

Analisa Topik Pendidikan Dalam Al-Quran dengan Pendekatan *Text Mining*

Bustami Yusuf^{1*}, Muhammad Zaeki², Hendri Ahmadian³, Khairan Ar⁴, Sri Wahyuni⁵

^{1,3,4,5}Prodi Teknologi Informasi, Fakultas Sains dan Teknologi, UIN Ar-Raniry Banda Aceh

²Prodi Pendidikan Teknologi Informasi, Fakultas Tarbiyah dan Keguruan, UIN Ar-Raniry Banda Aceh

*Koresponden email: bustamiyusof@ar-raniry.ac.id

Diterima: 17 Desember 2020

Disetujui: 26 Desember 2020

Abstract

Education is one of the sciences that makes humans much better by learning various scientific disciplines. Al-Quran is one of the sources of knowledge that is believed by Muslims around the world. Because technology has penetrated almost every domain of our lives, including the world of education. Thus, the authors make technology as tool for researching educational topics in Al-Quran by implementing text exploration. The research was carried out by making some basic words that were related to the subject of education as the keywords in this study. The keywords are "Ajar", "Bicara", "Cipta", "Dengar", "Ingat" and "Lihat". Then, the authors implemented the Naïve Bayes Classifier algorithm. To test and evaluate the results, the author used two methods, *i.e.* recall and precision. The study results are the keyword "cipta" by 3.05 %, "Ingat" 2.25 %, "Ajar" 1.96 %, "Lihat" 0.82 %, finally "Dengar" 0.62% and "Bicara" 0.34% with total weight of 3,516 words that have been filtered. The overall percentage of the results is 9.04% of the total number of words 38,761 in the Al-Quran. For the Naïve Bayes algorithm evaluation method, the recall and precision scores are 0.605 and 0.366, respectively.

Keywords: *text mining, education, Al-Quran, naïve bayes, recall, precision*

Abstrak

Pendidikan merupakan salah satu ilmu untuk menjadikan manusia lebih baik dengan belajar berbagai disiplin ilmu pengetahuan. Al-Quran menjadi salah satu sumber ilmu pengetahuan yang diyakini oleh umat muslim di seluruh dunia. Oleh karena teknologi sudah merambah ke seluruh aspek kehidupan termasuk ke dalam dunia pendidikan, penulis menjadikan teknologi sebagai bahan untuk meneliti topik pendidikan di dalam *Al-Quran* dengan mengimplementasikan *text mining*. Penelitian dilakukan dengan menjadikan *Al-Quran* sebagai objek penelitian dan menjadikan beberapa kata dasar yang dianggap memiliki hubungan dengan topik pendidikan sebagai kata kunci pada penelitian ini. Kata kuncinya adalah "Ajar", "Bicara", "Cipta", "Dengar", "Ingat" dan "Lihat". Kemudian penulis mengimplementasikan algoritma Naïve Bayes Classifier. Untuk menguji dan sebagai evaluasi hasil penulis menggunakan dua Metode yaitu recall dan precision. Hasil yang diperoleh dari penelitian ini adalah pertama kata kunci "cipta" 3,05 %, kemudian "Ingat" 2,25 %, selanjutnya "Ajar" 1,96 %, lalu "Lihat" 0,82 %, terakhir "Dengar" 0,62 % serta "Bicara" 0,34% dengan total bobot kata sebanyak 3.516 kata yang di filter. Hasil persentase keseluruhan adalah 9,04 % dari total 38.761 kata bersih dalam *Al-Quran*. Untuk metode evaluasi pada algoritma Naïve Bayes, recall mendapat *score* sebanyak 0,605 dan *precision* sebanyak 0,366.

Kata Kunci: *text mining, pendidikan, Al-Quran, naïve bayes, recall, precision*

1. Pendahuluan

Data mining mengacu pada proses untuk menambang (*mining*) pengetahuan dari sekumpulan data yang sangat besar dan dapat diterapkan pada setiap bidang [1]. Saat ini, penerapan teknologi *data mining* memiliki 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-Quran*. Contoh implementasi *data mining* pada *Al-Quran* adalah untuk perulangan kata dan mengkalkulasikan kata dalam *Al-Quran* [2], klasifikasi kata [3], proses teks dan mengukur frekuensi kata pada *Al-Quran* [4], dan menganalisa topik tertentu termasuk topik pendidikan dalam *Al-Quran* secara ilmiah, logis, dan akurat [5].

Mempelajari topik pendidikan dalam *Al-Quran* dapat dilakukan dengan pendekatan *text mining*. *Text Mining* merupakan bagian dari *data mining* yang mana proses kerjanya adalah fokus pada dokumen

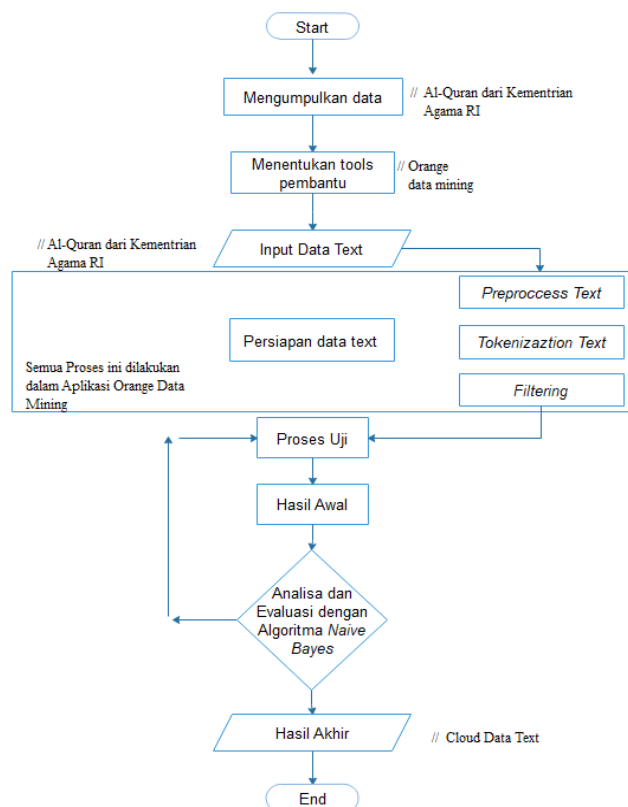
berbentuk teks. Tujuan utamanya adalah untuk menghasilkan kata kunci dan keterkaitan antar kata dalam dokumen. *Text Mining* yang diimplementasikan pada *Al-Quran* lebih dikenal dengan istilah *quran mining* [6]. Agar pencarian kata kunci pendidikan pada *Al-Quran* dapat dilacak (*tracer*) dengan baik, penulis menggunakan algoritma klasifikasi *naive bayes*, yang merupakan salah satu algoritma andal dalam proses klasifikasi kata.

Analisa kata-kata yang berafiliasi dengan topik pendidikan dalam *Al-Quran* 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-qu'ran* 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-Quran*. Hal ini disebabkan oleh banyaknya kata-kata lain dalam *Al-Quran* yang berafiliasi dengan topik pendidikan dalam *Al-Quran*, 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-Quran* 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-Quran* untuk dijadikan sebagai “*keyword*” dan kemudian dilakukan analisis kata tersebut yang berafiliasi dengan topik pendidikan [11].

Banyaknya perbendaharaan kata yang berkonotasi dengan istilah pendidikan pada *Al-Quran* menjadi tantangan tersendiri bagi penulis untuk menganalisis teks *Al-Quran* tersebut dengan cara mengimplementasikan metode *text mining* pada pencarian kata pendidikan pada *Al-Quran*. Hal tersebut menjadi tujuan utama penulis, sehingga penulis mengambil langkah-langkah analisis dengan 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 *naives bayes*, seperti yang telah penulis sampaikan sebelumnya.

2. Metode Penelitian

Pada penelitian ini metodologi yang akan digunakan adalah metodologi penelitian kuantitatif. Langkah-langkah penelitiannya digambarkan dengan diagram alir seperti yang terlihat pada **Gambar 1**.



Gambar 1. Tahapan penelitian

Proses Pengumpulan Data

Penelitian ini menggunakan *file* data *Al-Quran* 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/>.

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.

Penginputan Data

Data *Al-Quran* 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-Quran* adalah 134 surat. Data yang akan di input juga nantinya adalah dengan *import folder* dalam hal ini isi *folder* tersebut adalah semua data dari 134 surat yang telah disebutkan sebelumnya.

Persiapan data

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

a. Preprocessing

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

b. 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 di dalam sistem aplikasi akan diubah menjadi 3 token yaitu “saya”, “adalah”, “mahasiswa” [3].

c. 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. Contoh proses filtering dalam sebuah kalimat adalah sebagai berikut: “saya sering baca quran di teras rumah pada saat sore hari”

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

e. 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 daripada kata-kata utama (penting).

f. 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*) maupun akhiran (*sufiks*). Sebagai contoh cara kerjanya adalah sebagai berikut. Kata “Bersama”, “Kebersamaan”, “Menyamai” akan dianggap dan dijadikan menjadi satu kata dasar yaitu “Sama” [13].

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 [13].

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.

Alasan pemilihan beberapa kata yang telah dipaparkan pada tabel di atas adalah karena dilatarbelakangi 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 Drs. Saifuddin Mahmud, M.Pd, ahli bahasa Indonesia Universitas Syiah Kuala Banda Aceh.

g. Sinonim Kata

Sinonim merupakan istilah yang berasal dari Bahasa Yunani Kuno yang terdiri dari kata *syn* yang berarti “dengan” dan kata *anima* yang berarti “nama”. Jadi jika diartikan kedua kata tersebut memiliki makna yang berarti “Nama lain untuk benda yang sama” [14].

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.

h. Ayat Yang Mengandung

Penelitian ini memiliki keterkaitan dan implementasinya pada kitab suci *Al-Quran*. Oleh karena itu setiap item (kata kunci) pada penelitian ini juga akan di kaitkan dengan ayat-ayat yang ada dalam *Al-Quran*. Ada sekitar enam item yang dipakai pada penelitian ini. Adapun beberapa ayat yang terkandung adalah *Ar-Rahman* Ayat 2, 3 dan 4. *Nahl* 78 dan *Al-Baqarah* ayat 30.

Pengujian

Data yang telah dipersiapkan yang dimulai dari *Tokenization* sampai *Filtering* maka telah siap untuk dilakukan pengujian dengan mengimplementasikan *text mining* dengan menggunakan metode klasifikasi algoritma *naïve bayes*. Setelah mengeluarkan hasil dari pengujian dan penerapan *text mining* maka barulah masuk ke tahap analisa. Jika tidak mengeluarkan hasil yang diharapkan pada saat pengujian maka proses akan diulangi.

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 analisis untuk menilai keberhasilan dan mengambil keputusan untuk menetapkan hasil akhir dari penelitian.

Analisa dan Evaluasi dengan Algoritma Naïve Bayes

Pada proses analisa akan dilakukan proses penjabaran dari 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-Quran*. Sehingga hasil yang harapkan dapat dipertanggungjawabkan 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.

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 analisis persentase dan bentuk visualisasi dengan *cloud data text* diharapkan benar-benar dapat dipertanggungjawabkan dan dapat mewakili isi dari *Al-Quran* sesuai dengan topik penelitian yaitu pendidikan.

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*, *entropy* 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 [15].

Berikut adalah rumus dari *recall* [16].

$$\text{Recall} = \frac{\text{Jumlah dokumen relevan yang di temukan}}{\text{Jumlah semua dokumen relevan didalam koleksi}} \times 100\% \dots \dots \dots (1)$$

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 [15]. Berikut dibawah ini adalah rumus dari *precision* [16].

$$\text{Precision} = \frac{\text{Jumlah dokumen relevan yang di temukan}}{\text{Jumlah semua dokumen yang ditemukan}} \times 100\% \dots \dots \dots (2)$$

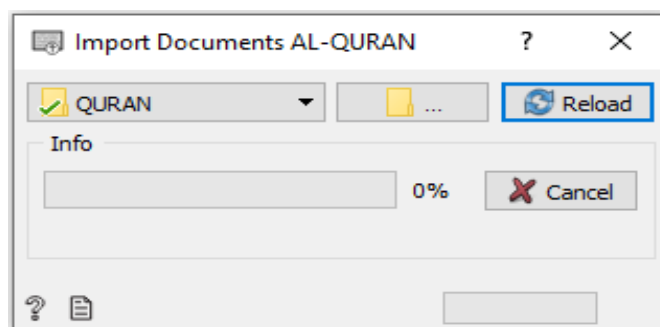
3. Hasil dan Pembahasan

Penelitian ini menggunakan *Al-Quran* digital yang diperoleh dari Website resmi Kementerian Agama RI. Untuk keperluan penelitian, penulis hanya mengambil *Al-Quran* terjemahan Bahasa Indonesia saja. Selain itu, penulis juga menggunakan aplikasi pendukung untuk menyelesaikan penelitian ini yaitu aplikasi *orange canvas data mining*.

Data *Al-Quran* yang telah penulis siapkan, 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.

Proses Input Data

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



Gambar 2. Import data Al-Quran ke aplikasi orange
Sumber: Hasil penelitian, 2020.

Input data berupa sebuah folder yang berisikan data file *al-quran*. Al-Quran 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-Quran*. Data file yang *al-quran* dipisah menjadi 30 bagian, selanjutnya diinputkan pada Aplikasi *Orange Canvas Data Mining*. Setelah proses input selesai maka barulah melanjutkan ke proses selanjutnya.

Proses Filtering

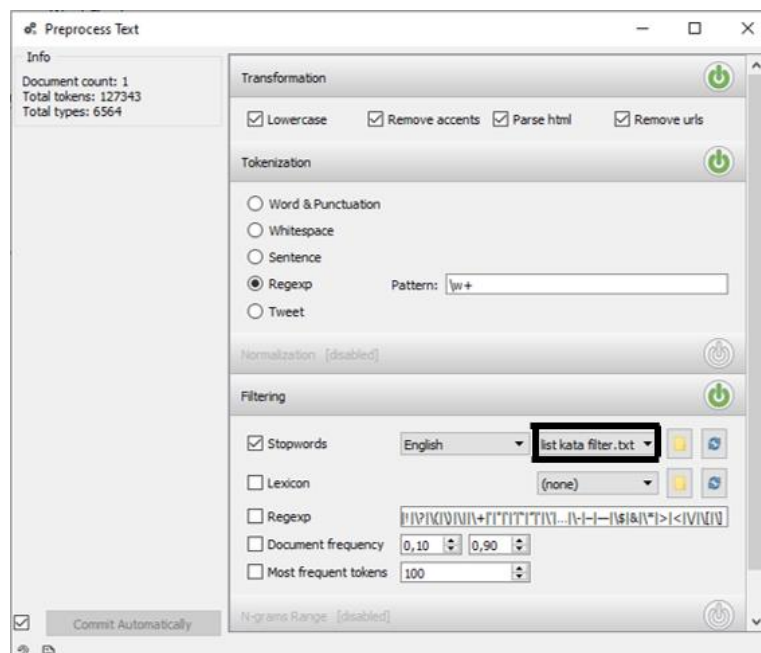
Proses *filtering* adalah *preprocess* atau hal yang dilakukan pada data sebelum diambil hasil atau menambahkan proses selanjutnya. Adapun proses *filter* yang dilakukan adalah membuang kata-kata yang tidak perlu, tanda baca, tidak membedakan huruf besar dan kecil, link web, angka dan lain sebagainya (*stop words*). Adapun beberapa kata yang dibuang dapat dilihat pada **Tabel 1**.

Tabel 1. Tabel *stopwords* pada data

No.	Kata	Angka
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

Sumber: Hasil penelitian, 2020.

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. Tujuan *filtering* kata karena kata-kata dan angka yang telah dijabarkan pada **Tabel 1** dan juga pada lampiran tidaklah memiliki makna yang mendalam, hanya sebagai kata sambung, kata petunjuk, numerik dan lain sebagainya.



Gambar 3. Proses *filtering* pada aplikasi *orange*

Sumber: Hasil penelitian, 2020

Berdasarkan **Gambar 3** dapat dijabarkan item-item yang di *filter*, antara lain adalah *lowercase*, *remove accent*, *parse html*, *remove url*. Adapun pengertiannya adalah sebagai berikut.

a. Filtering

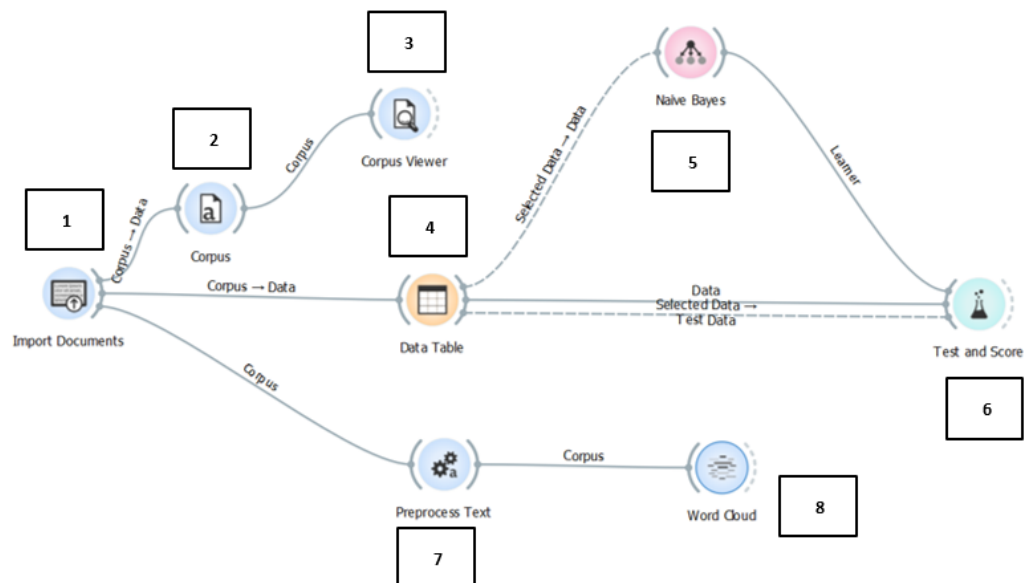
Filtering dilakukan dengan memasukkan kata-kata secara manual. Kata-kata yang telah dijabarkan dalam **Tabel 1** diinput ke aplikasi *orange canvas data mining*. Beberapa kata yang telah dijabarkan pada **Tabel 1** diketikkan pada *text editor* (pada penelitian ini peneliti menggunakan aplikasi *notepad*) dengan format *txt (text)*. Pada **Gambar 3** peneliti melingkari dengan garis warna hitam, yaitu sebagai langkah untuk melakukan input data kata-kata yang akan dibuang pada data.

Kata-kata yang telah diketikkan pada text editor notepad yang kemudian di input ke aplikasi *orange canvas data mining*. Sehingga aplikasi akan membaca kata-kata yang telah di input pada data nantinya, kata-kata yang telah di input tidak akan lagi dimunculkan pada saat mengeluarkan hasil nantinya.

Kata-kata yang telah diketikkan pada text editor notepad yang kemudian di input ke aplikasi *orange canvas data mining*. Sehingga aplikasi akan membaca kata-kata yang telah di input pada data nantinya, kata-kata yang telah di input tidak akan lagi dimunculkan pada saat mengeluarkan hasil nantinya.

b. Bentuk Kerangka Proses Menggunakan Aplikasi Orange

Setiap proses yang telah dijelaskan diatas, dimulai dari input data sampai hasil sebelumnya telah dijelaskan bahwa seluruh proses dilakukan dengan menggunakan aplikasi *orange canvas data mining*. Berikut dibawah ini peneliti lampirkan gambar kerangka proses yang dilakukan pada aplikasi *orange*.



Gambar 4. Bentuk rangkaian proses pada aplikasi *Orange*

Sumber: Hasil penelitian, 2020.

c. 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** 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 skor 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 5** dan hasil yang diambil hanyalah dua metode saja yaitu

Tabel 3. Tabel persentase setiap bobot kata serta penjabaran

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 Total 763	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.
2.	Bicara	<ul style="list-style-type: none"> - Perkataan = 70 - Berita = 64 Total 134	Beberapa kata-kata yang peneliti peroleh yang didasari pada kata dasar “Bicara” adalah sebanyak 134 Kata. Kata yang diperoleh adalah kata yang memiliki persamaan kata (sinonim) dengan kata dasar “bicara”.
3.	Cipta	<ul style="list-style-type: none"> - Jadikan = 100 - Menjadikan = 228 - Berbuat = 260 - Menciptakan = 144 - Kehendaki = 151 - Kerjakan = 181 - Melaksanakan = 59 - Tangan = 61 Total 1.184	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 - Berita = 64 - Mendengar = 118 Total 244	Perolehan hasil yang didapat pada kata dasar “Dengar” adalah sebanyak 244 kata saja.
5.	Ingat	<ul style="list-style-type: none"> - Ketahuilah = 64 - Perintah = 65 - Bertakwalah = 71 - Bertakwa = 139 - Kehendaki = 151 - Peringatan = 178 - Ingatlah = 205 Total 873	“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.
6.	Lihat	<ul style="list-style-type: none"> - Ketahuilah = 64 - Mata = 69 - Saksi = 77 - Berpaling = 108 Total 318	Kata dasar “Lihat” yang memiliki hubungan dengan panca indera penglihatan manusia ini mengeluarkan hasil sebanyak 318 kata.

Sumber: Hasil penelitian, 2020.

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-Quran*. Jumlah kata bersih yang diperoleh dari *Al-Quran* adalah sebanyak 38.761 kata dan jumlah ini yang akan dikalkulasikan dengan jumlah bobot masing-masing kata dalam **Tabel 3**. Berikut hasil dan penjabaran persentase setiap kata tersebut pada **Tabel 4**.

Tabel 4. Tabel persentase setiap bobot kata serta penjabaran

No.	Kata	Persentase
1.	Ajar	$763 / 38.761 * 100 = 1,96 \%$
2.	Bicara	$134 / 38.761 * 100 = 0,34 \%$
3.	Cipta	$1.184 / 38.761 * 100 = 3,05 \%$
4.	Dengar	$244 / 38.761 * 100 = 0,62 \%$
5.	Ingat	$873 / 38.761 * 100 = 2,25 \%$
6.	Lihat	$318 / 38.761 * 100 = 0,82 \%$

Sumber: Hasil penelitian, 2020.

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%. Persentase keseluruhan kata atau keenam kata-kata yang berafiliasi dengan topik pendidikan dalam *Al-Quran* yang telah disebutkan diatas dapat dilihat pada **Tabel 4**.

Hasil 9,04 % didapat dari hasil persentase keenam kata-kata yang berafiliasi dengan pendidikan dalam *al-quran* (berdasarkan data bersih yang diperoleh dari aplikasi *orange*). Nilai 9,04 % adalah persentase dari sejumlah 38.761 kata dalam *al-quran*, 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*. Hasil ini menunjukkan pentingnya pendidikan kepada umat manusia. Dalam *Al-Quran* 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.

4. Kesimpulan

Dari pembahasan pada bagian sebelumnya, dapat disimpulkan bahwa sebanyak 9, 04 % total dari seluruh data bersih di dalam *al-quran* sesuai dengan peneliti dapatkan. Hasil dari 9, 04 % tersebut diperoleh dari enam kata yang peneliti jadikan sebagai kata kunci (kata-kata yang berafiliasi dengan pendidikan) pada penelitian ini. Kemudian 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 metode yaitu *recall* dan *precision*. *Recall* berfungsi sebagai metode evaluasi untuk mengeluarkan data yang relevan untuk dikeluarkan kepada user sedangkan untuk *precision* adalah metode untuk mengevaluasi tingkat keberhasilan pengeluaran data yang tidak relevan untuk dikeluarkan. Hasil metode evaluasi *recall* menunjukkan *score* sebesar 0,604 dan *precision* mengeluarkan hasil *score* sebesar 0,366.

5. Ucapan Terima Kasih

Penulis mengucapkan terima kasih pada Program Studi Pendidikan Teknologi Informasi UIN Ar-Raniry Banda Aceh atas kesempatan serta kerja sama yang diberikan kepada penulis dalam rangka penyelesaian penelitian ini. Serta ucapan terima kasih tak terhingga kepada pihak *orange biolab* (<https://orange.biolab.si/>) atas penggunaan aplikasi pengolahan teks yang penulis telah gunakan.

6. Daftar Pustaka

- [1] D. Firdaus, “Penggunaan Data Mining dalam Kegiatan Sistem Pembelajaran Berbantuan Komputer,” *J. Format*, **6** (2), pp. 91–97, 2017.
- [2] A. M. Rusli, “Penerapan Data Mining Dengan Algoritma C4.5 Untuk Kalkulasi Data Dalam Al-Quran Terjemahan,” *Klik - Kumpul. J. Ilmu Komput.*, **6** (2), p. 152, 2019, doi: 10.20527/klik.v6i2.214.
- [3] E. Supriyati, M. Iqbal, “Pengukuran Similarity Tema Pada Juz 30 Al Qur'an Menggunakan Teks Klasifikasi,” *Simetris J. Tek. Mesin, Elektro dan Ilmu Komput.*, **9** (1), pp. 361–370, 2018, doi: 10.24176/simet.v9i1.1955.
- [4] M. Alhawarat, M. Hegazi, A. Hilal, “Processing the Text of the Holy Quran: a Text Mining Study,” *Int. J. Adv. Comput. Sci. Appl.*, **6** (2), 2015, doi: 10.14569/ijacs.2015.060237.
- [5] C. S. Algorithm, “Penerapan Algoritma Cosine Similarity pada Text Mining Terjemah Al-Quran Berdasarkan Keterkaitan Topik,” **22** (1), pp. 41–50, 2019, doi: 10.18196/st.221235.
- [6] M. Kom, “Implementasi Text Mining dengan Metode Naive Bayes pada Tafsir Al-Quran PROPOSAL,” 2017.

- [7] S. Asyafiq, "Berbagai Pendekatan Dalam Pendidikan Nilai Dan Pendidikan Kewarganegaraan," *J. Dimens. Pendidik. dan Pembelajaran*, **5** (1), pp. 29–37, 2016, doi: 10.2426/dpp.v4i1.56.
- [8] S. Wahidin, Unang, Ahmad, "Media Pendidikan Dalam Perspektif Pendidikan Islam" *J. Pendidik. Islam*, vol. 07, no. 1, pp. 47–66, 2018.
- [9] A. I. 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-Quran Kajian Pada Qs. Al-Rahman 1-4," *Inspiratif Pendidik.*, **7** (2), p. 281, 2018, doi: 10.24252/ip.v7i2.7911.
- [11] S. Chua, P. N. E. Nohuddin, "Relationship analysis of keyword and chapter in Malay-translated tafseer of Al-Quran," *J. Telecommun. Electron. Comput. Eng.*, **9** (2-10), pp. 185–189, 2017.
- [12] C. Narbuko and Abu achmadi, "Teori Metodologi Penelitian," *Teor. Metodol. Penelit.*, 2011.
- [13] D. Wahyudi, T. Susyanto, D. Nugroho, "Implementasi Dan Analisis Algoritma Stemming Nazief & Adriani Dan Porter Pada Dokumen Berbahasa Indonesia," *J. Ilm. SINUS*, **15** (2), pp. 49–56, 2017, doi: 10.30646/sinus.v15i2.305.
- [14] U. Ridlo, "Sinonim Dan Antonim Dalam Al-Quran," *J. Al Bayan J. Jur. Pendidik. Bhs. Arab*, vol. 9, no. 2, 2018, doi: 10.24042/albayan.v9i2.2253.
- [15] K. D. P. Novianti, N. A. Setiawan, S. S. Kusumawardani, "Peningkatan Nilai Recall dan Precision pada Penelusuran Informasi Pustaka Berbasis Semantik (Studi Kasus: Sistem Informasi Ruang Referensi Jurusan Teknik Elektro dan Teknologi Informasi UGM)," *Proceeding Konferensi Nasional Sistem dan Informatika*, 2015.
- [16] E. Musvida, "Recall dan Precision Pada Sistem Temu Kembali Informasi Bidang Ilmu Fikih Di Perpustakaan UIN Ar-Raniry Banda Aceh," 2017.