

IMPLEMENTASI RANDOM FOREST DALAM MELAKUKAN KLASIFIKASI KATA SARKASME PADA MEDIA SOSIAL FACEBOOK

Chikitha Syahrika Arsyah¹, Marina Elsera²

1,2) Sistem Informasi, Fakultas Teknik dan Komputer, Universitas Harapan Medan, Indonesia

Article Info

Article history:

Received: 07 Juni 2023

Revised: 15 Juni 2023

Accepted: 03 Juli 2023

ABSTRACT

Abstrak

Perkembangan teknologi informasi beberapa tahun belakangan ini berkembang dengan kecepatan yang sangat tinggi, teknologi sebagai sarana untuk mengakses dan menyebarkan informasi telah banyak digunakan, salah satunya menggunakan media sosial facebook. facebook dalam penelitian ini digunakan sebagai sumber data untuk menganalisis tweet berbahasa Indonesia terhadap pembelajaran online e-learning. Analisis sentimen pada Facebook digunakan untuk melihat kecenderungan respon masyarakat apakah berkecenderungan positif, netral atau negatif berdasarkan hasil tweet terhadap pembelajaran online e-learning. Analisis sentimen ini menggunakan metode random forest yang merupakan metode pengklasifikasian. Tahap klasifikasi dilakukan setelah melalui proses preprocessing, dimana hasil klasifikasi tweet berkecenderungan positif, netral atau negatif, menggunakan metode *random forest*. Akurasi yang didapat pada analisis sentimen terhadap pembelajaran online e-learning adalah 53.23%.

Kata Kunci: E-Learning, facebook, random forest

Abstract

The development of information technology in recent years has grown at a very high speed, technology as a means to access and disseminate information has been widely used, one of which is using social media Facebook. Facebook in this study is used as a data source to analyze Indonesian-language tweets on e-learning online. Sentiment analysis on Facebook is used to see the tendency of the community's response to be positive, neutral or negative based on the results of tweets on e-learning online learning. This sentiment analysis uses the random forest method which is a classification method. The classification stage is carried out after going through the preprocessing process, where the results of tweet classification tend to be positive, neutral or negative, using the random forest method. The accuracy obtained in the sentiment analysis of e-learning online learning is 53.23%.

Keywords: E-Learning, facebook, random forest

Djtechno: Jurnal Teknologi Informasi oleh Universitas Dharmawangsa Artikel ini bersifat open access yang didistribusikan di bawah syarat dan ketentuan dengan Lisensi Internasional Creative Commons Attribution NonCommercial ShareAlike 4.0 ([CC-BY-NC-SA](https://creativecommons.org/licenses/by-nc-sa/4.0/)).



Corresponding Author:
Email : chikithaarsya@gmail.com

1. PENDAHULUAN

Bahasa merupakan salah satu alat komunikasi antar manusia yang paling efektif dan banyak digunakan (Indrayanti, 2017). Pada era teknologi informasi ini komunikasi antar masyarakat sering dilakukan pada media sosial seperti facebook yang merupakan sarana komunikasi dalam dunia maya yang digandrungi oleh banyak kalangan yang kemudian menimbulkan opini opini beragam dimasyarakat (Asrial et al., 2019). dalam permasalahan ini facebook dapat digunakan dan dimanfaatkan untuk berbagai tujuan antara lain promosi, pembentukan opini oleh sebab itu facebook dapat digunakan sesuai dengan kebutuhannya. sehingga penggunaan kalimat atau kata pada media sosial facebook ada yang mengandung kalimat yang tidak santun bahkan mengarah pada sarkasme pada status media sosial facebook banyak ditemukan. sarkasme merupakan ungkapan yang memiliki nada ironi dan sindiran.

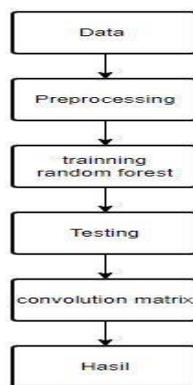
Permasalahan pada penelitian ini terdapat banyak kata sarkasme pada media sosial facebook kata sarkasme tersebut berasal dari pengguna facebook yang melakukan postingan status, penelitian ini akan mengambil data dari status facebook dengan kata kunci pembelajaran online kemudian akan melakukan klasifikasi sentimen terhadap status, klasifikasi dilakukan dengan tujuan untuk mengelompokan status yang mengandung kata sarkasme dan tidak mengandung kata sarkasme. Dalam melakukan pengambilan data status penelitian ini akan menggunakan teknik crawling dengan mengambil data sebanyak 2 tahun kemudian dilakukan preprocessing dan tahap selanjutnya menggunakan algoritma random forest, random forest merupakan algoritma dalam machine learning yang digunakan untuk pengklasifikasian data set dalam jumlah besar. Karena fungsinya bisa digunakan

untuk banyak dimensi dengan berbagai skala dan performa yang tinggi (Belgiu & Drăguț, 2016). Klasifikasi ini dilakukan melalui penggabungan tree dalam decision tree dengan cara training dataset yang Anda miliki .

Penelitian terdahulu yang dilakukan oleh (Alita & Isnain, 2020) dengan judul Pendeteksian Sarkasme pada Proses Analisis Sentimen Menggunakan Random Forest Classifier menyimpulkan bahwa algoritma random forest dapat melakukan klasifikasi terhadap kata sarkasme pada proses sentimen analisis dengan akurasi 95%. Adapun penelitian terdahulu yang dilakukan oleh (Septiani & Sibaroni, 2019) yang berjudul Sentiment Analysis Terhadap Status Bernada Sarkasme Berbahasa Indonesia menyimpulkan bahwa sentimen analisis sangat dipengaruhi oleh jumlah data dan kamus kata sarkasme, algoritma yang digunakan pada penelitian ini dapat melakukan deteksi kata sarkasme. Algoritma yang digunakan pada penelitian ini dapat melakukan deteksi kata sarkasme. penelitian terdahulu yang dilakukan oleh (Putra, 2021) yang berjudul Klasifikasi Komentar Terhadap Polisi Pada Media Sosial Youtube Menggunakan Support Vector Machine menyimpulkan bahwa akurasi yang dihasilkan oleh algoritma support vector machine sebesar 94% sehingga dapat melakukan klasifikasi komentar pada media sosial dengan baik.

2. METODE PENELITIAN

Pada rancangan penelitian merupakan rangkaian dan alur penelitian yang merupakan tahapan-tahapan dalam membangun sistem deteksi sarkasme. Berikut ini rancangan penelitian yang penulis buat yang terdapat pada gambar 3.1



Gambar 1. Kerangka Kerja Penelitian

Keterangan gambar 3.1 menjelaskan bahwa dalam kerangka kerja penelitian terdapat tahapan seperti berikut ini :

1. Data masukan diambil dari hasil crawling. Pada data latih aspek sudah ditentukan bersama sentimennya.
2. Kemudian data dilakukan preprocessing untuk menghilangkan tanda baca dan emoticon
3. Setelah didapatkan data dari hasil preprocessing kemudian dilakukan tahapan training dengan menggunakan data yang sudah di berikan label
4. Kemudian dilakukan tahapan testing dengan menggunakan parameter pengukuran convutional matrix

3. HASIL DAN PEMBAHASAN

Pada bagian ini akan dipaparkan hasil dari implementasi proses pembuatan dan implementasi sistem klasifikasi kata positif dan negative dengan Teknik sentiment analysis menggunakan algoritma Random Forest

1. Pengumpulan Data

Pengumpulan data yang digunakan pada penelitian ini adalah data facebook dari tahun 2020 sampai dengan 2022. Setelah dilakukan proses crawling maka didapatkan data sebanyak 25.000 data. Kode untuk melakukan crawling data dapat dilihat pada gambar berikut.

```
Untitled4.ipynb
File Edit View Insert Runtime Tools Help

+ Code + Text

import requests
import json

facebookUserAccessToken = "YOUR_USER_ACCESS_TOKEN"

def getFacebookCampaignData(pageID=None, ID=None, paid=None):
    # Let's segregate data collection into two
    # 1. Data Collected using AdvertID (Paid Campaigns)
    # 2. Data Collected using PostID (Unpaid Campaigns)

    advertIDFlag = False
    postIDFlag = False
    objectIDFlag = False

    # Allocating ID for Paid and Unpaid campaigns respectively
    if(paid==False):
        postID = ID
        advertID = None
        postIDFlag = True
    else:
        advertID = ID
        postID = None
        advertIDFlag = True

    # Creating structure to log data from facebook Servers
    facebookCampaignData = {}
    # The Base URL
    baseURL = "https://graph.facebook.com/v2.10/"

    if(advertIDFlag):
```

Gambar 2. Kode Crawling Data Facebook

Hasil dari pengumpulan data dari facebook yang diambil menggunakan proses crawling. Data diambil dari facebook menggunakan Bahasa pemrograman python. Data yang diambil hanya datetime , dan tweet. Data tersebut kemudian disimpan kedalam file csv seperti pada gambar di bawah ini untuk memudahkan tahapan atau proses selanjutnya

created_at	tweet	language	mentions
2020-11-18 23	@sunnova1324 @tanyainr lya sih... tapi maksud sender kuliah online kali	in	[]
2020-11-18 23	males kuliah online temennya sikit	in	[]
2020-11-18 23	tumbenan td kuliah online dosennya minta join zoom, trs temen ngcht nyuruh masuk krm yg lain udh join, tinggal gua doang yg blm, Okedeh.... cepet cepet,cuma pakek daster + ki in	in	[]
2020-11-18 22	@monsoulei nangis krm kecapean kuliah online https://t.co/U1RopoEqNI	in	[]
2020-11-18 22	Apa hanya aku yang merasa semenjak kuliah online, kinerja otakku menurun	in	[]
2020-11-18 22	Sumpah gak kebayang kalo kuliah online sampe tahun depan, memori kepehaman cuma hanya kebanyakan dokumen. 0y"	in	[]
2020-11-18 22	@safeerr Truseeee Kuliah online lebih ribet dan nyusahin banget. Bikin nangis tiap malam.	in	[]
2020-11-18 22	@samsungID Ga hanya seru utk hiburan. #GalaxyTabA7 ini jg menunjang utk produktivitas harian kita. Kuliah online / meeting online seharisan, gakan ada istilah "lowbat" karena s/in	in	["screen_
2020-11-18 22	Bakat kuliah online = jadi yutuber	in	[]
2020-11-18 22	hehe gimana udah ngga gadek bgt sm kuliah online: https://t.co/lanjPACjY	in	[]
2020-11-18 22	sumpah kuliah online anjim bgt sli tugasnya 70% belajarnya 20% sisanya kosong:) capenya dapet pt otaknya kosong0y"	in	[]
2020-11-18 22	Kemakalran gua selama kuliah online adalah cabut kelas tanpa rasa bersalah0y"	in	[]
2020-11-18 22	Tuhan, terima kasih. Kau sudah memperbolehkan hambamu ini untuk kuliah online mengenai Ewald Summatton. And Alex just called me beautiful 0y"	in	[]
2020-11-18 22	Tonton "JAM TIDUR KEBALUK? KULIAH ONLINE AMERIKA TAPI DI INDONESIA!! (UW Seattle)" di YouTube https://t.co/NT3VSpRHxH	in	[]
2020-11-18 22	@we1rdgrl Tapi menurutku ya, coba di tengok dari 2 sisi. penyelenggara rnya juga membuat tatib untuk meminimalisir kecurangan. Yg kita semua tau, selama kuliah online gini, cu in	in	[]
2020-11-18 22	nongkrong, rebahan, kuliah online, haha hihi gak kerasa aja tahun depan gue udah masuk semester 8 aja anjir.	in	[]
2020-11-18 22	Rasanya sia" saya tiap hari ikutan kuliah online, gk ada yang mau nyantol k otak0y"	in	[]
2020-11-18 22	#NAME?	in	[]
2020-11-18 21	aku - kuliah online ti suuk nani neuting si mamah - siza nu beah uue diajar	in	[]

Gambar 3. Pengumpulan Data Kedalam File CSV

2. Proses Akuisisi Data

Berikut ini tabel dari hasil crawling terhadap data status yang mengandung kata pembelajaran daring/online.

Tabel 1. Data Status tentang pembelajaran online

teks yang akan diolah masih bercampur dengan karakter karater lainnya masih melekat pada data tersebut.

4. SIMPULAN

Berdasarkan penelitian yang telah dilakukan, diperoleh bahwa penelitian tentang klasifikasi sentimen pengguna facebook terhadap topik pembelajaran e-learning dengan menggunakan metode random forest dengan kesimpulan sebagai berikut:

1. Metode random forest dapat diterapkan untuk melakukan analisis sentimen terhadap topik penelitian pembelajaran e-learning.
2. Kosakata tidak baku mempengaruhi hasil klasifikasi terhadap suatu kelas data uji apabila suatu kelas data latih memiliki data jumlah kata tidak baku lebih banyak dibandingkan kelas data latih lainnya.
3. Sistem belum mampu memfilter kata-kata asing pada tweet yang diambil berupa campuran bahasa Indonesia dan asing.
4. Hasil uji akurasi menghasilkan nilai akurasi terbesar, yaitu adalah 53,23%.

UCAPAN TERIMAKASIH

Puji dan syukur penulis mengucapkan kepada Tuhan Yang Maha Kuasa atas limpahan berkat, rahmat serta kemudahan yang telah diberikan sehingga penulis dapat menyelesaikan Skripsi ini yang merupakan syarat untuk mendapatkan gelar Sarjana Komputer pada Program Studi Sistem Informasi, Fakultas Teknik dan Komputer, Universitas Harapan Medan.

1. Rektor Universitas Harapan Medan Bapak Drs. Sriadhi, S.T., M.Pd, M.Kom, Ph.D
2. Dekan Fakultas Teknik dan Komputer Bpk. Dodi Siregar, S.T., M.Kom
3. Bapak Ahmad Zakir, S.T., M.Kom selaku Ketua Program Studi Sistem Informasi Fakultas Teknik dan Komputer Universitas Harapan Medan.
4. Bapak Andi Marwan Elhanafi, S.T., M.Kom selaku Sekretaris Program Studi Sistem Informasi, Fakultas Teknik dan Komputer Universitas Harapan Medan.

5. Bapak Dr. Arif Ridho Lubis, M.Sc.IT selaku pembimbing I dan Ibu Marina Elsera, S.T., M.Kom selaku pembimbing II yang telah meluangkan waktu membimbing penulis selama pengerjaan Skripsi ini.
6. Semua pihak yang tidak dapat disebutkan satu persatu, terima kasih atas segalanya.

PUSTAKA

- Algoritma K-Means. Techno. Com, 16(1), 1–8.
- Alita, D., & Isnain, A. R. (2020). Pendeteksian Sarkasme Pada Proses Analisis Sentimen Menggunakan Random Forest Classifier. *Jurnal Komputasi*, 8(2), 50– 58.
- Aprilia, R., & Sari, R. F. (2017). Implementation Of Pagerank Algorithm In Matlab. *Zero: Jurnal Sains, Matematika Dan Terapan*, 1(1), 33–40. <https://doi.org/10.30829/Zero.V1i1.1458>
- Asrial, A., Syahril, S., Kurniawan, D. A., & Amalina, N. (2019). Analisis Hubungan Kompetensi Bahasa Indonesia Terhadap Kompetensi Pedagogik Mahasiswa Pendidikan Guru Sekolah Dasar. *Premiere Educandum: Jurnal Pendidikan Dasar Dan Pembelajaran*, 9(1), 1–8.
- Belgiu, M., & Drăguț, L. (2016). Random Forest In Remote Sensing: A Review Of Applications And Future Directions. *Isprs Journal Of Photogrammetry And Remote Sensing*, 114, 24–31.
- Budi, S. (2017). Text Mining Untuk Analisis Sentimen Review Film Menggunakan
- Rahmah, S. A., & Antares, J. (2021, November). Application of K-Means Clustering Algorithm in Determining Prospective Students Receiving Foundation Scholarship. In *2021 International Conference on Computer Science and Engineering (IC2SE)* (Vol. 1, pp. 1-7). IEEE.