IJCIT (Indonesian Journal on Computer and Information Technology) 6 (1) (2021) 77-81
Diterima 1 April 2021; Revisi 21 April 2021; Disetujui 26 April 2021

IJCIT
(Indonesian Journal on Computer and Information Technology)
Journal Homepage: http://ejournal.bsi.ac.id/ejurnal/index.php/ijcit

Pendeteksi Lokasi Kejadian Covid-19 Menggunakan Social Media dengan Kombinasi Algoritma Stemming Bahasa Indonesia

Dwiki Jatikusumo[1], Herry Derajad Wijaya[2]

Teknik Informatika, Universitas Mercu Buana
Jakarta, Indonesia
e-mail: dwiki.jatikusumo@mercubuana.ac.id[1], herry.derajad@mercubuana.ac.id[2]

ABSTRAK

Berdasarkan dari data wabah *Corona* Virus Disease-19 atau COVID-19 ini dari kuartal pertama tahun 2020 di Indonesia, telah banyak yang mungkin terjangkit dengan penyakit ini. Dari kasus-kasus terakhir ini juga kita tidak mendapatkan informasi secara lengkap di mana lokasi yang terjangkit. Dari *website* pemerintah juga harus buka dahulu tiap-tiap halaman *website*, dan tidak secara langsung lokasi yang ada. Penelitian ini bertujuan untuk dapat membantu dalam mencari lokasi dengan cara mendapat informasi terkini dari *posting* sosial media salah satunya yaitu Twitter, dari sini kita bisa tahu titik lokasi *post* maupun dari kalimat yang ada dalam *posting*-an tersebut. Metode yang digunakan dalam penelitian ini menggunakan SDLC atau *System Development Life Cycle* dengan model *prototype*. Data berdasarkan dari Twitter, akan dikombinasikan dengan algoritma stemming Bahasa Indonesia. Penelitian ini menghasilkan persentase akurasi yang didapat dalam menggabungkan algoritma stemming tersebut bisa mencapai lebih dari 95%.
Katakunci: algoritma stemming, covid-19, sosial media

ABSTRACTS

Based on data from the Corona Virus Disease-19 or COVID-19 outbreak from the first quarter of 2020 in Indonesia, many people have been infected with this disease. We also don't get complete information from the latest cases where the infected locations are. You must first open each website page from the government website, not directly from the existing location. With this research the aim is to be able to assist in finding the location by getting the latest information from social media posts, one of which is Twitter, where we can know the point of the post location and from the sentences, in the post, The method used in this research uses SDLC or System Development Life Cycle with a prototype model. Data based on Twitter will be combined with the Indonesian stemming algorithm. Then it will be seen the percentage of accuracy obtained in combining the stemming algorithm, and the results can reach more than 95%.
***Keywords**: covid-19, social media, stemming algorithm*

1. PENDAHULUAN

Penyakit *corona* virus 2019 atau *Corona* Virus Disease-19 (covid-19) adalah infeksi saluran pernapasan yang disebabkan oleh jenis virus corona. Nama lain dari penyakit ini adalah *Severe Acute Respiratory Syndrome* Coronavirus-2 (SARS-COV2). Kasus COVID-19 pertama kali dilaporkan di Kota Wuhan, Provinsi Hubei, Tiongkok, pada Desember 2019. Dalam beberapa bulan saja, penyebaran penyakit ini telah menyebar ke berbagai negara, baik di Asia, Amerika, Eropa, dan Timur Tengah serta Afrika. Pada tanggal 11 Maret 2020, Organisasi Kesehatan Dunia atau *World Health Organization* (WHO) mendeklarasikan penyebaran covid-19 dikategorikan sebagai pandemi (Widiyani, 2020).

Peningkatan jumlah kasus corona terjadi dalam waktu singkat dan membutuhkan penanganan segera. Virus *corona* dapat dengan mudah menyebar dan menginfeksi siapapun tanpa pandang usia. Virus ini dapat menular secara mudah melalui kontak dengan penderita. Sayangnya hingga kini belum ada obat spesifik untuk menangani kasus infeksi virus *corona* atau COVID-19 (Mona, 2020).

Menerapkan perilaku sehat dalam pencegahan COVID-19, merupakan langkah ampuh untuk menangkal penyakit, namun dalam praktiknya, penerapan ini yang kesannya sederhana tidak selalu mudah dilakukan terutama bagi responden yang tidak terbiasa, kurangnya pengetahuan dan sedikitnya kesadaran berperilaku hidup sehat. Upaya yang dapat dilakukan untuk menghadapi COVID-19 adalah melakukan *physical distancing*, rajin mencuci tangan dengan sabun, menggunakan masker bila keluar rumah (Restuning Prihati et al., 2020).

Menyebarnya wabah COVID-19 ini hingga ke wilayah Indonesia. Seperti dapat dicermati dari pengalaman beberapa negara serta wilayah lain, penanganan covid-19 tidak mungkin dapat dilakukan oleh Pemerintah semata. Dari dasar ini, peneliti akan menjadikan isu COVID-19 untuk menjadikan objek suatu penelitian dengan cara mendapatkan lokasi kejadian menggunakan sosial media sebagai sumber data dan algoritma *stemming* Bahasa Indonesia sebagai metode yang digunakan dalam penelitian ini.

Dari hal tersebut, tujuannya dalam penelitian ini mendapatkan titik lokasi kejadian COVID-19 melalui sosial media, dan mendapatkan akurasi yang baik dari hasil *stemming* titik lokasi kejadian COVID-19 sekitar 80% lebih dari data sebanyak 5.000 sampai 10.000 data.

Dari beberapa referensi berkaitan dengan algoritma stemming itu adalah inti pemrosesan bahasa alami teknik Pengambilan Informasi yang efisien dan efektif, dan salah satu yang diterima secara luas oleh pengguna. Itu sudah biasa mengubah varian kata menjadi akar kata yang sama dengan menerapkannya dalam banyak kasus aturan morfologi (Winarti et al., 2017).

Misalnya, dalam pencarian teks, itu harus mengizinkan pengguna mencari dengan menggunakan istilah *query stemming* untuk menemukan dokumen yang memuat istilah *stemmer* dan *stems* karena semua memiliki akar kata yang sama. Ia juga memiliki aplikasi dalam mesin terjemahan, dokumen peringkasan, dan klasifikasi teks. Untuk bahasa Inggris, *stemming* dipahami dengan baik, dengan teknik seperti yang dimiliki Lovin dan Porter (Sharma, 2012) di digunakan secara luas. Namun, *stemming* untuk bahasa lain adalah kurang terkenal: sementara ada beberapa pendekatan tersedia untuk bahasa seperti Prancis, Malaysia, dan bahasa Indonesia.

Algoritman *Nazief-Adriani Stemming* adalah cara yang digunakan untuk meningkatkan performa Information Retrieval dengan cara mentransformasikan kata-kata dalam sebuah dokumen teks ke kata dasarnya. Proses stemming pada teks Bahasa Indonesia digunakan untuk menghilangkan sufiks, konfiks, dan prefiks. Hal ini berbeda dengan teks Bahasa Inggris, di mana stemming digunakan untuk mengilangkan sufiks (Nugroho, 2017). Berikut tahapan-tahapan mengenai stemming menggunakan algoritma Idris (Prasidhatama & Suryaningrum, 2018) adalah sebagai berikut.Kata yang belum di-stemming dicari pada kamus umum atau Kamus Besar Bahasa Indonesia (KBBI). Jika kata itu langsung ditemukan, berarti kata tersebut adalah kata dasar. Kata tersebut dikembalikan dan algoritma dihentikan. Jika semua langkah sudah dilakukan termasuk *recording* dan tidak juga ditemukan dalam kamus, maka algoritma ini akan menganggap kata semula sebagai kata dasar.

Dalam Algoritma Arifin dan Setiono ini didahului dengan pembacaan tiap kata dari data yang ada (Novitasari, 2016). Pemeriksaan dalam 12 kombinasi ini sangat diperlukan karena fenomena *overstemming* pada algoritma pemotongan imbuhan. Kelemahan ini berakibat pada pemotongan bagian kata yang sebenarnya adalah milik kata dasar itu sendiri yang kebetulan mirip dengan salah satu jenis imbuhan yang ada. Dengan 12 kombinasi itu, pemotongan yang sudah terlanjur tersebut dapat dikembalikan sesuai posisinya.

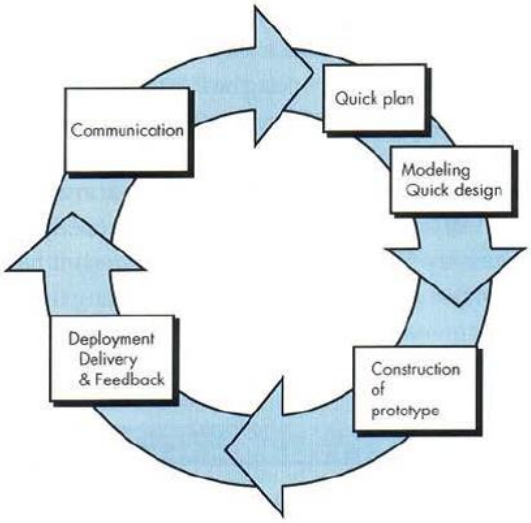
Dari beberapa referensi yang sudah dijelaskan, data yang digunakan adalah data dari twitter. Dalam hal ini twitter menjadi salah satu sosial media yang sering dipakai, dan menjadi sorotan karena selain sosial media, banyak orang mencari berita di twitter (Paramastri & Gumilar, 2019).

Dalam penggunaan *smartphone* di aplikasi twitter khususnya terdapat penggunaan GPS sebagai alat penentuan lokasi longitude dan latitude (Priambodo et al., 2019). Dari twitter ini juga mendapat lokasi longitude dan latitude dari smartphone GPS tersebut, dan dapat diambil menjadi data lokasi yang tepat (Fitrianah et al., 2020).

Hasil dari beberapa penelitian terdahulu yang sudah mendapatkan uji coba adalah sebagai berikut perbandingannya dengan akurasinya untuk memproses berbagai kalimat, Algoritma *Stemming* Nazief-Adriani sebesar 90.03% (Nugroho, 2017), Algoritma Idris sebesar 91.36% (Prasidhatama & Suryaningrum, 2018), dan Algoritma Arifin dan Setiono sebesar 95% (Novitasari, 2016).

2. METODE PENELITIAN

Tahapan pengembangan yang digunakan penulis dalam riset ini yaitu menggunakan model SDLC (System Development Life Cycle) dengan menggunakan model/metode prototype. Model ini dapat di lihat pada gambar 4.2 berikut.



Gambar 1. Kerangka Kerja Pengembangan Sistem Informasi (Protoype)
Sumber: Pressman & Maxim (2014)

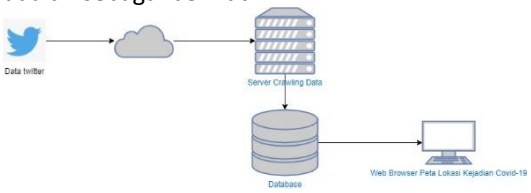
Metode *Prototype* dimulai dari tahap komunikasi. Tim pengembang perangkat lunak melakukan pertemuan dengan para stakeholder untuk menentukan kebutuhan perangkat lunak yang saat itu diketahui dan untuk menggambarkan area-area dimana definisi lebih jauh untuk iterasi selanjutnya.

Perencanaan iterasi pembuatan *prototype* dilakukan secara cepat. Setelah itu dilakukan pemodelan dalam bentuk "rancangan cepat". Pembuatan rancangan cepat berdasarkan pada representasi aspek-aspek perangkat lunak yang akan terlihat oleh para *end user* (misalnya rancangan antarmuka pengguna atau format tampilan). Rancangan cepat merupakan dasar untuk memulai konstruksi pembuatan *prototype*.

3. HASIL DAN PEMBAHASAN

3.1. Hasil Desain aplikasi

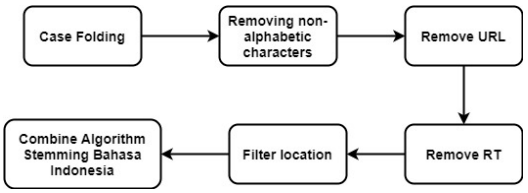
Desain aplikasi web yang akan digunakan adalah sebagai berikut:



Gambar 2. Arsitektur pendeteksi lokasi kejadian covid-19 (*Protoype*)

Pada gambar di atas yang dibutuhkan pada pendeteksi lokasi kejadian covid-19 adalah data dari twitter, kita ambil dari internet kemudian disimpan di dalam database dan akan diperlihatkan peta dari lokasi kejadian Covid-19.

Dalam pengesktrasian data dari data mentah menjadi data yang diolah dengan menentukan lokasi ditunjukkan pada gambar 3.

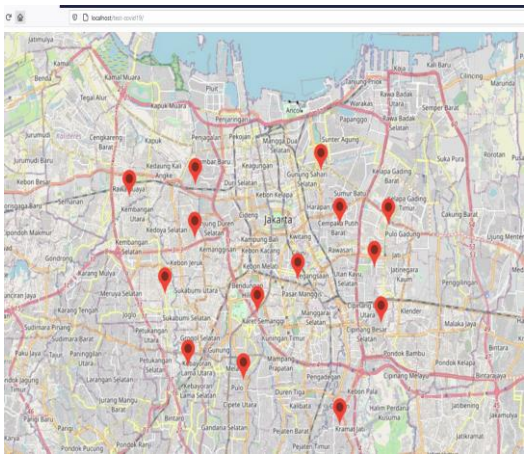


Gambar 3. *Flow preprocessing* pendeteksi lokasi kejadian covid-19

Gambar 3 menjelaskan Flow preprocessing pendeteksi lokasi kejadian covid-19 yang terdiri dari *Case Folding*, *Removing non-alphabetic characters, Remove* URL, *Remove* RT, Filter *location dan* Kombinasi Algoritma *Stemming* Bahasa Indonesia. *Case Folding* merupakan *l*angkah proses yang mengubah semua huruf dalam dokumen menjadi huruf kecil. Hanya huruf 'a' sampai 'z' yang diterima. Karakter selain huruf dihilangkan dan dianggap sebagai pembatas (delimiters). Selanjutnya proses r*emoving non-alphabetic characters* yaitu *m*enghilangkan tanda baca dan angka. Kemudian proses r*emove* URL, dimana kemunculan URL yang tinggi dari data twitter membuat data tersebut menjadi tidak efektif dan tidak berarti. Jadi diperlukan penghapusan URL. Munculnya alamat web atau url ini disebabkan banyaknya pengguna twitter yang mempromosikan suatu produk di situsnya. Proses selanjutnya yaitu r*emove* RT, Untuk menandai atau mengajak teman untuk berkomunikasi langsung di twitter dan website, dilakukan dengan menambahkan simbol "@"

sebelum *username* yang dimaksud. Sebuah penelitian tidak memperhatikan nama pengguna atau jumlah pengguna yang memberikan komentar. Peneliti hanya menggunakan data atau komentar pengguna, oleh karena itu perlu dihapus. Kemudian proses filter *location,* yaitu cek lokasi kata dari kalimat dengan cek lokasi *database* di Indonesia, misalnya kota, Jakarta, Bogor, Tangerang, Bekasi. Proses yang terakhir yaitu kombinasi algoritma *stemming* bahasa indonesia. Setelah filter lokasi, selanjutnya mengkombinasikan Algoritma *Stemming* Nazief-Adriani, Algoritma Idris, dan Algoritma Arifin dan Setiono menjadi teknik yang akan dilakukan untuk mencari kata-kata lokasi yang sesuai dengan hasil pencarian data dari twitter.

Hasil dari pemetaan dengan menggunakan openstreetmap.org dapat dilihat pada gambar 4. Jadi dari sini didapatkan letak posisi kejadian covid-19 dari cuitan twitter yang ada dari bulan Januari sampai Februari 2021.



Gambar 4. Hasil dari data proses menggunakan kombinasi algoritma stemming Bahasa Indonesia

3.2. Pembahasan

Data diproses menggunakan *preprocessing* seperti tahap pada metodologi riset ini, data yang diolah sebanyak 2458 data dari postingan twitter sebelumnya sudah disimpan di dalam database untuk keperluan riset ini, dari bulan Januari sampai Februari 2021 hasilnya adalah 1194 hasilnya adalah 1194 yang terdapat lokasi kejadian berupa kata lokasi yang dicari.

Selanjutnya adalah perbandingan yang merupakan kejadian atau tidak dengan ujicoba tanpa algoritma dan yang dikombinasikan dari ketiga Algoritma *Stemming* khususnya Bahasa Indonesia yaitu, Nazief-Adriani, Idris, dan Arifin-Setiono Di sini membedakan dari dua kata dengan "covid" dan "covid 19". Dari beberapa referensi untuk mengetahui akurasinya (Bhadoria & Kumar Patel, 2014; Winarti et al., 2017).

Tabel 1. Tanpa Algoritma

Kata	Kejadian	Tidak Kejadian	Total	Persentasi Akurasi
"covid"	435	82	517	84,13%
"covid 19"	564	113	677	83,31%

Tabel 2. Tabel menggunakan Algoritma Gabungan dari tiga Algoritma *Stemming* Bahasa Indonesia

Kata	Kejadian	Tidak Kejadian	Total	*Persentasi* Akurasi
"covid"	529	19	548	96,53%
"covid 19"	624	22	646	96,59%

Hasil perbandingan dari dua percobaan yang dilakukan dapat dilihat pada table 3.

Tabel 3. Perbandingan

Kata	Tanpa Algoritma	Kombinasi tiga Algoritma *Stemming* Bahasa Indonesia
"covid"	84,13%	96,53%
"covid 19"	83,31%	96,59%

Dari hasil tabel 3, persentase paling besar untuk mendapatkan kata "covid" dan "covid 19" adalah dikombinasikan dari ketiga Algoritma *Stemming* khususnya Bahasa Indonesia yaitu, Nazief-Adriani, Idris, dan Arifin-Setiono sebesar rata-rata 96,56%.

4. KESIMPULAN

Kesimpulan dari penelitian yang telah dilakukan, dari pengolahan data covid-19 didapat lokasi kejadiannya yang tidak menggunakan algoritma *stemming* terdapat besaran akurasi 90,57%. Kemudian untuk data yang diproses menggunakan algoritma *stemming* Bahasa Indonesia yang dikombinasikan dari tiga algoritma yaitu, Nazief-Adriani, Idris, dan Arifin-Setiono yaitu dengan besaran mencapai 96,42%. Dalam perhitungan dengan gabungan tiga algoritma tersebut bisa dikatakan setidaknya

lebih besar dari Arifin-Setiono. Penelitian selanjutnya memungkinkan untuk membuat algoritma terbaru khususnya untuk algoritma *stemming*.

5. REFERENSI

Bhadoria, S. S., & Kumar Patel, R. (2014). Web Text Content Extraction and Classification using Naïve Bayes Classifier Algorithm. International Journal of Scientific Research in Computer Science and Engineering, 2(5), 1–4. www.isroset.org

Fitrianah, D., Jatikusumo, D., & Nurhaida, I. (2020). D-Loc Apps: A Location Detection Application Based on Social Media Platform in the Event of A Flood Disaster. APIT 2020: Proceedings of the 2020 2nd Asia Pacific Information Technology Conference, 41–45. https://doi.org/10.1145/3379310.3381041

Mona, N. (2020). Konsep Isolasi Dalam Jaringan Sosial Untuk Meminimalisasi Efek Contagious (Kasus Penyebaran Virus Corona Di Indonesia). Jurnal Sosial Humaniora Terapan, 2(2), 117–125. https://doi.org/10.7454/jsht.v2i2.86

Novitasari, D. (2016). Perbandingan Algoritma Stemming Porter dengan Arifin Setiono untuk Menentukan Tingkat Ketepatan Kata Dasar. Jurnal STRING (Satuan Tulisan Riset Dan Inovasi Teknologi), 1(2), 120–129. https://doi.org/10.30998/string.v1i2.1031

Nugroho, H. T. (2017). Pengaruh Algoritma Stemming Nazief-Adriani Terhadap Kinerja Algoritma Winnowing Untuk Mendeteksi Plagiarisme Bahasa Indonesia. Jurnal ULTIMA Computing, 9(1), 36–40. https://doi.org/10.31937/sk.v9i1.572

Paramastri, N. A., & Gumilar, G. (2019). Penggunaan Twitter Sebagai Medium Distribusi Berita dan News Gathering Oleh Tirto.Id. Jurnal Kajian Jurnalisme, 3(1), 18. https://doi.org/10.24198/jkj.v3i1.22450

Prasidhatama, A., & Suryaningrum, K. M. (2018). Perbandingan Algoritma Nazief & Adriani Dengan Algoritma Idris Untuk Pencarian Kata Dasar. Jurnal Teknologi Dan Manajemen Informatika, 4(1), 1–4. https://doi.org/10.26905/jtmi.v4i1.1773

Pressman, R., & Maxim, B. (2014). Software Engineering : A Practitioner's Approach 8th ed. In Mc Grow Hill.

Priambodo, B., Ani, N., & Jumaryad, Y. (2019). Predict Next User Location to Improve Accuracy of Mobile Advertising. Journal of Physics: Conference Series, 1175(1). https://doi.org/10.1088/1742-6596/1175/1/012099

Restuning Prihati, D., K.Wirawati, M., & Supriyanti, E. (2020). Analisis Pengetahuan Dan Perilaku Masyarakat Di Kelurahan Baru Kotawaringin Barat Tentang Covid 19 Dyah. Concept and Communication, 2(23), 780–790. https://doi.org/10.15797/concom.2019..23.009

Sharma, D. (2012). Stemming Algorithms: A Comparative Study and their Analysis. International Journal of Applied Information Systems, 4(3), 7–12.

Widiyani, R. (2020). Latar Belakang Virus Corona, Perkembangan hingga Isu Terkini. Detik.Com. https://news.detik.com/berita/d-4943950/latar-belakang-virus-corona-perkembangan-hingga-isu-terkini

Winarti, T., Kerami, D., Lussiana, E. T. P., & Sudiro, S. A. (2017). Improving stemming algorithm using morphological rules. International Journal on Advanced Science, Engineering and Information Technology, 7(5), 1758–1764. https://doi.org/10.18517/ijaseit.7.5.1705