



IMPLEMENTASI CONTENT MODERATION DALAM SOCIAL MEDIA INSTAGRAM UNTUK DETEKSI CYBERBULLYING DENGAN MACHINE LEARNING BERBASIS CLOUD

Wahyu Hadikristanto¹

¹Program Studi Teknik Informatika, Fakultas Teknik, Universitas Pelita Bangsa.

wahyu.hadikristanto@pelitabangsa.ac.id

Jl. Inspeksi Kalimalang No.9, Cibatu, Cikarang Selatan, Kabupaten Bekasi, Jawa Barat

Keywords:

Content
Moderation,
CYberbullying,
Machine
learning

Kata Kunci:

Content
Moderation,
CYberbullying,
Machine
learning

Pendahuluan

Media sosial adalah platform di internet yang memungkinkan pengguna untuk berkomunikasi, berbagi informasi, dan terhubung dengan orang lain. Media sosial telah menjadi bagian penting dari kehidupan sehari-hari banyak orang di seluruh dunia[1]. Pengguna media sosial dapat berinteraksi dengan teman dan keluarga, bertemu orang baru, dan mengikuti akun yang berisi informasi yang mereka sukai.

Instagram adalah salah satu media sosial yang paling banyak digunakan di seluruh dunia. Platform ini memungkinkan pengguna untuk berbagi foto dan video dengan teman dan pengikut mereka, serta mengikuti akun yang berisi foto dan video yang mereka sukai. Instagram juga menyediakan fitur untuk

Abstrat

Social media is a platform on the internet that allows users to communicate, share information and connect with others. Instagram is one of the most used social media worldwide. Cyberbullying is the act of torturing, threatening or torturing someone online. Machine learning is a technology that enables computers to learn and develop capabilities without being given explicit instructions. Machine learning can be used in various fields, including content moderation to prevent cyberbullying. Overall, research on content moderation with cloud-based machine learning is quite efficient, considering the accuracy value obtained is 85%.

Abstrak

Media sosial adalah platform di internet yang memungkinkan pengguna untuk berkomunikasi, berbagi informasi, dan terhubung dengan orang lain. Instagram adalah salah satu media sosial yang paling banyak digunakan di seluruh dunia. *Cyberbullying* adalah tindakan menyiksa, mengancam, atau menyiksa seseorang secara online. *Machine learning* adalah teknologi yang memungkinkan komputer untuk mempelajari dan mengembangkan kemampuan tanpa diberi instruksi secara eksplisit. *Machine learning* dapat digunakan dalam berbagai bidang, termasuk *content moderation* untuk mencegah *cyberbullying*. Secara keseluruhan, penelitian tentang content moderation dengan machine learning berbasis *cloud* cukup efisien, mengingat nilai akurasi yang diperoleh sebesar 85%.

menambahkan *caption*, *tag*, dan *hashtag* ke postingan, sehingga mudah untuk menemukan konten yang sesuai dengan minat pengguna.

Sosial media telah menjadi bagian yang tidak terpisahkan dari kehidupan sehari-hari banyak orang di seluruh dunia. Namun, meskipun sosial media memiliki banyak manfaat, terdapat juga dampak buruk yang dapat ditimbulkan oleh penggunaannya. Salah satu dampak buruk yang paling menonjol adalah *cyberbullying*.

Cyberbullying adalah tindakan menyiksa, mengancam, atau menyiksa seseorang secara online. Ini dapat terjadi melalui pesan teks, posting di media sosial, atau melalui email. *Cyberbullying* dapat menjadi sangat menyakitkan bagi korban, terutama karena ia terjadi di dunia maya, yang dapat membuat

korban merasa terisolasi dan tidak memiliki tempat yang aman.

Dampak dari *cyberbullying* dapat sangat buruk bagi korban. Ini dapat menyebabkan depresi, kecemasan, dan bahkan bunuh diri. *Cyberbullying* juga dapat memiliki dampak yang luas pada keluarga dan teman-teman korban.

Machine learning adalah teknologi yang memungkinkan komputer untuk mempelajari dan mengembangkan kemampuan tanpa diberi instruksi secara eksplisit. *Machine learning* dapat digunakan dalam berbagai bidang, termasuk *content moderation* untuk mencegah *cyberbullying*.

Content moderation adalah proses mengelola konten yang diterbitkan di internet untuk memastikan bahwa konten tersebut sesuai dengan standar yang ditetapkan. Dalam konteks *cyberbullying*, *content moderation* dapat digunakan untuk memblokir atau menghapus pesan yang mengandung kekerasan, ancaman, atau tindakan menyiksa.

Machine learning dapat digunakan untuk membantu dalam proses *content moderation* dengan menganalisis konten yang diterbitkan secara otomatis dan mengidentifikasi pesan yang tidak sesuai dengan standar. Dengan menggunakan algoritma *machine learning*, komputer dapat mempelajari pola dan karakteristik pesan yang mengandung *cyberbullying* dan mengambil tindakan yang sesuai seperti memblokir atau menghapus pesan tersebut.

dalam penelitian ini akan dilakukan pembuatan *content moderation* untuk mendeteksi *cyberbullying* dalam komentar instagram menggunakan *machine learning* berbasis *cloud*.

Landasan Teori

Text mining

Text mining adalah proses mengekstrak informasi yang berguna dari data teks yang terstruktur atau tidak terstruktur. *Text mining* dapat digunakan dalam berbagai bidang, termasuk bisnis, politik, dan ilmu pengetahuan. Dengan menggunakan teknik *text mining*, kita dapat mengidentifikasi pola dan trend yang tersembunyi di dalam data teks, serta

mengekstrak informasi yang berguna untuk membuat keputusan atau mengambil tindakan.

Text mining menggunakan berbagai teknik, termasuk pembersihan data, pengenalan entitas, dan analisis sentimen, untuk mengolah data teks dan mengekstrak informasi yang berguna. *Text mining* juga dapat menggunakan algoritma *machine learning* untuk mempelajari pola dan karakteristik data teks dan mengambil tindakan yang sesuai.[2]

Pre-processing

Preprocessing merupakan tahap penting dalam proses *text mining*, di mana data teks yang akan dianalisis dibersihkan dan disiapkan untuk proses selanjutnya. Tahap ini meliputi penghapusan karakter yang tidak diinginkan, seperti tanda baca atau angka, serta penghapusan kata-kata yang tidak relevan atau tidak berguna, seperti kata umum atau *stop words*[3]

Case Folding

Case folding adalah proses mengubah semua huruf dalam suatu teks menjadi huruf kecil atau huruf besar. *Case folding* berguna untuk menghilangkan perbedaan antara huruf kecil dan besar dalam teks, sehingga teks dapat dianalisis dengan lebih mudah.[4]

Stemming

Stemming adalah proses penyederhanaan kata ke dasar katanya, sehingga kata-kata yang memiliki makna yang sama dapat dianggap sebagai satu kata. Contohnya, "melompat" dan "melompati" dianggap sebagai satu kata setelah di-stem. *Stemming* berguna untuk mengurangi variasi dalam kata-kata yang memiliki makna yang sama, sehingga data teks dapat dianalisis dengan lebih mudah.[5]

Cyberbullying

Cyberbullying adalah tindakan kekerasan yang dilakukan melalui media elektronik, seperti internet atau telepon seluler. *Cyberbullying* dapat terjadi kepada siapa saja, tetapi sering terjadi pada anak-anak dan remaja yang lebih mudah terpengaruh oleh lingkungan sekitarnya.

Di Indonesia, kasus *cyberbullying* semakin meningkat dari tahun ke tahun. Menurut data dari Komisi Perlindungan Anak Indonesia, terdapat lebih dari 10.000 kasus *cyberbullying*

yang dilaporkan pada tahun 2020. Kasus ini terjadi karena adanya perkembangan teknologi yang semakin pesat, sehingga anak-anak dan remaja mudah terpapar dan terpengaruh oleh media sosial. [6]

Hasil dan Pembahasan

Machine learning adalah teknologi yang memungkinkan komputer untuk mempelajari dan mengembangkan kemampuan tanpa diberi instruksi secara eksplisit. Machine learning dapat digunakan dalam berbagai bidang, seperti pengenalan wajah, pengenalan suara, atau pengenalan bahasa.[7]

Salah satu kelebihan dari machine learning adalah kemampuannya untuk mengelola dan mengolah data yang sangat besar. Namun, untuk mengelola dan mengolah data yang sangat besar, diperlukan komputer yang memiliki spesifikasi tinggi, yang tentunya tidak murah. Selain itu, pengelolaan dan pengolahan data yang besar juga dapat memakan waktu yang lama.

Untuk mengatasi masalah tersebut, muncullah machine learning tool berbasis cloud. Machine learning tool berbasis cloud merupakan aplikasi atau platform yang memungkinkan pengguna untuk mengelola dan mengolah data dengan menggunakan cloud computing. Dengan menggunakan machine learning tool berbasis cloud, pengguna tidak perlu memiliki komputer dengan spesifikasi tinggi, karena proses pengolahan data dilakukan di server yang terhubung dengan internet.[8]

Berikut ini adalah beberapa contoh penyedia machine learning berbasis cloud: Amazon Web Services (AWS), Google Cloud, dan Microsoft Azure dan juga nyckel.com. dalam penelitian ini akan menggunakan tool yang disediakan oleh nyckel.com



Gambar 1. Tampilan Home

Dataset adalah kumpulan data yang terorganisir dan disimpan secara sistematis, yang biasanya digunakan untuk keperluan analisis atau pembelajaran mesin. Dataset terdiri dari beberapa baris dan kolom, dimana setiap baris merupakan observasi, dan setiap kolom merupakan variabel.[9]

Dataset analisis sentimen dikumpulkan dari komentar sosial media Instagram dan dilabeli secara manual dengan label "positive" dan "negative". Jumlah total dataset terdapat 400 komentar dengan pembagian masing-masing kelas label sejumlah 200 komentar [10]. Sebelum dipublikasikan, dataset ini telah dilakukan beberapa praproses sebagai berikut:

Mengganti setiap username (@username) yang ada di teks komentar dengan tag <USERNAME>.

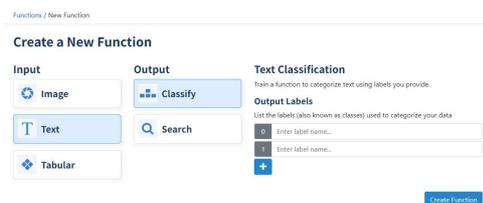
Kode regular expression untuk mendeteksi kemunculan username adalah:

```
(?<=^| (?<=[^a-zA-Z0-9_ \. ])) @ ([A-Za-z]+ [A-Za-z0-9_ ]+)
```

Machine learning berbasis cloud merupakan teknologi yang memungkinkan pengguna untuk mengelola dan mengolah data dengan menggunakan cloud computing. Dengan menggunakan machine learning berbasis cloud, pengguna dapat mengakses data dan hasil pengolahan data secara real-time, serta memiliki fitur-fitur yang memudahkan pengguna untuk mengelola dan mengolah data.

Langkah-langkah menggunakan machine learning berbasis cloud adalah sebagai berikut:

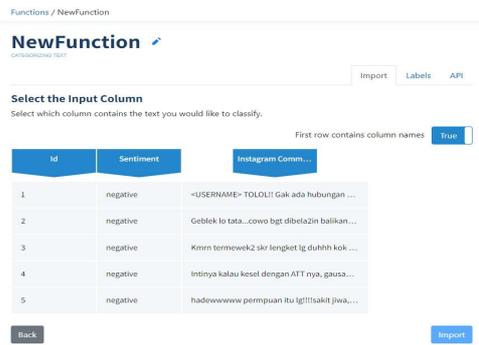
- 1. masuk dan pilih menu input tek dengan fungsi klasifikasi
- 2. masukkan label yang akan digunakan / dipakai. Dipenelitian ini menggunakan 2 label yaitu positif dan negative



Gambar 2. Pembuatan Fungsi

Dataset

3. kemudain upload dataset yang akan diolah dan kemudian pilih field mana yang akan jadi inputan.



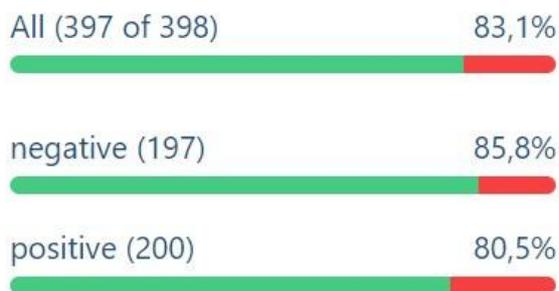
Gambar 3. Proses Upload Dataset

4. kemdian klik import dan system akan melakukan proses training selama beberapa menit.



Gambar 4. Hasil Training Data

Terakhir akan muncul hasil nilai akurasi klasifikasis sperti gambar dibawah ini;



Gambar 5. Hasil Akurasi Klasifikasi

Dari gambar diatas dapatdisimpulkan system cukup baik dalam mengenali / deteksi cyberbullying di label negative cukup tinggi yaitu sebsar 85 %.

Selanjutnya dapat diimplementasikan dalam sebuah program karena sudah disediakan API (Aplication programing Interfaces) yang bisa kita panggil dalam Bahasa pememrohraman apa saja.



Gambar 6. Contoh Input Data

Kesimpulan dan Saran

Dari pemhasan diatas dapat disimpulkan bahwa Penelitian tentang content moderation dengan machine learning berbasis cloud cukup efisien, mengingat nilai akurasi yang diperoleh sebesar 85%. Nilai akurasi yang tinggi menunjukkan bahwa hasil content moderation yang dilakukan dengan machine learning berbasis cloud cukup baik, dan dapat diandalkan dalam mengelola konten di media sosial atau platform lainnya.

Secara keseluruhan, penelitian tentang content moderation dengan machine learning berbasis cloud cukup efisien, mengingat nilai akurasi yang diperoleh sebesar 85%. Namun, penggunaan machine learning dalam content moderation juga memiliki beberapa kelemahan, seperti masalah bias. Oleh karena itu, perlu dipertimbangkan dengan seksama apakah penggunaan machine learning dalam content moderation sesuai dengan kebutuhan dan tujuan yang ingin dicapai.

Referensi

[1] Handayani, A. T., & Abdillah, L. A. (2019). Analisis Penerimaan Pengguna Dalam Memanfaatkan Media Sosial Terhadap Usaha Kecil Menengah Menggunakan Metode Technology Acceptance Model (Tam). In Bina Darma Conference on Computer Science (BDCCS) (Vol. 1, No. 6, pp. 2214-2221).

[2] Priyanto, A., & Ma'arif, M. R. (2018). Implementasi Web Scraping dan Text Mining untuk Akuisisi dan Kategorisasi

- Informasi dari Internet (Studi Kasus: Tutorial Hidroponik). *Indonesian Journal of Information Systems*, 1(1), 25-33.
- [3] Yusril, A. N., Larasati, I., & Aini, Q. (2020). Implementasi Text Mining untuk Advertising dengan Menggunakan Metode K-Means Clustering pada Data Tweets Gojek Indonesia. *Sistemasi: Jurnal Sistem Informasi*, 9(3), 586-596.
- [4] Gifari, O. I., Adha, M., Hendrawan, I. R., & Durrand, F. F. S. (2022). Analisis Sentimen Review Film Menggunakan TF-IDF dan Support Vector Machine. *Journal of Information Technology*, 2(1), 36-40.
- [5] Magriyanti, A. A. (2018). Analisis pengembangan algoritma porter stemming dalam bahasa indonesia.
- [6] Imani, F. A., Kusmawati, A., & Tohari, M. A. (2021). Pencegahan Kasus Cyberbullying Bagi Remaja Pengguna Sosial Media. *KHIDMAT SOSIAL: Journal of Social Work and Social Services*, 2(1), 74-83.
- [7] Chazar, C., & Erawan, B. (2020). Machine Learning Diagnosis Kanker Payudara Menggunakan Algoritma Support Vector Machine. *INFORMASI (Jurnal Informatika Dan Sistem Informasi)*, 12(1), 67-80.
- [8] Zahara, S., & Ilmiddafiq, M. B. (2019). Prediksi Indeks Harga Konsumen Menggunakan Metode Long Short Term Memory (LSTM) Berbasis Cloud Computing. *Jurnal RESTI (Rekayasa Sistem Dan Teknologi Informasi)*, 3(3), 357-363.
- [9] Indrawati, A. (2021). Penerapan Teknik Kombinasi Oversampling Dan Undersampling Untuk Mengatasi Permasalahan Imbalanced Dataset. *JIKO (Jurnal Informatika dan Komputer)*, 4(1), 38-43.
- [10] Luqyana, W., Cholissodin, I., & Perdana, R.S. (2018). Analisis Sentimen Cyberbullying pada Komentar Instagram dengan Metode Klasifikasi Support Vector Machine. *Jurnal Pengembangan Teknologi Informasi Dan Ilmu Komputer*, 2(11), 4704-4713. Diambil dari <http://j-ptiik.ub.ac.id/index.php/j-ptiik/article/view/3051>