

Pemberian Harakat Bahasa Arab Menggunakan Metode N-Gram dan C5.0

Ajib Hanani, Harry Soekotjo Dachlan, dan Purnomo Budi Santoso

Abstract – This research is about an Arabic diacritizer using N-Gram (Quad Gram) and C5.0. The input of application is an undiacritized Arabic sentence. Quad Gram and matching word to known patterns are used to diacritize the undiacritized Arabic sentence. C5.0 is used to simplify the rules for diacritizing the final diacritic mark of word. The output of application is a diacritized Arabic sentence based on morphological and syntactic rules. Then, the diacritized Arabic sentence is converted into speech and spoken by talking avatar using text to speech API.

The result of N-Gram (Quad Gram) and matching word to known patterns gives an accuracy of 96% based on morphological rules. The result of C5.0 gives an accuracy of 94% based on syntactic rules. Finally, the result of integrating N-Gram (Gram Quad) and C5.0 gives an accuracy of 90% based on morphological and syntactic rules.

Index Terms — Arabic Diacritics, C5.0, N-Gram, Text to Speech.

Abstrak – Penelitian ini bertujuan untuk membangun aplikasi pemberian harakat pada kalimat Bahasa Arab menggunakan N-Gram (Quad Gram) dan C5.0. Masukan berupa kalimat Bahasa Arab tanpa harakat. Quad Gram dan pengecekan bentuk kata digunakan untuk memberi harakat agar sesuai dengan kaidah *shorof*. C5.0 digunakan untuk menyederhanakan aturan dalam menentukan harakat akhir kata agar sesuai dengan kaidah *nahwu*. Keluaran dari aplikasi ini berupa kalimat Bahasa Arab berharakat yang sesuai dengan kaidah *shorof* dan *nahwu*. Kalimat tersebut kemudian dijadikan suara audio dan dapat diucapkan oleh animasi *talking avatar* menggunakan *text to speech API*.

Hasil pengujian sistem menunjukkan bahwa N-Gram (Quad Gram) dan pengecekan bentuk kata mampu memberi harakat sesuai kaidah *shorof* dengan tingkat akurasi 96%. C5.0 mampu menyederhanakan aturan berdasarkan data latih dan dapat memberi harakat akhir kata sesuai kaidah *nahwu* dengan tingkat akurasi 94%. Integrasi N-Gram (Quad Gram) dan C5.0 mampu memberi harakat sesuai kaidah *nahwu* dan *shorof* dengan tingkat akurasi 90%.

Kata Kunci — C5.0, Harakat Bahasa Arab, N-Gram, Text to Speech.

Ajib Hanani adalah Mahasiswa Program Magister Teknik Elektro, Fakultas Teknik, Universitas Brawijaya, Malang-Indonesia (email: xinformatika@gmail.com).

Harry Soekotjo Dachlan adalah Dosen Teknik Elektro, Universitas Brawijaya, Malang-Indonesia (email: harrysd@ub.ac.id).

Purnomo Budi Santoso adalah Dosen Teknik Industri, Fakultas Teknik, Universitas Brawijaya, Malang-Indonesia (email: budiakademika@yahoo.com).

I. PENDAHULUAN

BAHASA Arab adalah bahasa yang digunakan dalam Al-Qur'an, Al-Hadist, Tauhid, Fiqh dan berbagai sumber Agama Islam sehingga menguasai Bahasa Arab menjadi hal yang sangat penting dalam Agama Islam. Bahasa Arab merupakan bahasa yang sangat kompleks dan cukup rumit. Hal ini menjadi hambatan dalam mempelajarinya. Terlebih lagi media yang digunakan untuk pembelajaran Bahasa Arab juga masih terbatas. Salah satu hambatan dalam mempelajari Bahasa Arab adalah permasalahan pemberian harakat pada kalimat Bahasa Arab. Kesalahan dalam pemberian harakat akan berdampak pada perbedaan ucapan dan makna. Perbedaan makna ini berakibat pula pada pemahaman dalam menanggapi teks-teks berbahasa Arab. Dalam penulisan teks Bahasa Arab baik pada buku cetak maupun pada media elektronik, hampir semuanya menggunakan Bahasa Arab tanpa harakat. Bahasa Arab yang terkenal dengan kerumitan tata bahasa dan kekayaan kosa katanya menjadi tantangan tersendiri bagi masyarakat untuk menguasainya. Tidak hanya kalangan orang awam, bagi santri yang belajar Bahasa Arab di pondok pesantren pun mengalami kesulitan [1].

Bahasa Arab umumnya dipelajari dan digunakan oleh santri di pondok pesantren untuk membaca kitab kuning/kitab gundul. Disebut kitab kuning karena memang kitab-kitab tersebut kertasnya berwarna kuning dan warna kuning dianggap lebih nyaman serta mudah dibaca dalam keadaan yang redup ketika penerangan masih terbatas di masa lampau [2]. Selain dipelajari di pondok pesantren, Bahasa Arab juga dipelajari oleh mahasiswa di perguruan tinggi. Di Fakultas Syariah Universitas Islam Negeri Maulana Malik Ibrahim Malang terdapat matakuliah Praktikum *Qira'atul Kutub* yang merupakan program bimbingan membaca kitab dan matakuliah tersebut menjadi salah satu prasyarat ujian komprehensif [3]. Pada penerapannya, banyak mahasiswa yang kesulitan membaca kitab kuning.

Dari permasalahan kompleksitas Bahasa Arab yang cukup rumit dan media pembelajaran yang terbatas serta banyaknya Mahasiswa Fakultas Syariah UIN Maulana Malik Ibrahim Malang yang kesulitan mempelajari kitab kuning berbahasa Arab tanpa harakat, maka perlu adanya solusi untuk memudahkan dalam pembelajaran Bahasa Arab dengan membangun sebuah aplikasi pemberian harakat pada kalimat Bahasa Arab yang sesuai dengan kaidah *shorof* (perubahan bentuk kata)

dan kaidah *nahwu* (gramatika Bahasa Arab) menggunakan metode N-Gram (Quad Gram) dan Algoritma C5.0. Metode N-Gram (Quad Gram) yang dilanjutkan dengan pengecekan bentuk kata secara logika mampu untuk memberi harakat pada kalimat Bahasa Arab agar sesuai dengan kaidah *shorof*. Sedangkan Algoritma C5.0 mampu menyederhanakan aturan dan membentuk pohon keputusan dalam pemberian harakat akhir kata pada kalimat Bahasa Arab agar sesuai dengan kaidah *nahwu*.

Beberapa penelitian telah dilakukan terkait dengan Bahasa Arab, metode N-Gram dan C5.0, antara lain: Kategorisasi teks menggunakan N-Gram untuk dokumen berbahasa Indonesia [4]; *Automatic restoration of arabic diacritics: a simple, purely statistical* [5]; Pembetulan harakat akhir kata dalam kalimat Bahasa Arab menggunakan metode *decision tree* [6]; *A method for classification of network traffic based on C5.0 machine learning algorithm* [7]; *Pattern-based algorithm for part-of-speech tagging arabic text* [8].

Berdasarkan pembahasan pentingnya pembelajaran Bahasa Arab di atas, maka sangat perlu untuk melakukan penelitian dan pengembangan tentang pemberian harakat kalimat Bahasa Arab agar sesuai dengan kaidah *nahwu* dan *shorof*. Sistem ini akan menjadi alat bantu bagi mahasiswa, santri dan masyarakat dalam mempelajari Bahasa Arab.

II. LANDASAN TEORI

Landasan teori memuat teori tentang Bahasa Arab, *data mining*, N-Gram dan algoritma C5.0.

A. Bahasa Arab

Bahasa Arab termasuk rumpun Bahasa Semit yaitu bahasa yang dipakai bangsa-bangsa yang tinggal di sekitar Sungai Tigris dan Eufkrat, dataran Syria dan Jazirah Arabia (Timur Tengah) seperti Bahasa Finisia, Assyria, Ibrania, Arabia, Suryania dan Babilonia. Dari sekian banyak bahasa tersebut yang bertahan sampai sekarang adalah Ibrani.

Secara definisi ilmu *Nahwu* adalah ilmu yang mempelajari tentang jabatan kata dalam kalimat dan harakat akhirnya, baik secara *i'rab* (berubah) atau *bina* (tetap) [9].

Tashrif secara bahasa berarti merubah, secara istilah memindah hukum asal dari satu lafadz ke beberapa contoh lafadz yang berbeda-beda karena beberapa makna yang diinginkan [10].

B. Data Mining

Data mining adalah proses yang menggunakan teknik statistik, matematika, kecerdasan buatan dan *machine learning* untuk mengekstraksi dan mengidentifikasi informasi yang bermanfaat dan pengetahuan yang terkait dari berbagai *database* besar [11].

C. N-Gram

N-Gram adalah *substring* sepanjang n karakter dari sebuah *string*. Dalam definisi lain, N-Gram adalah potongan sejumlah n karakter dari sebuah *string*.

Sebagai contoh: kata "TEXT" dapat diuraikan ke dalam beberapa N-Gram berikut ("_" merepresentasikan *blank*):

- Uni-grams : T, E, X, T
- Bi-grams : _T, TE, EX, XT, T_
- Tri-grams : _TE, TEX, EXT, XT_
- Quad grams : _TEX, TEXT, EXT_
- Quint-grams : _TEXT, TEXT_

Salah satu keunggulan menggunakan N-Gram dan bukan suatu kata utuh secara keseluruhan adalah bahwa N-Gram tidak akan terlalu sensitif terhadap kesalahan penulisan yang terdapat pada suatu dokumen [12].

D. Algoritma C5.0

Algoritma C5.0 adalah salah satu algoritma klasifikasi *data mining* yang khususnya diterapkan pada teknik *decision tree*. C5.0 merupakan penyempurnaan algoritma sebelumnya yang dibentuk oleh Ross Quinlan pada tahun 1987, yaitu ID3 dan C4.5. Dalam algoritma ini pemilihan atribut yang akan diproses menggunakan *information gain*. Dalam memilih atribut untuk pemecah obyek dalam beberapa kelas harus dipilih atribut yang menghasilkan *information gain* paling besar. Atribut dengan nilai *information gain* tertinggi akan dipilih sebagai *parent* bagi *node* selanjutnya. Formula untuk *information gain* adalah [13]:

$$I(S_1, S_2, \dots, S_m) = - \sum_{i=1}^m p_i * \log_2(p_i) \quad (1)$$

Keterangan :

- S = himpunan kasus
- S_j = jumlah sampel
- p_i = proporsi kelas

Untuk mendapatkan informasi nilai *subset* dari atribut A tersebut maka digunakan formula di bawah ini :

$$E(A) = \sum_{j=1}^y \frac{S_{1j} + \dots + S_{mj}}{S} I(S_{1j}, \dots, S_{mj}) \quad (2)$$

Keterangan :

$\frac{S_{1j} + \dots + S_{mj}}{S}$ = jumlah *subset* j yang dibagi dengan jumlah sampel S.

Untuk mendapatkan nilai *gain*, digunakan formula di bawah ini :

$$Gain(A) = I(S_1, S_2, \dots, S_m) - E(A) \quad (3)$$

Keterangan :

- A = atribut
- S = himpunan kasus
- S_1 = jumlah sampel

III. METODE

Metode yang digunakan pada program aplikasi dapat dijelaskan sebagai berikut:

A. Koreksi Kata Bahasa Arab Tanpa Harakat

Sebelum kalimat Bahasa Arab diberi harakat, tiap kata pada kalimat dicek di *database* untuk dikoreksi bila ada kesalahan penulisan kata. Hal ini dilakukan agar tidak terjadi kesalahan dalam pemberian harakat.

B. Pemberian Harakat Kalimat Menggunakan Probabilitas Quad Gram (N-Gram)

Jenis N-Gram yang digunakan untuk memberi harakat kalimat Bahasa Arab adalah Quad Gram.

Teknik yang digunakan pada metode ini ada 2 tahapan. Tahap pertama adalah membuat daftar frekuensi Quad Gram (pola dari 4 huruf yang masing-masing huruf disertai dengan harakat) pada kalimat Bahasa Arab berharakat yang diambil dari teks Al-Qur'an dan Corpus Bahasa Arab *Maktabah Syamilah*.

Tahap kedua adalah menggunakan daftar tersebut untuk memberi harakat pada kalimat Bahasa Arab tanpa harakat. Spasi antar kata juga diperhitungkan dalam Quad Gram. Quad Gram dengan huruf yang sama tetapi harakatnya berbeda dikelompokkan. Probabilitas masing-masing kelompok akan dihitung dan Quad Gram dengan probabilitas yang paling tinggi dari masing-masing kelompok itulah yang akan dipilih. Probabilitas Quad Gram huruf dengan frekuensi paling tinggi dan kombinasi harakatnya dihitung dengan rumus:

$$P = \frac{F_h}{F_t} \quad (4)$$

dimana P adalah probabilitas huruf Quad Gram seperti "س ا ع د", F_h adalah frekuensi Quad Gram huruf dan harakat paling tinggi, sedangkan F_t adalah total frekuensi Quad Gram huruf dengan harakat yang berbeda.

Kalimat Bahasa Arab tanpa harakat yang dimasukkan akan diberi harakat menggunakan *database* Quad Gram. Misalnya dimasukkan $L_1, L_2, L_3, L_4, L_5, L_6, L_7, L_8$. Spasi juga dimasukkan ke dalam Quad Gram sehingga:

S, L_1, L_2, L_3
 L_1, L_2, L_3, L_4
 L_2, L_3, L_4, L_5
 L_3, L_4, L_5, L_6
 L_4, L_5, L_6, L_7
 L_5, L_6, L_7, L_8
 L_6, L_7, L_8, S

Setelah Quad Gram disusun, selanjutnya mencari Quad Gram yang sama dan yang sudah berharakat di *database*. Diasumsikan Quad Gram sudah ditemukan dan ditampilkan sebagai berikut:

S, $L_1, D_1, L_2, D_2, L_3, D_3, P_1$
 $L_1, D_4, L_2, D_5, L_3, D_6, L_4, D_7, P_2$
 $L_2, D_8, L_3, D_9, L_4, D_{10}, L_5, D_{11}, P_3$
 $L_3, D_{12}, L_4, D_{13}, L_5, D_{14}, L_6, D_{15}, P_4$
 $L_4, D_{16}, L_5, D_{17}, L_6, D_{18}, L_7, D_{19}, P_5$
 $L_5, D_{20}, L_6, D_{21}, L_7, D_{22}, L_8, D_{23}, P_6$
 $L_6, D_{24}, L_7, D_{25}, L_8, D_{26}, S, P_7$

dimana (P) adalah probabilitas pada masing-masing Quad Gram, (L) adalah *letter*/huruf dan (D) adalah *diacritic*/harakat pada masing-masing huruf.

Selanjutnya memberi harakat pada huruf satu persatu dengan menghitung semua Quad Gram dimana huruf tersebut berada. Pada huruf tertentu paling banyak memiliki 4 Quad Gram yang berharakat. Untuk masing-masing huruf di Quad Gram dihitung dengan menjumlahkan probabilitas huruf dengan harakat yang sama. Harakat pada huruf Quad Gram yang memiliki kelompok probabilitas paling tinggi dipilih untuk huruf tersebut. Alur pemilihan Quad Gram untuk masing-masing huruf disajikan pada gambar 1.

Jika suatu Quad Gram yang berharakat tidak tersedia

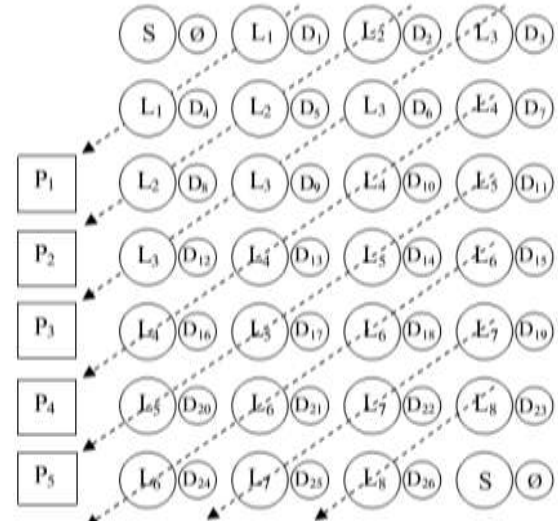
di *database*, data *dummy* akan dimasukkan di dalamnya. Sebagai contoh, jika urutan yang diharapkan adalah

ق ب ج د ن ل م ف ص ط

maka Quad Gram akan seperti:

ق ب ج د - ب ج د ن - ج د ن ل - د ن ل م - ن ل م ف -
 ل م ف ص - م ف ص ط

Jika Quad Gram kedua - ب ج د ن - tidak ada dalam *database*, sistem akan memasukkan Quad Gram kosong sebagai gantinya.



Gambar 1. Alur pemilihan Quad Gram dan harakat (D) untuk masing-masing huruf (L)

Rumus matematika untuk urutan yang sudah tersedia pada masing-masing huruf di atas adalah sebagai berikut:

- p_k adalah probabilitas k^{th} *sequence*.
- p_k^{dq} adalah probabilitas k^{th} *sequence* dengan q^{th} kandidat harakat yaitu d_q . Paling banyak 4 calon harakat untuk huruf (L) apapun.
- $P(G^{dq})$ adalah kumpulan probabilitas semua *sequence* dengan d_q . G^{dq} yang mewakili kelompok untuk d_q , yaitu semua *sequence* yang memiliki d_q .
- $P(G^{dq}) = \sum_{r=1}^t p_r^{dq}$ dimana t adalah nomor *sequence* yang memiliki d_q . (5)
- $I = \arg \max (P(G^{dq}))$ dimana n adalah nomor kandidat kelompok. (6)
 $q=1 \dots n$
- Harakat/*diacritic* yang dipilih (d_s) = d_I .

C. Pengecekan Kata Berharakat Dengan Bentuk Kata

Teknik yang digunakan pada metode ini ada 2 tahapan. Tahap pertama adalah membuat daftar bentuk kata dari frekuensi kata dari kalimat Bahasa Arab berharakat. yang diambil dari teks Al-Qur'an. Teks Al-Qur'an tersebut diuraikan tiap kata dan disimpan di tabel bentuk kata serta dihitung frekuensinya. Data *al-Qomus al-'Arobi* yang merupakan kamus berbagai macam bentuk kata dan kamus *Munawwir* juga dimasukkan ke tabel bentuk kata sebagai pelengkap.

Tahap kedua adalah menggunakan daftar tersebut untuk mengecek tiap kata pada kalimat berharakat hasil

Quad Gram. Kalimat berharakat hasil Quad Gram dipecah kata per kata dan disimpan dalam sebuah variabel *array*. Tiap kata tersebut dicocokkan dengan *database* untuk dicek apakah sesuai dengan bentuk katanya atau tidak. Jika kata berharakat hasil Quad Gram sesuai dengan yang ada dalam *database* ataupun dicek ternyata tidak ada di dalam *database*, maka kata berharakat hasil Quad Gram tersebut yang akan dipilih. Jika kata berharakat hasil Quad Gram tersebut dicek dan ternyata tidak sesuai dengan bentuk katanya, maka kata tersebut akan diganti dengan bentuk kata berharakat dengan jumlah frekuensi paling tinggi yang ada di dalam *database*. Setelah tiap kata dicek bentuk katanya, maka kata-kata tersebut disusun kembali seperti susunan kalimat semula.

D. Pemberian Harakat Akhir Kata Menggunakan Algoritma C5.0

Beberapa tahapan dalam proses ini antara lain adalah sebagai berikut:

- Proses Pemecahan Kalimat

Pada proses ini dilakukan pemecahan kata per kata dari suatu kalimat dan disimpan dalam sebuah variabel *array*.

- Identifikasi Kata

Setelah proses pemecahan kalimat menjadi per kata ke dalam variabel *array*, langkah selanjutnya adalah melakukan proses identifikasi kata. Dalam proses ini setiap kata akan diidentifikasi dengan diberikan kode unik sesuai dengan identitas masing-masing.

- Pemberian Harakat Akhir Kata

Setelah semua kata memiliki kode unik berdasarkan identitas kata masing-masing, tahap selanjutnya adalah pemberian harakat akhir kata. Berikut ini keterangan mengenai kode unik untuk susunan kalimat:

- KODE = kode yang terdapat pada kata itu sendiri
- KD1 = kode yang terdapat pada kata setelahnya
- KD2 = kode yang terdapat pada kata sebelumnya
- KD3 = kode yang terdapat pada kata dua langkah sebelumnya.

Untuk sampel susunan kalimat sebagai data latih disajikan dalam tabel I berikut ini.

TABEL I.
SAMPel SUSUNAN KALIMAT

ID	KD1	KODE	KD2	KD3	Harakat Akhir Kata
1	sf	al	al	hi	Dhammah
2	al	no	al	fa	Fathah
3	al	no	al	fm	Fathah
4	al	no	al	hi	Dhammah
5	al	no	al	no	Dhammah

sf = kata sifat, al = *isim alif lam*, hi = huruf *inna*, no = *isim* tanpa *alif lam*, fa = *fi'il madhi*, fm = *fi'il mudhori*

Berdasarkan susunan kalimat pada tabel I, akan dibuat pohon keputusan dan penyederhanaan aturan dalam menentukan harakat akhir kata agar sesuai dengan kaidah *nahwu* (gramatika Bahasa Arab) menggunakan C5.0 pada persamaan 1, 2 dan 3. Bagian kata yang diberi harakat akhir kata adalah pada kolom KODE (kata itu sendiri).

E. Suara Audio Menggunakan Google Text To Speech API

Kalimat Bahasa Arab berharakat yang sesuai dengan kaidah *nahwu* dan *shorof* dijadikan suara audio menggunakan Google *text to speech API*. Suara audio berfungsi untuk mengetahui bagaimana melafalkannya. Google *text to speech API* merupakan *API* Google yang mengubah teks menjadi suara audio secara gratis dan harus terhubung dengan internet [14]. Kalimat Bahasa Arab berharakat tersebut dikirim ke Server Google melalui Google *text to speech API* dan Server Google yang akan memproses teks menjadi suara audio. Hasil suara audio dikirim oleh Server Google ke program aplikasi dan disimpan dalam bentuk file audio mp3. File audio mp3 tersebut dapat di-*download* dan dapat didengarkan oleh pengguna.

F. Suara Audio Dan Animasi Talking Avatar Menggunakan Oddcast Text To Speech API

Kalimat Bahasa Arab berharakat yang sesuai dengan kaidah *nahwu* dan *shorof* juga dijadikan suara audio yang diucapkan oleh talking avatar menggunakan Oddcast *text to speech API*. Suara audio berfungsi untuk mengetahui bagaimana melafalkannya. Oddcast *text to speech API* merupakan *API* Oddcast yang mengubah teks menjadi suara audio yang diucapkan oleh *talking avatar*, harus terhubung dengan internet dan harus membayar atau *free trial* selama 15 hari melalui situs *sitepal.com* [15]. Peneliti menggunakan paket *silver* dan foto peneliti yang dijadikan sebagai *avatar*.

Kalimat Bahasa Arab berharakat tersebut dikirim ke Server Oddcast melalui Oddcast *text to speech API* dan Server Oddcast yang akan memproses teks menjadi suara audio. Server Oddcast mengirimkan *url* suara audio ke program aplikasi sehingga animasi *avatar* dapat berbicara dan mengeluarkan suara audio yang dapat didengarkan oleh pengguna.

IV. HASIL DAN PEMBAHASAN

A. Hasil

Penelitian ini menghasilkan sebuah program aplikasi antara lain adalah sebagai berikut:

- Laman Aplikasi Pemberian Harakat

Laman aplikasi pemberian harakat berfungsi untuk memberikan harakat pada kalimat Bahasa Arab tanpa harakat seperti pada gambar 2.



Gambar 2 Laman Aplikasi Pemberian Harakat

Hasil keluaran aplikasi berupa kalimat Bahasa Arab berharakat. Pengguna dapat mengubah tiap kata pada kalimat berharakat tersebut dengan cara mengklik kata

tersebut sehingga muncul bentuk kata yang lain dan pengguna dapat memilih kata yang sesuai dengan keinginannya. Tiap kata pada kalimat Bahasa Arab yang sudah berharakat dapat diidentifikasi jenisnya.

- Animasi *Talking Avatar*

Kalimat Bahasa Arab yang sudah berharakat juga dapat didengarkan melalui suara audio yang diucapkan oleh animasi *talking avatar*. Pengguna dapat memilih suara laki-laki atau perempuan. Suara laki-laki diucapkan oleh *talking avatar* yang menggunakan foto peneliti, sedangkan suara perempuan diucapkan oleh kartun wanita berjilbab. Animasi *talking avatar* disajikan pada gambar 3 di bawah ini.



Gambar 3 Animasi *Talking Avatar*

- Laman Proses C5.0

Laman proses C5.0 terdapat pada laman administrator yang berfungsi untuk melakukan proses perhitungan C5.0 sehingga menghasilkan pohon keputusan dan penyederhanaan aturan. Data latih berjumlah 194 aturan dapat disederhanakan menjadi 148 aturan melalui perhitungan C5.0.

B. Pembahasan

Pembahasan pengujian pada program aplikasi hasil penelitian adalah sebagai berikut:

- Akurasi

Pengujian aplikasi dilakukan dengan memasukkan lima puluh kalimat Bahasa Arab tanpa harakat yang diambil langsung dari kitab kuning. Kitab yang digunakan sebagai bahan uji adalah matan kitab *Safinatun Najah* [16]. Kitab ini dipilih karena memiliki penyajian bahasa yang mudah dan susunan kalimat yang ringan. Uji coba dilakukan oleh salah seorang Dosen Program Pengembangan Bahasa Arab (PPBA) UIN Maulana Malik Ibrahim Malang.

Dari hasil pengujian dapat diketahui bahwa:

Jumlah data = 50

Jumlah kaidah *nahwu* yang benar = 47

Jumlah kaidah *shorof* yang benar = 48

Jumlah kalimat yang benar = 45

Berdasarkan rumus :

$$\text{Akurasi} = \frac{\text{Ujicoba yang Benar}}{\text{Total Ujicoba}} \times 100\% \quad (7)$$

maka:

$$\text{Akurasi kaidah } \textit{nahwu} = \frac{47}{50} \times 100\% = 94\%$$

$$\text{Akurasi kaidah } \textit{shorof} = \frac{48}{50} \times 100\% = 96\%$$

$$\text{Akurasi kalimat} = \frac{45}{50} \times 100\% = 90\%$$

- Kuisisioner

Kuisisioner dilakukan guna mengetahui kepuasan

pengguna pada aplikasi yang telah dibangun dan kuisisioner ini dilakukan secara *online*. Responden diambil sampel sebanyak 50 mahasiswa UIN Maulana Malik Ibrahim Malang pada Jurusan yang terkait dengan matakuliah Bahasa Arab yaitu Jurusan Pendidikan Agama Islam (13 orang), Pendidikan Bahasa Arab (12 orang), *AL-Ahwal Al-Syakhshiyah* (9 orang), Hukum Bisnis Syariah (4 orang) serta Bahasa dan Sastra Arab (12 orang). Selain itu, responden juga diambil sampel sebanyak 14 dari dosen Program Pengembangan Bahasa Arab (PPBA) di UIN Maulana Malik Ibrahim Malang.

Skala pengukuran kuisisioner menggunakan skala *Likert*. Skala *Likert* digunakan untuk mengukur sikap, pendapat dan persepsi seseorang atau sekelompok orang tentang fenomena sosial [17]. Skala *Likert* adalah suatu skala psikometrik yang umum digunakan dalam kuisisioner dan merupakan skala yang paling banyak digunakan dalam riset berupa survei. Penilaian dalam skala *Likert* adalah sebagai berikut: sangat setuju (5), setuju (4), ragu-ragu (3), tidak setuju (2), sangat tidak setuju (1).

Untuk mendapatkan hasil interpretasi, harus diketahui terlebih dahulu nilai angka terendah (X) dan tertinggi (Y) dengan rumus sebagai berikut :

$$X = \text{Bobot terendah } \textit{Likert} \times \sum \text{responden} \quad (8)$$

$$Y = \text{Bobot tertinggi } \textit{Likert} \times \sum \text{responden} \quad (9)$$

Untuk menentukan interval nilai persen (I) dari jarak terendah 0% hingga tertinggi 100% menggunakan rumus:

$$I = \frac{100}{\text{Jumlah Bobot } (\textit{Likert})} \quad (10)$$

Rumus penilaian responden adalah sebagai berikut:

$$\text{Penilaian responden} = \sum \text{responden} \times \text{Bobot } \textit{Likert} \quad (11)$$

Penilaian interpretasi responden terhadap tiap pernyataan pada kuisisioner dihitung dengan menggunakan rumus index %:

$$\text{Index \%} = \frac{\text{Total Nilai}}{Y} \times 100\% \quad (12)$$

Hasil pengujian kuisisioner mahasiswa sebanyak 5 pernyataan adalah sebagai berikut:

- Responden mahasiswa sangat setuju bahwa aplikasi ini membantu dalam memberi harakat pada kalimat Bahasa Arab sehingga mudah untuk dibaca.
- Responden mahasiswa sangat setuju bahwa harakat pada setiap kata dalam kalimat yang dihasilkan sudah sesuai dengan perubahan bentuk katanya (kaidah *shorof*).
- Responden mahasiswa sangat setuju bahwa perubahan pada tiap-tiap akhir kata (*i'rob*) dalam kalimat yang dihasilkan sudah sesuai dengan gramatika Bahasa Arab (kaidah *nahwu*).
- Responden mahasiswa setuju bahwa aplikasi ini mampu mengidentifikasi tiap kata dalam kalimat dengan baik.
- Responden mahasiswa sangat setuju bahwa aplikasi ini sangat bermanfaat dan mudah untuk digunakan.

Sedangkan untuk hasil pengujian kuisisioner dosen

sebanyak 5 pernyataan adalah sebagai berikut:

- Responden dosen sangat setuju bahwa aplikasi ini membantu dalam memberi harakat pada kalimat Bahasa Arab sehingga mudah untuk dibaca.
- Responden dosen setuju bahwa harakat pada setiap kata dalam kalimat yang dihasilkan sudah sesuai dengan perubahan bentuk katanya (kaidah *shorof*).
- Responden dosen setuju bahwa perubahan pada tiap-tiap akhir kata (*i'rob*) dalam kalimat yang dihasilkan sudah sesuai dengan gramatika Bahasa Arab (kaidah *nahwu*).
- Responden dosen setuju bahwa aplikasi ini mampu mengidentifikasi tiap kata dalam kalimat dengan baik.
- Responden dosen sangat setuju bahwa aplikasi ini sangat bermanfaat dan mudah untuk digunakan.

V. KESIMPULAN

Beberapa hal yang dapat disimpulkan dari hasil penelitian ini adalah sebagai berikut:

- Untuk pemberian harakat pada kalimat Bahasa Arab dapat menggunakan N-Gram (Quad Gram) yang dilanjutkan dengan pengecekan bentuk kata. Metode N-Gram (Quad Gram) yang dilanjutkan dengan pengecekan bentuk kata mampu memberi harakat pada kalimat Bahasa Arab dengan tingkat akurasi 96% untuk kaidah *shorof*. Aplikasi ini menyediakan beberapa bentuk kata pada tiap kata sehingga pengguna dapat memilih bentuk kata yang lain sesuai dengan keinginannya.
- Untuk pemberian harakat akhir kata pada kalimat Bahasa Arab dapat menggunakan data aturan hasil perhitungan C5.0. Algoritma C5.0 mampu menyederhanakan aturan berdasarkan data latih dan dapat memberi harakat akhir kata pada kalimat Bahasa Arab dengan tingkat akurasi 94% untuk kaidah *nahwu*.
- Dalam pemberian harakat dan harakat akhir kata pada kalimat Bahasa Arab dapat menggunakan integrasi N-Gram (Quad Gram) dan C5.0. Integrasi N-Gram (Quad Gram) dan C5.0 mampu memberi harakat pada kalimat Bahasa Arab dengan tingkat akurasi 90% untuk kaidah *nahwu* dan *shorof*.
- Hasil kalimat Bahasa Arab yang sudah berharakat dapat dijadikan suara audio dan dapat diucapkan oleh animasi *talking avatar* sehingga pengguna dapat mengetahui bagaimana membaca/melafalkannya, tidak merasa jenuh dan akan terasa lebih menyenangkan dalam belajar Bahasa Arab.

REFERENSI

- [1] Machmudah U. dan Rosyidi A.W. 2008. *Active Learning* dalam pembelajaran Bahasa Arab. UIN-Malang Press. Malang.
- [2] Bruinessen, M. 1995. *Kitab Kuning, Pesantren dan Tarekat: Tradisi-Tradisi Islam di Indonesia*. Mizan. Bandung.
- [3] Tim Penyusun. 2010. *Pedoman Akademik Fakultas Syariah UIN Maulana Malik Ibrahim Malang Tahun Akademik 2010/2011*. UIN Maulana Malik Ibrahim. Malang.
- [4] Permadi, Y. 2008. *Kategorisasi Teks Menggunakan N-Gram untuk Dokumen Berbahasa Indonesia*. Ilmu Komputer Institut Pertanian. Bogor.
- [5] Alghamdi, Alhakami, Muzaffar. 2010. *Automatic restoration of Arabic Diacritics: a simple, purely statistical approach*. *The Arabian Journal for Science and Engineering*, 35 (2C).
- [6] Fuadi, M. 2010. *Skripsi UIN Maulana Malik Ibrahim Malang: Pembetulan Harokat Akhir Kata Dalam Kalimat Bahasa Arab Menggunakan Decision Tree*. Malang.
- [7] Bujlow T., Riaz T., Pedersen J.M. 2012. *A method for classification of network traffic based on C5.0 Machine Learning Algorithm*. *IEEE Workshop on Computing, Networking and Communications*: 237-241.
- [8] Alqrainy S., AlSerhan H.M., Ayesh A. 2008. *Pattern-based Algorithm for part-of-speech tagging Arabic text*. *IEEE* : 119-124.
- [9] Jamil Al-Luus. 1999. *Fushul Fi Al-Atsaqofah Al-Lughowiyah*. Darul Azminah. Amman.
- [10] Maksun, M. 2000. *Arrisalah Attashrifyyah*. Darul Hikmah. Jombang.
- [11] Turban, E., J.E. Aronson dan T.P. Liang. 2005. *Decision Support System and Intelligent Systems - 7th ed. Pearson Education, Inc.* Dwi Prabantini (penterjemah). 2005. *Sistem Pendukung Keputusan dan Sistem Cerdas*. Penerbit ANDI. Yogyakarta.
- [12] Hanafi, A. 2009. *Metode N-Gram*. <http://digilib.itelkom.ac.id/index.php?view=article&catid=20:informatika&id=531:metode-N-Gram&tmpl=component>. 24 Nopember 2013.
- [13] Kantardzic, M. 2003. *Data mining, concepts, models, methods, and algorithms*. IEEE Press.
- [14] Ruter, W. 2009. *Google Text-To-Speech (TTS)*. <https://weston.ruter.net/2009/12/12/google-tts>. 29 Oktober 2014.
- [15] Oddcast, Inc. 2009. *Sitepal User's Guide*. http://vhost.oddcast.com/support/docs/SitePal_UsersGuide.pdf. 28 Nopember 2014.
- [16] Nawawi, M. 1431 H. *Syarhu Kasyifatus Saja 'ala Safinat Naja fi Ushuliddini walfiqhi*. Karya Toha Putra. Semarang.
- [17] Nazir, M. 2003. *Metode Penelitian*. Ghalia Indonesia. Jakarta.