

LAPORAN AKHIR PENELITIAN



**METODE K-NEAREST NEIGHBOR UNTUK PEMBUATAN CHATBOT
LAYANAN AKADEMIK DI UNIVERSITAS KRISTEN DUTA WACANA**

TIM PENGUSUL

**Kristian Adi Nugraha, S.Kom.,MT
Danny Sebastian, S.Kom.,MM,MT**

DUTA WACANA

Informatika

UNIVERSITAS KRISTEN DUTA WACANA

November 2020

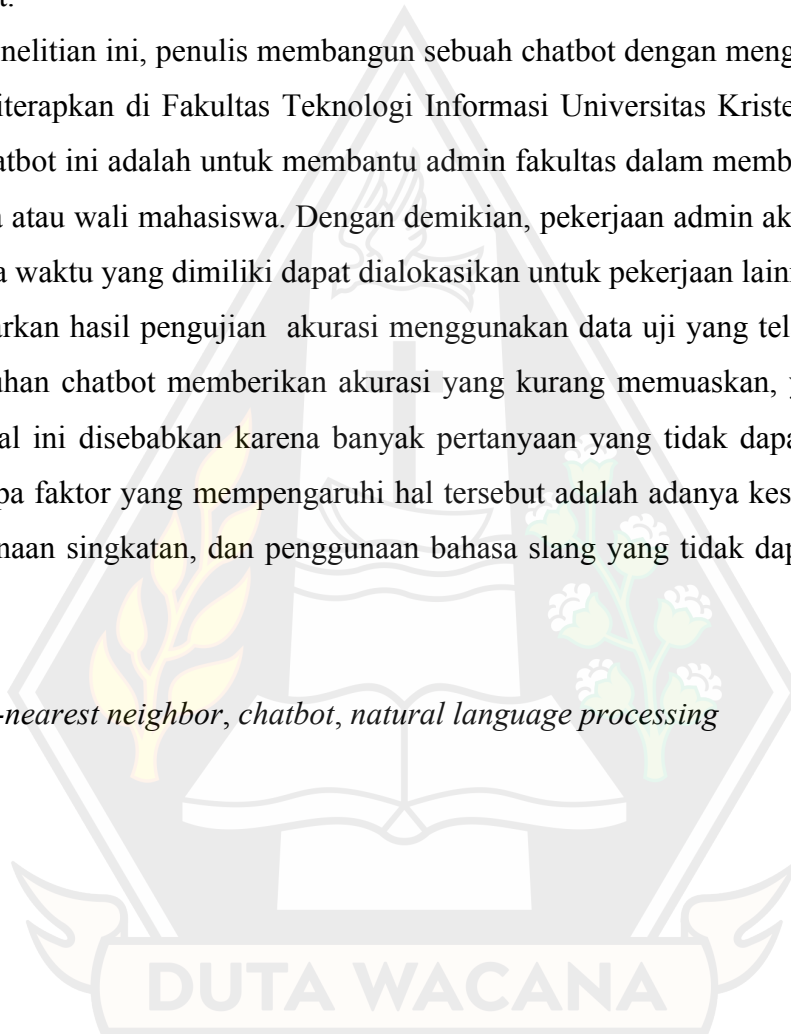
RINGKASAN

Chatbot merupakan sebuah teknologi yang digunakan untuk membalas pesan berbentuk teks dari seseorang secara otomatis. Hal ini dapat meringankan pekerjaan manusia dalam membalas pesan-pesan yang masuk karena dapat ditangani oleh chatbot. Demikian juga pada layanan kontak atau yang biasa disebut sebagai customer service yang seringkali menerima pesan dalam jumlah yang besar. Pekerjaan mereka akan menjadi lebih ringan apabila terdapat bantuan chatbot.

Pada penelitian ini, penulis membangun sebuah chatbot dengan menggunakan metode K-NN untuk diterapkan di Fakultas Teknologi Informasi Universitas Kristen Duta Wacana. Tujuan dari chatbot ini adalah untuk membantu admin fakultas dalam membalas pesan-pesan dari mahasiswa atau wali mahasiswa. Dengan demikian, pekerjaan admin akan menjadi lebih ringan sehingga waktu yang dimiliki dapat dialokasikan untuk pekerjaan lainnya.

Berdasarkan hasil pengujian akurasi menggunakan data uji yang telah dikumpulkan, secara keseluruhan chatbot memberikan akurasi yang kurang memuaskan, yaitu 15% untuk nilai $K = 3$. Hal ini disebabkan karena banyak pertanyaan yang tidak dapat dipahami oleh sistem. Beberapa faktor yang mempengaruhi hal tersebut adalah adanya kesalahan penulisan (typo), penggunaan singkatan, dan penggunaan bahasa slang yang tidak dapat diproses oleh sistem.

Kata Kunci: *k-nearest neighbor, chatbot, natural language processing*



LAPORAN AKHIR PENELITIAN



**METODE K-NEAREST NEIGHBOR UNTUK PEMBUATAN CHATBOT
LAYANAN AKADEMIK DI UNIVERSITAS KRISTEN DUTA WACANA**

TIM PENGUSUL

**Kristian Adi Nugraha, S.Kom.,MT
Danny Sebastian, S.Kom.,MM,MT**

DUTA WACANA

Informatika

UNIVERSITAS KRISTEN DUTA WACANA

November 2020

HALAMAN PENGESAHAN

Judul Penelitian : METODE K-NEAREST NEIGHBOR UNTUK PEMBUATAN
CHATBOT LAYANAN AKADEMIK DI UNIVERSITAS
KRISTEN DUTA WACANA

Kode Rumpun Ilmu : 458 Teknik Informatika

Ketua Peneliti

a. Nama Lengkap : Kristian Adi Nugraha, S.Kom., MT

b. Program Studi : 2232

c. Nomor HP : 08174126722

d. Alamat surel (e-mail) : adinugraha@staff.ukdw.ac.id

Anggota Peneliti

Nama Lengkap : Danny Sebastian, S.Kom., MM, MT

Anggota Mahasiswa

Nama Lengkap : Joshua Gibeon Mulyana

Nim : 71160004

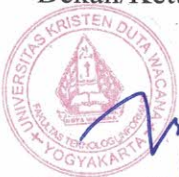
Biaya Penelitian

:	- Disetujui LPPM	Rp.10,000,000
:	- Disetujui Fakultas/Prodi	Rp.0
:	- Institusi lain	Rp.0
:	- Dana Pribadi	Rp.0

Yogyakarta, 16 November 2020

Mengetahui,
Dekan/Ketua

Ketua Peneliti



Restyandito, S.Kom., MSIS., Ph.D.

Kristian Adi Nugraha,
S.Kom., MT
154E402

Menyetujui
Ketua LPPM



Dr. – Ing., Wiyatiningsih, ST., MT.
964E238

DAFTAR ISI

HALAMAN PENGESAHAN	i
RINGKASAN	iii
DAFTAR ISI	iv
DAFTAR TABEL	vi
DAFTAR GAMBAR	vii
DAFTAR LAMPIRAN	viii
BAB 1 PENDAHULUAN	1
1.1 Latar Belakang	1
1.2 Rumusan Masalah	2
BAB 2 TINJAUAN PUSTAKA	3
2.1 Tinjauan Pustaka	3
2.2 Landasan Teori	3
2.2.1 Pengolahan Bahasa Natural	3
2.2.2 Chatbot	4
2.2.3 K-Nearest Neighbor	4
BAB 3 TUJUAN DAN MANFAAT PENELITIAN	5
3.1 Tujuan Penelitian	5
3.2 Manfaat Penelitian	5
3.3 Batasan Penelitian	5
3.4 Kontribusi dan Luaran Penelitian	5
BAB 4 METODE PENELITIAN	7
4.1 Roadmap Penelitian	7
4.2 Pengumpulan Data Awal	7
4.3 Pembangunan Sistem	8
4.4 Pengujian Sistem	8
4.5 Analisis Sistem	8
4.6 Penulisan Laporan	8

BAB 5	HASIL DAN LUARAN YANG DICAPAI	9
5.1	Artikel Publikasi Seminar Internasional IAIAI 2020.....	9
5.2	Artikel Publikasi IJASEIT	11
BAB 6	RENCANA TAHAPAN BERIKUTNYA.....	12
6.1	Rencana Tahapan Berikutnya.....	12
BAB 7	KESIMPULAN DAN SARAN	13
DAFTAR PUSTAKA		14
LAMPIRAN.....		30



DAFTAR TABEL

Tabel 3.1 Rencana Target Capaian5



DAFTAR GAMBAR

Gambar 4.1 Roadmap penelitian.....	7
Gambar 4.2 Langkah penelitian.....	7
Gambar 5.1 Contoh class dalam basis pengetahuan.....	9
Gambar 5.2 Tampilan simulasi chatbot.....	10



DAFTAR LAMPIRAN

Lampiran A: Artikel publikasi IAIAI 2020

Lampiran B: Artikel publikasi IJASEIT

Lampiran C: Rekapitulasi penggunaan biaya



BAB 1

PENDAHULUAN

1.1 Latar Belakang

Setiap perusahaan atau institusi yang memberikan layanan kepada publik selalu memiliki customer service yang selalu siap menjawab pertanyaan melalui telepon, SMS, atau chat dari pengguna layanan apabila terdapat hal-hal yang kurang jelas. Namun di beberapa perusahaan atau institusi dengan skala menengah ke bawah, seringkali posisi operator customer service dianggap tidak terlalu vital, sehingga seringkali posisi tersebut dirangkap oleh karyawan yang telah menjabat di posisi lain. Apabila jumlah pekerjaan utama karyawan tersebut sedang cukup banyak, hal ini akan berpengaruh terhadap menurunnya performa dari pelayanan customer service yang ditanganinya. Selain itu operator customer service hendaknya selalu siap 24 jam untuk menjawab pertanyaan-pertanyaan dari pengguna layanan, namun hal ini sulit untuk dilakukan oleh perusahaan atau institusi dengan skala menengah ke bawah. Salah satu cara untuk mengatasi permasalahan tersebut adalah dengan menggunakan metode FAQ atau Frequently Asked Question. Metode ini bekerja dengan cara merangkum pertanyaan-pertanyaan yang sering ditanyakan oleh pengguna layanan dan bersifat umum, kemudian menulis ulang pertanyaan-pertanyaan tersebut lengkap dengan jawaban dari setiap pertanyaan. FAQ dibuat untuk mengurangi beban kerja operator yang bertugas dalam menjawab pertanyaan yang bersifat umum. Akan tetapi FAQ dengan konten yang terlalu banyak seringkali menyulitkan pengguna dalam mencari jawaban dari pertanyaan yang dimiliki. Selain itu FAQ hanya dapat digunakan untuk mencari jawaban dari pertanyaan-pertanyaan yang bersifat umum, dengan demikian metode FAQ dianggap belum bisa menggantikan posisi operator customer service.

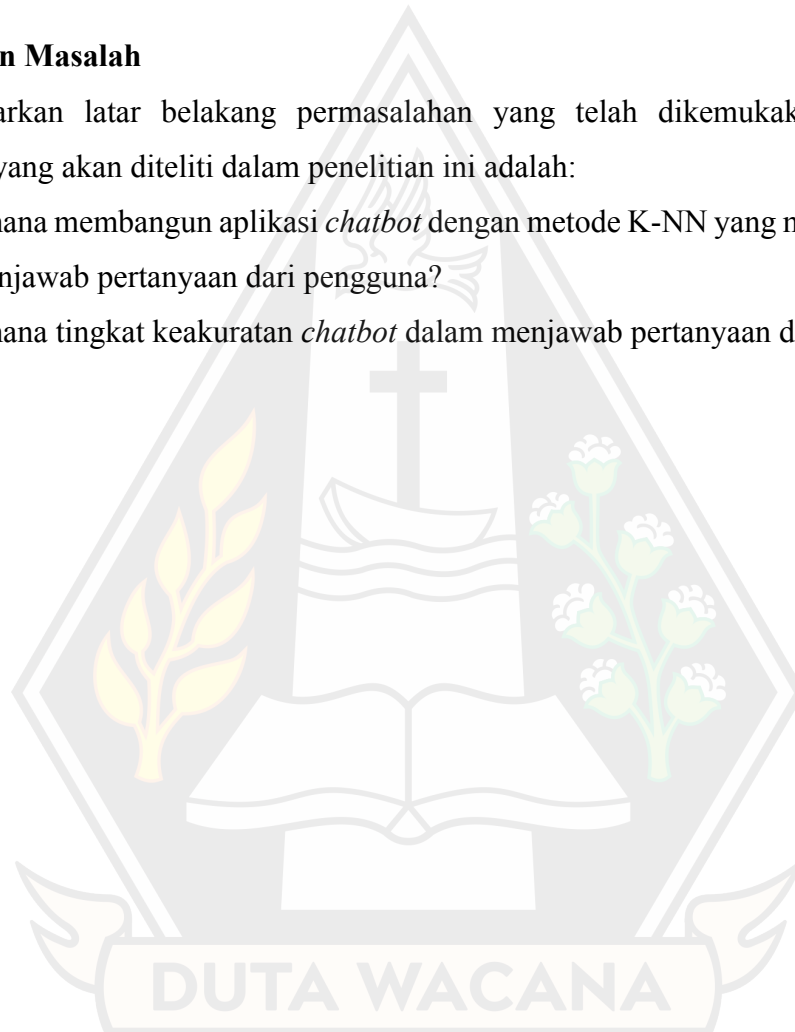
Berdasarkan permasalahan di atas, penulis memiliki ide untuk membuat mesin penjawab otomatis dalam bentuk chat (chatbot) yang dapat menjawab pertanyaan-pertanyaan dari pengguna layanan (Abdul-Kader & Woods, 2015). Pengguna layanan nantinya dapat mengajukan pertanyaan seolah sedang bertanya langsung kepada operator customer service, kemudian pertanyaan tersebut akan diolah dan dijawab oleh chatbot (Hill, Ford, & Farreras, 2015). Chatbot telah banyak diimplementasikan pada berbagai sektor, salah satunya adalah pada aplikasi chatbot yang digunakan untuk customer services (Xu, Liu, Guo, Sinha, & Akkiraju, 2017). Penulis akan mencoba mengimplementasikan chatbot di lingkup Universitas Kristen Duta Wacana, khususnya Fakultas Teknologi Informasi. Chatbot ini akan berfokus

pada pertanyaan-pertanyaan yang sering ditanyakan oleh mahasiswa kepada pegawai administrasi fakultas terkait dengan kegiatan akademik. Sistem chatbot tersebut akan dibangun dengan menggunakan metode K-Nearest Neighbor (K-NN), di mana metode ini telah banyak digunakan untuk kasus pengklasifikasian teks dan memberikan hasil yang cukup baik (Wajeed & Adilakshmi, 2011). Harapan penulis, dibangunnya sistem chatbot ini akan dapat meringankan tugas dari operator customer service, sehingga pelayanan yang diberikan oleh perusahaan atau institusi menjadi lebih optimal.

1.2 Rumusan Masalah

Berdasarkan latar belakang permasalahan yang telah dikemukakan sebelumnya, permasalahan yang akan diteliti dalam penelitian ini adalah:

1. Bagaimana membangun aplikasi *chatbot* dengan metode K-NN yang mampu mengolah dan menjawab pertanyaan dari pengguna?
2. Bagaimana tingkat keakuratan *chatbot* dalam menjawab pertanyaan dari pengguna?



BAB 7

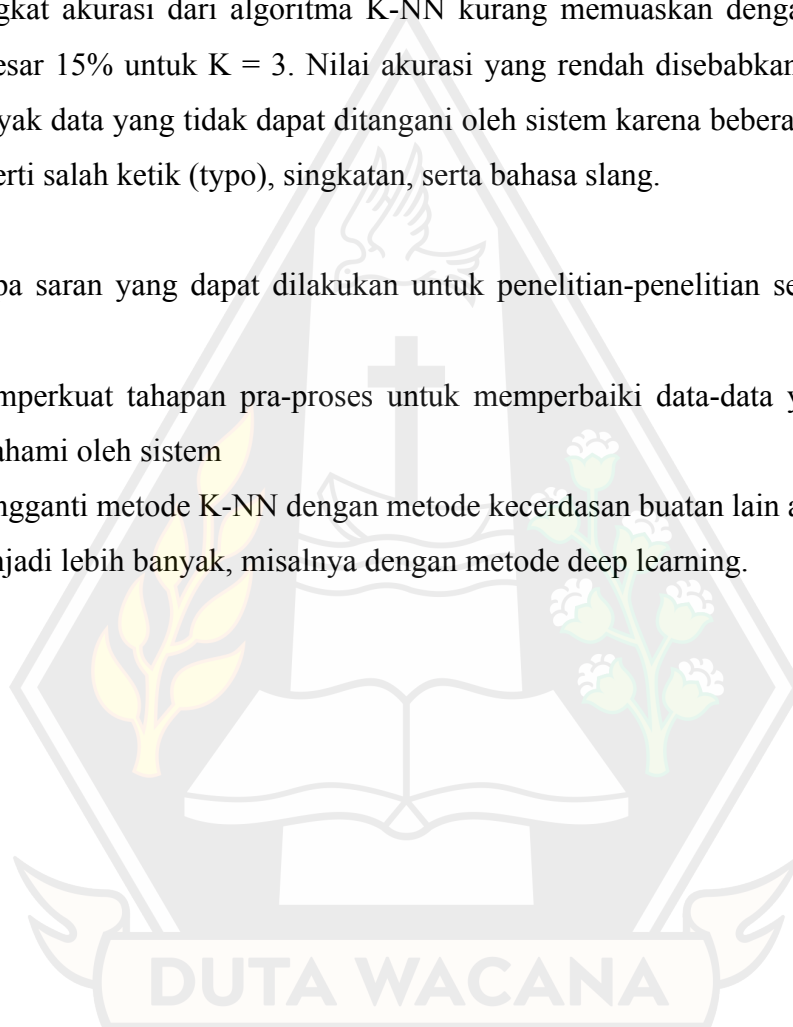
KESIMPULAN DAN SARAN

Setelah melakukan pengujian dan analisa, maka kesimpulan yang dapat diambil dari penelitian ini adalah:

1. Algoritma K-NN dapat diimplementasikan untuk chatbot dan mampu menjawab pertanyaan pengguna.
2. Tingkat akurasi dari algoritma K-NN kurang memuaskan dengan nilai tertinggi sebesar 15% untuk $K = 3$. Nilai akurasi yang rendah disebabkan karena terdapat banyak data yang tidak dapat ditangani oleh sistem karena beberapa permasalahan seperti salah ketik (typo), singkatan, serta bahasa slang.

Beberapa saran yang dapat dilakukan untuk penelitian-penelitian sejenis berikutnya adalah:

1. Memperkuat tahapan pra-proses untuk memperbaiki data-data yang tidak dapat dipahami oleh sistem
2. Mengganti metode K-NN dengan metode kecerdasan buatan lain agar jumlah class menjadi lebih banyak, misalnya dengan metode deep learning.



DAFTAR PUSTAKA

- Aggarwal, C. C., & Zhai, C. X. (2012). *Mining Text Data*. Boston, MA: Springer.
- Agnihotri, D., Verma, K., & Tripathi, P. (2014). Pattern and cluster mining on text data. *Fourth International Conference on Communication Systems and Network Technologies*.
- Agung, B. (2017, Desember 2017). *Pengguna Internet di Indonesia Akses Medsos 3 Jam Per Hari*. (CNN Indonesia) Retrieved Juli 26, 2018, from <https://www.cnnindonesia.com/teknologi/20171218192500-192-263281/pengguna-internet-di-indonesia-akses-medsos-3-jam-per-hari>
- Ahmad, S., & Varma, R. (2018). Information extraction from text messages using data mining techniques. *Malaya Journal of Matematik*, *S(1)*, 26-29.
- Al-Anzi, F., & AbuZeina, D. (2017). Toward an enhanced Arabic text classification using cosine similarity and latent semantic indexing. *Journal of King Saud University – Computer and Information Sciences*, *29(2)*, 189-195.
- Albayrak, N., Özdemir, A., & Zeydan, E. (2018). An overview of artificial intelligence based chatbots and an example chatbot application. *2018 26th Signal Processing and Communications Applications Conference (SIU)*. Izmir, Turkey: IEEE.
- Allahyari, M., Pouriyeh, S., Assefi, M., Safaei, S., Trippe, E., Gutierrez, J., & Kochut, K. (2017). A brief survey of text mining: Classification, clustering and extraction techniques. *arXiv preprint arXiv:1707.02919*.
- Bahtar, A., & Muda, M. (2016). The Impact of User--Generated Content (UGC) on Product Reviews towards Online Purchasing-A Conceptual Framework. *Procedia Economics and Finance*. *37*, pp. 337-342. Elsevier.
- Bhardwaj, B. (2016). Text Mining, its Utilities, Challenges and Clustering Techniques. *International Journal of Computer Applications*, *135(1)*.
- Chae, J., Thom, D., Bosch, H., Jang, Y., & Maciejewski, R. (2012). Spatiotemporal Social Media Analytics for Abnormal Event Detection and Examination using Seasonal-Trend Decomposition. *Visual Analytics Science and Technology (VAST)* (pp. 143-152). IEEE.
- Chen, X., Vorvoreanu, M., & Madhavan, K. (2014). Mining Social Media Data for Understanding Student's Learning Experiences. *IEEE Transactions on Learning Technologies*, *7(3)*, 246-259.
- Chrismanto, A., & Lukito, Y. (2016). Klasifikasi Sentimen Komentar Politik dari Facebook Page Menggunakan Naive Bayes. *Jurnal Informatika dan Sistem Informasi*, *2(2)*, 26-34.

- Cooley, R., Mobasher, B., & Srivastava, J. (1997). Web mining: Information and pattern discovery on the world wide web. *Tools with Artificial Intelligence, 1997. Proceedings., Ninth IEEE International Conference on. IEEE.*
- Crowston, K., & Fagnot, I. (2018). Stages of motivation for contributing user-generated content: A theory and empirical test. *International Journal of Human-Computer Studies, 109*, 89-101.
- Daugherty, T., Eastin, M. S., & Bright, L. (2008). Exploring consumer motivations for creating user-generated content. *Journal of interactive advertising, 2*(2), 16-25.
- Gomaa, W., & Fahmy, A. (2013). A Survey of Text Similarity Approaches. *International Journal of Computer Applications , 68*(13), 13-18.
- Gupta, V., & Lehal, G. S. (2009). A Survey of Text Mining Techniques and Applications. *Journal of Emerging Technologies in Web Intelligence, 1*(1), 60-76.
- He, W., Zha, S., & Li, L. (2013). Social media competitive analysis and text mining: A case study in pizza industry. *International Journal of Information Management, 33*(3), 464-472.
- Iglesias, J. A., Tiemblo, A., Ledezma, A., & Sanchis, A. (2016). Web news mining in an evolving framework. *Information Fusion, 28*, 90-98.
- Inzalkar, S., & Sharma, J. (2015). A survey on text mining-techniques and application. *International Journal of Research In Science & Engineering, 24*, 1-14.
- Jo, T. (2017). Using K Nearest Neighbors for text segmentation with feature similarity. *2017 International Conference on Communication, Control, Computing and Electronics Engineering (ICCCCEE)*. Khartoum, Sudan: IEEE.
- Kaplan, A. M., & Haenlein, M. (2010). Users of the world, unite! The challenges and opportunities of Social Media. *Business horizons, 53*(1), 59-68.
- Kaplan, A. M., & Haenlein, M. (2010). Users of the world, unite! The challenges and opportunities of Social Media. *Business Horizons, 53*, 59-68.
- Kim, A. J., & Johnson, K. K. (2016). Power of consumers using social media: Examining the influences of brand-related user-generated content on Facebook. *Computer in Human Behavior, 58*, 98-108.
- Kohavi, R. (2001). Mining E-Commerce Data: The good, the bad, and the ugly. *Proceedings of the seventh ACM SIGKDD international conference on Knowledge discovery and data mining*. ACM.
- Kompan, M., & Bielikova, M. (2010). Content-based news recommendation. *International conference on electronic commerce and web technologies* (pp. 61-72). Springer.

- Kosala, R., & Blockeel, H. (2000). Web mining research: A survey. *ACM Sigkdd Explorations Newsletter*, 2(1), 1-15.
- Lahitani, A. R., Permanasari, A. E., & Setiawan, N. A. (2016). Cosine similarity to determine similarity measure: Study case in online essay assessment. *2016 4th International Conference on Cyber and IT Service Management*. Bandung.
- Laksana, N. C. (2018, Maret 13). *Ini Jumlah Total Pengguna Media Sosial di Indonesia*. (Okezone) Retrieved Juli 26, 2018, from <https://techno.okezone.com/read/2018/03/13/207/1872093/ini-jumlah-total-pengguna-media-sosial-di-indonesia>
- Manap, K., & Adzharudin, N. (2013). The role of user generated content (UGC) in social media for tourism sector. *The 2013 WEI International Academic Conference Proceedings*, (pp. 52-58).
- Mathioudakis, M., & Koudas, N. (2010). Twittermonitor: trend detection over the twitter stream. *Proceedings of the 2010 ACM SIGMOD International Conference on Management of data* (pp. 1155-1158). ACM.
- Menaka, S., & Radha, N. (2013). Text classification using keyword extraction technique. *International Journal of Advanced Research in Computer Science and Software Engineering*, 3(12), 734-740.
- Muangkammuen, P., Intiruk, N., & Saikaew, K. R. (2018). Automated Thai-FAQ Chatbot using RNN-LSTM. *2018 22nd International Computer Science and Engineering Conference (ICSEC)*. Chiang Mai, Thailand, Thailand: IEEE.
- O'reilly, T. (2005). What is web 2.0.
- Pandit, R., & Naskar, S. K. (2015). A memory based approach to word sense disambiguation in Bengali using k-NN method. *2015 IEEE 2nd International Conference on Recent Trends in Information Systems (ReTIS)*. Kolkata, India: IEEE.
- Salloum, S., Al-Emran, M., Monem, A., & Shaalan, K. (2018). Using text mining techniques for extracting information from research articles. *Intelligent Natural Language Processing: Trends and Applications*, 373-397.
- Sawant, D., Jaiswal, A., Singh, J., & Shah, P. (2019). AgriBot - An intelligent interactive interface to assist farmers in agricultural activities. *2019 IEEE Bombay Section Signature Conference (IBSSC)*. Mumbai, India, India: IEEE.
- Sebastian, D. (2017). Rancang Bangun Website Klasifikasi Untuk Pencarian Produk Pasar Online Menggunakan Algoritma K-Nearest Neighbor. *Jurnal Teknik Informatika dan Sistem Informasi*, 3(3).

- Sebastian, D. (2017). Rancang Bangun Website Klasifikasi Untuk Pencarian Produk Pasar Online Menggunakan Algoritma K-Nearest Neighbor. *Jurnal Teknik Informatika dan Sistem Informasi*, 3(3).
- Tan, A.-H. (1999). Text Mining: The state of the art and the challenges. *Proceedings of the PAKDD 1999 Workshop on Knowledge Discovery from Advanced Databases*, 8.
- Tuten, T. L. (2008). *Advertising 2.0: social media marketing in a web 2.0 world: social media marketing in a web 2.0 world*. ABC-Clio.
- Udovychenko, Y., Popov, A., & Chaikovsky, I. (2015). Ischemic heart disease recognition by k-NN classification of current density distribution maps. *2015 IEEE 35th International Conference on Electronics and Nanotechnology (ELNANO)*. Kiev, Ukraine: IEEE.
- Vijayarani, S., Ilamathi, J., & Nithya. (2015). Preprocessing techniques for text mining-an overview. *International Journal of Computer Science & Communication Networks*, 5(1), 7-16.
- Yao, Z., & Ze-wen, C. (2011). Research on the construction and filter method of stop-word list in text preprocessing. *Fourth International Conference on Intelligent Computation Technology and Automation* (pp. 217-221). IEEE. doi:10.1109/ICICTA.2011.64
- Zaanen, M. V., & Kanters, P. (2010). Automatic Mood Classification Using TF*IDF Based on Lyrics. *11th International Society for Music Information Retrieval Conference (ISMIR 2010)*.
- Zong, Z., & Hong, C. (2018). On Application of Natural Language Processing in Machine Translation. *2018 3rd International Conference on Mechanical, Control and Computer Engineering (ICMCCE)*. Huhhot, China: IEEE.

