

Analisis Pengguna Facebook Terhadap Objek Wisata di Maluku Tengah Menggunakan Metode *K-Nearest Neighbor*

Novi Yanti Sangaji^{a,1,*}, Herman^{a,2}, Ihwana As'ad^{a,3}

^a *Fakultas Ilmu Komputer, Universitas Muslim Indonesia, Jalan Urip Sumoharjo, Makassar, 90231, Indonesia*
¹noviyantisangaji@gmail.com; ²herman@umi.ac.id; ³ihawana.asad@umi.ac.id
*corresponding author

INFORMASI ARTIKEL

Diterima : 28 – 07 – 2022
Direvisi : 15 – 08 – 2022
Diterbitkan : 31 – 08 – 2022

Kata Kunci:
Wisata
K-NN
Klasifikasi
Accuracy
Precision
Recall

ABSTRAK

Pemerintah dinas pariwisata kabupaten Maluku Tengah berupaya untuk mengembangkan dan memajukan wilayahnya dengan memanfaatkan wisata yang ada di wilayahnya, namun dalam upaya mengembangkan dan memajukan objek wisata terdapat problematika yang dihadapi pemerintah kabupaten Maluku Tengah di antaranya yaitu implementasi kebijakan pengembangan objek wisata yang masih belum optimal, tanggapan dan komentar dari pengunjung wisata tentu sangat diperlukan untuk pengembangan objek wisata, dengan adanya *feedback* dari pengunjung diharapkan dapat mempermudah pemerintah kabupaten Maluku Tengah untuk menentukan langkah apa yang harus dilakukan dalam pengembangan objek wisata yang ada. Penelitian ini bertujuan untuk menerapkan sentiment analisis pada data komentar dengan menggunakan metode K-NN, data komentar yang digunakan yaitu komentar dari pengunjung wisata di Maluku Tengah pada facebook terhadap objek wisata yang ada di Maluku Tengah. Teknik klasifikasi yang digunakan yaitu dengan menggunakan pengukuran performa (*accuracy*, *precision*, dan *recall*) menggunakan K-nearest neighbor. Imputan yang digunakan yaitu dataset sentiment public yang mengandung positif dan negative. Dataset sentiment public yang digunakan mengenai objek wisata di Maluku Tengah berjumlah 1000. Berdasarkan hasil pengujian performa maka didapatkan hasil *accuracy* 50% *precision* 56%, *recall* 50%, *f-measure* 34%.

This is an open access article under the [CC-BY-SA](#) license



I. Pendahuluan

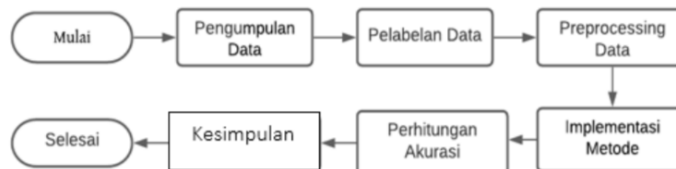
Maluku Tengah merupakan kabupaten yang ada di provinsi Maluku terdapat beberapa wilayah kabupaten ini yang terpisah seperti pulau Ambon, pulau Seram, serta pulau-pulau kecil lain. Maluku Tengah juga terkenal dengan objek wisata [1]. Wisata dan budaya merupakan salah satu sektor yang sangat berpengaruh untuk berlangsungnya pembangunan suatu daerah. Pemerintah dinas pariwisata kabupaten Maluku Tengah berupaya untuk mengembangkan dan memajukan wilayahnya dengan memanfaatkan wisata yang ada di wilayahnya, namun dalam upaya mengembangkan dan memajukan objek wisata terdapat problematika yang dihadapi pemerintah kabupaten Maluku Tengah di antaranya yaitu implementasi kebijakan pengembangan objek wisata yang masih belum optimal, tanggapan dan komentar dari pengunjung wisata tentu sangat diperlukan untuk pengembangan objek wisata, dengan adanya *feedback* dari pengunjung diharapkan dapat mempermudah pemerintah kabupaten Maluku Tengah untuk menentukan langkah apa yang harus dilakukan dalam pengembangan objek wisata yang ada [2]. Proses analisis dan pengklasifikasian data komentar yang dilakukan dengan cara membaca dan memilah satu per satu komentar yang bersifat negatif kemudian mengklasifikasikannya satu per satu dengan menggunakan Ms.Excel. Namun hal tersebut dianggap kurang efektif apabila data yang akan diproses dalam jumlah yang banyak, oleh karena itu penelitian ini bertujuan untuk menerapkan sentiment analisis pada data komentar dengan menggunakan metode K-NN, data komentar yang digunakan yaitu komentar dari pengunjung wisata di Maluku Tengah [3]. Analisis sentimen merupakan teknik pemrosesan data yang digunakan untuk menentukan apakah data tersebut positif, negatif, atau netral. *Fine-grained sentiment analysis* merupakan salah satu jenis yang paling umum fokusnya ada pada tingkat polaritas pendapat, tipe analisis sentimen ini akan mengelompokkan respon atau pendapat ke dalam beberapa kategori seperti sangat positif, agak positif, netral, agak negatif, dan negatif. Analisis sentimen dapat dilakukan pengolahan kata untuk melacak *mood* pengunjung wisata dari komentar pengunjung yang didapatkan dalam aplikasi facebook [2]. Opini berupa komentar-komentar dalam bentuk teks tersebut dapat dikumpulkan menggunakan teknik web crawling. Metode K-NN digunakan untuk proses klasifikasi dalam penelitian ini

karena memiliki kesederhanaan dimana prosesnya berdasarkan pada pendekatan pembobotan yang sederhana dan kemudahan dalam implementasi, adaptasi dan proses learning serta memiliki nilai akurasi yang tinggi [4].

Dengan adanya permasalahan di atas, maka penulis melakukan penelitian tentang “Analisis Pengguna Facebook Terhadap Objek Wisata Di Maluku Tengah Menggunakan Metode *K-Nearest neighbor*” yang bertujuan untuk mengklasifikasikan komentar positif dan negatif agar dapat mempermudah untuk mengetahui respon positif dan negatif dari pengguna facebook yang memberikan komentarnya terhadap objek wisata di Maluku Tengah sehingga dapat mempermudah pemerintah dinas pariwisata kabupaten Maluku Tengah dalam menentukan langkah apa yang harus dilakukan dalam pengembangan objek wisata yang ada di daerahnya.

II. Metode

Metode penelitian adalah metode yang akan digunakan dalam penelitian, pada bagian ini setiap tahapan dijelaskan pada Gambar 1:

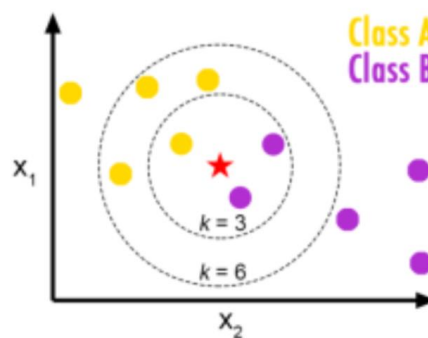


Template ini telah menetapkan sistem dan *style* penomoran dan *bullet* yang boleh digunakan. Berikut adalah petunjuk dan standar penulisan singkatan, penggunaan satuan, penulisan persamaan matematika, penulisan tabel dan grafik pada BUSITI.

A. Metode *K-Nearest Neighbor* (*K-NN*)

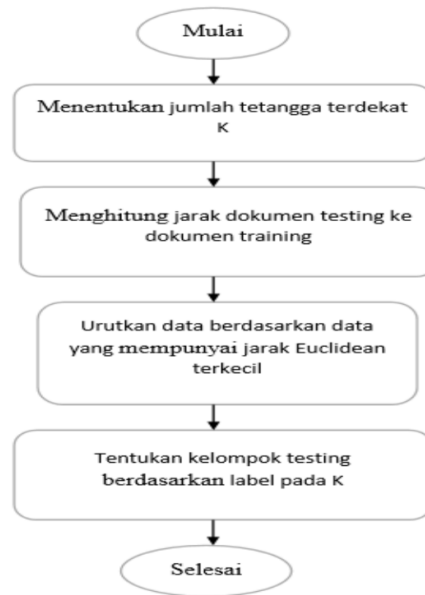
Algoritma *K-Nearest neighbor* (*KNN*) adalah sebuah metode untuk melakukan klasifikasi terhadap objek berdasarkan data pembelajaran yang jaraknya paling dekat dengan objek tersebut. Data pembelajaran diproyeksikan ke ruang berdimensi banyak, dimana masing- masing dimensi merepresentasikan fitur dari data. Ruang ini dibagi menjadi bagian-bagian berdasarkan klasifikasi data pembelajaran [5].

Algoritma ini bertujuan untuk mengklasifikasikan kumpulan data pembelajaran (data training) yang sudah diklasifikasikan sebelumnya. Algoritma ini melakukan klasifikasi berdasarkan jarak terdekat. *K-NN* adalah salah satu algoritma paling sederhana untuk memecahkan masalah klasifikasi algoritma ini sering digunakan untuk klasifikasi teks dan data. Pada metode ini dilakukan klasifikasi terhadap objek berdasarkan data yang jaraknya paling dekat dengan objek tersebut, tujuan dari algoritma ini adalah untuk mengklasifikasikan objek berdasarkan atribut dan training sample algoritma ini menggunakan klasifikasi ketetanggaan sebagai nilai prediksi dari query instance yang baru [6].



Gambar 2. Ilustrasi Metode *k-NN*

Berdasarkan ilustrasi pada Gambar 2 dapat dilihat bahwa nilai *k* yang terbaik untuk *K-NN* tergantung pada data. Secara umum nilai *k* yang tinggi akan mengurangi efek *noise* pada klasifikasi tetapi membuat batasan antara setiap klasifikasi menjadi lebih kabur. Nilai *k* yang bagus dapat dipilih dengan optimasi parameter, misalnya dengan menggunakan *cross validation*. Kasus khusus di mana klasifikasi diprediksikan berdasarkan data pembelajaran yang paling dekat (dengan kata lain, $k = 1$) disebut algoritma *nearest neighbor* [7]. Berikut adalah *flowchart* proses *K-NN* dapat dilihat pada Gambar 3.



Gambar 3. Flowchart Proses K-NN

B. Confusion Matrix

Confusion matrix merupakan tabel matrix yang terdiri dari 2 kelas yaitu positif dan negatif untuk menghitung nilai akurasi dalam proses data mining dan proses sistem pendukung keputusan [8]. Ada empat istilah untuk melakukan pengukuran kinerja dari proses algoritma klasifikasi yaitu:

- 1) *True positive (TP)*: data positif yang terprediksi benar.
- 2) *False positive (FP)*: data negative namun terprediksi positif.
- 3) *True negative (TN)*: data negative yang terprediksi benar.
- 4) *False negative (FN)*: data positif namun terprediksi negative.

Tabel 1. *Confussion Matrix*

		<i>TRUE VALUES</i>	
		<i>TRUE</i>	<i>FALSE</i>
<i>Prediction</i>	<i>TRUE</i>	TP <i>Correct Result</i>	FP <i>Unexpected Results</i>
	<i>FALSE</i>	FN <i>Missing Result</i>	TN <i>Correct Absence of Result</i>

Berdasarkan pada Tabel 1 dapat dilihat bahwa nilai dari TP, FP, TN, dan FN maka akan didapatkan nilai *precision*, *recall* dan akurasi. *Precision* merupakan tingkat ketepatan nilai prediksi positif dengan sistem. *recall* merupakan tingkat sebuah sistem berhasil dalam mendapatkan kembali informasi, dan akurasi adalah tingkat kedekatan nilai prediksi dengan nilai aktual atau total data yang diidentifikasi dengan data yang dinilai [8]. Rumus perhitungan *precision*, *recall* dan akurasi terlihat pada formula 1,2 dan 3.

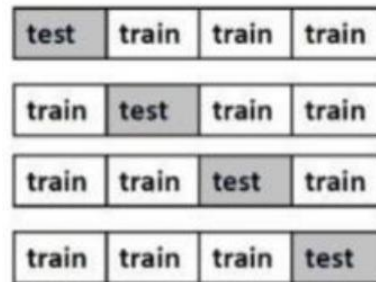
$$Precision = \frac{TP}{FP + TP} \times 100\% \tag{1}$$

$$Recall = \frac{TP}{FN + TP} \times 100\% \tag{2}$$

$$Akurasi = \frac{TP + TN}{TP + TN + FP + FN} \times 100\% \tag{3}$$

C. Cross Validation

Cross validation merupakan salah satu teknik untuk menilai atau memvalidasi keakuratan sebuah model yang dibangun berdasarkan dataset tertentu. Pembuatan model biasanya bertujuan untuk melakukan prediksi maupun klasifikasi terhadap suatu data baru yang boleh jadi belum pernah muncul di dalam Dataset. Data yang digunakan dalam proses pembangunan model disebut data latih atau data training, sedangkan data yang akan digunakan untuk memvalidasi model disebut sebagai data test. Salah satu metode *cross-validation* yang populer adalah *K-fold cross validation*. *K-fold* bekerja dengan melipat data sebanyak *K* dan mengulangi (men-iterasi) eksperimennya sebanyak *K* juga [3].



Gambar 4. Ilustrasi *Cross Validation*

Berdasarkan gambar 4 ditunjukkan bahwa nilai fold yang digunakan adakah 4-fold *cross validation*. Berikut diberikan langkah-langkah pengujian data dengan 4 fold *cross validation*.

- 1) Dataset yang digunakan dibagi menjadi 4 bagian, yaitu D1, D2, D3, dan D4. Di, t = (1, 2, 3, 4) digunakan sebagai data testing dan dataset lainnya sebagai data training.
- 2) Tingkat akurasi, *precision*, *recall*, dan *f-measure* dihitung pada setiap iterasi (iterasi-1, iterasi-2, iterasi-3, iterasi-4), kemudian dihitung rata-rata tingkat akurasi, *precision*, *recall*, dan *f-measure* dari seluruh iterasi untuk mendapatkan data keseluruhan.

D. Web Crawling

Web crawling merupakan sebuah aplikasi yang secara otomatis melintasi web dengan mengunduh dokumen dan mengikuti link dari halaman ke halaman sehingga *web crawler* dapat dijadikan sebuah alat untuk mengambil konten-konten yang ditampilkan oleh halaman website, kemudian konten-konten tersebut dikelompokkan dalam satu atau lebih atribut. *Web crawler* telah digunakan dalam berbagai penelitian untuk proses ekstraksi dan analisis data [9].

III. Hasil dan Pembahasan

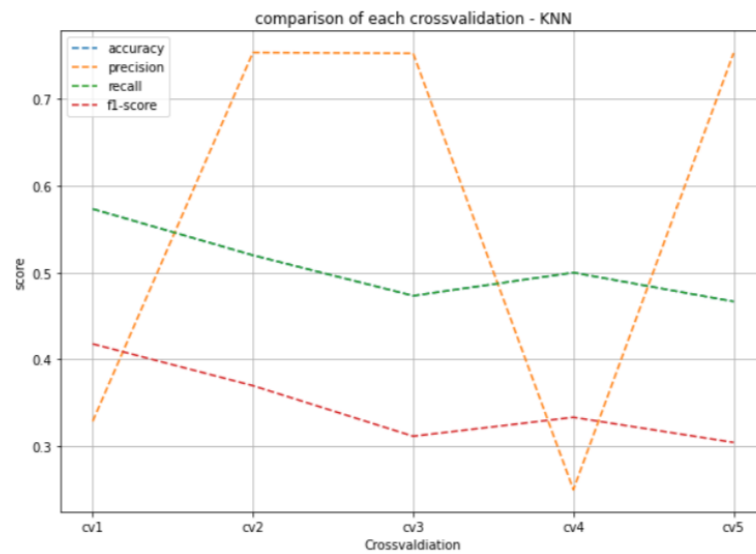
A. Hasil Penelitian

Berdasarkan pengujian performa metode K-NN pada data sentimen publik mengenai objek wisata di Maluku Tengah dengan melakukan pengujian performa akurasi, presisi, *recall*, dan *f-measure*, maka didapatkan hasil pengujian yang dapat dilihat pada Tabel 2.

Tabel 2. Nilai Pengujian *Cross Validation*

Iterasi	Performa			
	<i>Accuracy</i>	<i>Precision</i>	<i>Recall</i>	<i>F-measure</i>
1	0.57333333	0.32871111	0.57333333	0.41785311
2	0.52	0.75351351	0.52	0.36977444
3	0.47333333	0.75257271	0.47333333	0.31149445
4	0.5	0.25	0.5	0.33333333
5	0.46666667	0.75302013	0.46666667	0.30436339

Berikut ini adalah diagram yang menampilkan pengujian *cross validation* yang dapat dilihat pada Gambar 5.



B. Pembahasan

1) Pembahasan Dataset

Dataset yang digunakan pada penelitian ini adalah data yang diambil dari postingan dan komentar pada situs facebook mengenai objek wisata di Maluku Tengah. Terdapat 1000 data yang dikelolah dalam penelitian ini data yang diambil dari facebook merupakan data yang berbahasa Indonesia dengan menggunakan beberapa kata kunci mengenai objek wisata di Maluku Tengah beberapa kata kunci diantaranya adalah Pantai Ora, Pantai Liang, Gunung binaiya, dan Tebing Makariki. Kemudian data yang telah di crawling tersebut dilabeli dengan dua jenis kelas yaitu positif dan negatif jumlah data dari masing-masing kelas adalah 500 kelas positif dan 500 kelas negatif.

2) Pembahasan Hasil Penelitian

Untuk menggunakan algoritma K-NN perlu ditentukan banyaknya k tetangga terdekat yang digunakan untuk melakukan klasifikasi data baru algoritma ini mengklasifikasikan data berdasarkan similarity atau kedekatannya terhadap data lainnya, dalam menentukan nilai k, bila jumlah klasifikasi kita genap maka sebaiknya kita gunakan nilai k ganjil untuk menghindari ikatan dalam klasifikasi.

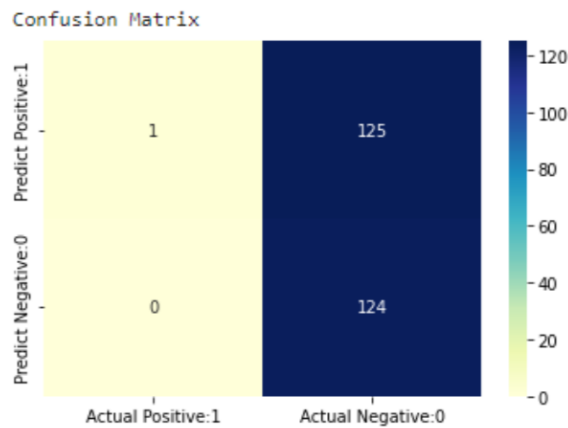
Pada penelitian ini kelas yang akan diklasifikasi adalah dua kelas yaitu negatif dan positif maka nilai k yang digunakan adalah $n=3$ KNN bekerja berdasarkan prinsip bahwa setiap titik data yang berdekatan satu sama lain akan berada di kelas yang sama. Dengan kata lain, KNN mengklasifikasikan titik data baru berdasarkan kemiripan. Model KNN akan menemukan fitur serupa dari kumpulan data baru berdasarkan fitur yang paling mirip kemudian memasukkannya ke dalam kategori positif atau negatif.

Model klasifikasi dapat memprediksi nilai pada kelas terbesar untuk semua prediksi dan bisa memberikan nilai akurasi yang tinggi dan tentu saja model yang dihasilkan juga dapat memprediksikan nilai yang salah, sehingga perlu metrik evaluasi lain yang dapat mengukur performa model klasifikasi yang kita buat. Metrik yang dimaksudkan adalah *Precision*, *Recall* dan *Confusion Matrix*. *Confusion Matrix* merepresentasikan prediksi dan kondisi sebenarnya dari data yang dihasilkan oleh algoritma ML. *Confusion matrix* digunakan untuk menghitung nilai akurasi, *presicion*, *recall*, dan *F1-score*. Berikut dapat dilihat hasil model dari klasifikasi menggunakan K-NN pada Tabel 3.

Tabel 3. Model K-NN

<i>Score Accuracy</i>	<i>Precision</i>	<i>Recall</i>	<i>F-measure</i>
50.0%	74.8995983935743%	50.39682539682539%	34.031369403221376%

Berikut dapat dilihat diagram dari hasil *confussion matrix* terlihat pada Gambar 6.



Gambar 6. Diagram *Confussion Matrix*

Cross validation merupakan suatu metode tambahan dari teknik data mining yang bertujuan untuk memperoleh hasil akurasi yang maksimal, fungsi dari penggunaan metode *cross validation* adalah untuk mengetahui performa dari suatu model algoritma dengan melakukan percobaan sebanyak k kali dan juga untuk meningkatkan tingkat performansi dari model tersebut.

Berdasarkan pada tabel 11 dapat dilihat bahwa data training pada penelitian ini dibagi menjadi k=5 iterasi yang artinya adalah melakukan percobaan sebanyak 5 kali tahapan. Percobaan pertama yaitu menjadikan bagian partisi pertama menjadi data testing dan partisi lainnya menjadi data training, percobaan kedua yaitu menjadikan bagian partisi kedua menjadi data *testing* dan partisi lainnya menjadi data *training*, percobaan ketiga yaitu menjadikan bagian partisi ketiga menjadi data testing dan partisi lainnya menjadi data *training* dan begitu seterusnya. Dari lima hasil percobaan ini, kita akan catat nilai evaluasi performa dari model tersebut kemudian dapat ditentukan nilai rata-rata dari setiap percobaan. Maka disitu akan ditemukan percobaan mana yang dapat dijadikan acuan dari penggunaan suatu model algoritma yang telah dipilih. Kemudian terlihat bahwa nilai dari setiap iterasi dapat berbeda nilainya hal itu dikarenakan data yang di validasi juga berbeda.

IV. Kesimpulan dan saran

Pengujian performansi *accuracy*, presisi, *recall*, dan *f-measure* menggunakan metode K-NN yang dilakukan pada dataset sentimen publik yang berjumlah 1000. Berdasarkan hasil pengujian performa didapatkan hasil *accuracy* 50%, presisi 56%, *recall* 50%, dan *f-measure* 34%. Berdasarkan hasil analisis dan kesimpulan yang diperoleh, maka dapat diberikan saran yaitu sistem yang di bangun hanya dapat mengklasifikasikan komentar yang mengandung positif dan negatif, maka diharapkan pada penelitian selanjutnya agar sistem dapat dikembangkan dengan jumlah kelas yang bervariasi, diharapkan agar penelitian menggunakan dataset yang lebih banyak, dan diharapkan pada penelitian selanjutnya menggunakan metode *supervised learning* seperti metode *support vector machine* (SVM), *Naïve bayes* dan metode yang lainnya.

Daftar Pustaka

- [1] A. M. Ferdinandus and I. A. Suryasih, "Studi Pengembangan Wisata Bahari Untuk Meningkatkan Kunjungan Wisatawan Di Pantai Natsepa Kota Ambon Provinsi Maluku," *Jurnal Destinasi Pariwisata*, vol. II, no. 2, pp. 1-14, 2014.
- [2] A. Imron, "Analisis Sentimen Terhadap Tempat Wisata di Kabupaten Rembang Menggunakan Metode Naive Bayes Classifier," pp. 10-13, 2019.
- [3] Z. U. Siregar, R. R. A. Siregar and R. Arianto, "Klasifikasi Sentiment Analysis Pada Komentar Peserta Diklat Menggunakan Metode K-Nearest Neighbor," *Jurnal KILAT*, vol. VIII, no. 1, pp. 81-92, 2019.
- [4] S. Saidah and J. Mayary, "Analisis Sentimen Pengguna Twitter Terhadap Dompot Elektronik Dengan Metode Lexicon Based Dan K – Nearest Neighbor," *Jurnal Ilmiah Informatika Komputer*, vol. XXV, no. 1, pp. 1-17, 2020.

-
- [5] W. Yustanti, "Algoritma K-Nearest Neighbour untuk Memprediksi Harga Jual Tanah," *JMSK*, vol. IX, no. 1, pp. 57-68, 2012.
- [6] R. R. Sani, J. Zeniarja and A. Luthfiarta, "Penerapan Algoritma K-Nearest Neighbor pada Information Retrieval dalam Penentuan Topik Referensi Tugas Akhir," *Journal of Applied Intelligent System*, vol. I, no. 2, pp. 123-133, 2016.
- [7] M. M. Baharuddin, T. Hasanuddin and H. Azis, "Analisis Performa Metode K-Nearest Neighbor Untuk Identifikasi Jenis Kaca," *ILKOM Jurnal Ilmiah*, vol. XI, no. 3, pp. 269-274, 2019.
- [8] M. S. Lydia, "Sistem Analisis Sentimen pada Fanpage Facebook Kandidat Presiden 2019-2024," 2020.
- [9] L. B. Ilmawan, "Membangun Web Crawler Berbasis Web Service Untuk Data Crawling Pada Website Google Play Store," *ILKOM Jurnal Ilmiah*, vol. X, no. 2, pp. 215-224, 2018.