis
1	Lestari dkk	Personality Types Classification for Indonesian Text in
2	Setiawati, dkk	Klasifikasi Terjemahan Ayat Al-Qur'an tentang ilmu sains menggunakan decision tree.
3	Effendi, dkk	Klasifikasi dokumen berbahasa indonesia menggunakan <i>naïve bayes classifier</i>
4	Hilwah, dkk	Klasifikasi <i>text mining</i> untuk terjemahan ayat-ayat Al-Qur'an menggunakan metode klasifikasi <i>naïve bayes</i> .

- | | | |
|---|---------------|---|
| 5 | Cui | A Chinese text classification system based on Naïve Bayes Algorithm. |
| 6 | Setiawan, dkk | Klasifikasi Artikel Berita
Menggunakan Metode Text Mining Dan Naive Bayes Classifier |

3.2 Material

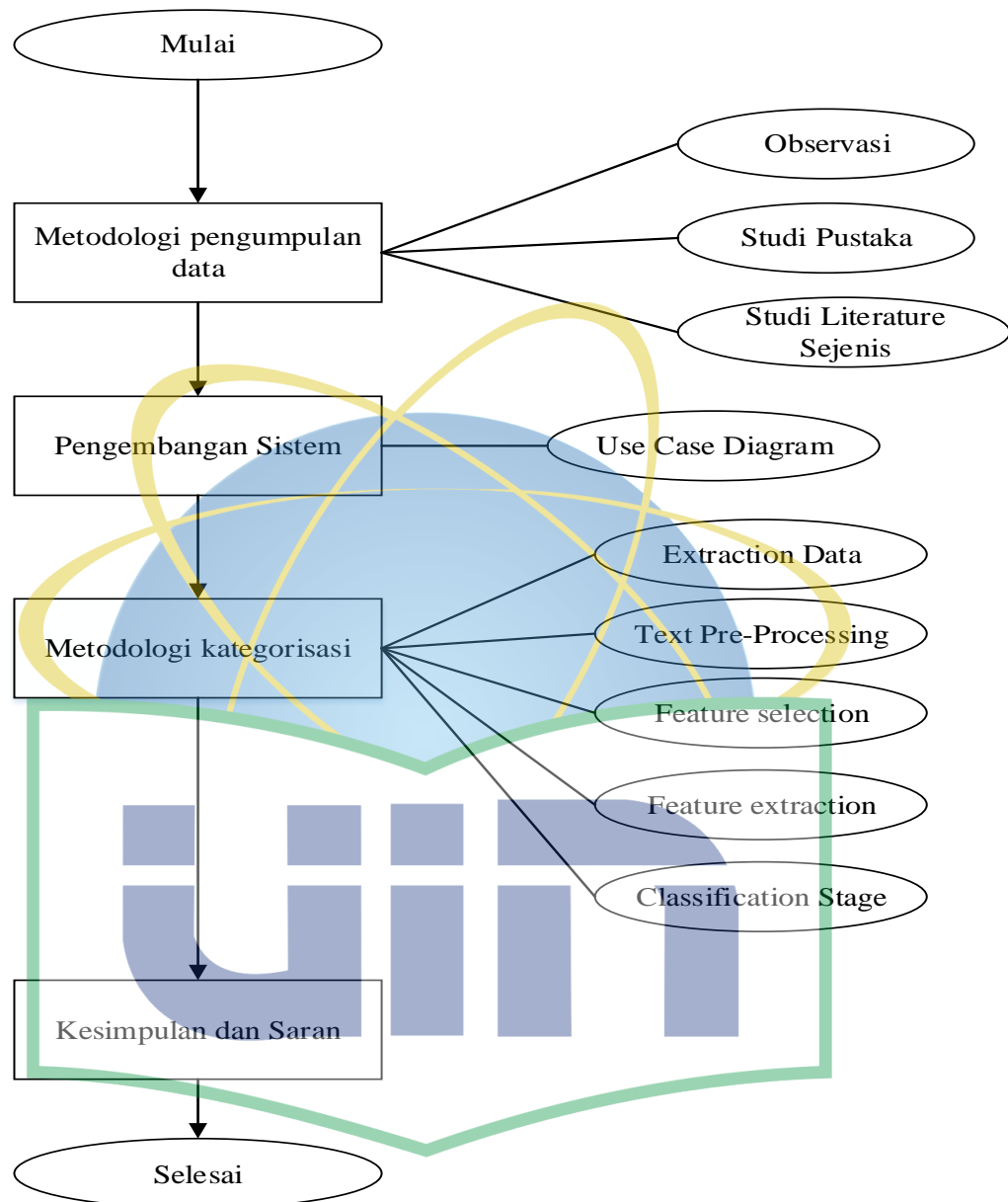
Eksperimen yang dilakukan dalam penelitian ini menggunakan material-material sebagai berikut:

1. Dataset yang digunakan adalah berasal dari situs Al- Qur'an terjemahan yaitu www.tanzil.net juz 1 sampai juz 4 yaitu surat Al-Baqarah, Ali 'Imran dan An-Nisa berupa *file .txt* dapat dilihat pada lampiran 4.
2. Data terjemahan Al- Qur'an adalah kata-kata umum yang sering muncul pada bahasa Indonesia (Tala, 2003).
3. Data *stoplabel* yang digunakan berjumlah 0. Hal ini dimaksudkan untuk mengetahui hasil asli dari label-label indeks klasifikasi yang di hasilkan.

3.3 Kerangka Berfikir

Penelitian ini dilakukan dalam 3 tahap yang dapat di lihat pada Gambar

3.1, antara lain: metodologi pengumpulan data dengan melakukan observasi, studi pustaka dan studi *literature* sejenis. Kemudian lanjut pada pengembangan model yaitu pembuatan *Use case Diagram*. Tahap selanjutnya yaitu metodologi kategorisasi yang menjelaskan mengenai 5 proses dari kategorisasi seperti *extraction data*, *text pre-processing*, *feature selection*, *feature extraction* dan *classification stage*. Tahap terakhir yaitu membuat kesimpulan dan saran setelah melakukan penelitian.



Universitas Islam Negeri
SYARIF HIDAYATULLAH JAKARTA

Gambar 3. 1 Kerangka Berfikir

3.4 Metodologi Kategorisasi

Proses kategorisasi dimulai dengan menerjemahkan, membuat *training* data dan data *test* dan mengekstraksi dokumen terjemahan kemudian *input document* kedalam *localhost Hadoop* dan dikelola oleh *apache Mahout* kemudian hasilnya diproses menggunakan algoritma *Naïve Bayes* sebagai berikut dapat dilihat pada Gambar 3.2 Metode Klasifikasi. Dalam metodologi kategorisasi terdapat 5 proses yaitu, *extraction data*, *text pre-processing*, *feature selection*, *feature extraction* dan *classification stage* dan terdapat 2 folder yaitu *folder training* dan *folder test*. Dalam klasifikasi diperlukan folder tersebut untuk membedakan mana folder yang akan di proses untuk klasifikasi dan mana folder yang hanya berisi data pelatihan saja. Isi dari kedua folder tersebut sama untuk *folder training* berfungsi sebagai data pelatihan yang diklasifikasikan juga dengan 5 proses klasifikasi dan nantinya akan menjadi *database training* kemudian akan digunakan sebagai panduan di dalam *folder test*. *Folder test* ini merupakan folder yang diklasifikasikan dengan proses klasifikasi dan melihat database pelatihan untuk menghasilkan *classification stage* berupa 5 kategori yaitu akhlak, akidah, ibadah, kisah dan muamalah.

3.4.1 Extraction Data

Extraction data yaitu membaca dan menerjemahkan input dataset terjemahan Al-Qur'an agar menjadi objek yang dimengerti sistem. terjemahan Al-Qur'an diekstrak berdasarkan foldernya masing-masing, untuk terjemahan Al-Qur'an *training* diekstrak pada *folder training* dan untuk terjemahan Al-Qur'an *test* diekstrak pada *folder test*

3.4.2 Text Pre-Processing

Fase ini merupakan proses awal untuk mempersiapkan teks menjadi data yang akan diolah lebih lanjut. Sekumpulan karakter yang bersambungan (teks) harus dipecah-pecah menjadi unsur yang lebih berarti. Hal ini dapat dilakukan dalam beberapa tingkatan yang berbeda. Data yang telah diterjemahkan di *tokenisasi* berdasarkan spasi dan huruf menghasilkan *token-token* kemudian hasil *token* tersebut di *filtering/transformation* dan dicatat nilai *tf-idf* (*term frequency-inverse document frequency*) dan selanjutnya *apache mahout* akan melakukan *stemming* yaitu mengubah kata menjadi kata dasar.

3.4.3 Feature Selection

Feature Selection bertujuan untuk mengurangi dimensi dari suatu kumpulan teks, atau dengan kata lain menghapus kata-kata yang dianggap tidak penting atau tidak menggambarkan isi dokumen sehingga proses pengklasifikasian lebih efektif dan akurat (Geng et

al., 2004). Dalam penelitian ini yang termasuk bagian dari *feature selection* adalah membuat *vektor* dari dokumen yang sudah di buat.

3.4.4 Feature Extraction

Fase ini, kata-kata yang sering muncul diekstrak berdasarkan data *tf-idf* (*term frequency-inverse document frequency*) yang telah dibuat pada fase sebelumnya. Kata-kata tersebut diseleksi berdasarkan ketentuan-ketentuan tertentu hingga menghasilkan label-label *index* yang dikandidatkan untuk menjadi label klasifikasi. Setelah itu membuat model untuk mengelola label *index* tersebut. Model yang di gunakan adalah Algoritma *Naïve Bayes* (Geng et al., 2004)

3.4.5 Classification Stage

Classification stage merupakan fase terakhir yaitu menghasilkan kategorisasi berdasarkan data test yang dibuat, hasil yang di dapat terdapat 5 topik kategorisasi yaitu, akhlak, akidah, ibadah, kisah dan muamalah.

3.5 Performance Metrics

Kualitas klasifikasi yang dihasilkan diuji menggunakan metode perhitungan *kappa*, *acuracy*, *reability*, *precision*, *recall* dan *F1*

1. *Kappa*

Ukuran seberapa baik klasifikasi yang tampil dibandingkan dengan seberapa baik hal itu akan dilakukan secara kebetulan. Dengan kata lain, model akan memiliki nilai *Kappa* yang tinggi jika ada perbedaan besar antara keakuratan dan tingkat kesalahan nol.

$$\text{Kappa} = \frac{\text{Total Accuracy} - \text{random Accuracy}}{1 - \text{random Accuracy}}$$

2. Accuracy

Untuk menghitung proporsi jumlah prediksi informasi yang benar yang dihasilkan oleh sistem. Adapun rumus Accuracy adalah sebagai berikut :

$$\text{Accuracy} = \frac{tp + tn}{(tp + tn + fp + fn)}$$

$$\text{Accuracy Random} = \frac{(tn + fp) \times (tn + fn) \times (fn + tp) \times (fp + tp)}{(tp + tn + fp + fn) \times (tp + tn + fp + fn)}$$

Dimana tp adalah *true positive*, tn adalah *true negative* fp adalah *false positive*, dan fn adalah *false negative*.

3. Precision & recall

Untuk menghitung berapa banyak informasi yang benar yang dihasilkan oleh sistem dan untuk mengukur seberapa banyak informasi relevan yang diekstraksi oleh sistem. Adapun rumus *precision* dan *recall* adalah sebagai berikut (Butt, 2013):

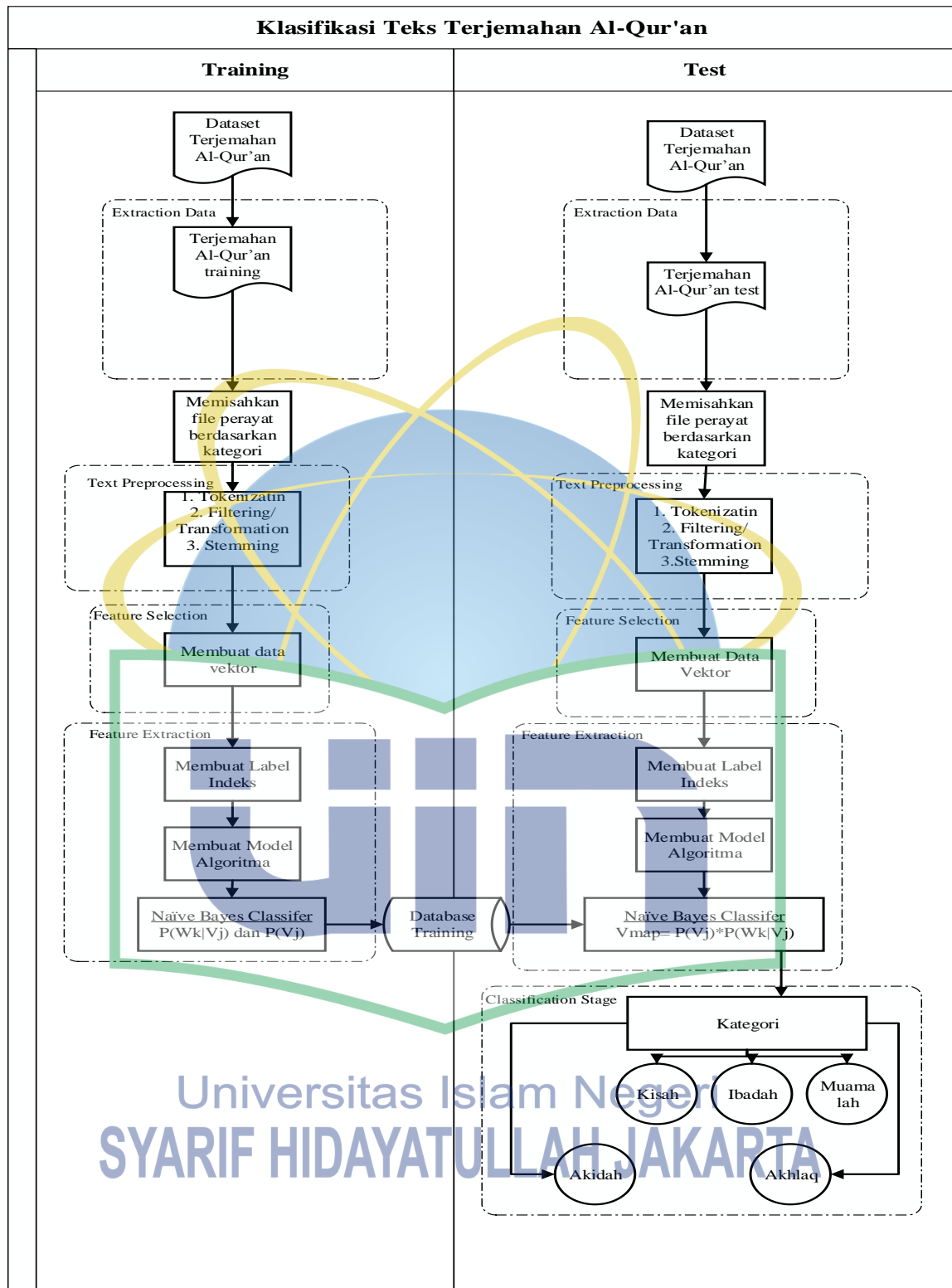
$$\text{Precision} = \frac{tp}{(tp + fp)}$$

$$\text{Recall} = \frac{tp}{(tp + fn)}$$

4. F1

Ini adalah rata-rata tertimbang dari tingkat positif (*recall*) yang benar dan presisi.

$$F1 = \frac{2}{\frac{1}{\text{Recall}} + \frac{1}{\text{Presisi}}}$$



Gambar 3. 2 Metode Klasifikasi

BAB 4

HASIL DAN PEMBAHASAN

Bab ini memaparkan mengenai hasil penelitian, mencakup penjelasan-penjelasan mengenai *use case diagram*, *extracting data*, *text preprocessing*, *feature selection*, *feature extraction*, *classification stage* dan *performance metrics*. Tujuannya secara tidak langsung memberikan penjelasan hasil berdasarkan tahap *classification* kepada para pembaca.

4.1 Hasil Penelitian

Berikut ini merupakan penjelasan mengenai hasil penelitian berdasarkan eksperimen data klasifikasi terjemahan Al-Qur'an.

4.1.1 Use case Diagram

Use case diagram ini menggambarkan interaksi antara *actor* dan sistem.

1. Identifikasi Aktor

Identifikasi aktor merupakan deskripsi dari aktor apa saja yang dapat menggunakan sistem. Berikut ini terdapat satu aktor yaitu admin yang dapat menggunakan sistem secara keseluruhan yang dapat dilihat pada tabel 4.1.

Tabel 4. 1 Identifikasi Aktor

No	Aktor	Deskripsi
1	Admin	Aktor yang menggunakan sistem secara keseluruhan. Dari login hingga melihat hasil klasifikasi.

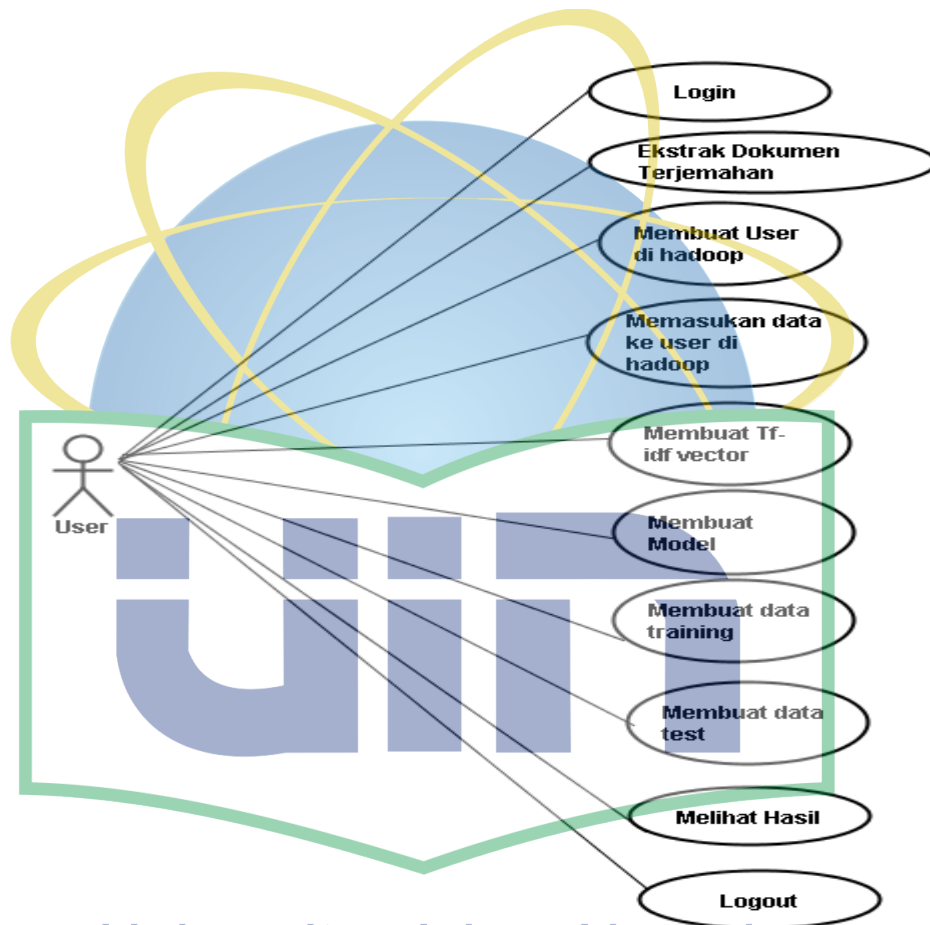
2. Identifikasi *Use case*

Identifikasi *use case* berisi mengenai deskripsi nama *use case* dan actor yang memiliki hak akses terhadap *use case* tersebut. Berikut ini pada tabel 4.2 merupakan nama *use case* beserta deskripsinya.

Tabel 4. 2 Identifikasi Use case

No	Nama Use Case	Deskripsi	Aktor
1	<i>Login</i>	<i>Use case</i> ini menggambarkan kegiatan untuk memasukan password hduser	Admin
2	Ekstrak dokumen terjemahan	<i>Use case</i> ini menjelaskan mengenai mengekstrak dokumen terjemahan ke dalam Hadoop.	Admin
3.	Membuat <i>user</i> di Hadoop	<i>Use case</i> ini menjelaskan mengenai membuat data user sebelum data akan masuk ke dalam <i>user</i> di Hadoop.	Admin
4.	Memasukan data ke <i>user</i> di Hadoop	<i>Use case</i> ini menjelaskan mengenai proses memasukan data ke <i>user</i> Hadoop	Admin
5.	Membuat <i>Tf-idf vector</i>	<i>Use case</i> ini menjelaskan mengenai proses <i>tf-idf</i> untuk memperoleh informasi mengenai teks.	Admin
6	Membuat Model	<i>Use case</i> ini menjelaskan mengenai membuat model <i>naïve bayes</i> dalam mahout.	Admin
7	Membuat data <i>training</i>	<i>Use case</i> ini menjelaskan mengenai training data.	Admin
8	Membuat data <i>test</i>	<i>Use case</i> ini menjelaskan mengenai data test berdasarkan training data yang dibuat.	Admin
9	Melihat Hasil	<i>Use case</i> ini menjelaskan mengenai hasil klasifikasi yang akan tampil.	Admin
10	Logout	<i>Use Case</i> ini menggambarkan kegiatan untuk keluar dari sistem	Admin

Pada Gambar 4.1 merupakan *use case* diagram yang menggambarkan sistem *Apache mahout*. Di dalam *Apache Mahout* memiliki 10 usecase dan 1 aktor yang dapat menggunakan sistem tersebut.

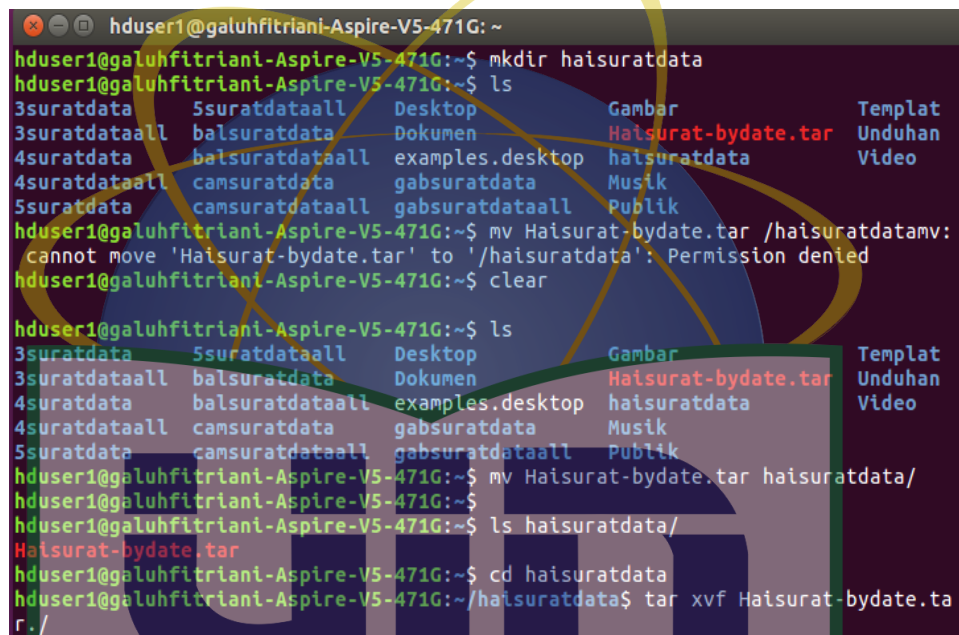


Gambar 4. 1 *Use Case Diagram*

4.1.2 *Extracing Data*

Sumber dokumen yang digunakan merupakan dokumen terjemahan dalam bentuk *.txt*. Dokumen terjemahan tersebut dipisahkan atau diekstrak berdasarkan surat yaitu surat Al- Baqarah, Ali ‘Imran dan An-

nisa, untuk prosesnya dapat dilihat pada Gambar 4.2. Setelah dipisahkan berdasarkan surat kemudian di pisahkan per-ayat yang nantinya akan dibagi ke dua data yaitu data training dan data test. Di dalam data training dan data test, dokumen per-ayat akan di bagi kedalam 5 topik Akhlaq, akidah, iman, ibadah, kisah, muamalah.



```

hduser1@galuhfitriani-Aspire-V5-471G: ~
hduser1@galuhfitriani-Aspire-V5-471G:~$ mkdir haisuratdata
hduser1@galuhfitriani-Aspire-V5-471G:~$ ls
3suratdata      5suratdataall  Desktop        Gambar          Templat
3suratdataall   balsuratdata   Dokumen        Haisurat-bydate.tar  Unduhan
4suratdata      balsuratdataall examples.desktop haisuratdata      Video
4suratdataall   camsuratdata   gabsuratdata   Musik
5suratdata      camsuratdataall gabsuratdataall Publik
hduser1@galuhfitriani-Aspire-V5-471G:~$ mv Haisurat-bydate.tar /haisuratdata/
cannot move 'Haisurat-bydate.tar' to '/haisuratdata': Permission denied
hduser1@galuhfitriani-Aspire-V5-471G:~$ clear

hduser1@galuhfitriani-Aspire-V5-471G:~$ ls
3suratdata      5suratdataall  Desktop        Gambar          Templat
3suratdataall   balsuratdata   Dokumen        Haisurat-bydate.tar  Unduhan
4suratdata      balsuratdataall examples.desktop haisuratdata      Video
4suratdataall   camsuratdata   gabsuratdata   Musik
5suratdata      camsuratdataall gabsuratdataall Publik
hduser1@galuhfitriani-Aspire-V5-471G:~$ mv Haisurat-bydate.tar haisuratdata/
hduser1@galuhfitriani-Aspire-V5-471G:~$
hduser1@galuhfitriani-Aspire-V5-471G:~$ ls haisuratdata/
Haisurat-bydate.tar
hduser1@galuhfitriani-Aspire-V5-471G:~$ cd haisuratdata
hduser1@galuhfitriani-Aspire-V5-471G:~/haisuratdata$ tar xvf Haisurat-bydate.tar
./

```

Gambar 4. 2 Ekstraktion Data

4.1.3 Text Preprocessing

Tokenisasi terhadap *input documents* dilakukan berdasarkan karakter spasi yang ditemukan. Tokenisasi dan *filtering/transformation* dilakukan langsung oleh *apache mahout*. Sebelum melakukan *token* di *mahout* buat data *seqdirectory* dengan menggunakan kode *bin/mahout seqdirectory -i /usr/balsurat -o /usr/balsuratdataseq*

Setelah membuat data *seqdirectory* kemudian hasil dari *seqdirectory* berupa *part-m-0000* dijadikan data yang akan di *tokenisasi* dengan menggunakan kode *bin/mahout seq2sparse -l/usr/balsuratdataseq/part-m-0000 -o /usr/balsuratvector -lnorm -mv -wt tfidf* sebagaimana dapat dilihat pada lampiran pertama.

4.1.4 Feature Selection

Fase ini menyeleksi kalimat dengan menghapus kata-kata yang tidak penting. Di *apache mahout* fase ini yaitu membuat *vector* dari data yang sudah di *tokenisasi* yang nantinya menghasilkan *library* dan langsung masuk ke dalam *training output* dan *test output* di *localhost hadoopnya*. Kode yang digunakan yaitu: *bin/mahout split -i /usr/balsuratvector/tfidf-vectors --trainingOutput /usr/balsuratdatatraining --testOutput /usr/balsuratdatatest --randomSelectionPct 40 --overwrite --sequenceFiles -xm sequential* sebagaimana dapat dilihat pada lampiran pertama.

4.1.5 Feature Extraction

Fase ini, kata-kata yang sering muncul diekstrak berdasarkan data *tf-idf* (*term frequency-inverse document frequency*) yang telah dibuat di fase *feature selection* setelah itu dibuatkan label-label *index* yang akan menjadi label klasifikasi. Dan setelah itu membuat model algoritma *Naïve Bayes* dengan kode *bin/mahout trainnb -i /usr/balsurattrain -o /usr/model -li /usr/labelindex -ow -c*. Dalam kode tersebut langsung menggunakan data *training* karena nantinya hasil dari label klasifikasi

data *training* akan digunakan di data *test* sebagaimana dapat dilihat pada lampiran pertama.

4.1.6 Classification Stage

Di fase ini, dari 635 *input document .txt* terdiri dari surat Al-Baqarah sampai surat An-Nisa, yang digunakan sebagai bahan eksperimen dikategorikan kedalam 5 kelas topik yang telah di buat pada *training data* dan data *test*. Data *test* berjumlah 343 *input document* terjemahan dan data *training* berjumlah 292 *input document* terjemahan.

1. Melakukan training dengan model algoritma Naïve Bayes

Setelah membuat model di *mahout*, kemudian melakukan training

data dengan perintah `bin/mahout trainnb -i /usr/balsuratdatatrain -o /usr/model -li /usr/labelindex -ow -c`

2. Melakukan test dengan model algoritma Naïve Bayes

Setelah melakukan *training* data dengan perintah `bin/mahout trainnb`

`-i /usr/balsuratdatatrain -o /usr/model -li /usr/labelindex -ow -c`

selanjutnya melakukan *test* data dengan model algoritma yang sama

yaitu *Naïve Bayes*, *test* data di lakukan berdasarkan data yang sudah

di *training*. *Test* data ini yang akan menghasilkan *result* dari

klasifikasi. Perintah melakukan test data yaitu `bin/mahout testnb -i`

`/usr/balsuratdatatest -m /usr/model -l /usr/labelindex -o /usr/results`

`-ow`.

```

=====
Summary
-----
Correctly Classified Instances      :      223      82,8996%
Incorrectly Classified Instances    :       46      17,1004%
Total Classified Instances         :      269

=====
Confusion Matrix
-----
a      b      c      d      e      <--Classified as
48      5      4      2      1      |      60      a      = Akhlak
14     122     7      9      0      |     152     b      = Akidah
0       0     10      0      0      |      10      c      = Ibadah
1       0      0     14      0      |      15      d      = Kisah
2       0      0      1     29      |      32      e      = Muamalah

=====
Statistics
-----
Kappa                                0,69
Accuracy                            82,8996%
Reliability                         74,0369%
Reliability (standard deviation)    0,3709
Weighted precision                   0,8702
Weighted recall                     0,829
Weighted F1 score                   0,8388

18/05/04 18:58:25 INFO MahoutDriver: Program took 29723 ms (Minutes: 0.495383333
33333334)
hduser1@galuhfitriani-Aspire-V5-471G: /usr/local/mahout$

```

Gambar 4. 3 Hasil Klasifikasi

Berikut ini merupakan hasil klasifikasi terjemahan Al-Qur'an berbahasa Indonesia berupa *Confusion Matrix*, terdiri dari *Summary* yang berisikan *Correctly Classified Instances* sebanyak 223 ayat dengan persentase 82,8996%, *Incorrectly Classified Instances* sebanyak 46 ayat dengan persentase 17,1004 % sehingga menghasilkan *Total Classified Instances* sebanyak 266 ayat.

Dalam gambar 4.3 terdapat *confusion matrix*, setiap baris matriks mewakili *instance* dalam kelas yang diprediksi sementara masing-masing kolom mewakili *instance* di kelas sebenarnya (atau sebaliknya). Berarti dalam klasifikasi ini berbentuk matriks 5x5 dengan kelas :

a = Akhlak, b = Akidah, c = Ibadah, d = Kisah, e = Muamalah

Jika sistem klasifikasi telah dilatih untuk membedakan ayat mana yang masuk kedalam Akhlak, Akidah, Ibadah, Kisah, dan Muamalah, matriks kebingungan akan merangkum hasil pengujian algoritma untuk pemeriksaan lebih lanjut. Dengan asumsi sampel 264 data – 60 akhlak, 152 akidah, 10 ibadah, 15 kisah, dan 32 muamalah. Berikut ini *confusion matrix* yang di ilustrasikan pada tabel 4.3 :

Tabel 4. 3 Hasil klasifikasi

a(Akhlak)	b(Akidah)	c(Ibadah)	d(Kisah)	e(Muamalah)	
48	5	4	2	1	60
14	122	7	9	0	152
0	0	10	0	0	10
1	0	0	14	0	15
2	0	0	1	29	32

Dalam matriks kebingungan ini, dari

- 1) 60 ayat mengenai akhlak, sistem tersebut memperkirakan bahwa 5 akidah, 4 ibadah, 2 kisah dan 1 muamalah.
- 2) 152 ayat mengenai akidah, sistem tersebut memperkirakan bahwa 14 akhlak, 7 ibadah dan 9 kisah.
- 3) 10 ayat mengenai ibadah
- 4) 15 ayat mengenai kisah, sistem tersebut memperkirakan bahwa 1 akhlak
- 5) 32 ayat mengenai muamalah, sistem tersebut memperkirakan bahwa 2 akhlak dan 1 akidah

Semua prediksi yang benar terletak di diagonal tabel, sehingga mudah untuk memeriksa tabel untuk kesalahan prediksi secara visual, karena akan

ditunjukkan oleh nilai di luar diagonal. Untuk surat dan apasaja yang diklasifikasikan dapat dilihat di lampiran 3.

4.1.7 Performance Metrics

Pengukuran kualitas kategori-kategori yang di hasilkan menunjukan nilai bagus, yakni dengan nilai *kappa* 0,69, *Accuracy* 82,8996%, *Reliability* 74,0369%, *Random Acuracy* 265,1922, *precision* 0,8702 dan *recall* 0,7263

1. Kappa

Nilai *kappa* pada klasifikasi terjemahan ayat Al-Qur'an ini mencapai 0,69 dengan perhitungan sebagai berikut :

$$\text{Kappa} = \frac{\text{Total Acuracy} - \text{random Acuracy}}{1 - \text{random Acuracy}}$$

$$\text{Kappa} = \frac{82,8996 - 265,1922}{1 - 265,1922}$$

$$\text{Kappa} = 0,69$$

2. Acuracy

Acuracy yang dihasilkan sistem untuk klasifikasi terjemahan Al-Qur'an ini mencapai 82,8996 dengan perhitungan sebagai berikut :

$$\text{Acuracy} = \frac{tp+tn}{(tp+tn+fp+fn)} \times 100\%$$

$$\text{Acuracy} = \frac{223}{269} \times 100 \%$$

$$\text{Acuracy} = 82,8996$$

3. Precision & recall

Untuk menghitung berapa banyak informasi yang benar yang dihasilkan oleh sistem dan untuk mengukur seberapa banyak

informasi relevan yang diekstraksi oleh sistem.

Adapun rumus *precision* dan *recall* adalah sebagai berikut (Butt, 2013):

$$Precision = \frac{tp}{(tp + fp)}$$

$$Recall = \frac{tp}{(tp + fn)}$$

$$Recall = \frac{223}{223 + 46}$$

$$Recall = 0,829$$

Tabel 4. 4 Perhitungan precision

Kategori	Tp	Fp	Fn	Precision
Akhlak	48	12	0	$\frac{48}{48 + 12} = 0,8$
Akidah	122	30	0	$\frac{122}{122 + 30} = 0,8026$
Ibadah	10	0	0	$\frac{10}{10} = 1$
Kisah	14	1	0	$\frac{14}{14 + 1} = 0,9333$
Muamalah	29	3	0	$\frac{29}{29 + 3} = 0,9062$
			Precision	$\frac{0,8 + 0,8026 + 1 + 0,9333 + 0,9062}{5} = 0,888$

4. F1

Ini adalah rata-rata tertimbang dari tingkat positif (*recall*) yang benar dan presisi.

$$F1 = \frac{2}{\frac{1}{Recall} + \frac{1}{Presisi}}$$

$$F1 = \frac{2}{\frac{1}{0,829} + \frac{1}{0,8702}}$$

$$F1 = 0,8388$$

Berikut ini pada tabel 4.5 merupakan nilai akurasi klasifikasi setiap kategori.

- 1) Akidah nilai *true* positif (tp) adalah 48, total dari *true* positif (tp) dan *false* positif (fp) adalah 60
- 2) Akhlak nilai *true* positif (tp) adalah 122, total dari *true* positif (tp) dan *false* positif (fp) adalah 152
- 3) Ibadah nilai *true* positif (tp) adalah 10, total dari *true* positif (tp) dan *false* positif (fp) adalah 10
- 4) Muamalah nilai *true* positif (tp) adalah 14, total dari *true* positif (tp) dan *false* positif (fp) adalah 15
- 5) Kisah nilai *true* positif (tp) adalah 29, total dari *true* positif (tp) dan *false* positif (fp) adalah 32

Tabel 4. 5 Perhitungan Nilai akurasi

Kategori	Nilai Akurasi
Akidah	$\frac{48}{60} \times 100\% = 80\%$
Akhlak	$\frac{122}{152} \times 100\% = 80,26\%$
Ibadah	$\frac{10}{10} \times 100\% = 100\%$
Muamalah	$\frac{14}{15} \times 100\% = 93,33\%$
Kisah	$\frac{29}{32} \times 100\% = 90,62\%$

4.2 Pembahasan

Berdasarkan hasil penelitian klasifikasi dengan algoritma naïve bayes ada beberapa point penting yang harus diperhatikan yaitu sebagai berikut:

1. Dalam proses klasifikasi diperlukannya 2 data yaitu data training dan data test. Isi dari data training harus lebih banyak daripada data test dikarenakan data yang di test akan melihat berdasarkan data training yang dibuat. Semakin banyak data training yang dibuat maka semakin baik hasil klasifikasi yang diberikan.
2. Setelah data training dan data test di buat proses klasifikasi selanjutnya adalah melakukan *text preprocessing* dengan tokenisasi dan menghilangkan kata dasar dari sebuah kata di dalam dokumen terjemahan.
3. Hasil akhir dari penelitian menunjukan algoritma naïve bayes bekerja dengan baik di dalam klasifikasi terjemahan Al-Qur'an .
sehingga layak untuk dilanjutkan ke penelitian klasifikasi dokumen lainnya.
4. Mengklasifikasikan dokumen dengan *tools mahout* sangatlah kurang efisien dikarenakan hasil akhir hanya berupa *confusion matriks* yang menggambarkan mengenai pengelompokan klasifikasinya saja tidak dapat menampilkan isi dari hasil pengelompokan klasifikasinya. seperti dalam penelitian ini hanya dapat mengelompokkan 5 klasifikasi yaitu akhlak, akidah, ibadah, kisah dan muamalah dengan jumlah surat yang ditampilkan tanpa memberitahu surat apa saja yang masuk ke dalam kategori akhlak, akidah, ibadah, kisah dan muamalah.

Maka diperlukan kembali tools lain yang lebih efektif dapat menampilkan hasil ayat yang diklasifikasikan.

5. Di dalam Al-Qur'an versi *the wisdom* terdapat 6 klasifikasi yaitu akhlak, akidah, ibadah, kisah, muamalah dan ilmu. Namun dalam penelitian ini klasifikasi mengenai ilmu tidak dicantumkan karena menurut peneliti di dalam terjemahan Al-Qur'an juz 1-4 untuk ayat yang membahas mengenai klasifikasi ilmu hanya beberapa ayat saja.



Universitas Islam Negeri
SYARIF HIDAYATULLAH JAKARTA

BAB 5

KESIMPULAN DAN SARAN

5.1 Kesimpulan

Berdasarkan uraian dan pembahasan pada bab-bab sebelumnya, maka dapat disimpulkan bahwa:

1. Penelitian ini peneliti membuat sebuah kategori terjemahan Al-Qur'an juz 1 sampai 4 dengan menggunakan algoritma *Naïve Bayes*.
2. Dengan total dokumen klasifikasi 269 ayat dengan hasil *correctly classified instances* sebanyak 223 ayat dan *incorrectly classified instances* sebanyak 46 ayat.
3. Memberikan hasil 5 klasifikasi yaitu akhlak, akidah, ibadah, muamalah dan kisah. Dengan jumlah ayat dari masing-masing masing klasifikasi yaitu sebagai berikut akhlak 60 ayat, akidah 152 ayat, ibadah 10 ayat, kisah 15 ayat dan muamalah 32 ayat.
4. Hasil eksperimen memberikan hasil yang memuaskan dengan nilai rata-rata *precision* 0,8702 dan *recall* 0,7263. Hal tersebut menunjukkan bahwa sistem mampu mengategorikan terjemahan Al-Qur'an secara menyeluruh dengan ketepatan yang tinggi.
5. Dengan hasil tersebut algoritma *Naïve Bayes* memiliki ketepatan dengan tingkat akurasi yang besar dalam proses klasifikasi.

5.2 Saran

Berdasarkan kesimpulan dan analisis yang dikemukakan, maka diajukan beberapa saran untuk penelitian berikutnya, yaitu:

1. Menggunakan algoritma lain untuk kategori teks seperti algoritma *Decision Tree* dan *K-Nearest Neighbour*.
2. Melanjutkan juz berikutnya untuk kategori terjemahan Al- Qur'an.
3. Dapat membuat sistem tampilan *user interface* secara umum mengenai klasifikasi terjemahan Al- Qur'an.



Universitas Islam Negeri
SYARIF HIDAYATULLAH JAKARTA

DAFTAR PUSTAKA

- Al-taani, A. T., & Al-gharaibeh, A. M. (2010). Searching Concepts and Keywords in the Holy Quran. *System*, (November), 1–3.
- Arifin, & Purnama, K. E. (2012). Classification of Emotions in Indonesian Texts Using K-NN Method. *International Journal of Information and Electronics Engineering*, 2(6), 7763.
<https://doi.org/10.7763/IJIEE.2012.V2.237>
- Butt, M. (2013). Precision and Recall, (January), 1–23.
https://doi.org/10.1007/978-0-387-30164-8_652
- Efendi, R., & Malik, R. F. (2012). Klasifikasi dokumen berbahasa indonesia menggunakan naive bayes classifier. *Research in Computer Science and Applications*, 1(I), 7–13.
- Geng, W., Cosman, P., Member, S., Berry, C. C., Feng, Z., & Schafer, W. R. (2004). Automatic Tracking , Feature Extraction and Classification of C . elegans Phenotypes. <https://doi.org/10.1109/TBME.2004.831532>
- Hilwah, N., Kudus, A., & Sunendiari, S. (n.d.). Klasifikasi Text Mining untuk Terjemahan Ayat-Ayat Al- Qur'an menggunakan Metode Klasifikasi Naive Bayes, 179–185.
- Ingersoll, Grant, S; Morton, Thomas, S; Farris, A. L. (2012). *Text, Taming*.
- Ingersoll, G. S., Morton, T. S., & Farris, A. L. (2013). *Taming text: how to find, organize, and manipulate it*. Shelter Island: Manning.
- Lestari, N. M. A., Putra, I. K. G. D., & Cahyawan, A. A. K. (2013). Personality Types Classification for Indonesian Text in Partners Searching Website

Using Naïve Bayes Methods. *International Journal of Computer Science (IJCSI)*, 10(1), 1–8.

Liao, S. H., Chu, P. H., & Hsiao, P. Y. (2012). Data mining techniques and applications - A decade review from 2000 to 2011. *Expert Systems with Applications*, 39(12), 11303–11311.

<https://doi.org/10.1016/j.eswa.2012.02.063>

Natalius, S. (2011). Metoda Naïve Bayes Classifier dan Penggunaannya pada Klasifikasi Dokumen. *Jurnal Institut Teknologi Bandung*, (3).

Owen, S., Anil, R., Dunning, T., & Friedman, E. (2011). *Mahout in Action*. Online. <https://doi.org/citeulike-article-id:7544201>

Putra, S. J., Mantoro, T., & Gunawan, M. N. (2017, November). Text mining for Indonesian translation of the Quran: A systematic review. In *Computing, Engineering, and Design (ICCED), 2017 International Conference on* (pp. 1–5). IEEE.

Putra, S. J., Gunawan, M. N., & Suryatno, A. (2018, May). Tokenization and N-Gram for Indexing Indonesian Translation of the Quran. In *2018 6th International Conference on Information and Communication Technology (ICoICT)* (pp. 158–161). IEEE.

Putra, S. J., Gunawan, M. N., Khalil, I., & Mantoro, T. (2017, December). Sentence boundary disambiguation for Indonesian language. In *Proceedings of the 19th International Conference on Information Integration and Web-based Applications & Services* (pp. 587–590). ACM.

Setiawan, I., & Nursantika, D. (2017). Klasifikasi Artikel Berita Menggunakan

Metode Text Mining Dan Naive Bayes Classifier. *Prosiding SENIATI*, 1–6.

Retrieved from <http://ejournal.itn.ac.id/index.php/seniati/article/view/790>

Setiawati, D., Taufik, I., Jumadi, & Budiawan, W. Z. (2016). Klasifikasi Terjemahan Ayat Al-Quran Tentang Ilmu Sains Menggunakan Algoritma Decision Tree Berbasis Mobile. *Jurnal Online Informatika (JOIN)*, 1(1), 24–27.

Sumathy, K.L.; Chidambaram, M. (2013). Text Mining: Concepts, Applications, Tools and Issues - An Overview. *International Journal of Computer Applications*, 80(4), 29–32. Retrieved from <http://www.ijcaonline.org/archives/volume80/number4/13851-1685>

Sumathy, K., & Chidambaram, M. (2013). Text Mining: Concepts, Applications, Tools and Issues—An Overview. *International Journal of Computer Applications*, 80(4), 29–32.

Syam, M. M., Fayed, Z. T., & Habib, M. B. (2006). AN INTELLIGENT SYSTEM FOR ARABIC TEXT CATEGORIZATION, 6(1).

Tala, F. Z. (2003). A Study of Stemming Effects on Information Retrieval in Bahasa Indonesia. *M.Sc. Thesis, Appendix D*, pp, 39–46.

Wei, C. (n.d.). A Chinese text classification algorithm based on granular computing.pdf, 5, 1–5.



Universitas Islam Negeri
SYARIF HIDAYATULLAH JAKARTA

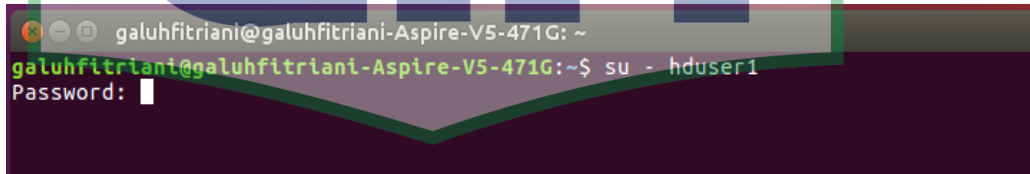
LAMPIRAN 1

INTERFACE MAHOUT

Dalam lampiran ini terdiri dari tampilan interface mahout

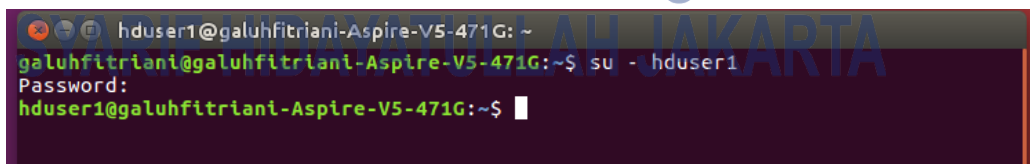
- 1) *Login User*
- 2) *Mahout Home*
- 3) *Localhost Mahout*
- 4) *Extracing Data*
- 5) *Proses Extracting Data*
- 6) *Text Preprocessing*
- 7) *Feature Selection*
- 8) *Feature Extraction*
- 9) *Feature Extraction*
- 10) Melakukan test dengan model algoritma *Naïve Bayes*

1. Login User

A terminal window with a dark background. The prompt is 'galuhfitriani@galuhfitriani-Aspire-V5-471G: ~'. The user enters 'su - hduser1' and is prompted for a password. A large, semi-transparent watermark of a blue dome with yellow orbits is visible in the background.

```
galuhfitriani@galuhfitriani-Aspire-V5-471G: ~  
galuhfitriani@galuhfitriani-Aspire-V5-471G:~$ su - hduser1  
Password: 
```

Universitas Islam Negeri

A terminal window showing the user switch from 'galuhfitriani' to 'hduser1'. The prompt changes to 'hduser1@galuhfitriani-Aspire-V5-471G: ~'. A large, semi-transparent watermark of a blue dome with yellow orbits is visible in the background.

```
hduser1@galuhfitriani-Aspire-V5-471G: ~  
galuhfitriani@galuhfitriani-Aspire-V5-471G:~$ su - hduser1  
Password:  
hduser1@galuhfitriani-Aspire-V5-471G:~$ 
```

LAMPIRAN 1

INTERFACE MAHOUT

2. Mahout Home

```
hduser1@galuhfitriani-Aspire-V5-471G: /usr/local/mahout
hduser1@galuhfitriani-Aspire-V5-471G:~$ cd $MAHOUT_HOME
hduser1@galuhfitriani-Aspire-V5-471G: /usr/local/mahout$
```

3. Mahout Localhost

```
hduser1@galuhfitriani-Aspire-V5-471G: /usr/local/mahout
hduser1@galuhfitriani-Aspire-V5-471G:~$ cd $MAHOUT_HOME
hduser1@galuhfitriani-Aspire-V5-471G: /usr/local/mahout$ start-dfs.sh
Starting namenodes on [localhost]
localhost: starting namenode, logging to /usr/local/hadoop-2.7.4/logs/hadoop-hduser1-namenode-galuhfitriani-Aspire-V5-471G.out
localhost: starting datanode, logging to /usr/local/hadoop-2.7.4/logs/hadoop-hduser1-datanode-galuhfitriani-Aspire-V5-471G.out
Starting secondary namenodes [0.0.0.0]
0.0.0.0: starting secondarynamenode, logging to /usr/local/hadoop-2.7.4/logs/hadoop-hduser1-secondarynamenode-galuhfitriani-Aspire-V5-471G.out
hduser1@galuhfitriani-Aspire-V5-471G: /usr/local/mahout$
```

```
hduser1@galuhfitriani-Aspire-V5-471G: /usr/local/mahout
hduser1@galuhfitriani-Aspire-V5-471G: /usr/local/mahout$ jps
2067 NameNode
3079 Jps
2746 NodeManager
2411 SecondaryNameNode
2620 ResourceManager
2191 DataNode
hduser1@galuhfitriani-Aspire-V5-471G: /usr/local/mahout$
```

Universitas Islam Negeri

```
hduser1@galuhfitriani-Aspire-V5-471G: /usr/local/mahout
hduser1@galuhfitriani-Aspire-V5-471G: /usr/local/mahout$ start-yarn.sh
starting yarn daemons
starting resourcemanager, logging to /usr/local/hadoop-2.7.4/logs/yarn-hduser1-resourcemanager-galuhfitriani-Aspire-V5-471G.out
localhost: starting nodemanager, logging to /usr/local/hadoop-2.7.4/logs/yarn-hduser1-nodemanager-galuhfitriani-Aspire-V5-471G.out
hduser1@galuhfitriani-Aspire-V5-471G: /usr/local/mahout$
```

LAMPIRAN 2

KLASIFIKASI DALAM AL-QUR'AN VERSI THE WISDOM

INDEKS AKHLAK (Lanjutan)

Dengki dan Dendam	Menghormati Orang Tua	Sikap Adil
Desersi	Mengingat Allah	Sumpah
Doa Ibrahim	Mengingkari Nazar	Su'uzann
Dosa	Mengolok	Syukur
Fatalisme Kafir	Mengumbar Janji	Tamak
Harta dan Anak	Menjaga Pandangan	Taubatan Nasuha
Hati yang Keras	Menyalahkan Orang	Tawakal
Ihsan pada Orang Tua	Nafsu Muntainnah	Tazkiyatun Nafs
Ikhlas	Orang Beriman	Teguh Pendirian
Insya Allah	Orang Tua Kufur	Tekad Berjihad
Interaksi Al-Qur'an	Orientasi Kehidupan	Teladan Nabi
Iri	Para Penyair	Teman Munafik
Islah	Patuh Pada Rasul	Terburu-buru
Istighfar	Pencaci Agama	Terpedaya Dunia
Istikamah	Penyelewengan Kitab Suci	Tidak Meminta-minta
Jangan Memaki Agama Lain	Perbuatan Sia-sia	Utamakan Nabi
Judi	Perbuatan yang Dilarang	Zina

LAMPIRAN 2

KLASIFIKASI DALAM AL-QUR'AN VERSI THE WISDOM

INDEKS AKIDAH

Agama Allah	Hawa Nafsu	Nabi Bukan Penyihir
Ahli Surga	Hidayah	Nabi Musa dan Zulkarnain
Akibat Mendustakan Rasul	Hidup Panca Mati	Neraka
Allah Mahakaya	Hikmah yang Sempurna	Nikmat Allah
Allah Penolong Kita	Ilmu Allah	Nuzul Al-Quran
Al-Qur'an dimudahkan	Isi Hati	Orang Beruntung
Apologi Kaum Kafir	Islam dan Agama Lain	Orang Kafir
Argumen Setan	Istighfar bagi Musyrik	Para Rasul Allah
Arsy Allah	Janji Allah	Pelajaran Dari Al- Qur'an
Asmaul Husna	Jihad	Penanggung Azab
Azab Orang Kafir	Jin	Perjanjian Primordial
Balasan Orang Bertaqwa	Kafir dan Azab Allah	Perlindungan Allah
Bantahan Trinitas	Kalimtu Sawa	Rekayasa Allah
Berhala	Karakter Orang Musyrik	Rezeki Semua Makhluk
Bida dari Surga	Kaum Nasrani	Rida Allah
Bukti Kebenaran Allah	Keadilan Allah	Seputar Siksa Allah
Buta Hati	Kebebasan Beragama	Setan Penyusup
Catatan yang Bicara	Keragaman Agama	Sifat Al-Qur'an
Cinta Allah	Kerajaan Allah	Sifat Jin

LAMPIRAN 2

KLASIFIKASI DALAM AL-QUR'AN VERSI THE WISDOM

INDEKS AKIDAH (Lanjutan)

Cinta Rasul	Keselamatan Jiwa	Sifat Kafir
Ciri Orang Beriman	Kesetaraan Spiritual	Siksa Neraka
Derajat Keyakinan	Keterperincian Al-Qur'an	Sukses Duniawi
Doa Malaikat	Ketetapan Umur	Syafaat
Eksistensi Rasul	Keyakinan Orang Kafir	Taklid Buta
Fitrah	Kiamat	Tanggung Jawab Diri
Fungsi Al-Qur'an	Kitab Catatan	Tantangan Al-Qur'an
Gangguan Kafir	Komparasi Kafir dan Mukmin	Tasbih Semesta
Golongan Surga	Laknat Allah	Tirhijab Al-Qur'an
Hakikat Rezeki	Logika Orang Kafir	Tiga Golongan Di Hari Kiamat
Hakikat Wahyu	Makna MasyaAllah	Tipologi Manusia
Hari Kebangkitan	Metafor Surga	Tipu Daya Setan
Hati Orang Kafir	Mispersepsi Kepada Nabi	

LAMPIRAN 2

KLASIFIKASI DALAM AL-QUR'AN VERSI THE WISDOM

INDEKS KISAH

Ancaman Fir'aun	Kurban Ismail	Para Utusan
Ashabul Kafhi	Lukman Al-Hakim	Pengaduan Sodara Yusuf
Azab Kaum Lut	Makar Fir'aun	Pengakuan Zulaikha
Bai'at Al-Ridwan	Malak Al-Karim	Penghiburan Nabi
Banjir Nuh	Maqam Ibrahim	Penyelamatan Nuh
Behala dan Nabi Ibrahim	Maryam	Penyesatan Samiri
Dakwah Syu'aib	Masa Kecil Musa	Perang Badar
Dialog Nabi Ibrahim dan Azar	Melecehkan Perempuan	Perintah Kurban
Fath Makkah	Mukjizat Baju Yusuf	Perjanjian Hudaibiyah
Fir'aun dan Azab Allah	Mukjizat Nabi Musa	Pertemuan Yusuf
Hidangan Langit	Nabi Ayyub	Pertobatan Bani Israil
Hikmah dan Ilmu Yusuf	Nabi dan Wahyu	Qabil dan Habil
Homoseksual	Nabi Dawud	Qarun
Ibrahim dan Ka'bah	Nabi Dawud dan Sulaiman	Rumor Seputar 'Aisyah
Ibunda Nabi Musa	Nabi Hud	Sahabat Nabi
Ismah Yusuf	Nabi Ishaq	Saudara Yusuf
Istri-istri Nabi	Nabi Musa Bertemu Allah	Takwil Mimpi
Jasad Fir'aun	Nabi Musa dan Fir'aun	Tantangan Mukjizat
Kapal Nuh	Nabi Musa dan Harun	Tauhid Ibrahim

LAMPIRAN 2

KLASIFIKASI DALAM AL-QUR'AN VERSI THE WISDOM

INDEKS KISAH (Lanjutan)

Kaum 'Ad	Nabi Musa dan Ilmu Sihar	Taurat
Kaum Hawariyyun	Nabi Musa dan Para Penyihir	Teladan Ibrahim
Kaum Nabi Lut	Nabi Musa Hijrah	Tongkat Nabi Musa
Kaum Saba	Nabi Musa Membunuh	Ujian Hari Sabat
Kemenangan Muslim	Nabi Nuh	Ulul Azmi
Kemenangan Nabi Musa	Nabi Saleh	Umat Nabi Lut
Kemenangan Persia	Nabi Sulaiman dan Bilqis	Unta Nabi Saleh
Kemuliaan Nabi Isa	Nabi Sulaiman dan Hud-hud	Ya'juj-Ma'juj
Kepasrahan Penyihir	Nabi Sulaiman dan Semut	Yusuf Penguasa
Kesombongan Fir'aun	Nabi Syu'aib dan Kaumnya	
Ketergelinciran Adam	Nabi Yunus	
Kisah Penyaliban	Nabi Yusuf dan Benyamin	
Kuda Nabi Sulaiman	Nabi Zakariya	

LAMPIRAN 2

KLASIFIKASI DALAM AL-QUR'AN VERSI THE WISDOM

INDEKS IBADAH

Amalan Umat	Pemakmur Masjid
Aurat	Perintah Bertasbih
Bekal Hari Esok	Puasa
Ciri Orang Iman	Ritual dan Sosial
Fungsi Salat	Salat
Haid	Salat dan Zakat
Haji	Salat Jumat
Hakikat Haji	Salat Tahajud
Hakikat Shalawat	Sifat Mukmin
Hijrah	Syariat
Infak	Waktu Salat
Khamar	Wudu
Mahar	Zakat
Membaca Al-Qur'an	Zikir

Universitas Islam Negeri
SYARIF HIDAYATULLAH JAKARTA

LAMPIRAN 2

KLASIFIKASI DALAM AL-QUR'AN VERSI THE WISDOM

INDEKS MUAMALAH

Alasan Perang	Menyusui
Anak Yatim	Pembiaran Ulama
Ashabiyah	Pembinaan Generasi
Diskriminasi Sosial	Perang
Etika Perang	Pernikahan
Fungsi Masjid	Qisas
Hakikat Keluarga	Retorika Kesesatan
Hukum Potong Tangan	Riba
Kejahatan Pemimpin	Suaka
Kemewahan	Suami Istri
Kepemimpinan	Suap
Keragaman Umat	Tabayyun
Khalifah	Tuduhan Zina
Makanan	Ulil Amri
Makanan dan Kedaruratan	Urgensi Ilmu Agama
Makanan Haram	Waris

LAMPIRAN 2

KLASIFIKASI DALAM AL-QUR'AN VERSI THE WISDOM

Tema Muamalah dalam Al-Qur'an	
1.	Tema muamalah berkaitan dengan relasi/hubungan sosial, misalnya keluarga dan pernikahan, perdagangan, kepemimpinan, sanksi sosial, hukuman atas kejahatan, hubungan sosial, dan lain-lain.
2.	Perintah utama dalam muamalah adalah menegakkan yang ma'ruf (benar, baik) dan mencegah yang mungkar (salah, jelek)—amar ma'ruf nahi mungkar.
3.	Kaidah umum dalam muamalah adalah semuanya diperbolehkan kecuali yang dilarang dengan pertimbangan memperoleh kemaslahatan (yang sebesar-besarnya dan menghindari kemudharatan sekecil-kecilnya).
4.	Penetapan hukum sosial dalam Islam berasaskan 5 tujuan pokok melindungi agama, (hifz al-din), jiwa (hifz al-nafs), akal (hifz al-'aql), kehormatan (hifz al-ardh), dan harta benda (hifz al-mal).
5.	Nilai-nilai utama dalam muamalah adalah keadilan (menolak kezaliman/ketidakadilan), kejujuran (menolak kecurangan), kemaslahatan (menolak kerusakan/kesia-siaan/ kemudharatan), kesetaraan, (menolak ketimpangan/diskriminasi), musyawarah/dialog (menolak pemaksaan dan otoritarianisme), kemusahhatan dan kepedulian (menolak kekikiran dan ketidakpedulian), persaudaraan universal (menolak rasialisme, fanatisme kesukuan/kelompok, dan berbagai interaksi yang saling mengembangkan potensi dan perbaikan kualitas hidup manusia (menolak eksploitasi dan dekadensi).
6.	Formalisasi dari nilai-nilai utama tersebut diwujudkan dalam bentuk penetapan dan penegakan hukum yang adil, serta pembentukan tata masyarakat yang beradab.
7.	Akhlak mulia yang dikembangkan dalam hubungan muamalah antara lain: berlaku adil, memelihara amanat, membina hubungan yang baik, berbuat kebajikan, saling mengingatkan dalam kebenaran dan kesabaran, menepati janji, bermusyawarah, saling menolong dalam kebaikan.
8.	Akhlak tercela yang dihindari dalam hubungan muamalah antara lain: menebar fitnah, sombong, berbantah-bantahan/debat kusir, mencuri, bercerai berai, berlaku curang, diskriminasi, berdusta, kikir, berkhianat, mengolok-olok, berprasangka buruk, melakukan riba, pamer/riya, berbuat kerusakan, berlaku sewenang-wenang kepada anak yatim.

LAMPIRAN 3

HASIL KLASIFIKASI SURAT DAN AYAT

Klasifikasi Akhlak			No	Surat ke-	Ayat ke-
No	Surat ke-	Ayat ke-			
1		9	28		82
2		10	29		94
3		11	30		98
4		13	31		14
5		14	32		16
6		15	33		18
7		16	34		28
8		17	35		32
9		18	36		36
10		19	37		50
11		24	38	4	54
12	2	28	39		58
13		39	40		62
14		42	41		64
15		57	42		118
16		59	43		120
17		63	44		127
18		83	45		128
19		84	46		129
20		156	47		142
21		186	48		172
22		194			
23		16			
24		36			
25	3	38			
26		76			
27		78			

Universitas Islam Negeri
SYARIF HIDAYATULLAH JAKARTA

Klasifikasi Kisah		
No	Surah ke-	Ayat ke-
1	3	52
2		58
3		60
4		68
5		140
6		142
7		146
8		166
9		168
10	4	154
11		158
12		160
13		164
14		166

Klasifikasi Ibadah		
No	Surah ke-	Ayat ke-
1	2	128
2		150
3		158
4		184
5		196
6		198
7		200
8		202
9		238
10	3	96

Klasifikasi Muamalah		
No	Surah ke-	Ayat ke-
1	2	180
2		182
3		188
4		226
5		228
6		230
7		232
8		236
9		240
10		254
11		262
12		264
13		266
14		268
15		270
16		272
17		274
18		276
19		278
20		280

No	Surat ke-	Ayat ke-
21	3	130
22		180
23		186
24	4	02
25		06
26		08
27		10
28		12
29		20
30		24
31		130
32		176