

Chatbot Layanan Akademik Menggunakan K-Nearest Neighbor

Kristian Adi Nugraha¹⁾, Danny Sebastian²⁾

¹⁾²⁾ Program Studi Informatika, Fakultas Teknologi Informasi, Universitas Kristen Duta Wacana
Jalan Dr. Wahidin Sudirohusodo No. 5-25, Yogyakarta 55224

¹⁾ adinugraha@ti.ukdw.ac.id

²⁾ danny.sebastian@staff.ukdw.ac.id

Abstrak

Perusahaan atau institusi yang bergerak di bidang pelayanan publik pasti memiliki layanan *customer service* untuk menjawab pertanyaan dari konsumen. Namun perusahaan atau institusi dengan skala menengah ke bawah seringkali tidak sanggup untuk menyediakan karyawan khusus untuk menangani pekerjaan tersebut, sehingga pekerjaan tersebut dirangkap oleh karyawan di posisi lain. *Chatbot* dapat digunakan untuk menyelesaikan permasalahan yang berkaitan dengan layanan tanya jawab, khususnya bagi perusahaan atau institusi yang tidak memiliki sumber daya khusus untuk menangani pekerjaan tersebut. Dengan adanya *chatbot*, pertanyaan-pertanyaan konsumen yang bersifat redundan dapat ditangani secara otomatis. Pada penelitian ini, penulis membangun sistem *chatbot* untuk layanan tanya jawab seputar kegiatan akademik dengan menggunakan metode *K-Nearest Neighbor*. Berdasarkan hasil pengujian yang telah dilakukan, sistem dapat memberikan nilai akurasi sebesar 53.48% untuk nilai $K = 3$.

Kata kunci: *chatbot, k-nearest neighbor*, pengolahan bahasa natural

Abstract

To address customers' questions, company involved in public services must provide customer service. Nevertheless, company of a medium to lower scale are frequently unable to provide special personnel to perform the task, so that the job is held simultaneously by staff in other roles. Chatbot can be used to resolve issues related to question and answer services, especially for businesses or organizations that do not have specific resources to deal with this work. Therefore, redundant customer queries can be answered automatically with the chatbot. In this research, using the *K-Nearest Neighbor* process, the authors developed a chatbot framework for question and answer services about academic activities. The device can have an accuracy value of 53.48% for the value of $K=3$ based on the results of the tests that have been conducted.

Keywords: *chatbot, k-nearest neighbor, natural language processing*

1. PENDAHULUAN

Perusahaan maupun institusi yang bergerak di bidang layanan publik memiliki *customer service* yang bertujuan untuk menjawab pertanyaan pengguna apabila terdapat hal-hal yang kurang jelas. Perusahaan biasanya menyediakan beberapa media komunikasi seperti telepon, pesan singkat (*chat*) atau *email*. Pada perusahaan dengan skala kecil, posisi operator *customer service* seringkali dirangkap oleh karyawan yang telah menduduki posisi lain. Jika beban pekerjaan utama yang dilakukan oleh karyawan tersebut sedang berada dalam jumlah yang cukup banyak, hal ini akan membuat karyawan tersebut kesulitan sehingga akan berpengaruh terhadap menurunnya performa dari pelayanan *customer service* yang ditanganinya. Di samping itu, operator *customer service* dituntut untuk selalu siap dalam menjawab pertanyaan-pertanyaan dari pengguna tanpa mengenal waktu. Hal tersebut sulit untuk dilakukan oleh perusahaan skala kecil karena harus mempekerjakan karyawan khusus dalam dua *shift*, yaitu pagi dan malam. Salah satu

solusi untuk mengatasi permasalahan dalam menjawab pertanyaan-pertanyaan dari pengguna adalah dengan membuat daftar pertanyaan yang sering ditanyakan beserta jawabannya, atau yang lebih dikenal dengan istilah *FAQ (Frequently Asked Question)*. *FAQ* berisi rangkuman pertanyaan-pertanyaan yang cukup sering ditanyakan dan bersifat umum, serta dilengkapi dengan jawaban untuk masing-masing pertanyaan. *FAQ* dibuat dengan tujuan untuk mengurangi beban pekerjaan dari *customer service*, sehingga *customer service* hanya perlu menjawab pertanyaan yang tidak terdapat pada *FAQ*. Namun *FAQ* dengan pertanyaan dalam jumlah besar seringkali membuat pengguna kesulitan saat mencari daftar pertanyaan yang sesuai dengan pertanyaan yang akan ditanyakan. Pengguna harus mencari pertanyaan yang sesuai dengan pertanyaan miliknya, hal ini akan sulit dilakukan jika daftar pertanyaan yang disediakan terlalu banyak. Pada beberapa implementasi, disediakan fitur pencarian berdasarkan kata kunci yang dimasukkan oleh pengguna. Namun fitur ini juga tidak cukup membantu, karena meskipun kata kunci yang dimasukkan oleh pengguna terdapat pada sebuah pertanyaan, namun konteks dari pertanyaan tersebut belum tentu sama dengan yang hendak ditanyakan oleh pengguna. Dengan demikian, fitur pencarian hanya akan efektif apabila pengguna bisa memasukkan kata kunci yang bersifat unik dan hanya dimiliki oleh satu pertanyaan saja.

Penulis memiliki gagasan untuk mengatasi permasalahan di atas dengan membangun sebuah mesin penjawab pesan otomatis (*chatbot*) yang sanggup menjawab pertanyaan dari pengguna secara otomatis [1]. Pengguna dapat mengirimkan pertanyaan melalui aplikasi perpesanan biasa, di mana pengguna seolah-olah sedang bertanya langsung seperti biasa kepada operator *customer service* menggunakan bahasa bebas tanpa *format* tertentu. Langkah berikutnya, pertanyaan akan diproses oleh mesin *chatbot* untuk mendapatkan data pertanyaan yang paling sesuai dalam basis data, sehingga dapat memberikan jawaban yang tepat [2]. *Chatbot* banyak diimplementasikan untuk berbagai bidang kebutuhan, di antaranya adalah aplikasi *chatbot* yang digunakan untuk *customer service* [3]. Penulis mencoba mengimplementasikan *chatbot* di lingkup kampus Universitas Kristen Duta Wacana, khususnya Fakultas Teknologi Informasi. *Chatbot* akan berfokus pada pertanyaan yang sering ditanyakan oleh mahasiswa atau wali kepada pegawai administrasi fakultas, terkait dengan kegiatan akademik. *Chatbot* akan dibangun dengan menggunakan metode *K-Nearest Neighbor (K-NN)*, di mana metode ini telah banyak diimplementasikan untuk menyelesaikan permasalahan terkait klasifikasi teks dan dapat memberikan hasil yang cukup baik [4]. Pengguna cukup mengetikkan pertanyaan yang hendak ditanyakan pada admin, kemudian sistem akan mengolah pertanyaan tersebut dan mencari pertanyaan yang paling serupa dengan pertanyaan tersebut dengan menggunakan metode *K-NN*, kemudian sistem akan mengirimkan jawaban berdasarkan pertanyaan serupa dari basis data. Harapan penulis, dengan dibangunnya *chatbot* ini akan dapat mengurangi beban tugas operator *customer service*, sehingga kualitas pelayanan yang dapat diberikan oleh institusi menjadi semakin optimal.

2. TINJAUAN PUSTAKA

Chatbot telah banyak diimplementasikan pada berbagai sektor seperti sektor kesehatan, *e-commerce*, maupun segala bentuk *customer call centers* [5]. Salah satu contoh dari implementasi *chatbot* adalah mesin bernama *Agribot* yang bertujuan untuk menangani keperluan agrikultur, di mana sistem ini akan digunakan oleh para petani [6]. *Agribot* digunakan para petani untuk mengetahui jenis tumbuhan yang paling optimal untuk ditanam, dengan memperhitungkan kondisi tanah, kondisi lingkungan, dan keadaan cuaca atau iklim. Penelitian lain yang sejenis adalah penelitian dalam membangun *chatbot* yang berfungsi untuk menjawab pertanyaan-pertanyaan bersifat umum menggunakan bahasa Thailand [7]. Berdasarkan hasil pengujian yang telah dilakukan, nilai akurasi *chatbot* dalam memahami pertanyaan adalah 86.36%, sedangkan nilai akurasi dalam memberikan jawaban adalah 93.2%.

Chatbot merupakan kombinasi antara bidang ilmu pengolahan teks dan bidang ilmu kecerdasan buatan, sehingga *chatbot* dapat dibangun dengan menggunakan kombinasi berbagai macam metode. Metode *K-NN* adalah salah satu metode kecerdasan buatan untuk bidang klasifikasi yang

cukup sering digunakan dalam menangani pengolahan teks, contohnya pada penelitian untuk memotong kalimat pada sebuah paragraf [8]. *K-NN* pada penelitian tersebut bekerja dengan cara menganalisa relasi antara kalimat yang satu dengan kalimat yang lain dalam sebuah paragraf memiliki keterkaitan atau tidak. Dengan menggunakan metode *K-NN*, nilai akurasi yang dihasilkan pada penelitian tersebut dapat memberikan hasil yang cukup baik. Penelitian lain dalam bidang pengolahan teks menggunakan metode *K-NN* adalah penelitian yang bertujuan untuk mengetahui arti sebenarnya dari sebuah kalimat yang ditulis dalam bahasa Bengali [9]. Penelitian tersebut mencoba untuk menyelesaikan permasalahan *Word Sense Disambiguation (WSD)* dengan menggunakan metode *K-NN*, sehingga arti yang sebenarnya dari sebuah kalimat dapat diketahui dengan memperhatikan konteks yang ada pada kalimat tersebut. Tingkat akurasi yang dihasilkan oleh penelitian tersebut terbilang cukup baik, yaitu dengan nilai persentase yang dihasilkan sebesar 71%.

Berdasarkan hasil luaran dari penelitian-penelitian sebelumnya yang cukup baik dengan metode *K-NN* untuk pengolahan teks, maka penulis melakukan penelitian dengan menggunakan metode *K-NN* untuk membangun sebuah mesin *chatbot* untuk menangani pertanyaan seputar layanan akademik di Universitas Kristen Duta Wacana, khususnya di Fakultas Teknologi Informasi.

2.1 Pengolahan Bahasa Natural

Pengolahan bahasa natural (*natural language processing*) merupakan salah satu turunan dari bidang ilmu komputer (*computer science*) yang merupakan kombinasi antara bidang pengolahan teks (*text processing*) dan bidang kecerdasan buatan (*artificial intelligence*). Metode tersebut dapat mengubah sebuah komputer agar memiliki kemampuan dalam memahami percakapan (lisan maupun teks) yang dilakukan oleh manusia. Jika dibandingkan dengan mesin pengolahan berbasis kecerdasan buatan di bidang lain, misalnya pengolahan citra, proses yang terjadi dalam pengolahan bahasa natural tidak jauh berbeda [10]. Dengan demikian, tingkat performa yang dihasilkan oleh mesin pengolahan bahasa natural sangat bergantung pada jenis algoritma kecerdasan buatan yang diimplementasikan di dalamnya.

2.2 Chatbot

Chatbot adalah perangkat lunak yang memiliki kemampuan untuk melakukan komunikasi dengan manusia secara otomatis dengan format pesan singkat tertulis (*chat*). *Chatbot* adalah salah satu produk dari bidang ilmu pengolahan bahasa natural yang dikembangkan dengan menggunakan metode kecerdasan buatan agar dapat memproses informasi dan memberikan jawaban yang paling tepat sesuai dengan diharapkan pengguna [11]. *Chatbot* diharapkan dapat menjadikan komputer atau sebuah mesin agar sanggup menggantikan peran manusia sebagai lawan bicara dari manusia yang lain.

2.3 Pra-pemrosesan Teks

Pra-pemrosesan teks (*text preprocessing*) merupakan sebuah tahapan yang dilakukan sebelum dokumen teks diolah lebih lanjut. Tahapan ini bertujuan untuk mengubah dokumen teks ke dalam bentuk teks lain yang lebih mudah diolah, sehingga akan memberikan hasil akhir yang lebih optimal [12, 13]. Pra-pemrosesan teks dapat memiliki satu atau lebih sub tahapan, di antaranya:

1. Tokenisasi

Proses untuk mengambil setiap kata dari sebuah kalimat. Masing-masing kata tersebut disebut sebagai *token*. Contohnya pada kalimat 'saya pergi ke sekolah' memiliki empat buah *token* yaitu: saya, pergi, ke, sekolah.

2. *Stopwording*

Proses untuk menghilangkan kata-kata yang dianggap tidak memiliki arti pada sebuah kalimat. Kata ini biasanya tergolong kata sambung dan kata hubung, contohnya: dari, ke, ini, pada.

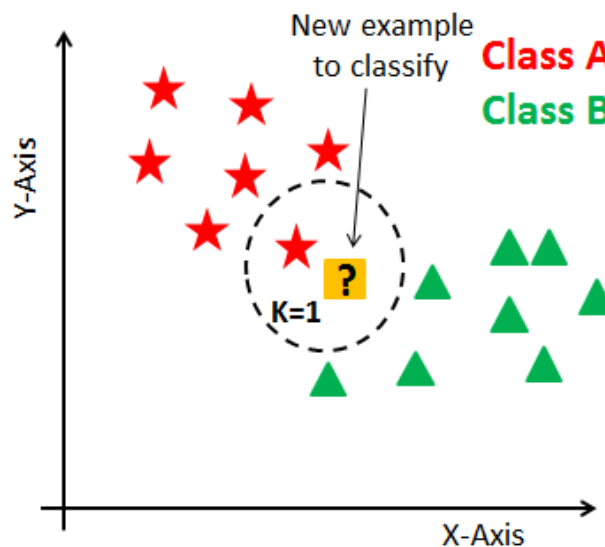
3. *Stemming*

Proses mengambil kata dasar dari sebuah kata berimbuhan, contohnya kata 'belajar' menjadi 'ajar', 'menyapu' menjadi 'sapu'.

Dengan menggunakan ketiga pra-pemrosesan di atas, apabila terdapat kalimat berupa 'Agus dan Budi pergi ke sekolah menggunakan sepeda', maka hasil pra-pemrosesan akhirnya menjadi: 'Agus', 'Budi', 'pergi', 'sekolah', 'guna', 'sepeda'. Hasil tersebut nantinya akan digunakan pada proses berikutnya, yaitu proses klasifikasi dengan menggunakan metode *K-NN*.

2.4 K-Nearest Neighbor

K-Nearest Neighbor atau *K-NN* adalah salah satu metode kecerdasan buatan bertipe *supervised learning*, yaitu kecerdasan buatan dengan pembelajaran terarah, yang digunakan untuk melakukan klasifikasi atau identifikasi terhadap sekumpulan data ke dalam kelas-kelas yang telah didefinisikan [14, 15]. Metode *K-NN* melakukan klasifikasi dengan cara mencari sejumlah *k*-data yang memiliki jarak terdekat dari data yang sedang diujikan, berikutnya menghitung jumlah kelas yang paling banyak muncul dari *k*-data tersebut seperti yang ditunjukkan pada gambar 1.



Gambar 1. Ilustrasi cara kerja *K-NN* [16]

Data yang sedang diujikan dianggap satu kelompok dengan kelas yang paling banyak muncul dari sejumlah *k*-data tersebut. Maka dengan mengacu pada ilustrasi di gambar 1, data yang sedang diujikan (data dengan simbol "?") dengan nilai $K = 1$ akan dianggap satu jenis dengan data bintang berwarna merah, karena satu data yang paling dekat dengan data yang diujikan adalah data bintang berwarna merah. Apabila nilai K dinaikkan, maka jumlah data terdekat yang harus dihitung akan ditambah sesuai dengan nilai K tersebut.

Perhitungan jarak pada metode *K-NN* umumnya digunakan dengan menggunakan persamaan berbasis spasial seperti *euclidean distance*. Namun pada kasus pemrosesan teks, perhitungan dilakukan dengan menggunakan persamaan *cosine similarity* seperti ditunjukkan pada persamaan (1).

$$\text{cos_similarity}(\mathbf{x}, \mathbf{y}) = \frac{\sum_{i=1}^n x_i \times y_i}{\sqrt{\sum_{i=1}^n x_i^2} \times \sqrt{\sum_{i=1}^n y_i^2}} \quad (1)$$

Keterangan:

n = jumlah atribut

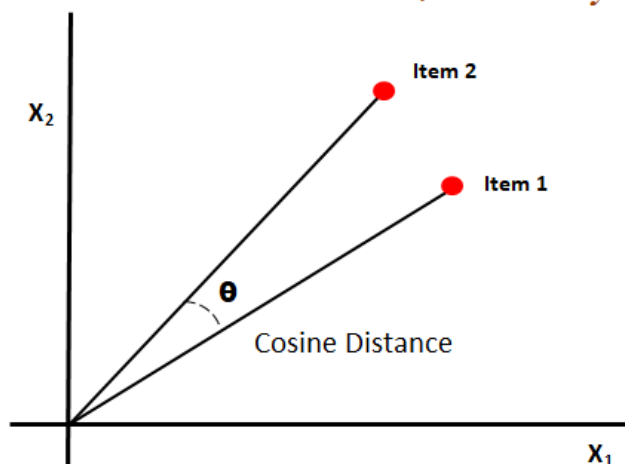
\mathbf{x} = data uji

\mathbf{y} = data target

Cosine similarity merupakan sebuah metode untuk mencari kemiripan dua buah data berdasarkan kedekatan sudut (*cosine*) seperti yang ditunjukkan pada gambar 2. Metode tersebut cocok digunakan untuk mencari kemiripan dua buah dokumen yang memiliki format teks, karena

panjang isi dari masing-masing dokumen tidak selalu sama, sehingga akan menjadi sulit jika dihitung menggunakan persamaan berbasis spasial seperti *euclidean distance* [17].

Cosine Distance/Similarity



Gambar 2. Ilustrasi perhitungan jarak pada persamaan *cosine similarity* [18]

3. METODE PENELITIAN

Penelitian dilakukan melalui beberapa tahapan yang disusun secara runtut, yaitu pengumpulan data, pembangunan sistem, pengujian, dan evaluasi.

3.1 Pengumpulan Data

Tahap pengumpulan data dilakukan dengan cara mengumpulkan data percakapan *chat* dari perangkat telepon seluler milik fakultas yang selama ini digunakan sebagai sarana layanan tanya jawab dari pihak mahasiswa atau wali dengan admin. Selain itu, data juga dikumpulkan melalui kuisisioner yang disebarakan kepada mahasiswa dan wali. Kuisisioner tersebut meminta responden untuk memasukkan daftar pertanyaan yang biasa mereka tanyakan kepada pihak kampus. Dari tahap ini, terkumpul 534 data pertanyaan yang akan digunakan untuk proses berikutnya.

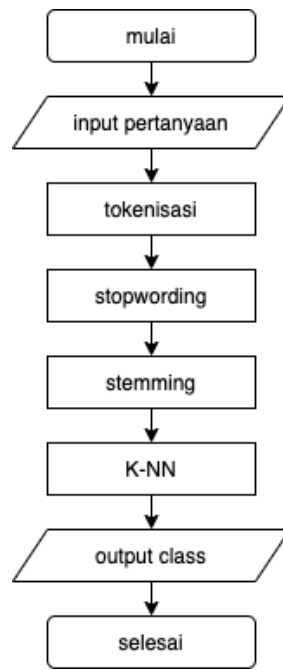
Data yang berhasil dikumpulkan disimpan dalam sebuah file teks dengan format JSON (JavaScript Object Notation) sebagai basis data pengetahuan atau *knowledge database*. Struktur file tersebut terdiri dari attribute 'knowledge' yang di dalamnya berisi class-class hasil pengelompokan seluruh data yang telah terkumpul. Setiap class memiliki attribute 'class' untuk menyimpan nama class, 'patterns' untuk menyimpan daftar pertanyaan terkait class tersebut, dan 'responses' berisi daftar jawaban terkait class tersebut. Salah satu contoh struktur sebuah class di file JSON adalah seperti berikut ini:

```
{
  "class": "beasiswa",
  "patterns": [
    "Bagaimana cara mendaftar beasiswa?",
    "Apakah saya bisa minta informasi beasiswa?"
  ],
  "responses": [
    "Informasi beasiswa bisa diakses melalui
    ukdw.ac.id/beasiswa"
  ],
}
```

Gambar 3. Struktur JSON dari sebuah data latihan

3.2 Pembangunan Sistem

Sistem *chatbot* dibangun dengan menggunakan bahasa pemrograman Python versi 3.8. Beberapa pustaka pihak ketiga turut ditambahkan pada sistem *chatbot* untuk mempermudah proses pengolahan data, di antaranya adalah pustaka Natural Language Toolkit (NLTK) untuk keperluan tokenisasi, serta pustaka Sastrawi untuk keperluan *stemming* dan *stopwording*. Diagram alir mengenai cara kerja sistem dapat dilihat pada gambar 4.



Gambar 4. Diagram alir cara kerja sistem *chatbot*

3.3 Pengujian Sistem

Pengujian dilakukan dengan cara meminta bantuan dari 20 partisipan yang merupakan mahasiswa atau wali mahasiswa untuk mencoba mengajukan beberapa pertanyaan pada sistem *chatbot* yang telah dibangun. Setiap partisipan diminta untuk mengajukan 5 buah pertanyaan terkait dengan kegiatan akademik dengan tema yang telah ditentukan sebelumnya berdasarkan *class* yang berhasil dibangun, sehingga total pertanyaan yang berhasil dikumpulkan adalah 100 pertanyaan. Dari 100 pertanyaan tersebut, penulis melakukan seleksi untuk membuang pertanyaan-pertanyaan yang dianggap bias sehingga tidak dapat diproses oleh sistem. Setelah melalui proses tersebut, jumlah pertanyaan yang valid dan dapat diproses oleh sistem berjumlah 86 pertanyaan.

4. PEMBAHASAN

Sistem *chatbot* yang dibangun oleh penulis memiliki basis data yang terdiri dari 10 *class*, di mana masing-masing *class* memiliki paling sedikit 5 buah pertanyaan sebagai data latih. Keterangan untuk masing-masing *class* dapat dilihat pada tabel 1.

Tabel 1. *Class* pada basis data

<i>Nama</i>	<i>Keterangan</i>
beasiswa	Pertanyaan seputar pendaftaran beasiswa
registrasi	Pertanyaan seputar registrasi perkuliahan
pembayaran	Pertanyaan seputar tata cara pembayaran kuliah
wisuda	Pertanyaan seputar prosedur pendaftaran wisuda

poin	Pertanyaan seputar poin keaktifan
kuliah	Pertanyaan seputar agenda perkuliahan
jadwal	Pertanyaan seputar jadwal kegiatan program studi selama satu semester
administrasi	Pertanyaan seputar layanan administrasi fakultas
dosen	Pertanyaan seputar jadwal konsultasi setiap dosen
info	Pertanyaan seputar informasi terbaru dari fakultas

Pengujian akurasi metode *K-NN* dilakukan dengan nilai $K = 1$ sampai dengan $K = 5$ dengan pertimbangan jumlah data latih paling sedikit dari sebuah *class* adalah 5, sehingga dengan menggunakan nilai $K = 5$ masih terdapat peluang agar seluruh *class* yang menjadi target dapat terpilih seluruhnya. Melalui pengujian yang dilakukan menggunakan data sebanyak 86 pertanyaan, didapatkan nilai tingkat akurasi metode *K-NN* seperti ditunjukkan pada tabel 2.

Tabel 2. Nilai akurasi pengujian

<i>K</i>	Nilai akurasi (%)
1	52.33%
2	52.33%
3	53.48%
4	48.84%
5	41.86%

Berdasarkan hasil pengujian seperti yang ditunjukkan pada tabel 2, maka dapat disimpulkan bahwa nilai K tertinggi didapatkan pada $K = 3$ yaitu sebesar 53.48%. Beberapa kendala yang mengakibatkan sistem *chatbot* memberikan hasil yang tidak tepat disebabkan oleh adanya kata-kata penting dalam kalimat yang masuk ke dalam dua atau lebih *class*, sehingga sistem *chatbot* kesulitan untuk mengklasifikasikan pertanyaan tersebut secara tepat. Contohnya seperti ditunjukkan pada tabel 3, di mana ketiga pertanyaan tersebut memiliki susunan kata yang mirip, hanya berbeda pada bagian akhir saja.

Tabel 3. Nilai akurasi pengujian

<i>N</i>	Pertanyaan	Class Target	Token
1	bagaimana tata cara pendaftaran wisuda?	wisuda	tata, cara, daftar, wisuda
2	bagaimana tata cara pendaftaran beasiswa?	beasiswa	tata, cara, daftar, beasiswa
3	bagaimana tata cara pendaftaran mata kuliah?	registrasi	tata, cara, daftar, mata, kuliah

Pada contoh di tabel 3, *token* yang dihasilkan oleh ketiga pertanyaan tersebut memiliki susunan kata yang hampir serupa, ketiganya sama-sama memiliki tiga buah kata: tata, cara, dan daftar. Sedangkan kata kunci yang dapat membedakan pertanyaan tersebut dibanding pertanyaan lainnya hanya satu atau dua kata, lebih sedikit jika dibandingkan dengan jumlah kata yang mirip (tiga). Hal ini menyebabkan sistem kesulitan dalam mengklasifikasikan pertanyaan tersebut secara tepat. Selain itu, terdapat beberapa kata yang menggunakan bentuk tidak baku singkatan seperti 'sy' (saya) dan 'bgmn' (bagaimana) yang tidak dapat dikenali oleh sistem. Kemudian terdapat kata tidak baku *slang*/informal seperti 'gimana' (bagaimana) dan 'kalo' (kalau) yang juga tidak dapat dikenali oleh sistem setelah melalui tahap tokenisasi. Kedua hal tersebut turut mengakibatkan sistem tidak dapat memberikan nilai akurasi yang maksimal dalam mengklasifikasikan pertanyaan.

5. KESIMPULAN

Sistem *chatbot* yang dibangun dapat bekerja dengan baik dan memberikan nilai akurasi maksimal sebesar 53.48% untuk nilai $K = 3$. Beberapa kendala yang dialami oleh sistem adalah adanya pertanyaan-pertanyaan dari *class* berbeda tetapi memiliki susunan kata yang serupa, sehingga sistem sulit untuk mengklasifikasikan pertanyaan tersebut dengan tepat. Kemudian

terdapat kata-kata yang tidak dapat dikenali oleh sistem karena kata tersebut tergolong sebagai jenis kata tidak baku, sementara basis data yang dimiliki oleh sistem hanya terdiri dari kata-kata dalam bentuk baku. Beberapa hal yang dapat dilakukan untuk dapat meningkatkan nilai akurasi dari sistem adalah dengan menghilangkan kata-kata yang memiliki irisan pada beberapa *class*, sehingga masing-masing *class* hanya berisi daftar kata yang unik. Selain itu pra-pemrosesan tambahan dapat diimplementasikan untuk mengolah kata-kata tidak baku menjadi baku agar dapat dikenali oleh sistem.

DAFTAR PUSTAKA

- [1] S. A. Abdul-Kader and J. Woods, "Survey on Chatbot Design Techniques in Speech Conversation Systems," *International Journal of Advanced Computer Science and Applications*, vol. 6, no. 7, pp. 72-80, 2015.
- [2] J. R. Hill, W. R. Ford and I. Farreras, "Real conversations with artificial intelligence: A comparison between human-human online conversations and human-chatbot conversations," *Computers in Human Behavior*, vol. 49, pp. 245-250, 2015.
- [3] A. Xu, Z. Liu, Y. Guo, V. Sinha and R. Akkiraju, "A New Chatbot for Customer Service on Social Media," in *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*, Denver, 2017.
- [4] K. A. Nugraha, "Deteksi Area Parkir Mobil Berbasis Marker Menggunakan Moment Invariants dan K-NN," *Jurnal Teknik Informatika dan Sistem Informasi*, vol. 5, no. 1, pp. 112-121, 2019.
- [5] N. Albayrak, A. Özdemir and E. Zeydan, "An overview of artificial intelligence based chatbots and an example chatbot application," in *26th Signal Processing and Communications Applications Conference (SIU)*, Izmir, Turkey, 2018.
- [6] D. Sawant, A. Jaiswal, J. Singh and P. Shah, "AgriBot - An intelligent interactive interface to assist farmers in agricultural activities," in *2019 IEEE Bombay Section Signature Conference (IBSSC)*, Mumbai, India, 2019.
- [7] P. Muangkammuen, N. Intiruk and K. R. Saikaew, "Automated Thai-FAQ Chatbot using RNN-LSTM," in *2018 22nd International Computer Science and Engineering Conference (ICSEC)*, Chiang Mai, Thailand, 2018.
- [8] T. Jo, "Using K Nearest Neighbors for text segmentation with feature similarity," in *2017 International Conference on Communication, Control, Computing and Electronics Engineering (ICCCCEE)*, Khartoum, Sudan, 2017.
- [9] R. Pandit and S. K. Naskar, "A memory based approach to word sense disambiguation in Bengali using k-NN method," in *2015 IEEE 2nd International Conference on Recent Trends in Information Systems (ReTIS)*, Kolkata, India, 2015.
- [10] Z. Zong and C. Hong, "On Application of Natural Language Processing in Machine Translation," in *2018 3rd International Conference on Mechanical, Control and Computer Engineering (ICMCCE)*, Huhhot, China, 2018.
- [11] N. Albayrak, A. Özdemir and E. Zeydan, "An overview of artificial intelligence based chatbots and an example chatbot application," in *2018 26th Signal Processing and Communications Applications Conference (SIU)*, Izmir, Turkey, 2018.
- [12] K. A. Nugraha and D. Sebastian, "Pembentukan Dataset Topik Kata Bahasa Indonesia pada Twitter Menggunakan TF-IDF & Cosine Similarity," *Jurnal Teknik Informatika dan Sistem Informasi*, vol. 4, no. 3, pp. 376-386, December 2018.
- [13] D. Sebastian and K. A. Nugraha, "Text normalization for indonesian abbreviated word using crowdsourcing method," in *2019 International Conference on Information and Communications Technology (ICOIACT)*, Yogyakarta, 2019.

- [14] Y. Udovychenko, A. Popov and I. Chaikovsky, "Ischemic heart disease recognition by k-NN classification of current density distribution maps," in *2015 IEEE 35th International Conference on Electronics and Nanotechnology (ELNANO)*, Kiev, Ukraine, 2015.
- [15] L. Li, Y. Yu, S. Bai, Y. Hou and X. Chen, "An Effective Two-Step Intrusion Detection Approach Based on Binary Classification and k -NN," *IEEE Access*, vol. 6, pp. 12060 - 12073, 2017.
- [16] A. Navlani, "KNN Classification using Scikit-learn," 2 Agustus 2018. [Online]. Available: <https://www.datacamp.com/community/tutorials/k-nearest-neighbor-classification-scikit-learn>. [Accessed 17 Maret 2021].
- [17] K. A. Nugraha and D. Sebastian, "Analisis Trend Akun Media Sosial Twitter Menggunakan TF-IDF dan Cosine Similarity," in *Prosiding Nasional Rekayasa Teknologi Industri dan Informasi XIII*, 2018.
- [18] P. Dangeti, "Cosine similarity," O'Reilly, [Online]. Available: <https://www.oreilly.com/library/view/statistics-for-machine/9781788295758/eb9cd609-e44a-40a2-9c3a-f16fc4f5289a.xhtml>. [Accessed 17 Maret 2021].

Biodata Penulis

Kristian Adi Nugraha, S.Kom., M.T., lahir di Yogyakarta pada tanggal 04 Oktober 1989. Meraih gelar Sarjana Komputer (S.Kom.) di Program Studi Informatika Universitas Kristen Duta Wacana pada tahun 2012 dan gelar Magister Teknik (M.T.) di Universitas Atma Jaya Yogyakarta pada tahun 2013. Saat ini berprofesi sebagai dosen di Program Studi Informatika Universitas Kristen Duta Wacana.

Danny Sebastian, S.Kom., M.M., M.T., lahir di Pekalongan pada tanggal 26 November 1988. Meraih gelar Sarjana Komputer (S.Kom.) di Program Studi Informatika Universitas Kristen Duta Wacana pada tahun 2011, gelar Magister Manajemen (M.M.) di Universitas Pelita Harapan pada tahun 2014 dan gelar Magister Teknik (M.T.) di Universitas Atma Jaya Yogyakarta pada tahun 2016. Saat ini berprofesi sebagai dosen di Program Studi Informatika Universitas Kristen Duta Wacana.