

**ANALISA TOPIK PENDIDIKAN DALAM AL-QUR'AN DENGAN
PENDEKATAN TEXT MINING**

SKRIPSI

Diajukan Oleh :

MUHAMMAD ZAEKI

NIM. 160212047

Mahasiswa Fakultas Tarbiyah dan Keguruan (FTK)
Prodi Pendidikan Teknologi Informasi



**FAKULTAS TARBIYAH DAN KEGURUAN
UNIVERSITAS ISLAM NEGERI AR – RANIRY BANDA ACEH
PROGRAM STUDI PENDIDIKAN TEKNOLOGI INFORMASI
DARUSSALAM 2020 M/1441 H**

LEMBAR PERNYATAAN KEASLIAN ILMIAH

Yang bertanda tangan dibawah ini

Nama : Muhammad Zaeki
NIM : 160212047
Program Studi : Pendidikan Teknologi Informasi
Fakultas : Tarbiyah dan Keguruan

Dengan ini menyatakan bahwa dalam penulisan SKRIPSI ini, saya:

- 1. Tidak menggunakan ide orang lain tanpa mampu mengembangkan dan mempertanggungjawabkan.*
- 2. Tidak melakukan plagiasi terhadap naskah karya orang lain.*
- 3. Tidak menggunakan karya orang lain tanpa menyebutkan sumber asli atau tanpa pemilik karya.*
- 4. Tidak melakukan pemanipulasian dan pemalsuan data.*
- 5. Mengerjakan sendiri karya ini dan mampu bertanggung jawab atas karya ini.*

Bila di kemudian hari ada tuntutan dari pihak lain atas karya saya, dan telah melalui pembuktian yang dapat dipertanggungjawabkan dan ternyata memang ditemukan bukti bahwa saya telah melanggar pernyataan ini, maka saya siap untuk dicabut gelar akademik saya atau diberikan sanksi lain berdasarkan aturan yang berlaku di Fakultas Tarbiyah dan Keguruan UIN Ar-Raniry Banda Aceh.

Demikian pernyataan ini saya buat dengan sesungguhnya.

Banda Aceh, 1 September 2020

Yang Menyatakan,



Muhammad Zaeki

**ANALISA TOPIK PENDIDIKAN DALAM AL-QURAN DENGAN
PENDEKATAN TEXT MINING**

SKRIPSI

Diajukan Kepada Fakultas Tarbiyah dan Keguruan (FTK) Universitas Islam Negeri
Ar-Raniry Darussalam Banda Aceh Sebagai Beban Studi Untuk Memperoleh Gelar
Sarjana dalam Ilmu Pendidikan Islam



Pembimbing I,

Bustami, M.Sc

NIP. 198604082014031001

NIDN. 2008048601

Pembimbing II,

Hendri Ahmadian, S. Si., M. I. M

NIP. 198301042014031002

NIDN. 2004018303

ANALISA TOPIK PENDIDIKAN DALAM AL-QUR'AN DENGAN
PENDEKATAN TEXT MINING

Skripsi .

Telah Diuji Oleh Panitia Sidang Munaqasyah Skripsi Fakultas Tarbiyah dan
Keguruan Universitas Islam Negeri Ar-Raniry Banda Aceh dan
Dinyatakan Lulus Serta Diterima Sebagai Salah Satu Beban
Studi Program Sarjana (S-1) dalam Ilmu Pendidikan Islam

Pada Hari/Tanggal

14 Juli 2020

23 Dzulqaidah 1441 H

Panitia Sidang Munaqasyah Skripsi

Ketua,

Bustamhi, M. Sc
NIP. 198604082014031001

Sekretaris,

Izzah Al-Fikry, M. Pd
NIP.

Penguji I,

Hendri Ahmadian, S. Si., M. I. M
NIP. 1983010420140310

Penguji II,

Basrul, M. S
NIDN. 2027038701

Mengetahui

Dekan Fakultas Tarbiyah dan Keguruan UIN Ar-Raniry
Darussalam Banda Aceh



Dr. Muslim Razali, Sh., M. Ag
NIP. 195903091989031001

ABSTRAK

Penelitian ini bertujuan untuk meneliti Topik Pendidikan yang terkandung dalam *Al-Qur'an*. Langkah awal yang peneliti lakukan adalah dengan menjadikan beberapa kata dasar yang dianggap memiliki hubungan dengan topik pendidikan sebagai kata kuncinya. Kata kunci tersebut ada enam, yaitu “Ajar”, “Bicara”, “Cipta”, “Dengar”, “Ingat” dan “Lihat”. Kemudian peneliti mengimplementasikan sebuah algoritma dalam proses penelitian ini, algoritma yang dimaksud adalah *Naïve Bayes Classifier*. Pada tahap pengujian dan evaluasi peneliti melanjutkan dengan menggunakan metode *Recall* dan *Precision*. Hasil yang diperoleh adalah terdapat total bobot kata yang berafiliasi dengan pendidikan sebanyak 3.516 kata yang telah di *filter* dengan persentase keseluruhan sebesar 9.04% dari total 38.761 bobot kata bersih (di *filter*) dalam *al-qur'an*. Adapun rincian keenam kata kunci tersebut adalah “ciptat” 3,05 %, “Ingat” 2,25 % , “Ajar” 1,96 %, “Lihat” 0,82 %, “Dengar” 0,62 % dan “Bicara” 0,34%. Sedangkan untuk metode evaluasi pada algoritma *Naïve Bayes*, *Recall* mendapat *score* sebanyak 0.605 dan *Precision* sebanyak 0.366.

Kata Kunci: *Text Mining*, Pendidikan, *Al-Qur'an*, Algoritma, *Naïve Bayes*

KATA PENGANTAR

Segala puji dan syukur hanya bagi Allah, k karena atas kehendaknyalah segala sesuatu dapat terjadi dimuka bumi ini. Begitu juga halnya terhadap diri peneliti, dengan kehendaknyalah skripsi yang berjudul **ANALISA TOPIK PENDIDIKAN DALAM AL-QUR'AN DENGAN PENDEKATAN TEXT MINING** ini dapat peneliti selesaikan. Shalawat dan Salam kepada junjungan kita Nabi besar Muhammad SAW, penghulu segala Nabi dan Rasul.

Menyadari bahwa penelitian ini merupakan karya ilmiah yang pertama kalinya dari peneliti, tentu masih banyak kejanggalan atau kekurangannya, baik dalam segi teknis penulisan maupun dari hasil penelitian. Oleh karena itu segala kritikan dan bimbingan yang bersifat membangun dengan senang hati akan peneliti terima demi kesempurnaan penelitian dan peneliti sendiri.

Begitu juga peneliti menyadari skripsi ini dapat terselesaikan berkat bimbingan dan pengarahan serta bantuan dari berbagai pihak, disamping pengetahuan peneliti yang pernah peneliti peroleh selama belajar pada Progtam Studi Pendidikan Teknologi Informasi Fakultas Tarbiyah dan Keguruan Universitas Islam Negeri Ar-Raniry.

Maka pada kesempatan ini, izinkanlah peneliti menyampaikan ucapan terimakasih yang tidak terhingga kepada:

1. Ayahanda dan Ibunda dari Peneliti (Muhammad Zaeki) yang telah bersusah payah mendidik dan membesarkan peneliti dengan penuh kasih sayang dan kesabaran.
2. Pak Dr. Muslim Razali, Sh., M.Ag selaku Dekan Fakultas Tarbiyah dan Keguruan Universitas Islam Negeri Ar-Raniry Banda Aceh.
3. Pak Yusran, M. Pd selaku ketua Program Studi Pendidikan Teknologi Informasi Fakultas Tarbiyah dan Keguruan Universitas Islam Negeri Ar-Raniry Banda Aceh.
4. Bapak Bustami, M. Sc, selaku pembimbing I yang telah membimbing dari awal hingga akhir dan Bapak Hendri Ahmadian, M. I. M selaku pembimbing II yang telah membimbing peneliti dalam penyelesaian skripsi ini.

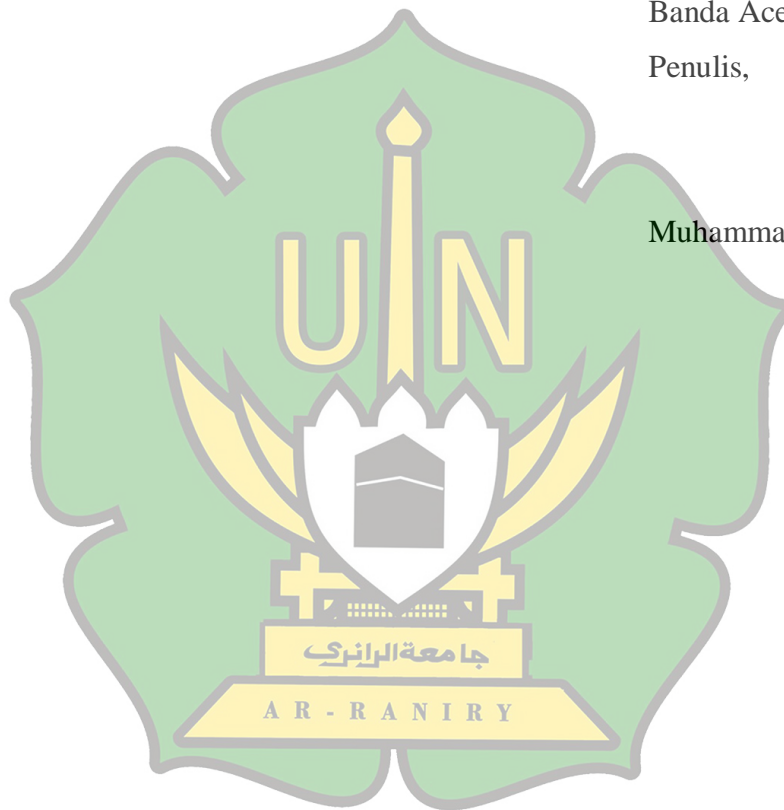
5. Semua kawan-kawan seperjuangan yang telah memberikan motivasi dan bantuan dalam menyelesaikan skripsi ini.

Sebagai penutup kata peneliti mengucapkan terimakasih yang tak terhingga kepada semua pihak yang telah membantu, semoga Allah SWT, memberikan Rahmat, taufiq dan hidayahnya kepada kita semua dan hanya kepadanya kita berserah diri.

Banda Aceh, 14 Juli 2020

Penulis,

Muhammad Zaeki



DAFTAR ISI

LEMBARAN JUDUL	i
LEMBAR PERNYATAAN	ii
LEMBAR PENGESAHAN PEMBIMBING	iii
LEMBAR PENGESAHAN SIDANG.....	ii
ABSTRAK.....	v
KATA PENGANTAR.....	vi
DAFTAR ISI.....	viii
DAFTAR GAMBAR.....	xi
DAFTAR TABEL	xii
DATAR RUMUS	xiii
DAFTAR LAMPIRAN.....	xiv
BAB I PENDAHULUAN.....	1
1.1 Latar Belakang.....	1
1.2 Rumusan Masalah.....	2
1.3 Tujuan	3
1.4 Manfaat	3
1.4.1 Manfaat Bagi Peneliti.....	3
1.4.2 Manfaat Bagi Pembaca	3
1.4.3 Manfaat Bagi Pendidikan.....	4
1.5 Batasan Masalah	4
BAB II TINJAUAN PUSTAKA	5
2.1 Data Mining.....	5
2.1.1 Text Mining	5
2.1.2 Perbedaan Data Mining dengan Text Mining.....	5
2.1.3 Metode Dalam Data Mining	6
2.1.4 Metode Klasifikasi	6
2.1.5 Algoritma <i>Naive Bayes</i>	7

2.2	<i>Al-Qur'an</i>	8
2.2.1	<i>Qur'an Mining</i>	8
2.3	Pendidikan.....	9
2.3.1	Pendidikan Dalam Pandangan Al-Qur'an.....	9
2.4	Metode Evaluasi Penelitian	10
2.5	Penelitian Terdahulu	11
2.5.1	Muhammad Alhawarat, 2015	11
2.5.2	M. Didik R. Wahyudi, 2019	12
2.5.3	Endang Supriyanti, 2018.....	13
BAB III METODELOGI PENELITIAN.....		14
3.1	Metodelogi Penelitian	14
3.2	Waktu dan Tempat Penelitian	16
3.3	Proses Pengumpulan Data.....	16
3.4	Pemilihan <i>Tools</i>	16
3.5	Penginputan Data.....	16
3.6	Persiapan data.....	17
3.6.1	<i>Preprocessing</i>	17
3.6.2	<i>Word Parsing dan Tokenization</i>	17
3.6.3	<i>Filtering</i>	17
3.7	Pengujian.....	25
3.8	Hasil Awal.....	26
3.9	Analisa dan Evaluasi dengan Algoritma <i>Naïve Bayes</i>	26
3.10	Hasil Akhir	26
BAB IV HASIL DAN PEMBAHASAN.....		28
4.1	Proses Input Data.....	28
4.2	Proses <i>Filtering</i>	29
4.3	Evaluasi dengan Algoritma <i>Naïve Bayes</i>	38
4.4	Hasil <i>Data Cloud</i>	39
4.5	Hasil Persentase	44

BAB V PENUTUP	48
5.1 Kesimpulan.....	48
5.2 Saran	49
DAFTAR PUSTAKA	50



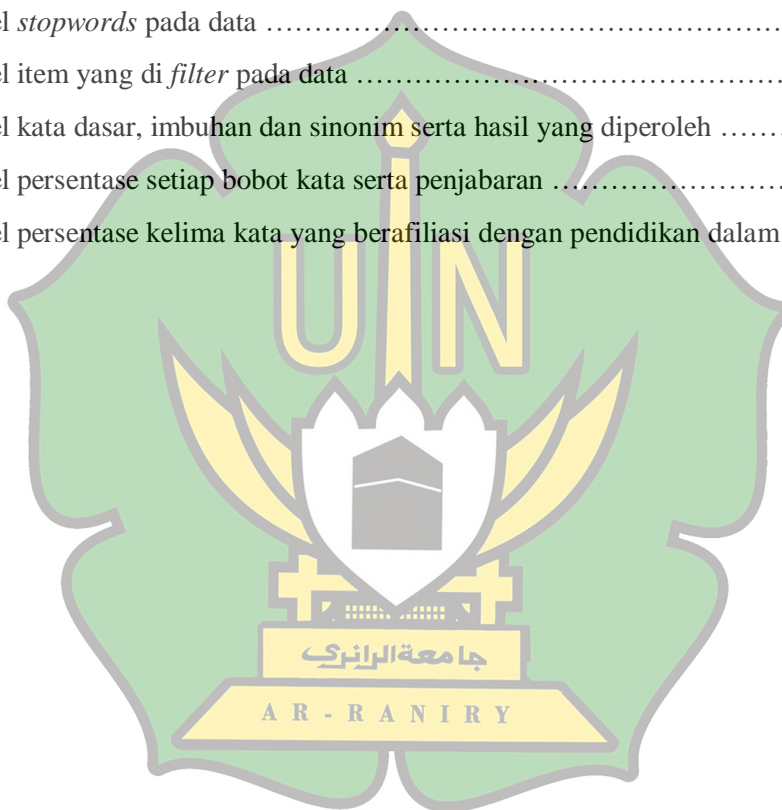
DAFTAR GAMBAR

Gambar	Halaman
3.1 : <i>Flowchart</i> Penelitian	15
4.1 : <i>Import</i> data al-quran ke aplikasi <i>orange</i>	28
4.2 : Proses <i>filtering</i> pada aplikasi <i>orange</i>	31
4.3 : Bentuk rangkaian proses pada aplikasi <i>orange</i>	33
4.4 : Bentuk <i>Viewer Corpus</i> pada aplikasi <i>orange</i>	34
4.5 : Bentuk gambaran <i>data table</i>	35
4.6 : Bentuk <i>Flowchart Naïve Bayes</i>	36
4.7 : Bentuk hasil dari Algoritma <i>Naïve Bayes</i>	37
4.8 : Bentuk visualisasi hasil <i>data cloud al-quran</i> yang telah di <i>filter</i>	39
4.9 : Bentuk Grafik <i>Chart</i> jumlah Bobot Kata Pada Setiap Item Kata	43



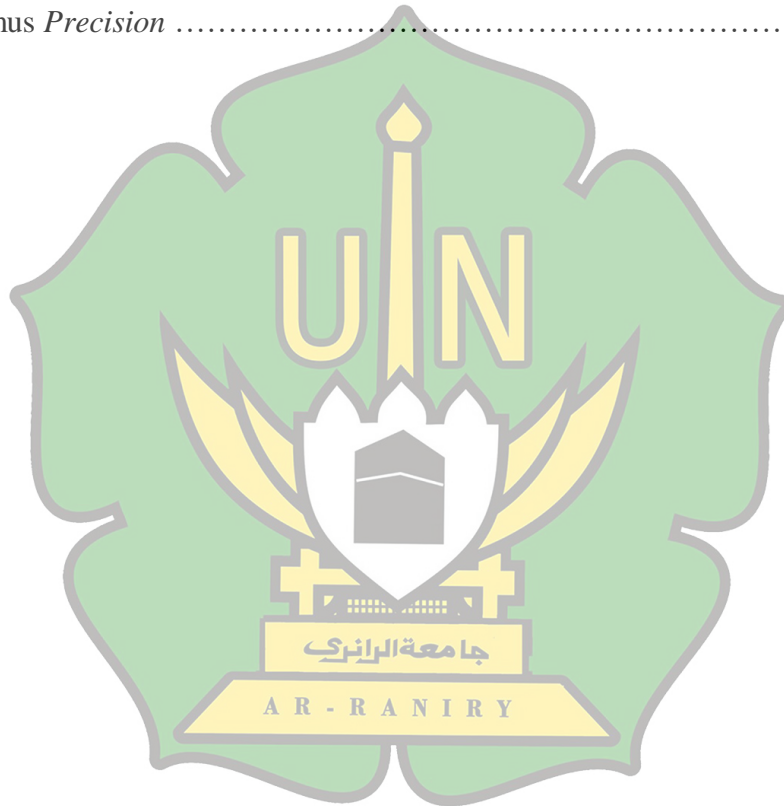
DAFTAR TABEL

Tabel	Halaman
3.1 : Tabel kata dasar dan kata dasar yang berimbuhan	19
3.2 : Tabel kata kunci beserta sinonim	21
3.3 : Tabel kata kunci beserta ayat yang mengandung setiap item	24
4.1 : Tabel <i>stopwords</i> pada data	29
4.2 : Tabel item yang di <i>filter</i> pada data	31
4.3 : Tabel kata dasar, imbuhan dan sinonim serta hasil yang diperoleh	40
4.4 : Tabel persentase setiap bobot kata serta penjabaran	44
4.5 : Tabel persentase kelima kata yang berafiliasi dengan pendidikan dalam <i>al-quran</i> ..	46



DATAR RUMUS

Rumus	Halaman
2.1 : Rumus <i>Prior</i>	7
2.2 : Rumus <i>Posterior</i>	7
2.3 : Rumus <i>Recall</i>	9
2.4 : Rumus <i>Precision</i>	10



DAFTAR LAMPIRAN

Lampiran	Halaman
1. Halaman <i>Link Stopwords</i> pada akun <i>Google Drive</i>	53



BAB I

PENDAHULUAN

1.1. Latar Belakang

Seiring dengan perkembangan zaman dan teknologi informasi kehidupan manusia dipermudah sedemikian rupa dengan bantuan teknologi informasi. Teknologi informasi merangka pada berbagai bidang dan hampir semua bidang telah dimasukkan bantuan dan olahan teknologi. Salah satu bidangnya adalah penyampaian informasi yang akurat, cepat dan sistematis. Demikian juga pada bidang bisnis pelayanan publik, keamanan data, perbankan, pendidikan dan lain sebagainya. Salah satu kegunaan teknologi pada setiap bidangnya ialah pada proses analisis data dengan menggunakan *data mining*. *Data Mining* adalah salah satu dari bagian teknologi yang mana *Data Mining* sendiri merupakan suatu langkah dalam *Knowledge Discovery in Database* (KDD). *Data mining* mengacu pada proses untuk menambang (*mining*) pengetahuan dari sekumpulan data yang sangat besar dan dapat diterapkan pada setiap bidang [1].

Penerapan teknologi *data mining* memiliki sangat banyak kegunaan pada setiap bidang dikarenakan *data mining* memberikan pemaparan hasil yang optimal dan dapat dipertanggung jawabkan. *Data mining* mengimplementasikan *sains* untuk setiap prosesnya pada objek penelitian. Salah satu penerapan *data mining* dapat dilakukan pada *Al-Qur'an*. Contoh implementasi *data mining* pada *al-qur'an* adalah untuk perulangan kata dan mengkalkulasikan kata dalam *al-qur'an* [2], Klasifikasi kata [3], Proses teks dan mengukur frekuensi kata pada *al-qur'an* [4] dan menganalisa topik tertentu termasuk topik pendidikan dalam *al-qur'an* secara ilmiah, logis dan akurat [5].

Mempelajari topik pendidikan dalam *al-qur'an* dapat dilakukan dengan pendekatan *text mining*. *Text Mining* merupakan bagian dari *data mining* yang mana proses kerjanya adalah fokus pada dokumen berbentuk teks. Tujuan utamanya adalah

untuk menghasilkan kata kunci dan keterkaitan antar kata dalam dokumen. *Text Mining* yang diimplementasikan pada *al-qur'an* lebih dikenal dengan istilah *qur'an mining* [6].

Analisa kata-kata yang berafiliasi dengan topik pendidikan dalam *al-qur'an* dengan pendekatan *text mining* didasarkan pada beberapa kata kunci yang memiliki hubungan dengan pendidikan. Berdasarkan definisi umum, pendidikan adalah usaha sadar manusia yang terencana untuk mengembangkan potensi dalam dirinya untuk memiliki kemampuan spiritual, karakter, kecerdasan serta keterampilan [7]. Beberapa kata kunci yang memiliki keterkaitan dan rujukan mengenai konsep pendidikan dalam *al-qur'an* diantaranya *At-Tarbiyyah* (membina), *Rabbani* (sempurna ilmu) [8] dan *Ta'lim* (Mengajar) [9]. Namun, ketiga kata tersebut tidak cukup kuat jika dijadikan rujukan yang mewakili konsep pendidikan dalam *al-qur'an*. Hal ini disebabkan oleh banyaknya kata-kata lain dalam *al-qur'an* yang berafiliasi dengan topik pendidikan dalam *al-qur'an*, misal kata “bacalah” Q.S. Al-‘Alaq (95:1) [9], “mengajar” Q.S. Ar-Rahman (55:2) [10], “menciptakan” Ar-Rahman (55:3) [10] dan lain sebagainya. Oleh karena *al-qur'an* mengandung sangat banyak kata-kata dengan makna yang berbeda-beda, maka masih terbuka peluang untuk mendalaminya. Salah satu caranya adalah pemilihan beberapa kata dalam *al-qur'an* untuk dijadikan sebagai “*keyword*” dan kemudian dilakukan analisa terhadap tersebut yang berafiliasi dengan topik pendidikan [11].

Oleh karena itu penelitian ini akan membahas bagaimana menganalisa teks *al-qur'an* yang berafiliasi dengan topik pendidikan dengan menggunakan pendekatan *text mining*, dimana metode yang akan digunakan adalah algoritma klasifikasi *naive bayes*.

1.2. Rumusan Masalah

Berdasarkan latar belakang masalah yang telah peneliti paparkan di atas, maka rumusan masalah yang akan peneliti ambil adalah sebagai berikut :

- 1 Bagaimana proses penerapan *text mining* dengan menggunakan algoritma *naive bayes* dalam menganalisa kata-kata yang berafiliasi dengan pendidikan dalam *al-qur'an* ?
- 2 Bagaimana melakukan analisa kata-kata yang berafiliasi dengan pendidikan dalam *al-qur'an* berdasarkan hasil dari proses penelitian yang menggunakan pendekatan *Text Mining* dengan algoritma *Naive Bayes* ?

1.3. Tujuan

Tujuan dari penelitian ini adalah untuk menganalisa teks *al-qur'an* dengan cara mengimplementasikan *Text Mining* pada *Al-Qur'an*. Kemudian langkah yang akan dilakukan adalah memilih kata-kata yang berafiliasi dengan pendidikan untuk dianalisa dengan menggunakan pendekatan *text mining*. Prosesnya dilakukan dengan pendekatan *text mining* dengan metode klasifikasi dan algoritma *naive bayes*. Kemudian hasil dari proses yang akan dilakukan analisa untuk diambil kesimpulan penelitian.

1.4. Manfaat

Ini adalah penelitian pemula pertama tentang analisa *al-qur'an* dari sudut pandang Teknologi Informasi dilingkungan kampus UIN Ar-Raniry Banda Aceh. Semoga penelitian ini membawa manfaat bagi semua kalangan.

1.4.1. Manfaat Bagi Peneliti

Semoga penelitian ini dapat menambah pengetahuan kepada peneliti dibidang *data mining* yang berdasarkan pengalaman pada saat melakukan penelitian ini. kemudian juga dapat menemukan ide-ide baru dalam melakukan penelitian lain tentang teknologi informasi yang diharapkan oleh peneliti bermanfaat bagi khalayak banyak, terutama bagi peneliti sendiri.

1.4.2. Manfaat Bagi Pembaca

Harapan kepada para pembaca penelitian ini adalah agar dapat dijadikan sebagai referensi tentang penelitian yang akan dilakukan selanjutnya oleh para

peneliti yang lain dibidang data mining. Kemudian peneliti juga berharap kedepan ada peneliti selanjutnya yang melakukan inovasi pada penelitian ini.

1.4.3. Manfaat Bagi Pendidikan

Semoga penelitian ini bisa berguna dikemudian hari bagi peneliti lainnya yang membahas tentang *text mining* baik pada *al-qur'an* dan lain sebagainya. Kemudian juga semoga dapat bermanfaat pada dunia pendidikan dan *sains* sebagai bahan belajar dan bahan kajian belajar akademik.

1.5. Batasan Masalah

Penelitian ini menggunakan *al-qur'an* terjemahan Bahasa Indonesia yang diperoleh dari Kementrian Agama RI. Pada penelitian ini menggunakan beberapa kata kunci yang dijadikan sebagai acuan penelitian yang menggunakan pendekatan *text mining* dengan menggunakan metode klasifikasi *naive bayes*. Kata kunci yang dimaksud adalah beberapa kata dalam Bahasa Indonesia yang berafiliasi dengan pendidikan.

Tempat, lokasi dan waktu untuk melakukan penelitian ini adalah disepertaran kota Banda Aceh dan Kabupaten Pidie. Tidak adanya tempat khusus untuk melakukan proses penelitian ini. Peneliti membuat diberbagai kesempatan dan tempat baik itu dikampus Universitas Islam Negeri Ar-Raniry Banda Aceh atau bahkan diperpustakaan dan warung kopi.

BAB II

TINJAUAN PUSTAKA

2.1. Data Mining

Data Mining adalah sebuah ilmu yang dipakai untuk melakukan dan membuat sebuah kecerdasan buatan (*interesting knowledge*). Proses *data mining* merupakan ilmu yang dipakai untuk proses *interesting knowledge* pada sejumlah data dengan jumlah yang besar. *Data mining* memperoleh data dari *database*, *data warehouse* dan berbagai tempat penyimpanan data lainnya. *Data Mining* dapat mencakup apa saja antara lain adalah *data*, *image*, *text* dan lain sebagainya. Pada penelitian ini fokus penelitian yang dipakai adalah *text mining* [6].

2.1.1. Text Mining

Text Mining merupakan bagian dari pada *data mining* dimana *text mining* juga disebut sebagai *Text Data Mining* (KDT). Secara definisi umum *text mining* merupakan proses penambangan data yang berbentuk teks. Penambangan data dengan *Text Mining* dipakai pada data yang berbentuk teks dan diperoleh dari berbagai aspek seperti pada dokumen, buku, paper dan lain sebagainya yang mana fungsinya untuk mencari informasi yang berbentuk teks pada data yang dimaksud. Tujuannya untuk mencari informasi pada data teks atau *keywords* (kata kunci) yang mana benar-benar mewakili isi dari dokumen sehingga didapatkan satu atau lebih kesimpulan pada data [6].

2.1.2. Perbedaan Data Mining dengan Text Mining

Perbedaan antara *Data Mining* dengan *Text Mining* adalah proses analisis terhadap suatu data. *Data Mining* adalah proses analisa sejumlah data yang berukuran besar yang di simpan pada *Database*, *Data Warehouse* atau tempat penyimpanan informasi lainnya. Sementara *Text Mining* merupakan

analisa untuk mengumpulkan *Keywords* atau *Terms* (istilah) yang paling sering muncul dan dimana tujuannya adalah untuk menemukan hubungan pada *Keywords* tersebut [6].

2.1.3. Metode Dalam Data Mining

Beberapa metode *Data Mining* Menurut Rijois Iboy Erwin Saragih dan kawan-kawan dalam papernya yang berjudul “Penerapan Algoritma Aposiasasi Data Mining Untuk Mengetahui Kecurangan Skripsi”, *Data Mining* di bagi menjadi 6 dan antara lain adalah sebagai berikut [12]:

1. Deskripsi : Menggambarkan Pola dalam data.
2. Estimasi : Perkiraan, penilaian atau pendapat dan proses ini lebih cenderung dengan menggunakan metode Numerik.
3. Prediksi : Metode untuk memprediksi atau membaca keadaan yang tepat dan ilmiah di masa yang akan datang.
4. Klasifikasi : Mengkategorikan Variabel.
5. *Clusterring* : Pengelompokkan variabel berdasarkan kriteria, misalnya kemiripan.
6. Asosiasi : Menemukan atribut yang muncul dalam satu waktu.

Setiap metode yang tersebut diatas memiliki kelebihan dan kekurangan. Untuk penerapan metode itu sendiri dikategorikan pada bentuk penelitian. Sedangkan pada penelitian ini metode *data mining* yang akan dipakai adalah Klasifikasi. Karena bentuknya adalah untuk memilah topik pendidikan (variabel) dalam *al-qur'an*.

2.1.4. Metode Klasifikasi

Metode Klasifikasi adalah mengkategorikan data berdasarkan variabel. Data akan dipilah menjadi beberapa kelompok data sesuai dengan kemiripan dan memiliki hubungan. Pada tulisan ini bentuk klasifikasinya adalah mencari

kata-kata yang berafiliasi dengan topik pendidikan, maka hasilnya nanti adalah klasifikasi kata-kata yang berafiliasi dengan topik pendidikan dalam teks *al-qur'an* [13].

2.1.5. Algoritma *Naïve Bayes*

Algoritma *Naïve Bayes* atau sering juga disebut dengan nama *naïve Bayes Classifier*. *Naïve Bayes* merupakan algoritma yang dipakai untuk mencari data yang memiliki nilai dengan kemungkinan yang paling tinggi. Sehingga sekelompok data dapat dipilah atau dikelompokkan sesuai dengan kemiripan dan juga tepat. *Naïve Bayes* juga merupakan metode dalam Data Mining yang termasuk kedalam sepuluh metode klasifikasi yang paling populer diantara algoritma lainnya. Algoritma ini memanfaatkan metode probabilitas dan statistik dan juga sangat sederhana untuk mengasumsikan klasifikasi atribut [14].

Rumus Probabilitas pada algoritma *Naïve Bayes* adalah probabilitas *Posterior* dan *Prior*. Rumusnya adalah sebagai berikut [15].

Persamaan dari *Prior* :

$$P(H) = \frac{N_j}{N} \quad \text{Rumus 2.1 Rumus Probabiliti } \textit{Prior}$$

Dimana:

N_j : Jumlah Data pada suatu Class

N : Jumlah total data

Persamaan dari *Posterior* :

$$P(H|X) = \frac{P(X|H).P(H)}{P(X)} \quad \text{Rumus 2.2 Rumus Probabiliti } \textit{Posterior}$$

Dimana:

X : Data dengan class yang belum diketahui

- H : Hipotesis data merupakan suatu class spesifik
- $P(H|X)$: Probabilitas Hipotesis H berdasarkan kondisi X (Positerior Probabilitas)
- $P(H)$: Probabilitas Hipotesis H (*Prior* Probabilitas)
- $P(X|H)$: Probabilitas pada Hipotesis X berdasarkan kondisi pada Hipotesis H
- $P(X)$: Probabilitas X

2.2. *Al-Qur'an*

Al-Qur'an merupakan kitab suci umat Islam diseluruh dunia yang diturunkan Allah kepada Rasulullah sekaligus sebagai mukjizat yang terbesar diantara mukjizat-mukjizat lain yang pernah Allah berikan kepada Rasulullah Muhammad SAW. *Al-Qur'an* sebagai kitab suci yang paling terakhir Allah turunkan untuk dijadikan sebagai petunjuk dan pedoman hidup manusia sampai akhir zaman. Pada *Al-Qur'an* terkandung nilai-nilai yang luhur yang mencakup seluruh aspek kehidupan manusia dalam berhubungan dengan tuhan maupun dengan manusia dan juga antara manusia dengan alam disekitarnya [16].

Oleh karena pentingnya *Al-Qur'an* kepada umat Muslim di seluruh dunia, maka patutlah *Al-Qur'an* dipelajari, difahami dan dimanifestasikan dalam amalan kehidupan sehari-hari. Karena begitu besarnya pesan didalam *Al-Qur'an* yang membahas tentang seluruh kehidupan dunia [16].

2.2.1. *Qur'an Mining*

Qur'an mining adalah gabungan antara *al-qur'an* dengan *text mining*. *Al-Qur'an* adalah kitab suci umat Muslim yang diyakini dan dijadikan sebagai pedoman hidup umat Muslim. *Al-Qur'an* diturunkan kepada Nabi Muhammad SAW dan *al-qur'an* memiliki nilai-nilai yang berhubungan dengan keimanan, akhlak, syariat dan berbagai ilmu pengetahuan lainnya. Sementara *Text Mining* merupakan disiplin ilmu yang dapat menjelajahi dan memproses dataset yang

sangat besar untuk menggali informasi dalam bentuk teks pada data dokumen teks yang dimaksud [13].

Oleh karena itu dapat disimpulkan bahwa *qur'an mining* adalah gabungan antara *al-qur'an* dengan konsep ilmu pengetahuan *text mining*. Dimana gabungan keduanya menciptakan sebuah konsep untuk mencari informasi dalam bentuk teks pada *al-qur'an* yang mana tujuannya adalah untuk mencari informasi tertentu pada dengan menggunakan disiplin ilmu sains sehingga dapat diambil kesimpulan dengan ilmiah dan dapat dipresentasikan dengan ilmiah juga [13].

2.3. Pendidikan

Pendidikan merupakan kegiatan yang kompleks dimana meliputi berbagai komponen yang berkaitan erat satu dengan yang lain. Pendidikan adalah sarana yang sangat efektif untuk mencerdaskan kehidupan bangsa, hal ini juga merupakan salah satu wujud dari tujuan Negara Kesatuan Republik Indonesia. Secara definisi pendidikan adalah suatu usaha sadar dan terencana untuk mewujudkan suasana belajar dan proses pembelajaran yang mana tujuannya adalah untuk menjadikan para peserta didik menjadi aktif, mengembangkan potensi para peserta didik, memberikan pemahaman spiritual keagamaan, kepribadian, kecerdasan dan lain sebagainya [7].

Definisi pendidikan dari para ahli (John Dewey) yang menyatakan bahwa pendidikan adalah proses pembentukan manusia dalam kecakapan secara fundamental, menciptakan manusia menjadi pribadi yang berintelektual dan memiliki emosi dengan alam dan sesama manusia [10]. Oleh karena itu pendidikan adalah pelopor untuk menjadikan manusia sebagai pribadi yang lebih baik, berinovasi dan memiliki kompetensi yang handal.

2.3.1. Pendidikan Dalam Pandangan Al-Qur'an

Dalam *al-qur'an* pendidikan dibahas secara luas dan merangkap ke berbagai aspek. Tetapi dalam *al-qur'an* Bahasa atau kata-kata yang digunakan

adalah istilah dan kiasan. Dalam *al-qur'an* pendidikan disebut dengan istilah *Rabbah (tarbiyah)*, *allama (ta'lim)* dan *addaba (ta'dib)*. Ketiga istilah tersebut adalah istilah baku pendidikan yang ada dalam *al-qur'an*. Oleh karena pada tulisan ini *al-qur'an* yang digunakan adalah terjemahan Bahasa Indonesia maka akan diambil kesimpulan oleh peneliti untuk menggunakan kata-kata umum yang berhubungan dengan pendidikan dan jumlah yang akan di ambil adalah enam, antara lain adalah Ajar, Bicara, Cipta, Dengar, Ingat dan Lihat [17].

2.4. Metode Evaluasi Penelitian

Pada setiap penelitian dipastikan akan mengeluarkan hasil setelah semua proses dilakukan. Hasil yang diperoleh haruslah dapat diperhitungkan keberhasilannya. Adapun hal yang bisa dilakukan adalah dengan evaluasi sebuah penelitian produk, ada beberapa cara dan metode yang dapat digunakan untuk melakukan proses evaluasi antara lain adalah *recall*, *precision*, *entrphy* dan lain sebagainya.

Pada penelitian ini metode yang akan digunakan untuk melakukan proses evaluasi adalah dengan menggunakan *recall* dan *precision*. Adapun *recall* itu adalah nilai yang menunjukkan tingkat perolehan hasil yang dikembalikan. Nilai yang diperoleh dengan membandingkan jumlah item relevan yang dikembalikan. Nilai *recall* yang semakin besar tidak dapat menunjukkan suatu sistem baik atau tidak. Nilai tertinggi dari *recall* adalah 1 yang berarti bahwa seluruh dokumen dalam koleksi berhasil ditemukan [18].

Berikut adalah rumus dari *recall* [19].

$$Recall = \frac{\text{Jumlah dokumen relevan yang di temukan}}{\text{Jumlah semua dokumen relevan didalam koleksi}} \times 100 \%$$

Rumus 2.3 Rumus *recall*

Precision adalah kemampuan sebuah sistem untuk tidak mengembalikan nilai yang tidak relevan. Dapat diartikan juga bahwa *precision* adalah meminimalisir

gangguan. Hal ini digunakan agar mengetahui jumlah data salah yang dikembalikan [18].

Berikut dibawah ini adalah rumus dari *precision* [19].

$$Precision = \frac{\text{Jumlah dokumen relevan yang di temukan}}{\text{Jumlah semua dokumen yang ditemukan}} \times 100 \%$$

Rumus 2.4 Rumus *precision*

2.5. Penelitian Terdahulu

2.5.1. Muhammad Alhawarat, 2015, Pemrosesan teks *al-qur'an* dengan implementasi *text mining*

Penelitian ini bertujuan untuk pemrosesan Bahasa Arab yang alami. Kemudian juga pada penelitian ini memberikan informasi statistik tentang fakta-fakta pada dan ketentuan dalam *al-qur'an* yang akan bermanfaat bagi orang banyak. Penelitian ini menggunakan metodologi penelitian kuantitatif dimana menganalisa setiap hasil dari prosesnya dan *Term Frequency* (TF) sebagai metode untuk perhitungan frekuensi kata-kata [4].

Hasil dari penelitian ini menunjukkan karakteristik dari *al-qur'an* seperti kata-kata yang penting dan dominan pada *al-qur'an* kemudian juga memberitahukan frekuensi yang tinggi terhadap kata-kata tertentu dan juga mengetahui kata-kata yang tegas bunyinya (makna). Kelebihan penelitian ini adalah adanya penjelasan logis pada item perhitungan frekuensi kata-kata dimana diterapkannya metode *Term Frequency* (TF) sebagai metode untuk perhitungannya dan juga pemaparan yang jelas dan detail pada *al-qur'an*. Sementara kekurangan dari penelitian ini hampir tidak ada, tetapi disini peneliti mengambil satu kesimpulan pada kekurangannya adalah pada saat menghasilkan *data text cloud*. Karena tidak dijelaskan makna yang dalam pada kata-kata yang dominan atau kata yang memiliki frekuensi yang tinggi. Ayat *al-qur'an* yang

ditulis pada penelitian ini juga menggunakan Bahasa Arab yang tidak ada barisnya [4].

2.5.2. M. Didik R. Wahyudi, 2019, Penerapan Algoritma Cosine Similarity pada Text Mining Terjemah Al-Qur'an Berdasarkan Keterkaitan Topik

Pada penelitian ini dibahas mengenai pencarian ayat dan surat pada *al-qur'an* berdasarkan topik. Tujuannya adalah untuk memudahkan yang mana sebelumnya untuk mencari ayat pada *al-qur'an* masih menggunakan cara konvensional atau manual. Kemudian diciptakan sebuah sistem yang berbasis *text mining* yang menggunakan algoritma *Cosine Similarity* untuk menjawab masalah tersebut. Metodologi penelitian yang digunakan adalah kuantitatif dengan cara eksperimen atau penelitian terapan dengan menerapkan algoritma *Cosine Similarity* untuk mengklasifikasikan (mengelompokkan) terjemahan ayat *al-qur'an* yang sesuai dengan tingkat kemiripan [5].

Hasil pada penelitian ini menunjukkan bahwa tingkat kemiripan antar ayat *al-qur'an* sebesar 20% memberikan hasil terbaik pada pengelompokan index *Al-Qur'an* rata-rata sebesar 46,42%. Tingkat kemiripan antar terjemah ayat *Al-Qur'an* sebesar 40% memberikan rata-rata sebesar 15,39% pada pengelompokan index *Al-Qur'an*. Untuk tingkat kemiripan antar ayat diatas 40%, ada kelompok similaritas ayat yang tidak masuk dalam index *Al-Qur'an* [5].

Kelebihan pada penelitian ini menggunakan terlalu banyak langkah logis dan ilmiah secara runut dan teratur sehingga tidak akan bisa diragukan keilmiahannya. Kekurangannya yang diambil pada penelitian ini adalah proses dan hasil. Dimana *al-qur'an* terjemahan Bahasa Indonesia yang diambil adalah index dan tema *al-qur'an* yang kemudian mengambil sebuah kesimpulan dan dijabarkan dalam sebuah tabel dimana isinya adalah index dan nama index yang kemudian dijadikan dataset dan baru dilakukan proses olah data. Sedikit susah

dalam memahami langkah kerja pada penelitian ini. Sehingga pembaca sedikit dibuat kebingungan dalam memahami prosesnya. Tetapi disamping proses yang sangat banyak dan susah dimengerti penelitian ini membahasnya secara detail [5].

2.5.3. Endang Supriyanti, 2018, Pengukuran Similarity Tema Juz 30 Al-Qur'an Menggunakan Teks Klasifikasi

Penelitian ini bertujuan untuk membahas masalah kedekatan tema pada ayat-ayat *al-qur'an* pada juz 30 yang berkaitan dengan hari kiamat, hisab, surga dan neraka. Metode Penelitian yang digunakan pada penelitian ini adalah kuantitatif dimana langkah-langkah yang dilakukan adalah pengumpulan data, proses data dan berakhir dengan klasifikasi. Klasifikasi dilakukan dengan menggunakan beberapa algoritma dan algoritmanya adalah *Decision Tree*, *Support Vector Machine* (SVM) dan *Naive Bayes* [3].

Hasil Klasifikasi dari penelitian ini mengatakan bahwa diperoleh hasil terbaik dengan klasifikasi *decision tree* akurasi 74,34%, *SVM* akurasi 75,84% dan *Naive Bayes* akurasi 66,29% [3].

Kelebihan pada penelitian ini adalah penjelasan yang bagus, teratur, runut dan dapat dipertanggung jawabkan keilmiahannya. Setiap prosesnya dilakukan analisis dengan angka-angka dan juga memaparkan setiap hasil pada setiap algoritma sehingga pembaca mengetahui hasil dari setiap algoritma yang dipakai. Sementara kekurangannya adalah pada hasil *data text cloud* tidak dilakukan proses filtering kata sambung atau penghubung seperti “Dan”, “Yang”. Terlepas dari kekurangan itu penelitian ini banyak hal positif yang dapat diambil dan dijadikan referensi pada penelitian ini [3].

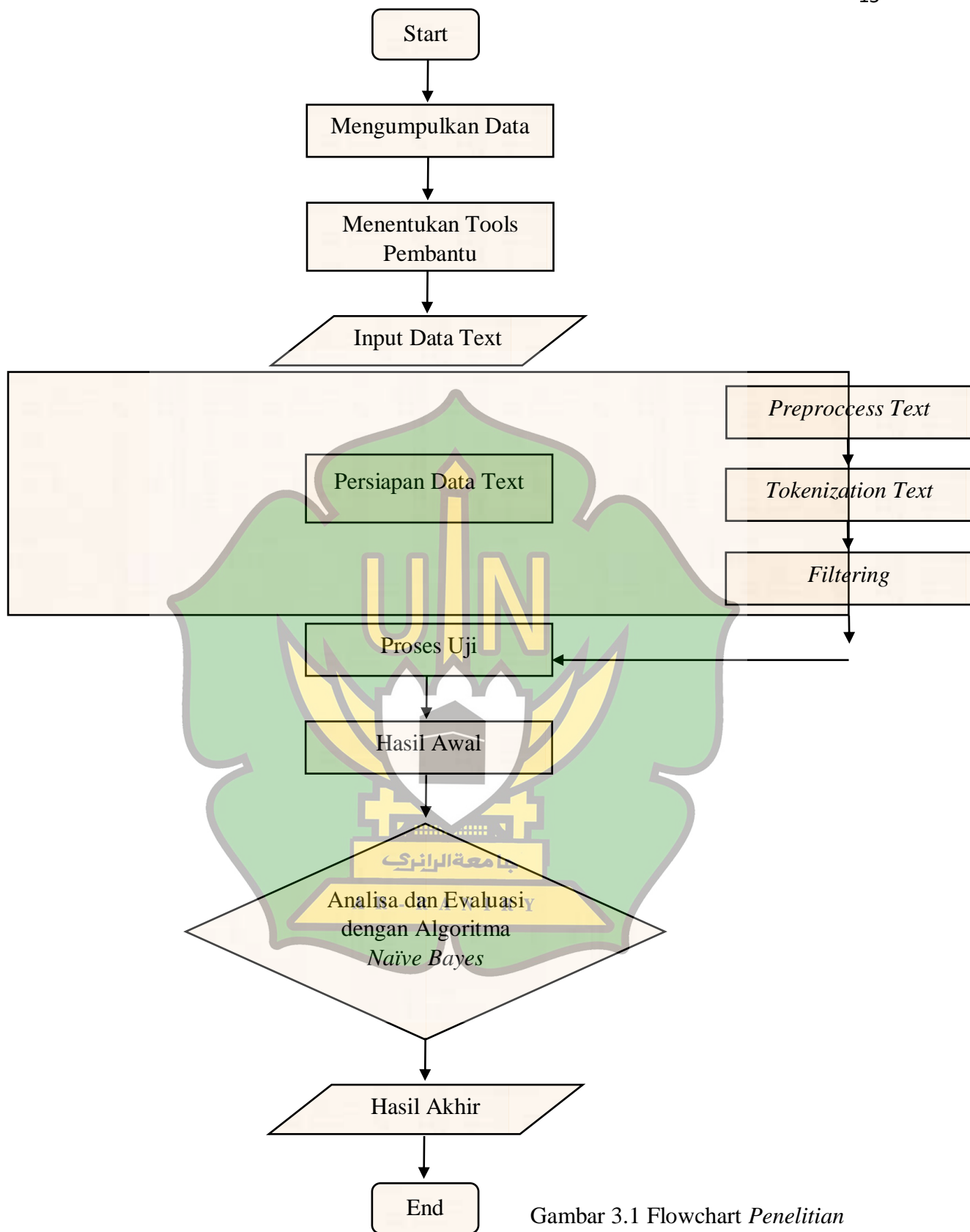
BAB III

METODELOGI PENELITIAN

3.1. Metodologi Penelitian

Metodologi penelitian adalah sebuah ilmu yang mempelajari tentang bagaimana cara melakukan pengamatan dengan langkah-langkah yang logis dan secara ilmiah. Cara-cara yang diterapkan pada metodologi penelitian haruslah berurutan dan logis untuk mencapai tujuan memecahkan masalah dan memberi bukti-bukti yang akurat pada setiap tahapan. Pada metodologi penelitian ada beberapa model penelitian yang bisa digunakan untuk diterapkan pada proses penelitian. Model penelitian yang paling sering digunakan adalah kualitatif dan kuantitatif. Kualitatif lebih kepada model penelitian kajian pustaka atau bacaan buku dan karya ilmiah yang sebelumnya. Sedangkan kuantitatif lebih melakukan penelitian eksperimen (tindakan). Model penelitian digunakan sesuai dengan kebutuhan penelitian dan ada juga penelitian yang menggunakan kedua model penelitian tersebut atau disebut juga model penelitian *mix* (campuran) [20].

Oleh karenanya pada penelitian ini metodologi yang akan digunakan adalah metodologi penelitian kuantitatif. Tujuan dari pada penerapan model penelitian ini dikarenakan proses penelitian pada penelitian ini harus melakukan eksperimen pada setiap tahapannya. Artinya selain didasarkan pada kajian ilmiah sebelumnya atau kualitatif penelitian ini juga harus melakukan langkah penelitian secara kuantitatif (eksperimen). Langkah-langkah penelitiannya digambarkan dengan menggunakan diagram alir, tentunya dengan bentuk-bentuk algoritma yang logis. Berikut gambar diagram alir seperti yang ditunjukkan pada gambar diagram dibawah ini.



Gambar 3.1 Flowchart Penelitian

3.2. Waktu dan Tempat Penelitian

Penelitian ini dilakukan disepetaran Kota Banda Aceh, Provinsi Nangroe Aceh Darussalam. Penelitian dimulai pada bulan Januari sampai dengan bulan Mei. Tempat untuk melakukan penelitian ini tidak terlalu spesifik pada satu tempat, antara lain adalah dilingkungan kampus dan berbagai tempat lainnya.

3.3. Proses Pengumpulan Data

Penelitian ini menggunakan *file* data *al-qur'an* digital terjemahan Bahasa Indonesia yang diperoleh dari laman *website* resmi Kementerian Agama RI. Proses download adalah dengan cara *request* pada layanan aplikasi website Kementerian, kemudian *file* data akan dikirimkan ke email. Berikut alamat *website*-Nya <https://quran.kemenag.go.id/>.

3.4. Pemilihan Tools

Proses penelitian ini tidak dilakukan secara manual tetapi tetap menggunakan *tools* atau *software* pembantu dalam mengolah data. *Tools* yang akan digunakan pada penelitian ini adalah aplikasi *Orange Canvas Data Mining*. Aplikasi ini bisa didapatkan secara gratis dan diinstall pada komputer tanpa harus membeli lisensi penggunaan aplikasi.

3.5. Penginputan Data

Data *al-qur'an* digital terjemahan Bahasa Indonesia yang telah disiapkan sebelumnya diinput ke dalam aplikasi. Data yang diinput berbentuk TXT (*Text*). Setiap surat dipisah-pisah untuk dijadikan satu file. Untuk seluruh dokumen maka data yang harus ada adalah 134 file, dikarenakan jumlah surat dalam *al-qur'an* adalah 134 surat. Data yang akan diinput juga nantinya adalah dengan import folder dalam hal ini isi folder tersebut adalah semua data dari 134 surat yang telah disebutkan sebelumnya.

3.6. Persiapan data

Ada beberapa persiapan yang harus dilakukan sebelum proses penelitian dengan *tools* (aplikasi pembantu). Adapun beberapa hal yang harus dipersiapkan adalah sebagai berikut.

3.6.1. *Preprocessing*

Preprocessing adalah proses merubah teks menjadi term index dari kata perkata. Tujuannya agar setiap kata pada data teks terindikasi array untuk dikenali oleh sistem *tools* untuk dilakukan proses implementasi *text mining* [3].

3.6.2. *Word Parsing dan Tokenization*

Merupakan proses mengubah teks menjadi kumpulan term dengan cara menghapus seluruh karakter tanda baca dalam data. *Parsing* merupakan proses pemecahan data menjadi beberapa bagian unit-unit kecil. Misal satu paragraf teks akan dipecah menjadi bagian-bagian teks saja. Contoh kalimat “saya adalah mahasiswa”, maka secara otomatis didalam sistem aplikasi akan diubah menjadi 3 token yaitu “saya”, “adalah”, “mahasiswa” [3].

3.6.3. *Filtering*

Proses *filtering* ini adalah proses pemilihan kata-kata penting setelah dilakukan proses *Tokenisasi*. Kata-kata penting yang dimaksud adalah dengan membuang kata-kata yang tidak penting seperti kata penghubung, kata imbuhan dan kata sambung contohnya adalah “di”, “dan”, “atau”, “dia”, “itu” dan lain sebagainya. Misal contoh dalam kalimat adalah seperti berikut dibawah ini :

Contoh kalimat : “saya sering baca qur’an di teras rumah pada saat sore hari”

Tokenization : “saya”, “sering”, “baca”, “quran”, “di”, “teras”, “rumah”, “pada”, “saat”, “sore”, “hari”

Filtering : “saya”, “sering”, “baca”, “quran”, “teras”, “rumah”, “sore”, “hari” [3].

Tujuan dari pada penerapan proses ini adalah untuk menyaring dan membuang kata-kata yang tidak penting. Karena jika tidak pada saat mengeluarkan hasil nanti akan sangat banyak mengeluarkan kata-kata yang demikian dan bahkan akan menjadi sangat dominan dari pada kata-kata utama (penting).

3.6.4. *Stemming*

Stemming merupakan sebuah proses yang dilakukan untuk mendapatkan sebuah kata dasar dalam sebuah kalimat, dokumen atau teks. Dimana cara kerjanya adalah dengan cara memisahkan kata dasar dengan imbuhan, baik itu awalan (*prefix*) ataupun akhiran (*suffix*). Sebagai contoh cara kerjanya adalah sebagai berikut. Kata “Bersama”, “Kebersamaan”, “Menyamai” akan dianggap dan dijadikan menjadi satu kata dasar yaitu “Sama” [21].

Stemming juga menerapkan yang namanya algoritma untuk proses sistem kerjanya. Untuk setiap Bahasa memiliki teknik *stemming* yang berbeda-beda. Hal ini menjadikan penerapan algoritma *stemming* sendiri tidaklah boleh dipakai sembarangan. Sebagai contoh dalam Bahasa Inggris memiliki morfologi yang berbeda dengan Bahasa Indonesia, sehingga penggunaan teknik *stemming* untuk kedua Bahasa ini juga pasti akan berbeda [21].

Sebagai contoh penerapan kata kunci untuk penelitian ini adalah Cipta, Bicara, Ajar, Ingat, Dengar dan Lihat. Semua ini adalah kata dasar yang menjadi kata kuncinya. Tetapi dalam Bahasa Indonesia (Teks, Dokumen, File DLL) tidaklah menerapkan Bahasa yang hanya menggunakan kata dasarnya saja, tetapi disertakan juga dengan Imbuhan (awalan dan akhiran). Oleh karena itu dalam hal mendapatkan hasil penelitian nantinya juga diharuskan mencari setiap kata yang memiliki kata dasar yang telah disebutkan di atas. Hal ini tentu dengan pemilihan semua kata yang memiliki kata dasar yang sama walaupun sudah disertakan dengan imbuhan pada masing-masing kata.

Kemudian barulah akan diperhatikan lagi pada pandangan Sinonim kata yang mana berfungsi untuk melihat makna dari pada setiap kata tersebut. Kata dasar yang disebutkan di atas akan dijadikan sebagai kata kunci pada penelitian ini. Berikut ini beberapa pemilihan kata dasar yang juga disertai dengan kata imbuhan.

Tabel 3.1 Tabel kata dasar dan kata dasar yang berimbuhan [22].

No.	Kata Dasar	Kata Dasar Berimbuhan
1.	Cipta	<ul style="list-style-type: none"> - Menciptakan - Diciptakan - Tercipta - dll
2.	Bicara	<ul style="list-style-type: none"> - Bicara - Berbicara - Dibicarakan - Membicarakan - dll
3.	Ajar	<ul style="list-style-type: none"> - Ajarkan - Mengajarkan - Belajar - Diajarkan - Terpelajar - dll
4.	Dengar	<ul style="list-style-type: none"> - Mendengar - Didengar - Pendengar - Terdengar

		<ul style="list-style-type: none"> - Dengarkan - dll
5.	Lihat	<ul style="list-style-type: none"> - Melihat - Dilihat - Lihatlah - Terlihat - dll
6.	Ingat	<ul style="list-style-type: none"> - Ingatlah - Ingatkan - Teringat - Mengingat - dll

Berikut yang disebutkan di Tabel 3.1 adalah beberapa kata kunci yang akan digunakan pada penelitian ini dan disertakan juga dengan beberapa imbuhan pada setiap katanya yang memiliki potensi akan ditemukan nantinya pada saat melakukan penelitian. Semua kata yang memiliki imbuhan tersebut semuanya tetap terpaku pada satu kata dasar yang sama.

Alasan pemilihan beberapa kata yang telah dipaparkan pada tabel di atas adalah karena dilatarkan pada sebuah jurnal referensi yang ditulis oleh Syukur Abu Bakar. Pada jurnal tersebut membahas masalah pendidikan dalam dalam *al-quran* dan menjadikan surat *ar-rahman* ayat 1-4 sebagai objek kajian. Oleh karena itu peneliti mengambil kesimpulan berdasarkan jurnal tersebut dan mengambil beberapa kata dasar pada surat *ar-rahman* ayat 1-4. Kata-kata yang dimaksud adalah ajar, cipta dan bicara [10].

Untuk kata-kata Ingat, Lihat, dan Dengar Peneliti memperolehnya dari hasil wawancara dengan Bapak Drs. Saifuddin Mahmud, M.Pd sebagai Ahli Bahasa Indonesia Universitas Syiah Kuala Banda Aceh.

3.6.5 Sinonim Kata

Sinonim merupakan istilah yang berasal dari Bahasa Yunani Kuno yang terdiri dari kata *syn* yang berarti “dengan” dan kata *anoma* yang berarti “nama”. Jadi jika diartikan kedua kata tersebut memiliki makna yang berarti “Nama lain untuk benda yang sama” [23].

Dengan Bahasa yang lebih sederhana dapat dikatakan juga bahwa Sinonim merupakan penyebutan kepada sebuah benda, nama, istilah atau lain sebagainya dengan sebutan yang berbeda. Misal jika kita ingin mengatakan “cerdas” kepada seseorang, kita juga bisa dengan hanya cara menyebutkan “Pintar”. Dikarenakan kedua kata tersebut memiliki persamaan arti yang sama.

Pada penelitian ini beberapa kata kunci yang digunakan juga memiliki sejumlah kata-kata lain yang memiliki persamaan kata (sinonim) yang sama. Misalnya dalam penelitian ini mengandung kata kunci “cipta”, kata “cipta” ini memiliki persamaan kata juga dengan kata “membuat”, “menjadikan”, “Menghasilkan” dan lain sebagainya. Berikut dibawah ini peneliti memaparkan semua kata kunci yang digunakan pada penelitian ini yang juga disertai dengan persamaan kata-Nya (sinonim).

Tabel 3.2 Tabel kata kunci beserta sinonim [22], [24].

No.	Kata Kunci	Sinonim / Persamaan Kata	
1.	Ajar	Asuh	Bercermin
		Bimbing	Berguru
		Didik	Menelaah
		Gasak	Memahirkan
		Latih	Mengkaji
		Membentuk	Menggali

		Membimbing	Beradap
		Mendidik	Berbudi
		Mengarahkan	Berilmu
		mengasuh	Berpendidikan
		Menunjuki	Berpengetahuan
		Menuntun	Membaca
		Berlatih	Membiasakan
2.	Bicara	Cakap	Perdebatan
		Kata	Perkataan
		Suara	Permusyawaratan
		Ucapan	Berbahasa
		Ujaran	Bercerita
		Membahas	Berdiskusi
		Membincangkan	Berkata
		Mempercakapkan	Berkisah
		Memperkatakan	Bersuara
		Mendiskusikan	Mengobrol
		Menyebut	Menjelaskan
		Perbincangan	Berpendapat
		Percakapan	
3.	Cipta	Agan-angan	Mengarang
		Mengadakan	Memproduksi
		Menjadikan	Membuat
		Buatan	Membikin
		Karangan	Membentuk
		Kreasi	Membangun
		Pembentukan	Melahirkan
		Pembuatan	Mengadakan
		Pendirian	Merupakan
		Penemuan	Mewujudkan
		Melahirkan	Menemukan
		Menyusun	
		Menghasilkan	
4.	Dengar	Akal Budi	Membahas
		Berkata	Berbincang

		Bercakap	Berpendapat
		Berbahasa	Suara
		Mempercakapkan	Tutur
		Memperkatakan	Ucapan
		Merundingkan	Ujaran
		Berpidato	Penalaran
		Percakapan	
5.	Ingat	Terkenang	Sadar
		Bermaksud	Berhajat
		Berniat	Hirau
		Mau	Pendapat
		Sangka	Memikirkan
		Tahu	Mempertimbangkan
6.	Lihat	Amati	Mengintai
		Lihat-lihat	Mengincar
		Lirik	Memeriksa
		Perhati	Mengawasi
		Pandang	Mengecek
		Tampak	Membesuk
		Tatap	Menjenguk
		Tengadah	Mengetahui
		Tinjau	Mempertunjukkan
		Tonton	menunjukkan
		Memandang	Tampak
		Memperhatikan	Terpandang
		Menyaksikan	Bertatapan
		Menatap	
		Melirik	
		Mengamati	

3.6.7. Rujukan Ayat Setiap Item Kata Dasar

Penelitian ini memiliki keterkaitan dan implementasinya pada kitab suci *Al-Qur'an*. Oleh karena itu setiap item (kata kunci) pada penelitian ini juga

akan di kaitkan dengan ayat-ayat yang ada dalam *Al-Qur'an*. Ada sekitar 6 (enam) item yang dipakai pada penelitian ini seperti yang telah disebutkan pada tabel 3.2. Oleh karena itu peneliti juga memaparkan beberapa jumlah ayat yang memiliki hubungan dengan setiap item-Nya. Berikut pemaparannya.

Tabel 3.3 Tabel kata kunci beserta ayat yang mengandung setiap item

No.	Kata Kunci	Ayat Yang Mengandung	Artinya	Surat
1.	Ajar	عَلَّمَ الْقُرْآنَ ٢	Yang Telah Mengajarkan Al-Qur'an.	Ar-Rahman Ayat 2 [10].
2.	Bicara	عَلَّمَهُ الْبَيَانَ ٤	Mengajarkannya Pandai Berbicara.	Ar-Rahman Ayat 4 [10].
3.	Cipta	خَلَقَ الْإِنْسَانَ ٣ جامعة الرانيري AR - RANIRY	Dia Menciptakan Manusia.	Ar-Rahman Ayat 3 [10]
4.	Dengar	وَاللَّهُ لَمُرْسِلِكُمْ مِنْ أَطْرَافِكُمْ أَنْتُمْ لَا تَعْلَمُونَ حَتَّىٰ يَأْتِيَكُمُ الْكُفْرُ مِنَ الْأَنْفُسِ وَالْأَنْفُسُ وَالْأَفْئِدَةُ لَعَلَّكُمْ تُشْكُرُونَ ﴿٧٨﴾	Dan Allah mengeluarkan kamu dari perut ibumu dalam keadaan tidak mengetahui sesuatupun, dan	An-Nahl Ayat 78 [25].

			Dia memberi kamu pendengaran, penglihatan dan hati, agar kamu bersyukur	
5.	Ingat	وَإِذْ قَالَ رَبُّكَ لِلْمَلَائِكَةِ	Ingatlah ketika Tuhanmu berfirman kepada para Malaikat.	Al-Baqarah Ayat 30 [26].
6.	Lihat	وَاللَّهُ أَخْرَجَكُمْ مِنْ بُطُونِ أُمَّهَاتِكُمْ لَا تَعْلَمُونَ شَيْئًا وَجَعَلَ لَكُمُ السَّمْعَ وَالْأَبْصَارَ وَالْأَفْئِدَةَ لَعَلَّكُمْ تَتَّقُونَ ﴿٧٨﴾	Dan Allah mengeluarkan kamu dari perut ibumu dalam keadaan tidak mengetahui sesuatupun, dan Dia memberi kamu pendengaran, penglihatan dan hati, agar kamu bersyukur	An-Nahl Ayat 78 [25].

3.7. Pengujian

Jika semua data sudah dipersiapkan, maka data telah siap untuk dilakukan pengujian dengan mengimplementasikan *text mining* dengan menggunakan metode

klasifikasi algoritma *naïve bayes clasifier*. Setelah mengeluarkan hasil dari pengujian dan penerapan *text mining* maka barulah masuk ketahap analisa. Jika tidak mengeluarkan hasil yang diharapkan pada saat pengujian maka proses akan diulangi.

3.8. Hasil Awal

Hasil awal berupa hasil dari tahap setelah pengujian. Dimana pada tahapan ini akan dipertimbangkan keberhasilannya berdasarkan hasil yang diharapkan sebelumnya. Setelah hasil ini didapatkan baru kemudian akan dilakukan proses analisa untuk menilai keberhasilan dan mengambil keputusan untuk menetapkan hasil akhir dari penelitian.

3.9. Analisa dan Evaluasi dengan Algoritma *Naïve Bayes*

Pada proses analisa akan dilakukan proses penjabaran dari pada hasil uji coba pada saat penerapan *text mining*. Proses penjabaran adalah mendeskripsikan dan memberi keterangan lanjutan pada hasil yang diperoleh dari pengujian. Hasil pengujian seperti *Cloud Data Text* akan diberi penguatan, penjelasan dan hasil persentase untuk setiap item kata yang berafiliasi dengan pendidikan dalam *al-qur'an*. Sehingga hasil yang harapkan dapat dipertanggung jawabkan dengan logis dan real.

Kemudian untuk evaluasi dengan menggunakan Algoritma *Naïve Bayes* pada penelitian ini dilakukan juga dalam aplikasi *Orange Canvas Data Mining*. Hasil *score* yang dikeluarkan berdasarkan dengan algoritma yang ditentukan. Oleh karena penentuan algoritma pada penelitian ini adalah *Naïve Bayes*, maka pada saat melakukan evaluasi juga dilakukan berdasarkan dengan hasil *score* algoritma *naïve bayes*. Sistem aplikasi akan mengeluarkan hasil evaluasi *score* dengan otomatis pada saat dilakukan proses (relasi) dengan algoritma yang ditentukan.

3.10. Hasil Akhir

Hasil akhir yang diharapkan dari penelitian ini adalah hasil dari proses pengujian yang logis, analisa dan penguatan yang bagus serta akurat. Kemudian juga

hasil dari *Cloud Data Text* yang menjadi visualisasi hasilnya benar-benar mendapatkan hasil yang sesuai dengan harapan. Kata-kata yang keluar dari proses uji serta analisa akan dilakukan proses persentase dari setiap item kata. Hasil analisa persentase dan bentuk visualisasi dengan *cloud data text* diharapkan benar-benar dapat dipertanggung jawabkan dan dapat mewakili isi dari *al-qur'an* sesuai dengan topik penelitian yaitu pendidikan.



BAB IV

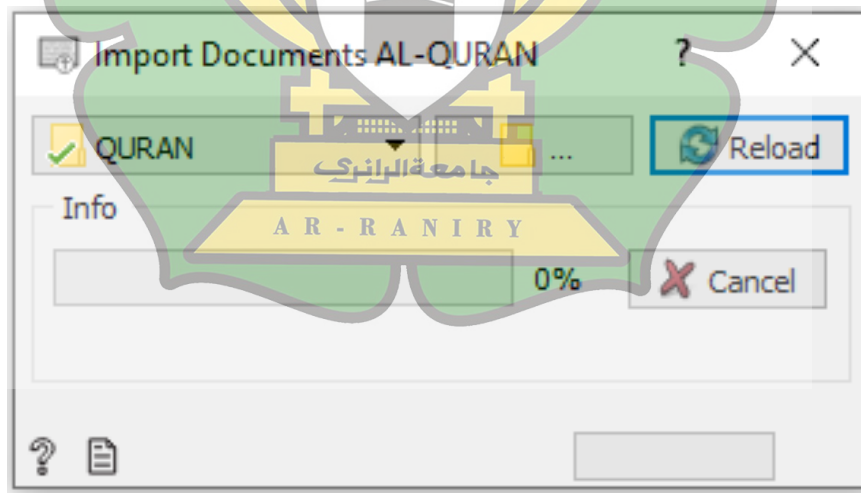
HASIL DAN PEMBAHASAN

Penelitian ini menggunakan *Al-Qur'an* digital yang diperoleh dari Website resmi Kementerian Agama RI. Kemudian peneliti mengambil hanya *Al-Qur'an* terjemahan Bahasa Indonesia saja. Lalu menggunakan aplikasi pendukung untuk menyelesaikan penelitian ini yaitu aplikasi *orange canvas data mining*.

Data *Al-Qur'an* yang telah disiapkan diinput ke aplikasi *orange canvas data mining* barulah dilakukan proses olah data untuk mendapatkan hasil. Hasil akhir akan dipaparkan menjadi dua bagian. Berikut proses untuk mendapatkan hasil dari penelitian ini. Sebagai catatan kata-kata yang akan dijadikan fokus penelitian adalah cipta, bicara, ajar [10], dengar, lihat dan ingat.

4.1. Proses Input Data

Persiapan data yang dilakukan semuanya dilakukan pada aplikasi *orange canvas data mining*. Dimulai dari *tokenization* dan *filtering*.



Gambar 4.1 *Import* data al-qur'an ke aplikasi *orange*

Input data berupa sebuah folder yang berisikan data file *al-qur'an*. *Al-Qur'an* diinput dengan format TXT (*Text Format*). Pada aplikasi *orange canvas data mining* proses dilakukan dengan teratur dan harus bertahap. Proses Input data ke Aplikasi *Orange* tidaklah dilakukan sekaligus. Tetapi dilakukan bertahap, yaitu Per *Juz Al-Qur'an*. *Al-Qur'an* mengandung sebanyak 30 *juz*, dan data yang dipisah menjadi 30 bagian dan barulah dilakukan proses input ke Aplikasi *Orange Canvas Data Mining*. Setelah proses input selesai maka barulah melanjutkan ke proses selanjutnya.

4.2. Proses *Filtering*

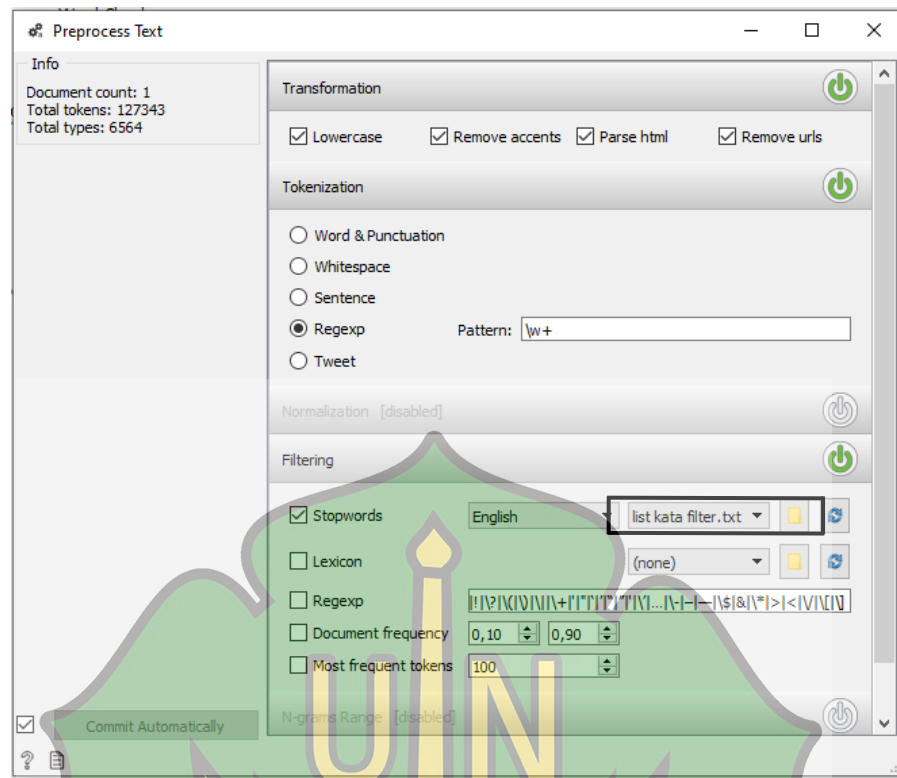
Proses *filtering* adalah *preprocess* atau hal yang dilakukan pada data sebelum diambil hasil atau menambahkan proses selanjutnya. Adapun proses *filter* yang dilakukan adalah membuang kata-kata yang tidak perlu, tanda baca, tidak membedakan huruf besar dan kecil, link web, angka dan lain sebagainya (*stopwords*). Adapun beberapa kata yang dibuang dapat dilihat pada tabel dibawah ini :

Tabel 4.1 Tabel *stopwords* pada data

No.	Kata	Angka dan Karakter
1.	Tidak	0
2.	Telah	1
3.	Dalam	2
4.	Untuk	3
5.	Yaitu	4
6.	Yang	5
7.	Di	6
8.	Ini	7
9.	Aku	8
10.	Ke	9
11.	Dia	10
12.	Itu	11

13.	Dan	12
14.	Dari	13
15.	Juga	14
16.	Akan	15
17.	Dengan	16
18.	Agar	17
19.	Lagi	18
20.	Atau	.
21.	Siapa	,
22.	Bagaimana	?
23.	Apa	>
24.	Dimana	;
25.	Kapan	‘
26.	Kenapa	“
27.	Mengapa	–
28.	Ya	-
29.	Adalah	+
30.	Sebagainya	=

Adapun beberapa kata yang dibuang adalah kata-kata sambung, 5 W + 1 H (*what, where, when, who, why and how*) karakter pada data, angka yang telah dijabarkan diatas dan ada lain sebagainya. Oleh karena banyaknya jumlah kata-kata yang harus di *filter* peneliti tidak dapat menjabarkan semuanya tetapi peneliti melampirkannya dalam sebuah lampiran yang bisa di akses publik (*lampiran I*). Tujuan *filtering* kata karena kata-kata dan angka yang telah dijabarkan pada tabel diatas dan juga pada lampiran tidaklah memiliki makna yang mendalam, hanya sebagai kata sambung, kata petunjuk, numerik dan lain sebagainya.



Gambar 4.2 Proses *filtering* pada aplikasi *orange*

Pada gambar diatas dapat peneliti jabarkan yaitu item-item yang di *filter*, antara lain adalah *lowercase*, *remove accent*, *parse html*, *remove url*. Adapun pengertiannya adalah sebagai berikut.

Tabel 4.2 Tabel item yang di *filter* pada data

No.	Item	Pengertian
1.	<i>Lowercase</i>	Merupakan huruf kecil. Pengertian dihapusnya <i>lowercase</i> adalah tidak lagi membedakan <i>lowercase</i> atau <i>non-lowercase</i> (huruf kecil dan huruf besar). Semuanya akan dinggap sama setelah

		masuk kedalam sistem aplikasi <i>orange data mining</i> .
2.	<i>Accent</i>	<i>Accent</i> yang dimaksud disini bukanlah <i>Accent</i> (logat bahasa) tetapi <i>accent</i> yang dimaksud adalah beberapa simbol yang ada pada huruf. Misalnya tanda koma atas (´) yang dibubuhkan pada huruf. Contoh á, Â, Ñ dan lain sebagainya.
3.	<i>Parse Html</i>	Menghapus seluruh struktur codingan atau beberapa perintah yang berhubungan dengan HTML (<i>HyperText Markup Language</i>). Contohnya adalah “<>”, “<html>”, “&”, kode unik dan lain sebagainya. Yang diterima setelah dilakukan <i>Remove Parse HTML</i> hanyalah kata-kata yang baik, benar dan baku saja.
4.	<i>Url</i>	<i>URL (Uniform Resource Locator)</i> , merupakan alamat atau link <i>website</i> . Link <i>website</i> akan dibuang dan tidak diterima pada sistem aplikasi.

4.2.1. *Filtering*

Filtering dilakukan dengan memasukkan kata-kata secara manual. Kata-kata yang telah dijabarkan dalam tabel 4.1 diinput ke aplikasi *orange canvas data mining*. Beberapa kata yang telah dijabarkan pada tabel 4.1 diketikkan pada *text editor* (pada penelitian ini peneliti menggunakan aplikasi *notepad*) dengan format *txt (text)*. Pada gambar 4.2 peneliti melingkari dengan

Beberapa proses penting telah dijelaskan diatas. Untuk lebih merinci lagi maka peneliti akan memberikan keterangan lanjutan pada proses yang telah dilakukan pada aplikasi *orange*. Sebagai catatan setiap proses telah peneliti berikan tanda dengan nomor. Pembahasan akan peneliti bahas dibawah berdasarkan nomor. Berikut penjelasannya dibawah ini.

Pertama adalah proses input data. Data yang dimaksud adalah *al-qur'an* translate Bahasa Indonesia. Data diinput ke aplikasi *orange canvas data mining*.

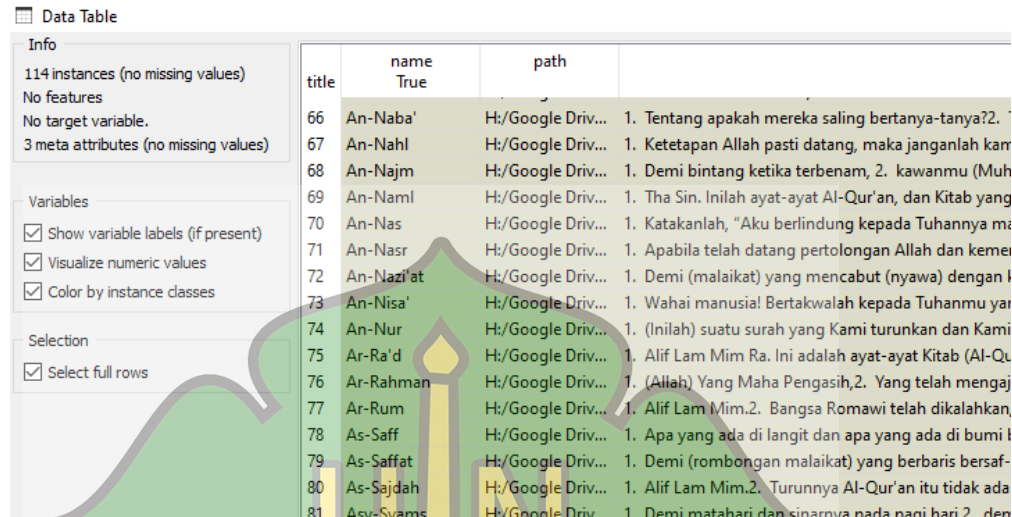
Kedua adalah *curpus*. Dimana fungsinya adalah agar data yang telah diinput dapat dibaca oleh user melalui aplikasi *orange*. Dengan adanya *corpus* ini maka user tidak perlu lagi keluar dari aplikasi *orange* untuk mengecek isi data yang diinput.

Ketiga *corpus viewer* adalah menu untuk membaca file yang ada di dalam *corpus*. *Corpus* telah menjadikan data yang diinput agar dapat dibaca oleh user sehingga tidak perlu lagi keluar dari aplikasi untuk mengecek isi file atau data yang telah diinput. Tetapi untuk melihat isi dari data tersebut user membutuhkan *corpus viewer*.



Gambar 4.4 Bentuk *Viewer Corpus* pada aplikasi *orange*

Keempat adalah *Data Table*. Data yang telah diinput ke dalam aplikasi akan dibaca oleh *data table* terlebih dahulu. Berikut dibawah ini peneliti lampirkan bentuk gambarannya.

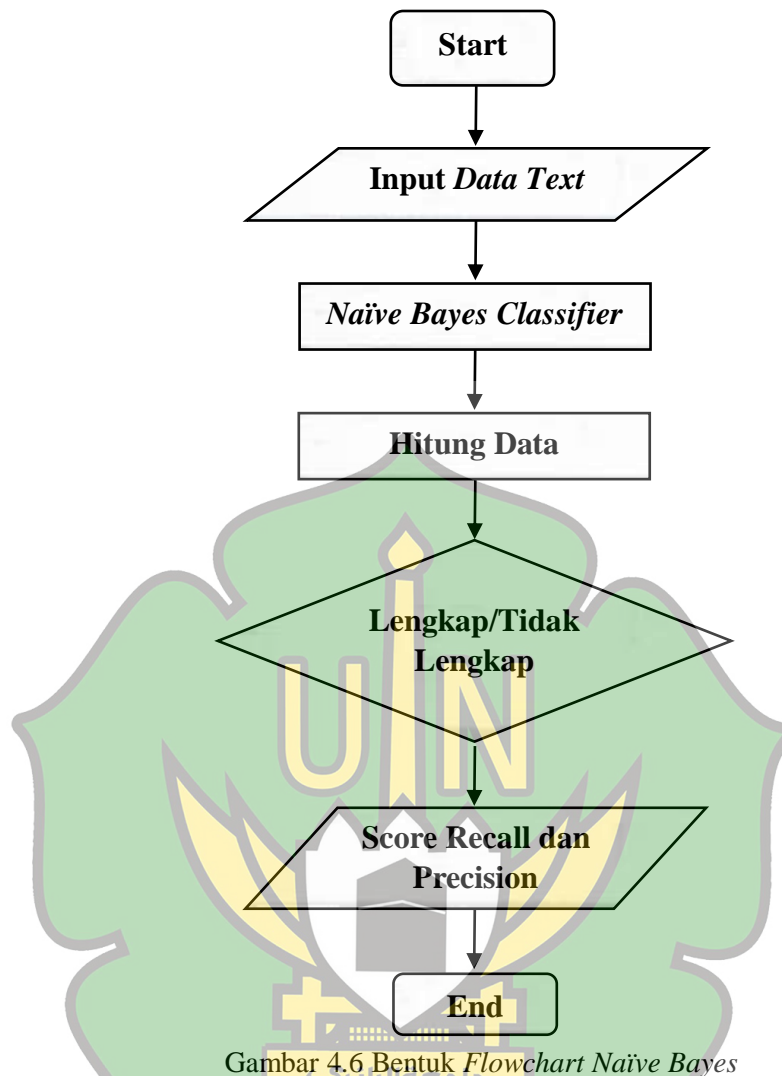


	title	name	path	
		True		
66	An-Naba'		H:/Google Driv...	1. Tentang apakah mereka saling bertanya-tanya?2.
67	An-Nahl		H:/Google Driv...	1. Ketetapan Allah pasti datang, maka janganlah kar
68	An-Najm		H:/Google Driv...	1. Demi bintang ketika terbenam, 2. kawanmu (Muh
69	An-Naml		H:/Google Driv...	1. Tha Sin. Inilah ayat-ayat Al-Qur'an, dan Kitab yang
70	An-Nas		H:/Google Driv...	1. Katakanlah, "Aku berlindung kepada Tuhannya ma
71	An-Nasr		H:/Google Driv...	1. Apabila telah datang pertolongan Allah dan keme
72	An-Naziat		H:/Google Driv...	1. Demi (malaikat) yang mencabut (nyawa) dengan l
73	An-Nisa'		H:/Google Driv...	1. Wahai manusia! Bertakwalah kepada Tuhanmu ya
74	An-Nur		H:/Google Driv...	1. (Inilah) suatu surah yang Kami turunkan dan Kami
75	Ar-Ra'd		H:/Google Driv...	1. Alif Lam Mim Ra. Ini adalah ayat-ayat Kitab (Al-Qu
76	Ar-Rahman		H:/Google Driv...	1. (Allah) Yang Maha Pengasih,2. Yang telah mengaj
77	Ar-Rum		H:/Google Driv...	1. Alif Lam Mim.2. Bangsa Romawi telah dikalahkan,
78	As-Saff		H:/Google Driv...	1. Apa yang ada di langit dan apa yang ada di bumi l
79	As-Saffat		H:/Google Driv...	1. Demi (rombongan malaikat) yang berbaris bersaf-
80	As-Sajdah		H:/Google Driv...	1. Alif Lam Mim.2. Turunnya Al-Qur'an itu tidak ada
81	Asu-Svams		H:/Google Driv...	1. Demi matahari dan sinarnya pada pagi hari 2. den

Gambar 4.5 Bentuk gambaran *data table*

Pada *data table* data akan dibaca berdasarkan setiap nama data, alamat penyimpanan data pada computer dan yang terakhir adalah *value*. *Value* adalah nilai pada setiap data yang masuk. Misal pada gambar diatas, terdapat beberapa nama surat, untuk value akan dibaca sebagai isi dari surat yaitu semua ayat-ayat pada surat tersebut.

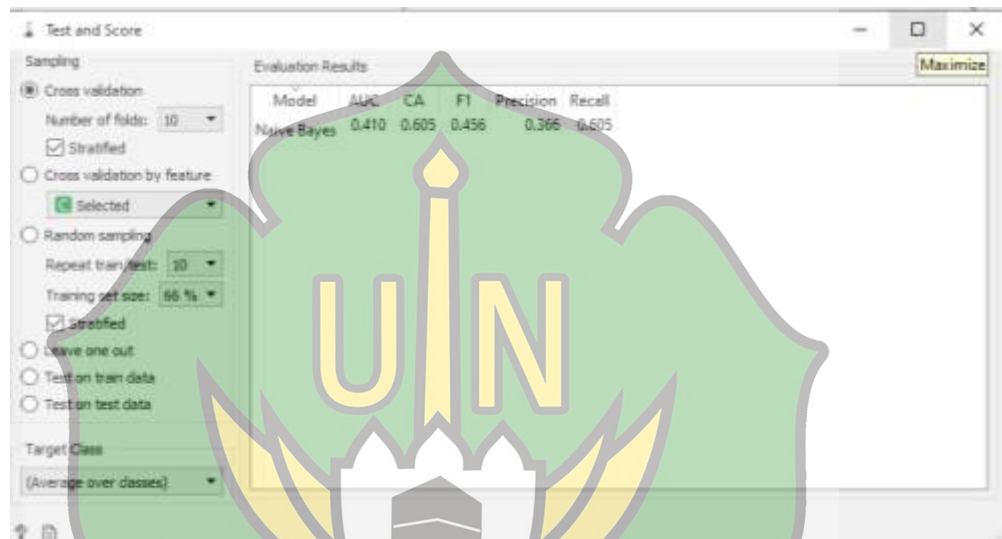
Kelima adalah algoritma *Naïve Bayes*. Sesuai dengan pembahasan diatas bahwa penelitian ini akan menggunakan algoritma *naïve bayes* untuk proses mendapatkan hasil dan bahan Evaluasi *score*. Algoritma *Naive Bayes* sendiri juga telah peneliti jelaskan diatas definisi dari algoritma ini berfungsi untuk mendapatkan hasil yang paling relevan dengan tingkat kemungkinan yang paling tinggi. Untuk langkah sistem kerja Algoritma *Naïve Bayes* pada penelitian ini akan dijelaskan pada diagram alir dibawah ini.



Gambar 4.6 Bentuk *Flowchart Naïve Bayes*

Data yang telah diinput ke aplikasi *Orange Canvas Data Mining* yang pada akhir kesimpulan adalah diimplementasikannya algoritma *naïve bayes*. Kemudian data akan dihitung atau dikalkulasikan dengan algoritma *naïve bayes* secara otomatis pada sistem aplikasi sehingga akan mengeluarkan sebuah tabel *score*. *Score* adalah hasil yang diharapkan untuk dikeluarkan oleh sistem aplikasi kepada *user* untuk dilakukan evaluasi pada hasilnya. *Score* yang dimaksud adalah metode evaluasi algoritma *naïve bayes* yaitu *Recall* dan *Precision*. Untuk evaluasi dibahas pada poin selanjutnya dibawah ini.

Keenam adalah proses *test and score*. Data yang telah dimasukkan yang kemudian dikaitkan dengan sebuah algoritma yang mana algoritma yang dimaksud adalah *naive bayes* maka langkah terakhir adalah untuk melihat *score*. *Score* ini bertujuan sebagai bahan evaluasi tingkat keberhasilan sebuah proses penelitian. Untuk lebih jelas peneliti melampirkan sebuah gambar dibawah ini. berikut gambarnya.



Model	AUC	CA	F1	Precision	Recall
Naive Bayes	0.410	0.605	0.456	0.366	0.605

Gambar 4.7 Bentuk hasil dari Algoritma *Naïve Bayes*

Oleh karena pada penelitian ini menggunakan algoritma *naïve bayes* maka hasil yang dikeluarkan pada proses evaluasi *test and score* juga hanya dengan algoritma *naïve bayes clasifier*. Untuk hasil evaluasi, seperti yang telah dibahas pada bab sebelumnya penelitian ini hanya akan mengambil dua metode saja yaitu *recall* dan *precision*. Hasil *recall* menunjukkan atau mengeluarkan hasil sebesar 0.605 dan *precision* mengeluarkan hasil angka sebesar 0.366. Artinya bahwa *recall* berhasil mengeluarkan hasil seluruh data yang paling relevan secara lebih dominan dari pada *precision* (tidak mengeluarkan data yang tidak relevan).

Ketujuh *Preprocess Text* adalah dimana semua proses penyaringan data atau pembersihan data dilakukan. Untuk memvisualisasikan data yang berupa sebuah gambar *word cloud* dengan jumlah kata yang sangat banyak maka di perlukan sebuah proses untuk membersihkan terlebih dahulu. *Preprocess Text* dapat dilihat pada gambar 4.2 dan tabel 4.1 yang menunjukkan kegunaan dan langkah pada *Preprocess Text*.

Kedelapan adalah *word cloud* dimana bentuk visualisasi data yang dikeluarkan dalam bentuk sebuah gambar yang mana isinya adalah tumpukan kata-kata dimulai dari yang terbesar (menunjukkan paling banyak jumlahnya/dominan) sampai yang terkecil (jumlah yang paling sedikit).

4.3. Evaluasi dengan Algoritma *Naïve Bayes*

Pada penelitian ini dilakukan sebuah evaluasi penelitian yang didasari pada sebuah algoritma, dimana algoritma yang dimaksud adalah *Naïve Bayes*. Algoritma *Naïve Bayes Classifier* memiliki keunggulan untuk mendapatkan hasil yang paling relevan dengan tingkat kemungkinan yang paling tinggi dan lebih tepat dipakai untuk klasifikasi Teks, seperti pada penelitian ini. Perhatikan pada gambar 4.3 point ke 5 menunjukkan penerapan proses algoritma *Naïve Bayes* dan point ke 6 yang menunjukkan proses *Test and Score* atau hasil evaluasi. Antara point ke 5 dan 6 adalah proses relasi antara algoritma *Naïve Bayes* dengan proses *Test and Score* yang mana ini adalah proses untuk mengeluarkan hasil atau *score* yang berhasil diperoleh.

Kemudian *score* yang dikeluarkan oleh algoritma *Naïve Bayes* juga didasari pada beberapa metode evaluasi, sesuai dengan penjelasan sebelumnya bahwa pada penelitian ini akan digunakannya metode evaluasi dan metode yang digunakan pada penelitian ini adalah *recall* dan *precision*. Algoritma *Naïve Bayes* memberikan hasil *score* dengan beberapa metode evaluasi seperti yang ditunjukkan pada gambar 4.6 dan hasil yang di ambil hanyalah dua metode saja yaitu *recall* dan *precision*. Untuk hasil

4.4.1. Analisis data cloud

Berdasarkan kata-kata yang telah dipilih dan dijabarkan diatas maka peneliti memberikan penjabaran secara definisi berdasarkan hasil dari *data cloud*. Adapun kata-kata pokok untuk melakukan penelitian adalah kata-kata yang berafiliasi dengan pendidikan. Beberapa kata-kata tersebut adalah ajar, cipta, bicara, dengar dan lihat.

Tabel 4.3 Tabel kata dasar, imbuhan dan sinonim serta hasil yang diperoleh

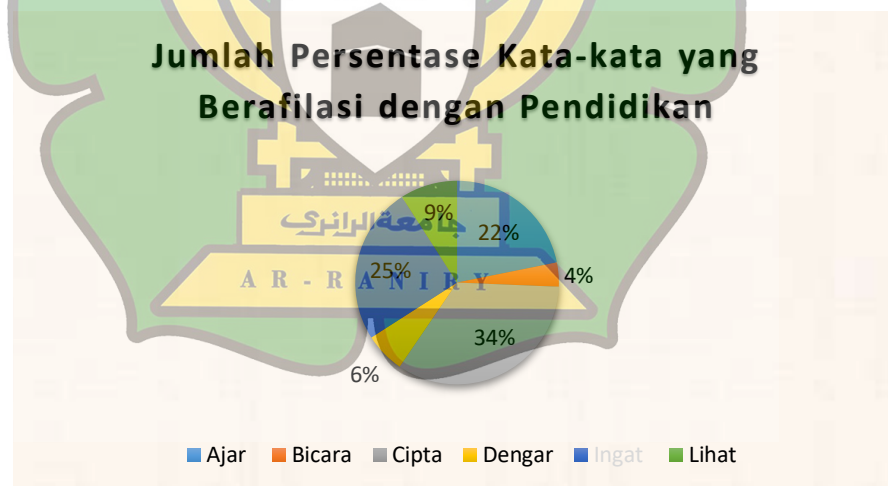
No.	Item	Jumlah Return	Penjabaran
1.	Ajar	<ul style="list-style-type: none"> - Mengajarkan = 11 - Pelajaran = 70 - Memperhatikan = 62 - Ilmu = 62 - Mengikuti = 110 - Petunjuk = 278 - Ketahuilah = 64 - Memperoleh = 106 <p>Total 763</p>	<p>Kata dasar “Ajar” beserta dengan beberapa kata-kata yang memiliki persamaan kata dengan kata dasar “ajar” (sinonim ajar) yang berhasil diperoleh setelah melakukan beberapa tahap proses adalah sebanyak 763 kata. Dimana kata yang terdiri adalah beberapa kata yang memiliki imbuhan (ada awalan dan akhiran pada kata dasarnya), bersinonim dengan kata dasar “Ajar” atau sama arti.</p>
2.	Bicara	<ul style="list-style-type: none"> - Perkataan = 70 - Berita = 64 <p>Total 134</p>	<p>Beberapa kata-kata yang peneliti peroleh yang didasari pada kata dasar “Bicara” adalah sebanyak 134 Kata. Kata yang</p>

			diperoleh adalah kata yang memiliki persamaan kata (sinonim) dengan kata dasar “bicara”. Ada beberapa kata lain yang memiliki hubungan dan persamaan kata (sinonim) dengan kata dasar “bicara” tetapi peneliti menganggap beberapa kata tersebut tidak terlalu relevan sehingga diputuskan untuk tidak dihitung. Serta tidak ada referensi ilmiah yang mendukung.
3.	Cipta	<ul style="list-style-type: none"> - Jadikan = 100 - Menjadikan = 228 - Berbuat = 260 - Menciptakan = 144 A R - Kehendaki = 151 - Kerjakan = 181 - Melaksanakan = 59 - Tangan = 61 <p>Total 1.184</p>	Kata dasar “cipta” mendapatkan jumlah hasil terbanyak, yaitu 1.184 kata total dari keseluruhan. Keseluruhan kata ini diperoleh dari beberapa kata yang memiliki imbuhan dengan kata dasar dan memiliki arti yang sama (sinonim) dengan kata dasar “cipta”.
4.	Dengar	<ul style="list-style-type: none"> - Memperhatikan = 62 	Perolehan hasil yang didapat pada kata dasar “Dengar”

		<ul style="list-style-type: none"> - Berita = 64 - Mendengar = 118 <p>Total 244</p>	<p>adalah sebanyak 244 kata saja. Sama dengan beberapa hasil dari kata dasar yang sebelumnya, kata dasar “dengar” juga diperoleh dari beberapa kata yang berimbuhan (memiliki awalan dan akhiran) dan memiliki persamaan kata (sinonim) dengan kata dasar “dengar”.</p>
5.	Ingat	<ul style="list-style-type: none"> - Ketahuilah = 64 - Perintah = 65 - Bertakwalah = 71 - Bertakwa = 139 - Kehendaki = 151 - Peringatan = 178 - Ingatlah = 205 <p>Total 873 جامة</p>	<p>“Ingat” mendapatkan hasil sejumlah 873 kata. Perolehan kata dasar “ingat” tergolong banyak juga, hal ini mungkin dikarenakan dengan ketegasan dan peringatan dari Allah SWT. Dalam firman-firmannya yang ada di dalam Al-Qur’an.</p>
6.	Lihat	<ul style="list-style-type: none"> - Ketahuilah = 64 - Mata = 69 - Saksi = 77 - Berpaling = 108 <p>Total 318</p>	<p>Kata dasar “Lihat” yang memiliki hubungan dengan panca indera penglihatan manusia ini mengeluarkan hasil sebanyak 318 kata. Sama halnya dengan semua proses yang dilakukan pada semua kata-kata dasar</p>

			sebelumnya, yaitu kesimpulan yang di ambil adalah pada kata dasar yang memiliki imbuhan dan persamaan arti kata dengan kata dasar “lihat”.
--	--	--	--

Berdasarkan hasil *output* dari *text data cloud* jumlah kata-kata yang berafiliasi dengan pendidikan dalam *al-qur'an* adalah seperti yang telah dijabarkan pada tabel di atas (tabel 4.3.). Dari ke-enam kata tersebut memiliki jumlah yang berfariasi, definisi yang berbeda dan ketegasan yang berbeda juga. Grafik dibawah ini menunjukkan jumlah kata pada setiap item. Persentase yang ditunjukkan pada grafik dibawah adalah persentase dari jumlah seluruh bobot dari keenam item kata yang disebutkan pada tabel 4.3. beserta jumlah bobot seluruhnya yang mana adalah 3.516 Kata.



Gambar 4.9 *Chart* jumlah kata pada setiap item

Grafik diatas menunjukkan dominannya atau jumlah kata yang paling banyak mengeluarkan hasil di antara 6 kata dasar yang dijadikan sebagai kata

kunci. Dari ke-enam kata dasar di atas, kata dasar “cipta” mendapatkan hasil yang paling banyak yaitu 34 % total dari 3.516 kata yang berafiliasi dengan pendidikan.

4.5. Hasil Persentase

Hasil dengan persentase ini merupakan jumlah persentase (%) setiap kata yang telah ditetapkan diatas. Persentase setiap kata diacukan kepada jumlah total data bersih pada *al-qur'an*. Jumlah kata bersih yang diperoleh dari *al-qur'an* adalah sebanyak **38.761** kata dan jumlah ini yang akan dikalkulasikan dengan jumlah bobot masing-masing kata pilihan diatas. Berikut hasil dan penjabaran persentase setiap kata tersebut.

Tabel 4.4. : Tabel persentase setiap bobot kata serta penjabaran

No.	Kata	Persentase	Penjabaran
1.	Ajar	$763 / 38.761 * 100$ = 1,96 %	Jumlah persentase yang berhasil didapatkan dari kata dasar “Ajar” adalah sebanyak 1,96 %. Jumlah persentase ini diperoleh dari keseluruhan kata (data bersih) dari <i>al-qur'an</i> .
2.	Bicara	$134 / 38.761 * 100$ = 0,34 %	Jumlah persentase kata dasar “Bicara” mendapatkan hasil yang paling rendah dari ke-enam kata dasar yang lain. Jumlah persentase yang diperoleh adalah sebanyak 0,34 %.
3.	Cipta	$1.184 / 38.761 * 100$ = 3,05 %	Kata “Cipta” mendapatkan hasil 3,05% hasil penjumlahan persentase dari keseluruhan data. Bobot kata dasar “Cipta” dalam <i>al-qur'an</i> mendapatkan hasil yang terbanyak di antara ke-enam

			kata dasar lainnya yang dijadikan sebagai kata kunci yang berafiliasi dengan pendidikan.
4.	Dengar	$244 / 38.761 * 100 = 0,62 \%$	Kata dasar “Dengar” mendapat jumlah persentase sebanyak 0,62 % dari keseluruhan data bersih yang diperoleh. Kata dasar “Dengar” menjadi kata dasar yang termasuk sedikit hasilnya, lebih banyak 0,28 % dari kata dasar “Bicara” yang menempati posisi paling sedikit.
5.	Ingat	$873 / 38.761 * 100 = 2,25 \%$	Kata dasar “Ingat” mendapatkan jumlah persentase sebesar 2,25 %. Bobot ini diperoleh berdasarkan hasil persentase dari keseluruhan kata (data bersih) dalam <i>al-qur'an</i> . Kata ini mendapatkan jumlah kedua terbanyak setelah kata dasar “Cipta”.
6.	Lihat	$318 / 38.761 * 100 = 0,82 \%$	Kata dasar yang berhubungan dengan indera penglihatan manusia ini mendapatkan hasil sejumlah 0,82 %. Keseluruhan bobot sama saja seperti kata dasar yang lainnya yaitu dikalkulasikan dengan jumlah data bersih dalam <i>al-qur'an</i> yang peneliti peroleh dengan aplikasi <i>orange</i> yaitu sebanyak 38.761 kata bersih. Kemudian dipersentasekan dengan

			setiap kata dasar, termasuk kata dasar “Lihat”.
--	--	--	---

Berdasarkan jumlah bobot dan persentase maka setiap kata-kata yang berafiliasi dengan pendidikan yang telah dijabarkan diatas, maka dapatlah diambil kesimpulan bahwa setiap kata memiliki jumlah persentase masing-masing. Ada yang lebih banyak dan ada juga yang lebih sedikit. Dari keenam kata tersebut kata “cipta” mendapatkan skor persentase yang paling banyak dengan mendapatkan skor 3,05 %. Ini adalah yang tertinggi dari keenam kata lainnya. Kemudian disusul oleh kata “Ingat” dengan jumlah skor 2,25 %, kata “Ajar” 1,96 %, kata “Lihat” 0,82 %, kata “Dengar” 0,62 % dan kata dasar “Bicara” yang mendapat peringkat yang paling sedikit yaitu 0,34%.

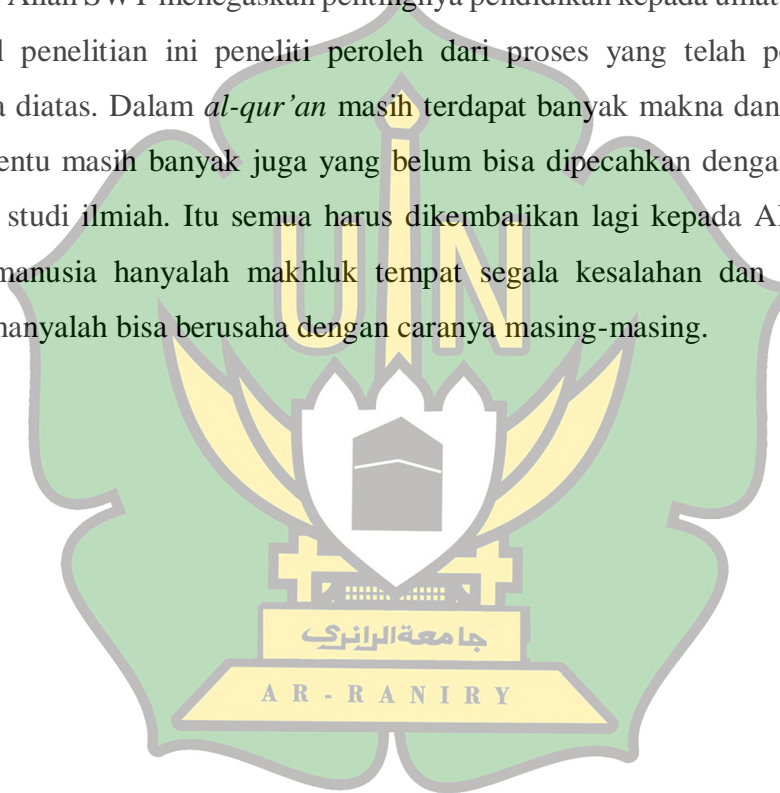
Untuk persentase keseluruhan kata atau keenam kata-kata yang berafiliasi dengan topik pendidikan dalam *al-qur'an* yang telah disebutkan diatas dapat dilihat pada tabel di bawah (tabel 4.5) ini yang juga disertakan dengan persentasenya. Berikut penjelasannya di bawah ini.

Tabel 4.5 Tabel persentase kelima kata yang berafiliasi dengan pendidikan dalam *al-qur'an*

No.	Kata dan Jumlah Persentase setiap kata yang berafiliasi dengan topik Pendidikan dalam Al-Qur'an			
	Item	Bobot	Persentase	Persentase Keseluruhan
1.	Ajar	763	$763 / 3.516 * 100$	1,96 %
2.	Bicara	134	$134 / 3.516 * 100$	0,34 %
3.	Cipta	1.184	$1184 / 3.516 * 100$	3,05 %
4.	Dengar	244	$244 / 3.516 * 100$	0,62 %
5.	Ingat	873	$873 / 3.516 * 100$	2,25 %
6.	Lihat	318	$318 / 3.516 * 100$	0,82 %
Total		3.516	$3.516 / 38.761 * 100$	9,04%

Hasil 9,04 % didapat dari hasil persentase keenam kata-kata yang berafiliasi dengan pendidikan dalam *al-qur'an* (berdasarkan data bersih yang diperoleh dari aplikasi *orange*). 9,04 % adalah persentase dari sejumlah 38.761 kata dalam *al-qur'an*, sebagaimana yang telah didapat dari hasil proses dengan menggunakan aplikasi *orange canvas data mining*.

Hasil yang diperoleh dapat dikatakan besar dengan mendapatkan hasil sebesar 9,04 % total dari data bersih dalam *al-qur'an*. Hampir mencapai 10 % (9,04 %) dalam *al-qur'an* Allah SWT menegaskan pentingnya pendidikan kepada umat manusia. Tentu saja hasil penelitian ini peneliti peroleh dari proses yang telah peneliti jelaskan semuanya diatas. Dalam *al-qur'an* masih terdapat banyak makna dan kata-kata yang tersirat, tentu masih banyak juga yang belum bisa dipecahkan dengan berbagai cara termasuk studi ilmiah. Itu semua harus dikembalikan lagi kepada Allah SWT. Kita sebagai manusia hanyalah makhluk tempat segala kesalahan dan keburukan dan manusia hanyalah bisa berusaha dengan caranya masing-masing.



BAB V PENUTUP

5.1. Kesimpulan

Dari pembahasan yang telah peneliti bahas pada bab-bab sebelumnya, dapat disimpulkan bahwa Analisa Topik Pendidikan dalam *Al-Qur'an* dengan Pendekatan *Text Mining* untuk mendapatkan hasil dan menggunakan algoritma *naïve bayes clasifier* serta aplikasi *orange canvas data mining* sebagai aplikasi pembantu untuk proses penelitian pada *al-qur'an*, yang merupakan tujuan dari pada penelitian ini telah tercapai. Dengan menghasilkan sejumlah analisa data, hitungan persentase data, mendapatkan hasil dan bentuk visualisasi data menyatakan bahwa penelitian telah berhasil.

Beberapa hasil yang berhasil diperoleh adalah sebagai berikut.

- Peneliti berhasil memperoleh hasil sebesar 9,04 % kata-kata yang berafiliasi dengan Pendidikan dari keseluruhan isi dalam Al-Qur'an.
- Bobot seluruh kata yang berafiliasi dengan Pendidikan (3.516) dipersentasekan dengan bobot total isi *al-qur'an* (38.761) sehingga mendapatkan hasil sebesar 9,04%.
- Adapun hasil dari keenam kata dasar atau kata-kata kunci tersebut dijabarkan sebagai berikut. Kata pertama adalah “cipta” mendapatkan skor persentase yang paling banyak dengan mendapatkan skor 3,05 %. Ini adalah yang tertinggi dari keenam kata lainnya. Kemudian disusul oleh kata “Ingat” dengan jumlah skor 2,25 %, kata “Ajar” 1,96 %, kata “Lihat” 0,82 %, kata “Dengar” 0,62 % dan kata dasar “Bicara” yang mendapat peringkat yang paling sedikit yaitu 0,34%.
- Sebagai metode evaluasi keberhasilan, peneliti mengambil dua (2) metode yaitu *recall* dan *precision*. *Recall* berfungsi sebagai metode evaluasi untuk membandingkan data yang relevan untuk dikeluarkan kepada user

sedangkan untuk *precision* adalah metode untuk mengevaluasi tingkat keberhasilan pengeluaran data yang tidak relevan untuk dikeluarkan.

- Hasil untuk metode Evaluasi *Recall* mengeluar hasil *score* sebesar 0.604 dan *Precision* mengeluarkan hasil *score* sebesar 0.366.

5.2. Saran

Ini adalah penelitian perdana dan pemula tentang penerapan *text mining* (sains) pada *al-qur'an* dilingkungan Universitas Islam Negeri Ar-Raniry Banda Aceh. Diharapkan kedepannya ada lagi yang akan meneruskan penelitian ini dengan tahapan yang lebih luas. Jika penelitian ini hanya terpaku pada beberapa kata kunci mungkin di masa depan ada lagi yang lain dengan konsep yang berbeda dan juga lebih bagus.

Misalnya dengan menggunakan kalimat atau frasa sebagai objek penelitian. Harapan ini diharapkan agar ada perkembangan tulisan ilmiah dilingkungan Universitas Islam Negeri Ar-Raniry Banda Aceh terkhusus adalah mahasiswa Prodi Pendidikan Teknologi Informasi dan Teknologi Informasi yang memiliki kemampuan di bidang teknologi dan sains.



DAFTAR PUSTAKA

- [1] D. Firdaus, "Penggunaan Data Mining dalam Kegiatan Sistem Pembelajaran Berbantuan Komputer," *J. Format*, vol. 6, no. 2, pp. 91–97, 2017.
- [2] A. M. Rusli, "Penerapan Data Mining Dengan Algoritma C4.5 Untuk Kalkulasi Data Dalam Al-Qur'an Terjemahan," *Klik - Kumpul. J. Ilmu Komput.*, vol. 6, no. 2, p. 152, 2019.
- [3] E. Supriyati and M. Iqbal, "Pengukuran Similarity Tema Pada Juz 30 Al Qur'an Menggunakan Teks Klasifikasi," *Simetris J. Tek. Mesin, Elektro dan Ilmu Komput.*, vol. 9, no. 1, pp. 361–370, 2018.
- [4] M. Alhawat, M. Hegazi, and A. Hilal, "Processing the Text of the Holy Quran: a Text Mining Study," *Int. J. Adv. Comput. Sci. Appl.*, vol. 6, no. 2, 2015.
- [5] C. S. Algorithm *et al.*, "Penerapan Algoritma Cosine Similarity pada Text Mining Terjemah Al- Qur ' an Berdasarkan Keterkaitan Topik," vol. 22, no. 1, pp. 41–50, 2019.
- [6] M. Kom, "Implementasi Text Mining dengan Metode Naive Bayes pada Tafsir Alquran PROPOSAL," 2017.
- [7] S. Asyafiq, "Berbagai pendekatan dalam pendidikan nilai dan pendidikan kewarganegaraan," *J. Dimens. Pendidik. dan Pembelajaran*, vol. 5, no. 1, pp. 29–37, 2016.
- [8] S. wahidin, unang; Ahmad, "Jurnal Edukasi Islami Jurnal Pendidikan Islam Vol . 07 / No . 1 , Teori-Teori Pendidikan ... Teori-Teori Pendidikan ...," *J. Pendidik. Islam*, vol. 07, no. 1, pp. 47–66, 2018.
- [9] Ahmad Islahud Daroini, "Tafsir Ayat Pendidikan Dalam Q.S. Al-'Alaq Ayat 1-5 Menurut Quraish Shihab," UIN Raden Intan Lampung, 2018.
- [10] A. S. Abu Bakar, "Pendidikan Dalam Al-Qur'an Kajian Pada Qs. Al-Rahman 1-4," *Inspiratif Pendidik.*, vol. 7, no. 2, p. 281, 2018.
- [11] S. Chua and P. N. E. Nohuddin, "Relationship analysis of keyword and chapter

- in Malay-translated tafseer of Al-Quran,” *J. Telecommun. Electron. Comput. Eng.*, vol. 9, no. 2–10, pp. 185–189, 2017.
- [12] R. I. E. Saragih and H. Sembiring, “Penerapan Algoritma Apriori Data Mining Untuk Mengetahui Kecurangan Skripsi,” *J. Tek. Inform.*, no. November, 2017.
- [13] E. Learning *et al.*, “Jurnal Antartika,” vol. 08, 2018.
- [14] G. F. Dimas, “Klasifikasi Dokumen Terjemahan Menggunakan Algoritma Naïve Bayes Pada Al- Qur ’ an Versi Berbahasa Indonesia (Juz 1 Sampai Juz 4) Klasifikasi Dokumen Terjemahan Menggunakan Algoritma Naïve Bayes Pada Al- Qur ’ an Versi Berbahasa Indonesia (Juz 1 Sampai,” pp. 1–112, 2018.
- [15] R. Y. Hayuningtyas, “Penerapan Algoritma Naïve Bayes untuk Rekomendasi Pakaian Wanita,” *J. Inform.*, vol. 6, no. 1, pp. 18–22, 2019.
- [16] M. R. Daulay, “Studi Pendekatan Alquran,” *J. Thariqah Ilm.*, vol. 01, no. 01, pp. 31–45, 2014.
- [17] N. M. Ud, “AL-Qur’an Sebagai Sumber Dan Ideologi Pendidikan Islam,” *J. AL-Muta’alimah STAI Darul Kamal NW Kembang Karang*, vol. 1, no. 1, pp. 140–161, 2017.
- [18] J. Teknik and T. Informasi, “Peningkatan Nilai Recall dan Precision pada Penelusuran Informasi Pustaka Berbasis Semantik (Studi Kasus : Sistem Informasi Ruang Referensi,” pp. 9–10, 2015.
- [19] E. Musvida, “Recall dan precision pada sistem temu kembali informasi bidang ilmu fikih di perpustakaan uin ar-raniry banda aceh,” 2017.
- [20] C. Narbuko and Abu achmadi, “Teori Metodologi Penelitian,” *Teor. Metodol. Penelit.*, 2011.
- [21] D. Wahyudi, T. Susyanto, and D. Nugroho, “Implementasi Dan Analisis Algoritma Stemming Nazief & Adriani Dan Porter Pada Dokumen Berbahasa Indonesia,” *J. Ilm. SINUS*, vol. 15, no. 2, pp. 49–56, 2017.
- [22] Y. L. Sabda, “Kamus Sabda,” 2012. [Online]. Available: [https://kamus.sabda.org/kamus/Ingat+/. \[Accessed: 09-Apr-2020\].](https://kamus.sabda.org/kamus/Ingat+/)
- [23] U. Ridlo, “Sinonim Dan Antonim Dalam Al-Quran,” *J. Al Bayan J. Jur.*

Pendidik. Bhs. Arab, vol. 9, no. 2, 2018.

- [24] KBBI, “kamus Besar Bahasa Indonesia.” [Online]. Available: <https://kbbi.web.id/sinonim>. [Accessed: 09-Apr-2020].
- [25] T. Inora, “PEMBELAJARAN SISTEM INDRA BERBASIS QUR’ANI DENGAN MENGGUNAKAN BULLETIN BOARD UNTUK MENINGKATKAN HASIL BELAJAR SISWA DI KELAS XI MIPA SMAN 5 BANDA ACEH,” Universitas Islam Negeri Ar-Raniry, 2018.
- [26] M. Makmun abha, “Pendidikan Dalam Perspektif Al-Qur’an Muhamad,” *J. Al-Afkar*, vol. II, no. II, pp. 23–36, 2013.



Lampiran I

Oleh karena tidak dapat disebutkan secara keseluruhan setiap item kata dan karakter yang akan di buang (*filter*) pada data, maka peneliti memuat sebuah link tautan yang berisikan setiap kata-kata yang telah di *filter* atau di buang pada data. Berikut Link aksesnya di sertakan bawah ini.

[https://drive.google.com/file/d/12dhwd19-LHT4VPtjACc7QTAiK-gLCiP3/view?usp=sharing.](https://drive.google.com/file/d/12dhwd19-LHT4VPtjACc7QTAiK-gLCiP3/view?usp=sharing)

Tujuan dari pada peneliti yang memberikan akses melalui Link tersebut agar para pembaca dapat melihat secara langsung semua kata-kata yang harus dibuang (*stopwords*) dalam data.

