

ANALISIS SENTIMEN MENGGUNAKAN SPEECH RECOGNITION DENGAN ALGORITMA NAIVE BAYES CLASSIFIER

Sentiment Analysis Using Speech Recognition with The Naive Bayes Classifier Algorithm

Fikram Tamrin¹, Sakina Sudin², Gamaria Mandar³

^{1,2,3}Universitas Muhammadiyah Maluku Utara, Jl. KH Ahmad Dahlan No.100, Kel. Sasa, Kota Ternate
¹fikramtamrin1866@gmail.com, ²sakinahsudin80@yahoo.com, ³gamariamandar20@gmail.com

ABSTRAK

Analisis sentimen atau disebut juga opini adalah proses pemahaman, penggalian, dan analisis data tekstual secara otomatis untuk mendapatkan informasi sentimen yang terkandung dalam sebuah kalimat opini terhadap suatu isu oleh seseorang, yang cenderung memiliki opini negatif atau positif. Tujuan dari penelitian ini adalah mengklasifikasikan data mahasiswa dan mahasiswi di lingkungan sekitar kampus Muhammadiyah Maluku Utara menjadi 2 kategori, positif dan negatif. Pada penelitian ini, teks yang kami gunakan berbahasa Indonesia yang memuat data responden dari kalangan umum dan mahasiswa. Adapun pengembangan website layanan masyarakat untuk menampung opini dari masyarakat dan mahasiswa yang menggunakan fitur speech recognition untuk mengubah ucapan menjadi tulisan secara langsung (real-time) dalam bahasa Indonesia dengan input microphone. Opini publik yang terdapat pada website layanan publik dapat dijadikan sebagai bahan analisis apakah masyarakat dan mahasiswa memiliki sikap negatif terhadap keberadaan kampus Universitas Muhammadiyah Maluku Utara. Data yang digunakan terdiri dari 300 data yang terbagi menjadi 2 yaitu 220 untuk data latih dan 80 data untuk data uji. Klasifikasi data sentimen menggunakan text mining dengan Naive Bayes Classifier. Sebelum klasifikasi dilakukan beberapa langkah pengolahan kata, seperti: Case Folding, Cleaning, Stopword, Tokenizing dan Stemming. Hasil dari 80 data pengujian yang diklasifikasikan adalah 43 data sentimen positif dan 7 data negatif. Dapat diartikan bahwa 80 data uji yang terklasifikasi termasuk dalam kategori sentimen positif karena data negatif lebih kecil dari data dengan sentimen positif. Dapat diartikan bahwa data uji 80 termasuk dalam kategori sentimen karena data negatif lebih kecil dari sentimen positif. Akurasi pada algoritma Naive Bayes Classifier memiliki nilai sebesar 62,5%.

Kata kunci: Analisis Sentimen, Website, Speech Recognition

ABSTRACT

Sentiment analysis or also called opinion is the process of understanding, extracting, and analyzing textual data automatically to obtain sentiment information contained in a sentence of opinion on an issue by someone who tends to have a negative or positive opinion. The purpose of this research is to classify data on students and female students in the environment around the Muhammadiyah North Maluku campus into 2 categories, positive and negative. In this study, the text we used was in Indonesian which included data on respondents from the general public and students. As for the development of a community service website to accommodate opinions from the public and students who use the speech recognition feature to convert speech into writing directly (real-time) in Indonesian with a microphone input. Public opinion contained on the public service website can be used as material for analysis whether the community and students have a negative attitude towards the existence of the North Maluku Muhammadiyah University campus. The data used consists of 300 data which is divided into 2, namely 220 for training data and 80 data for test data. Sentiment data classification using text mining with Naive Bayes Classifier. Before classification, several word processing steps are carried out, such as: Case Folding, Cleaning, Stopword, Tokenizing and Stemming. The results of the 80 classified test data are 43 positive sentiment data and 7 negative data. It can be interpreted that the 80 classified test data are included in the positive sentiment category because negative data is smaller than data with positive sentiment. It can be interpreted that the 80 test data is included in the sentiment category because negative data is smaller than positive sentiment. The accuracy of the Naive Bayes Classifier algorithm has a value of 62.5%.

Keywords: Sentiment Analysis, Website, Speech Recognition

Pendahuluan

Analisis sentimen atau disebut juga opini adalah proses pemahaman, penggalian, dan analisis data tekstual secara otomatis untuk mendapatkan informasi sentimen yang terkandung dalam sebuah kalimat opini terhadap suatu isu oleh seseorang, yang cenderung memiliki opini negatif atau positif. Tujuan dari penelitian ini adalah mengklasifikasikan data mahasiswa dan mahasiswi di lingkungan sekitar kampus Muhammadiyah Maluku Utara menjadi 2 kategori, positif dan negatif. Pada penelitian ini, teks yang kami gunakan berbahasa Indonesia yang memuat data responden dari kalangan umum dan mahasiswa. Adapun pengembangan website layanan masyarakat untuk menampung opini dari masyarakat dan mahasiswa yang menggunakan fitur speech recognition untuk mengubah ucapan menjadi tulisan secara langsung (real-time) dalam bahasa Indonesia dengan input microphone. Opini publik yang terdapat pada website layanan publik dapat dijadikan sebagai bahan analisis apakah masyarakat dan mahasiswa memiliki sikap negatif terhadap keberadaan kampus Universitas Muhammadiyah Maluku Utara.

Data yang digunakan terdiri dari 300 data yang terbagi menjadi 2 yaitu 220 untuk data latih dan 80 data untuk data uji. Klasifikasi data sentimen menggunakan text mining dengan Naïve Bayes Classifier. Sebelum klasifikasi dilakukan beberapa langkah pengolahan kata, seperti: Case Folding, Cleaning, Stopword, Tokenizing dan Stemming.

Hasil dari 80 data pengujian yang diklasifikasikan adalah 43 data sentimen positif dan 7 data negatif. Dapat diartikan bahwa 80 data uji yang terklasifikasi termasuk dalam kategori sentimen positif karena data negatif lebih kecil dari data dengan sentimen positif. Dapat diartikan bahwa data uji 80 termasuk dalam kategori sentimen karena data negatif lebih kecil dari sentimen positif. Akurasi pada algoritma Naïve Bayes Classifier memiliki nilai sebesar 62,5%.

Metode

2.1. Teori yang berkaitan dengan penelitian

Seperti Penelitian pada tahun 2018 yang dilakukan oleh Adhi Viky Sudiantoro dan Eri Zuliarso tentang Analisis Sentimen Twitter Menggunakan Text Mining dengan Algoritma Naive Bayes Classifier, ada isu terkait pilkada menimbulkan banyak pendapat, tidak hanya opini positif juga opini negatif. Dari hasil uji Algoritma Naive Bayes Classifier sangat efektif digunakan sebagai proses klasifikasi tweet yang dibutuhkan dalam sistem analisis sentimen ini dimana nilai yang didapatkan dalam pengujian sampai 84%.

Penelitian lainnya pada tahun 2018 yang dilakukan oleh Muhammad Elra Ghifary R dan Richi Dwi Agustia tentang Pembangunan Aplikasi Transkrip Wawancara

Menggunakan Metode Speech Recognition. Wawancara yang dilakukan berbentuk wawancara mendalam (In-depth Interview) dan diskusi kelompok terarah (Focus Group Discussion) yang mana terdapat moderator sebagai pewawancara, responden sebagai terwawancara, dan notulis sebagai pencatat isi wawancara.

Penelitian yang dilakukan oleh (Gamaria dan Gunawan, 2018) mereka melakukan penelitian pada artikel berita dengan menerapkan metode Cross Latent Semantic Analysis (CLSA). Pada penelitian ini membandingkan metode LSA dan CLSA dengan menggunakan data yang berasal dari 240 artikel www.kompas.com.

Penelitian selanjutnya pada tahun 2022 yang dilakukan oleh Selva Indah Nurhafidah dan Falentino Sembiring tentang Analisis Sentimen Aplikasi Novel Online Di Google Play Store Menggunakan Algoritma Support Vector Machine. Membaca menurut Poerwadarminta adalah kegiatan yang meliputi melihat apa yang tertulis dan memahami atau mampu mengucapkan apa yang tertulis.

Pada tahun 2022 sebuah penelitian dari I Komang Setia Buana yang berjudul Implementasi Aplikasi Speech to Text untuk Memudahkan Wartawan Mencatat Wawancara dengan Python. Berita yang baik adalah berita yang dihasilkan dari sebuah penggalian informasi yang terampil. Oleh karena permasalahan tersebut, dibutuhkan alat perekam untuk merekam jawaban dari narasumber.

2.2. Multinomial Naive Bayes Classifier

Multinomial Naive Bayes Classifier merupakan model pengembangan dari algoritma bayes yang cocok dalam pengklasifikasian teks atau dokumen. Pada formula Multinomial Naive Bayes Classifier, kelas dokumen tidak hanya ditentukan dengan kata yang muncul tetapi juga jumlah kemunculannya (Witten & Frank, 2005).

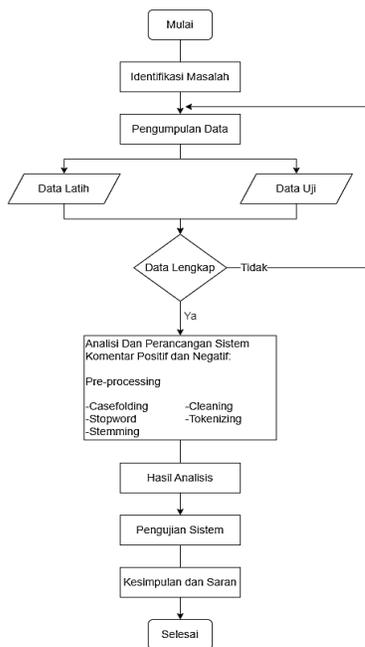
Inverse Document Frequency (IDF) adalah hubungan antara banyak dokumen yang memiliki kata (Document Frequency) dengan jumlah dokumen kalimat (D), sedangkan W adalah nilai bobot dari setiap kata pada sebuah dokumen, hasil dari pembobotan kata adalah sebagai dataset yang akan digunakan untuk tahap Naive Bayes Classifier (Gamaria, 2017).

2.3. Speech Recognition

Speech Recognition adalah sebuah proses dan teknologi yang berkaitan untuk mengubah sebuah percakapan atau speech ke dalam bentuk kata-kata yang sesuai atau entitas melalui algoritma tertentu yang diimplementasikan di dalam perangkat atau komputer. Aplikasi speech recognition sudah termasuk voice dialing, call routing, interactive voice response, data entry and dictation, voice command and control, gaming, structured document, appliance control by voice, computer-aided language learning, content-based spoken audio search, dan robotics (Hüb-Umbach dkk., 2015).

2.4. Metode Penelitian

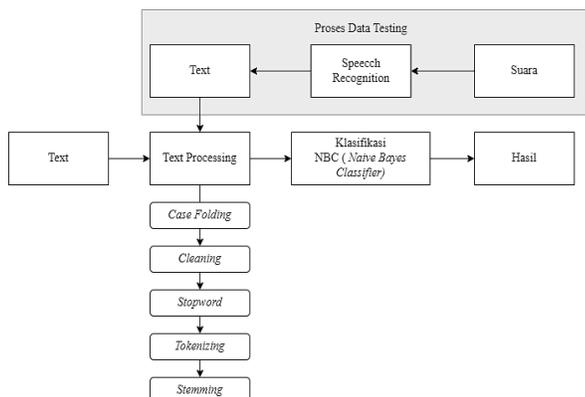
Penelitian Analisis Sentimen Menggunakan Speech Recognition Dengan Algoritma Naive Bayes Classifier Studi Kasus Universitas Muhammadiyah Maluku Utara dapat dilihat pada gambar skema di bawah ini:



Gambar 1. Metode Penelitian

2.5. Sistem yang diusulkan

Perancangan sistem yang diusulkan berupa aplikasi analisis sentimen dimana sistem tersebut menggunakan teknologi *Speech Recognition* untuk memproses suara dari responden menjadi sebuah text yang akan diproses dan mendapatkan hasil sentimen dari responden. Dengan menggunakan sistem ini dapat membantu dalam mengolah sebuah informasi secara langsung (*real time*) dan otomatis untuk merubah ucapan menjadi tulisan sehingga mempercepat untuk mendapatkan informasi.



Gambar 1. Perancangan Sistem Pengolahan Data Analisis Sentimen

Hasil Dan Pembahasan

Pembangunan aplikasi Analisis Sentimen yang menggunakan pengenalan ucapan atau *Speech Recognition* dengan menggunakan algoritma Naive Bayes Classifier studi kasus Universitas Muhammadiyah Maluku Utara, yang didasarkan pada rancangan yang telah dibahas sebelumnya.

3.1. Sisi Pengguna

3.1.1. Halaman Utama

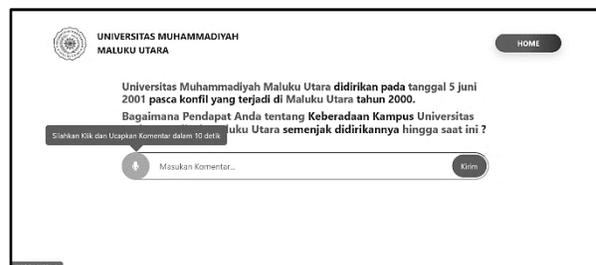
Tampilan utama dari sisi pengguna atau user pengunjung seperti masyarakat dan mahasiswa ataupun publik yang berkepentingan untuk memberikan pengaduan melalui website layanan publik Universitas Muhammadiyah Maluku Utara.



Gambar 2. Halaman Utama Pengguna

3.1.2. Halaman Pengaduan

Halaman Pengaduan memuat latar belakang singkat tentang Universitas Muhammadiyah Maluku Utara dan pertanyaan mengenai pendapat dari pengunjung pada website pelayanan publik tentang keberadaan kampus Universitas Muhammadiyah Maluku Utara, terdapat tombol untuk fitur *speech recognition* atau pengenalan ucapan untuk memudahkan pengunjung agar dapat memberikan pendapat atau komentar menggunakan



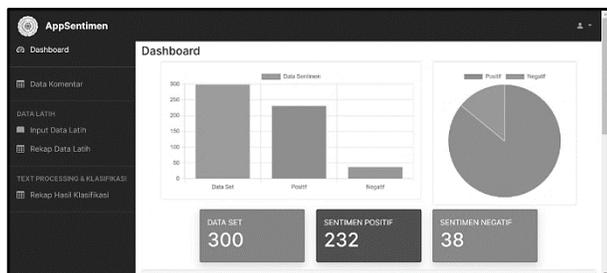
suara sebagai inputan.

Gambar 3. Fitur Speech Recognition

3.2. Sisi Pengelola

3.2.1. Halaman Utama

Halaman Utama terdapat informasi berupa visualisasi dalam bentuk grafik batang dan grafik pie yang memvisualisasikan data positif, negatif dan dataset, terdapat pula informasi berupa numerik yang menampilkan jumlah dataset, positif dan negatif.



Gambar 4. Tampilan Halaman Utama Pengelola

3.2.2. Halaman Hasil Klasifikasi

Halaman Hasil Klasifikasi dalam bentuk tabel, pada data tersebut hasil komentar telah diolah melalui Text Processing dan juga Classification menggunakan algoritma Naive Bayes Classifier untuk mengidentifikasi dan mengelompokkan data sentimen dari responden, pada halaman tersebut juga terdapat fitur *Print* dan *Search*.

The screenshot shows a table titled 'Hasil Text Processing dan Klasifikasi' with columns for 'ID', 'Komentar', and 'Kategori'. It lists 8 comments and their corresponding sentiment categories.

ID	Komentar	Kategori
1	sangat bagus segi bangun didik mahasiswa nya	positif
2	sangat lucu anak bangsa	positif
3	sangat bagus luas tingkat mutu didik	positif
4	bagus	positif
5	punya komentar	negatif
6	sangat bagus	positif
7	sangat manfaat sekali orang ingin kejar ilmu daerah malu, sekitar perlu pergi jauh	positif
8	terru sudah cukup	positif

Gambar 5. Hasil Klasifikasi

3.3. Pengujian

Hasil pengujian Algoritma Naive Bayes Classifier sebagai tahap klasifikasi analisis sentimen pelayanan publik mengenai keberadaan kampus Universitas Muhammadiyah Maluku Utara sangat diperlukan dalam sistem analisis sentimen ini dimana nilai yang didapat pada pengujian tersebut adalah sampai dengan 62,5%, berdasarkan data uji dari 80 responden, 53, 75% termasuk dalam kategori positif, 8,75% termasuk dalam kategori negatif, artinya dari 80 data uji sebanyak 43 data responden termasuk dalam kategori positif, 7 data responden dalam kategori negatif, sedangkan 30 data responden tidak termasuk dalam kategori positif atau negatif.

Simpulan Dan Saran

Telah dikembangkan sistem analisis sentimen berbasis website sebagai sarana pelayanan publik di Universitas Muhammadiyah Maluku Utara dengan menggunakan bahasa pemrograman Python dan MySQL. Berdasarkan hasil analisis, perancangan, dan pengujian yang didapatkan, sistem analisis sentimen pelayanan publik di Universitas Muhammadiyah Maluku Utara menggunakan speech recognition untuk memudahkan pengunjung dalam memberikan komentar atau sentimen tentang keberadaan Kampus Universitas Muhammadiyah Maluku Utara.

Analisis sentimen ini dapat diperluas tidak hanya terbatas pada masyarakat atau mahasiswa yang berada di sekitar Kampus Universitas Muhammadiyah Maluku Utara. Algoritma untuk proses analisis ini juga dapat menggunakan algoritma data mining lainnya, sehingga diperoleh perbandingan hasil analisis antara Naive Bayes Classifier dengan algoritma lain seperti *Support Vector Machine* atau SVM.

Pustaka Acuan

- Anwar, A. (2019). *Pengertian dan Sejarah Bahasa Pemrograman Python*. ilmuti.org.
- Buana, I. K. S. (2020). Implementasi Aplikasi Speech to Text untuk Memudahkan Wartawan Mencatat Wawancara dengan Python. *Jurnal Sistem Dan Informatika (JSI)*. <https://doi.org/10.30864/jsi.v14i2.293>
- Chauhan, N., Singh, M., Verma, A., Parasher, A., & Budhiraja, G. (2019). Implementation of database using python flask framework: college database management system. *International Journal of Engineering and Computer Science*. <https://doi.org/10.18535/ijecs/v8i12.4390>
- Ghifary, M. E., & Agustia, R. D. (2018). Pembangunan Aplikasi Transkrip Wawancara Menggunakan Metode Speech Recognition.
- Hüb-Umbach, R., Li, J., Deng, L., & Gong, Y. (2015). *Robust Automatic Speech Recognition: A Bridge to Practical Applications*. Elsevier Science.
- Kadir, A. (2018). *Dasar pemrograman internet untuk proyek berbasis arduino*. Andi.
- Kristanto, A. (2018). *Perancangan sistem informasi dan aplikasinya*. Gava Media.

- Lane, H., Howard, C., Hapke, H. M., & Hapke, H. (2019). *Natural Language Processing in Action: Understanding, Analyzing, and Generating Text with Python*. Manning.
- Manning, C. D., Raghavan, P., & Schütze, H. (2008). *Introduction to information retrieval* (P. Raghavan & H. Schütze, Eds.). Cambridge University Press.
- Mandar Gamaria, Gunawan (2017). *Peringkasan Dokumen Berita Bahasa Indonesia Menggunakan Metode Cross Latent Semantic Analysis*. Register Vol 3 No 2 . DOI: 10.26594/register.v3i2.1161
- Mandar Gamaria, Gunawan (2018). *Kajian Pemanfaatan Metode Klasifikasi Data Mining Pada Pelayanan Perpustakaan*. JTIFA Vol 1 No 1 DOI: <https://doi.org/10.52046/j-tifa.v1i1.117>
- Mandar Gamaria, Abdul Haris Muhaamad dkk (2020). *Klasifikasi Berita Indonesia Menggunakan Naiuve Bayes dengan Porter Stemmer*. J-TIFA Vol 3 No 2 . DOI: <https://doi.org/10.52046/j-tifa.v3i2.1121>
- Nurhafidah, S. I., & Sembiring, F. (2022). Analisis Sentimen Aplikasi Novel Online Di Google Play Store Menggunakan Algoritma Support Vector Machine (SVM). *J-SAKTI (Jurnal Sains Komputer & Informatika)*. <http://dx.doi.org/10.30645/j-sakti.v6i1.447>
- Rachman, F. H. (2020). *Komputasi Bahasa Alami*. Media Nusa Creative (MNC Publishing).
- Raj, A. (2018, February 9). *Interfacing MQ137 Sensor with Arduino to Measure Ammonia in PPM*. Circuit Digest. Retrieved April 9, 2022, from <https://circuitdigest.com/microcontroller-projects/arduino-mq137-ammonia-sensor>
- Stubbs, A., & Pustejovsky, J. (2012). *Natural Language Annotation for Machine Learning*. O'Reilly Media, Incorporated.
- Sudiantoro, A. V., & Zuliarso, E. (2018). ANALISIS SENTIMEN TWITTER MENGGUNAKAN TEXT MINING DENGAN ALGORITMA NAÏVE BAYES CLASSIFIER. *Prosiding SINTAK*.
- Witten, I. H., & Frank, E. (2005). *Data Mining: Practical Machine Learning Tools and Techniques, Second Edition*. Elsevier Science.
- XHEMALI, D., HINDE, C. J., & STONE, R. G. (2009). Naïve Bayes vs. Decision Trees vs. Neural Networks in the Classification of Training Web Pages. *International Journal of Computer Science Issues*.