

Analisis Sentimen Opini Masyarakat Terhadap Pembangunan Infrastruktur Kota Malang Melalui Twiter Dengan Menggunakan Metode Support Vector Machine

Nadhifa Tiara Putri¹, Indra Dharma Wijaya², Ariadi Retno Tri Hayati Ririd³

^{1,2,3}Program Studi Teknik Informatika, Jurusan Teknologi Informasi, Politeknik Negeri Malang

¹ nadhifatiaraputri@gmail.com, ²indra.dharma@gmail.com, ³faniri4education@gmail.com

Abstrak— Percepatan pertumbuhan ekonomi di Kota Malang masih terganjal oleh infrastruktur. Pemerintah daerah tidak menutup mata dengan kondisi tersebut. Namun pengambilan keputusan maupun kebijakan yang dilakukan oleh pemerintah tentunya harus didasari oleh informasi yang akurat agar kebijakan yang diambil tepat sasaran. Opini dari masyarakat dibutuhkan untuk menjadi acuan pengukuran kinerja kota. Pada penelitian ini dilakukan analisis sentimen terhadap pembangunan infrastruktur di Kota Malang dengan memanfaatkan media sosial Twitter. Dataset berjumlah 1980 yang didapatkan dari *tweet* yang menyebutkan akun pemerintah Kota Malang dan Dinas Pekerjaan Umum Kota Malang pada tahun 2018 sampai 2019. Klasifikasi opini dibagi menjadi dua, yaitu positif dan negatif dengan menggunakan algoritma *Support Vector Machine*. Data *tweet* yang didapat mengenai pembangunan infrastruktur lebih banyak bersentimen negatif daripada positif. Hasil akurasi klasifikasi metode *Support Vector Machine* dengan menggunakan 70, 80, dan 90 persen data *training* adalah 73.10, 72.62 dan 86.42 persen.

Kata kunci— Analisis Sentimen, Support Vector Machine, Twitter, Infrastruktur, Kota Malang

I. PENDAHULUAN

Pemerintah Indonesia saat ini sedang gencar membangun infrastruktur secara masif dan menyebar ke seluruh wilayah Indonesia. Pembangunan infrastruktur yang ada di Jawa Timur juga sangat masif dikarenakan penduduknya besar dan ekonominya bagus. Di Kota Malang percepatan pertumbuhan ekonomi pada tahun 2019 masih terganjal oleh infrastruktur. Salah satunya adalah faktor kemacetan akibat kurangnya infrastruktur jalan dan transportasi publik [1]. Pemerintah daerah sendiri tidak menutup mata dengan kondisi saat ini. Mulai dari permasalahan kemacetan, banjir, persampahan dan lainnya akan selalu dicarikan solusi terbaiknya. Namun pengambilan keputusan maupun kebijakan yang dilakukan oleh pemerintah tentunya harus didasari oleh informasi yang akurat agar kebijakan yang diambil tepat sasaran sehingga mampu menjadi sebuah solusi tanpa menimbulkan permasalahan baru.

Salah satu wadah untuk berbagi informasi mengenai keadaan Kota Malang untuk pemerintah dan warga asli

maupun pendatang di Malang adalah melalui media sosial yaitu Twitter. Twitter adalah layanan jejaring sosial dan mikroblog daring yang memungkinkan penggunanya untuk mengirim dan membaca pesan berbasis teks hingga 280 karakter yang dikenal dengan sebutan *tweet*. Informasi yang ditemukan di Twitter selain berupa pelaporan, keluhan, apresiasi suatu hasil kerja dan kebijakan. Informasi positif dan negatif adalah sentimen masyarakat yang layak dianalisis agar pantas dijadikan acuan pengukuran kinerja kota.

Dengan adanya sentimen analisis ini, diharapkan dapat meminimalisir masalah infrastruktur yang ada di Kota Malang. Dibuatnya sentimen analisis ini bertujuan untuk mengetahui respon masyarakat tentang masukan terhadap infrastruktur yang ada di Kota Malang secara akurat. Data yang didapat akan diproses menggunakan *text mining*, kemudian dilanjutkan dengan mengklasifikasikan *tweet* ke dalam dua kelas, yaitu positif dan negatif. Klasifikasi ini menggunakan metode *Support Vector Machine*.

II. TINJAUAN PUSTAKA

A. Penelitian yang Terkait

Berdasarkan penelitian terdahulu Zaky Reyhana telah melakukan analisis sentimen tentang pendapat masyarakat terhadap pembangunan infrastruktur di Kota Surabaya melalui Twitter dengan membandingkan metode *Support Vector Machine* dan *Neural Network*, dengan hasil ketetapan klasifikasi metode SVM sebesar 92,67% sedangkan Neural Network 80% [2].

Berdasarkan penelitian Miftahul Agtamas Fidyawan telah melakukan penelitian mengenai analisis sentimen menggunakan metode *Support Vector Machine* terhadap *review* film pada Twitter. Dalam penelitian ini dicantumkan beberapa kesimpulan, yaitu algoritma *Support Vector Machine* dapat digunakan untuk klasifikasi *tweet* tentang *review* film. Dan dari hasil pengujian yang telah dilakukan, semakin banyak data *training* yang digunakan, maka nilai *accuracy* dan *precision* cenderung mengalami peningkatan [3].

B. Text Mining

Text mining adalah proses ekstraksi informasi dari data sumber yang belum terstruktur. Data yang belum terstruktur akan diolah menggunakan teknik dan metode tertentu menghasilkan informasi yang berguna untuk pengguna. *Text mining* memiliki tujuan dan menggunakan proses yang sama dengan data mining, namun memiliki input yang berbeda. Input untuk *text mining* adalah data yang tidak (atau kurang) terstruktur, seperti dokumen word, PDF, kutipan teks, sedangkan input untuk data mining adalah data yang terstruktur [4].

C. Analisis Sentimen

Analisis sentimen atau *opinion mining* merupakan bidang studi yang menganalisis opini, sentimen, evaluasi, penilaian, sikap, dan emosi orang terhadap suatu entitas seperti produk, layanan, organisasi, individu, masalah, peristiwa, topik, dan atributnya. Analisis sentimen dilakukan untuk melihat pendapat dan kecenderungan opini terhadap sebuah masalah atau objek oleh seseorang, apakah cenderung berpandangan atau beropini negatif atau positif [5].

D. Web Scraping

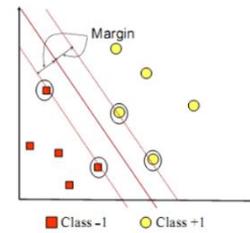
Web scraping adalah proses pengambilan dokumen semi-terstruktur dari internet, umumnya laman web dalam Bahasa marup HTML atau XHTML, digunakan untuk mengekstraksi data tertentu dari dokumen dan digunakan konteks lain. Web scraping hanya berkaitan tentang memperoleh data melalui pengambilan dan ekstraksi. Dapat melibatkan kumpulan data dengan ukuran yang bervariasi [6].

E. Twitter

Twitter adalah layanan jejaring sosial dan mikroblog daring yang memungkinkan penggunanya untuk mengirim dan membaca pesan berbasis teks hingga 280 karakter yang dikenal dengan sebutan *tweet* [7].

F. Support Vector Machine

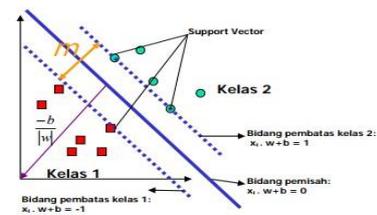
Support Vector Machine adalah algoritma pembelajaran terbimbing (supervised) digunakan untuk klasifikasi, regresi, dan anomali atau pendeteksian pencilan (outlier). SVM berusaha menemukan garis pemisah (hyperplane) yang terbaik pada ruang input space. Prinsip dasar SVM adalah linear classifier, mengingat masalah klasifikasi biner, jika kita memiliki data latihan yang tiap titik data atau observasi tergolong pada kelas yang spesifik, algoritma SVM bisa dilatih berdasarkan data tersebut dan bisa menetapkan titik data yang akan datang ke dalam salah satu dari dua kelas kategori. Namun, SVM juga bisa melakukan klasifikasi non-linear dengan pendekatan menarik yang dikenal dengan sebutan trik kernel, dimana fungsi kernel digunakan untuk mengoperasikan ruang fitur berdimensi tinggi yang terpisah secara non-linear. Algoritma SVM mengambil satu set titik data latihan dan mendirikan sebuah hyperplane dari koleksi hyperplane untuk ruang fitur berdimensi tinggi. Semakin besar batas hyperplane, semakin bagus pemisahannya, dan dengan demikian hal ini akan mengurangi kesalahan generalisasi alat yang mengklasifikasi (classifier) [8].



Gambar 1 Hyperplane terbaik yang memisahkan kedua kelas -1 dan +1
Pada klasifikasi, SVM dapat dibedakan menjadi dua, yaitu SVM linear dan SVM non-linear.

a. Linear Support Vector Machine

SVM menggunakan decision boundary (batas keputusan) yang akan menentukan klasifikasi dari data-data pelatihan sehingga dapat dibentuk sebuah model linear atau hyperplane yang paling optimal untuk mengklasifikasikan data data tersebut.



Gambar 2 Linear Support Machine

SVM menggunakan model linear sebagai decision boundary dalam bentuk sebagai berikut:

$$y(x) = w t\phi(x) + b \quad (1)$$

Dimana x adalah vektor input, w adalah parameter bobot, (x) adalah fungsi basis, dan b adalah suatu bias. Sehingga, sebuah hyperplane yaitu (x) dengan memaksimalkan margin didapatkan fungsinya. Margin maksimum adalah pilihan yang paling aman karena terjadi paling sedikit kesalahan pada klasifikasi. Apabila data dapat terklasifikasi secara linear maka data x ke n dapat dicari jaraknya dengan

$$\frac{tny(xn)}{\|w\|} = \frac{tn(w t\phi(xn) + b)}{\|w\|} \quad (2)$$

Margin adalah jarak decision boundary dengan data terdekat oleh karena itu memaksimalkan margin dapat menggunakan

b. Non Linear Support Vector Machine

Pada dunia nyata data-data pengukuran umumnya tidak dapat diklasifikasi dengan menggunakan metode SVM linear, namun dapat diklasifikasi dengan menggunakan SVM non-linear. Oleh karena itu SVM dimodifikasi dengan memasukkan fungsi kernel. Fungsi kernel adalah fungsi k yang mana untuk semua vektor input x dan z akan memenuhi kondisi:

$$\operatorname{argmax} \left\{ \frac{1}{\|w\|} \min [t_n (w^t \phi(x_n) + b)] \right\} \quad (3)$$

Dimana ϕ adalah fungsi pemetaan dari ruang input ke ruang fitur. Pada umumnya kernel-kernel yang dapat dipakai pada SVM adalah sebagai berikut:

TABEL 1 KERNEL YANG UMUM DIGUNAKAN

Jenis Kernel	Definisi
<i>Polynomial</i>	$K(x_i, x_j) = (x_i, x_j + 1)^p$ (4)
<i>Gaussian RBF</i>	$K(x_i, x_j) = \exp\left(-\frac{\ x_i - x_j\ ^2}{2\sigma^2}\right)$ (5)
<i>Sigmoid</i>	$K(x_i, x_j) = \tanh(\alpha x_i, x_j + \beta)$ (6)

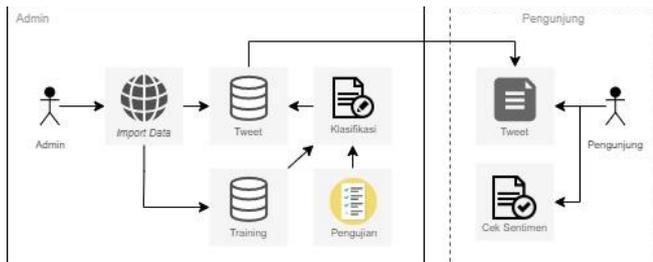
III. ANALISIS DAN PERANCANGAN

A. Analisis Sistem

Analisis sentimen Twitter pada pembangunan infrastruktur Kota Malang merupakan aplikasi berbasis *website* yang digunakan untuk klasifikasi sentimen berdasarkan *tweet* yang diperoleh dari Twitter. *Tweet* yang digunakan berkaitan dengan 5 jenis infrastruktur. Metode yang akan digunakan merupakan salah satu *Machine Learning*, yaitu *Support Vector Machine*. Klasifikasi sentimen dibagi menjadi dua, yaitu positif dan negatif. Hasil akhir dari aplikasi adalah klasifikasi sentimen mengenai pembangunan infrastruktur dan tingkat akurasi dari penggunaan metode *Support Vector Machine* dalam klasifikasi *tweet*.

1) Deskripsi Sistem

Sistem terbagi menjadi dua sisi yaitu admin dan pengunjung. Pada sisi admin dapat melakukan pengolahan aplikasi seperti melakukan *import* data (*tweet* dan *training*), pengolahan *tweet*, pengklasifikasian dan pengujian. Sedangkan pada sisi pengunjung untuk menampilkan hasil dari klasifikasi *tweet* dan melakukan cek sentimen.



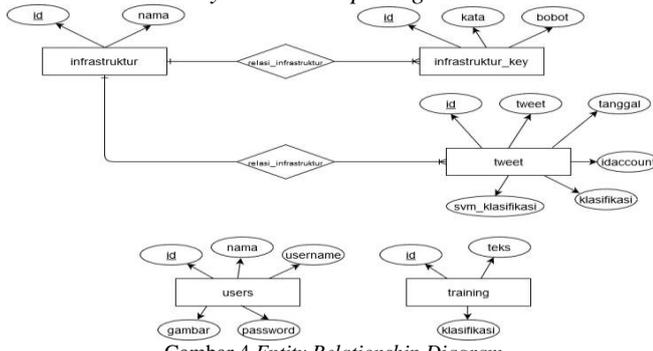
Gambar 3 Gambaran Sistem

B. Perancangan

Pada bagian ini diuraikan dengan jelas rancangan sistem meliputi perancangan basis data, *context diagram* dan *data flow diagram*.

1) Perancangan Basis Data

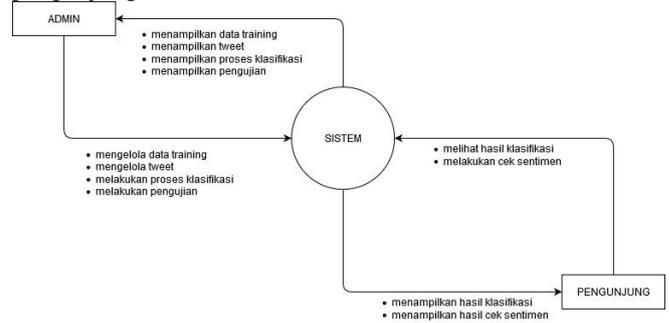
Berikut perancangan basis data yang akan digambarkan dalam bentuk *Entity Relationship Diagram*.



Gambar 4 Entity Relationship Diagram

2) Context Diagram

Berikut penggambaran sistem secara umum pada aplikasi yang digambarkan dalam bentuk *context diagram*. Pada *context diagram* memiliki dua entitas yaitu, admin dan pengunjung.



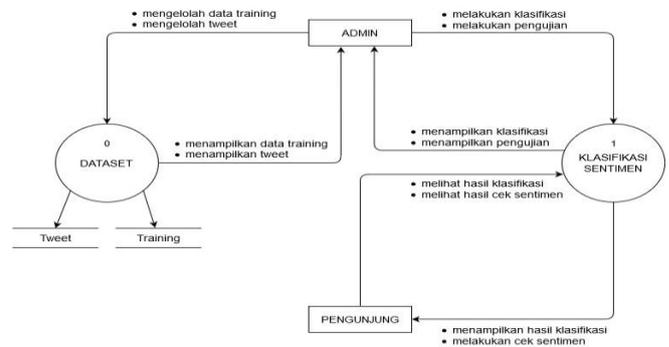
Gambar 5 Context Diagram

3) Data Flow Diagram

Berikut penggambaran sistem lebih detail dari penjelasan context diagram di atas. Dari context diagram diatas didapatkan 2 DFD yaitu:

a) DFD Level 1

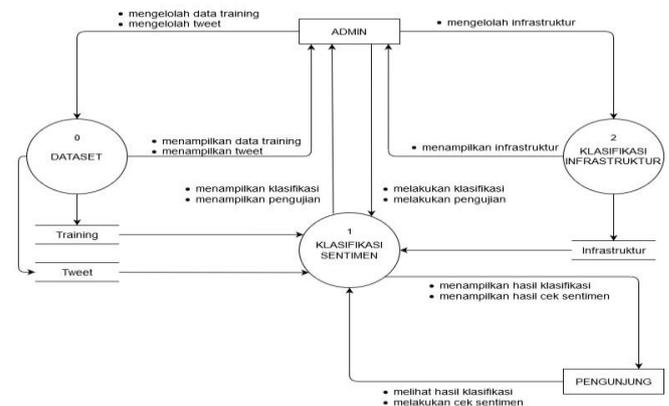
DFD level 1 memiliki dua proses yaitu dataset dan klasifikasi sentimen. Dan memiliki dua *datastore* yaitu *training* dan *tweet*.



Gambar 6 DFD Level 1

b) DFD Level 2

DFD level 2 memiliki tiga proses yaitu dataset, klasifikasi sentimen dan klasifikasi infrastruktur. Dan memiliki tiga *datastore* yaitu *training*, *tweet* dan *infrastruktur*.

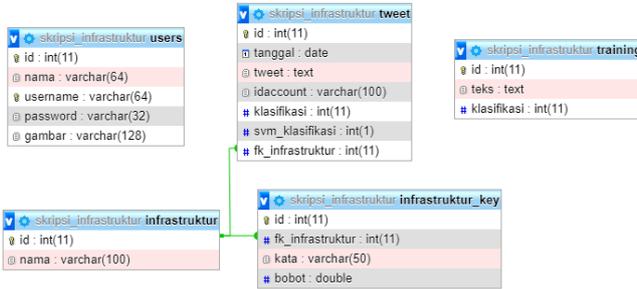


Gambar 7 DFD Level 2

IV. IMPLEMENTASI DAN PENGUJIAN

A. Implementasi Database

Implementasi database dengan nama skripsi_infrastruktur yang memiliki lima tabel, yaitu tabel *tweet*, *infrastruktur*, *infrastruktur_key*, *training* dan *users* sesuai dengan analisis dan perancangan sebagai berikut:

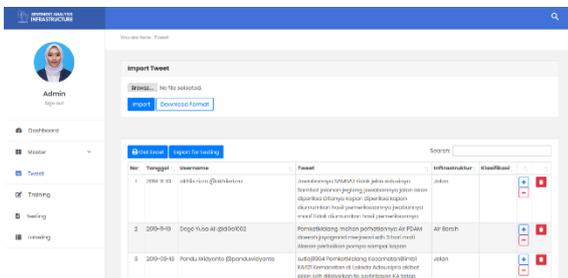


Gambar 8 Implementasi Database

B. Implementasi Tampilan

1) Tweet

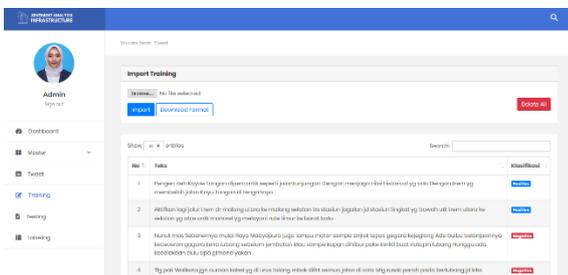
Interface menu *tweet* menampilkan data *tweet* yang didapatkan dari proses *import*. Pada halaman ini dapat mengimport, menghapus dan melakukan klasifikasi secara manual.



Gambar 9 Interface Menu Tweet

2) Training

Interface menu *training* digunakan untuk mengimport file data *training* berupa *xlsx* dan menampilkan hasil *import file*.



Gambar 10 Interface Menu Training

3) Testing

Interface menu *testing* berisi informasi dari dataset yang digunakan untuk proses *testing* dan hasil dari klasifikasi. Interface menu *testing* juga menyediakan informasi tentang tingkat akurasi yang dihasilkan klasifikasi.



Gambar 11 Interface Menu Testing (Bagian Preprocessing)



Gambar 12 Interface Menu Testing (Bagian hasil dan akurasi)

4) Home

Interface halaman *home* merupakan tampilan awal yang dapat diakses oleh pengunjung untuk melihat hasil *tweet* dan melakukan cek sentimen.



Gambar 13 Interface Home (Bagian Filter Data)



Gambar 14 Interface Home (Bagian Cek Sentimen)

C. Pengujian

Pengujian sistem informasi dilakukan dengan cara melakukan pengujian *Blackbox* yaitu pengujian yang ditujukan untuk memeriksa apakah fitur berjalan sesuai dengan yang ditentukan. Sedangkan untuk pengujian akurasi analisis sentimen dengan menggunakan metode *Support Vector Machine* dapat dilakukan dengan cara membandingkan klasifikasi manual dengan klasifikasi otomatis. Klasifikasi manual dilakukan dengan bantuan 2 orang.

1) Pengujian Fungsionalitas Sistem

Pengujian fungsionalitas sistem dilakukan dengan cara menjalankan tiap fitur dalam aplikasi dan melihat kesesuaian hasil yang terjadi dengan hasil yang diharapkan.

2) Pengujian Akurasi Sistem

Pengujian akurasi sistem dilakukan dengan cara menghitung nilai dari *accuracy*, *precision* dan *recall*. Rumus untuk menghitung nilai *accuracy* sebagai berikut:

$$\frac{TP + TN}{TP + FP + FN + TN} * 100 \quad (7)$$

Sedangkan untuk menghitung nilai *precision* menggunakan rumus sebagai berikut:

$$\frac{TP}{TP + FP} * 100 \quad (8)$$

Dan untuk menghitung nilai *recall* menggunakan rumus sebagai berikut:

$$\frac{TP}{TP + FN} * 100 \quad (9)$$

Keterangan

- TP : Jumlah hasil positif benar
- TN : Jumlah hasil negatif benar
- FP : Jumlah hasil positif salah
- FN : Jumlah hasil negatif salah

Pengujian dilakukan dengan jumlah data *training* yang berbeda. Tabel 2 akan menyajikan hasil pengujian akurasi sistem.

TABEL 2 PENGUJIAN AKURASI SISTEM

Data Training 70%		
Accuracy	Precision	Recall
73.10 %	89.29 %	37.09 %
Data Training 80%		
Accuracy	Precision	Recall
72.62 %	84.14 %	45.92 %
Data Training 90%		
Accuracy	Precision	Recall
86.42 %	96.43 %	22.73 %

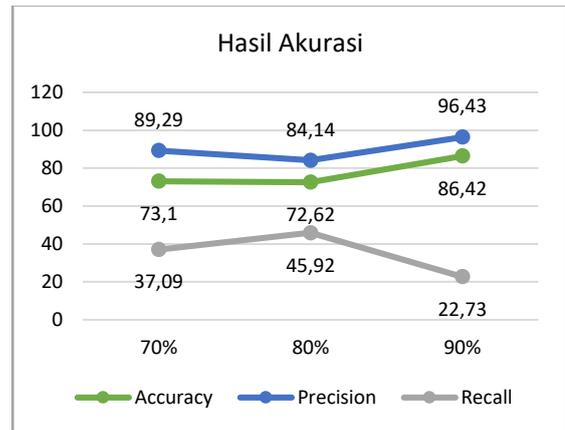
Berdasarkan tabel 2, terdapat perbedaan nilai *accuracy*, *precision* dan *recall* dari beberapa pengujian. Hal tersebut menunjukkan bahwa semakin besar data *training* yang digunakan, maka nilai *accuracy*, *precision* dan *recall* cenderung mengalami peningkatan. Namun pada hasil pengujian data *training* 80% nilai *accuracy* dan *precision* justru mengalami penurunan.

V. HASIL DAN PEMBAHASAN

A. Pengaruh Support Vector Machine

Pengaruh *Support Vector Machine* terhadap akurasi klasifikasi *tweet* berdasarkan hasil pengujian yang dilakukan dengan cara menghitung nilai *accuracy*, *precision* dan *recall*.

Pada penelitian ini jumlah data yang valid digunakan untuk melakukan pengujian sebanyak 1623 data. Data tersebut dilakukan pengujian sebanyak tiga kali dengan jumlah data *training* yang berbeda yaitu 70%, 80% dan 90%. Hasil dari pengujian tersebut direpresentasikan dalam bentuk grafik sebagai berikut.

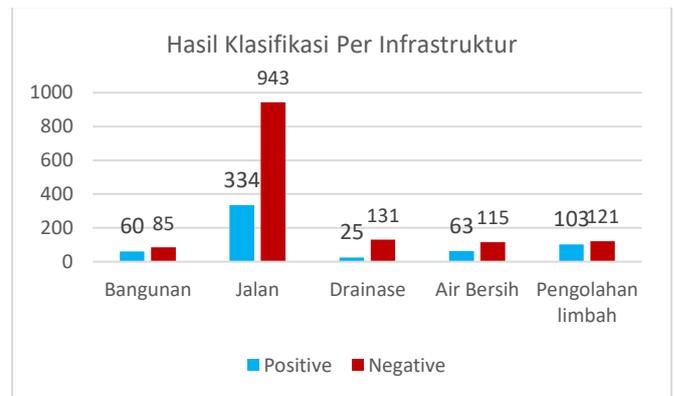


Gambar 15 Hasil Akurasi

Berdasarkan grafik di atas diperoleh kesimpulan bahwa nilai *accuracy* dan *precision* mengalami penurunan, namun nilai *recall* mengalami peningkatan pada data *training* 80%. Lalu pada data *training* 90%, nilai *accuracy* dan *precision* mengalami kenaikan berbanding terbalik dengan nilai *recall* yang mengalami penurunan yang cukup drastis.

B. Klasifikasi Infrastruktur

Data keseluruhan yang telah diambil dari bulan Januari 2018 sampai Desember 2019 didapatkan 1980 data *tweet* mengenai infrastruktur di Kota Malang. Dari data tersebut dilakukan klasifikasi sentimen menggunakan metode *Support Vector Machine* dan menghasilkan data sebagai berikut.



Gambar 16 Hasil Klasifikasi Per Infrastruktur

Berdasarkan bagan di atas diperoleh kesimpulan bahwa dari data yang diambil pada tahun 2018 sampai 2019 tentang pembangunan infrastruktur lebih banyak bersentimen negatif daripada positif.

VI. KESIMPULAN

Berdasarkan hasil penelitian dan pengujian yang telah dilakukan dapat ditarik kesimpulan sebagai berikut:

- a. Metode *Support Vector Machine* dapat digunakan untuk melakukan klasifikasi sentimen *tweet* tentang pembangunan infrastruktur.
- b. Pengujian dengan menggunakan jumlah data *training* yang berbeda dapat mempengaruhi nilai *accuracy*, *precision* dan *recall*. Hasil akurasi pengklasifikasian dengan metode *Support Vector Machine* untuk data

training sebesar 70%, 80% dan 90% yaitu 73.10%, 72.62% dan 86.42%.

- c. Data *tweet* yang diambil pada tahun 2018 sampai 2019 tentang pembangunan infrastruktur lebih banyak bersentimen negatif daripada positif. Dengan jumlah data yang bersentimen negatif sebanyak 1395 data dan 585 untuk data positif.

DAFTAR PUSTAKA

- [1] Ratri, N. (2019, April 30). *Darurat Infrastruktur Ekonomi, Wali Kota Malang: Langkah Pertama Urai Kemacetan*. Retrieved from Malang Times: <https://www.malangtimes.com/baca/38843/20190430/183500/darurat-infrastruktur-ekonomi-wali-kota-malang-langkah-pertama-urai-kemacetan>
- [2] Reyhana, Z. (2018). Analisis Sentimen Pendapat Masyarakat Terhadap Pembangunan Infrastruktur Kota Surabaya Melalui Twitter dengan Menggunakan Support Vector Machine dan Neural Network. *Masters thesis, Institut Teknologi Sepuluh Nopember*.
- [3] Fidyawan, M. A. (2018). Implementasi Twitter Sentiment Analysis Untuk Review Film Menggunakan Algoritma Support Vector Machine. *Jurnal Informatika Polinema*.
- [4] Feldman, R, Sanger, J., “*The Text Mining Handbook : Advanced Approaches in Analyzing Unstructured Data*”. New York : Cambridge University Press, 2007.
- [5] Liu, B. (2012). Sentiment Analysis and Opinion Mining. Dalam B. Liu, *Sentiment Analysis and Opinion Mining* (hal. 1). Morgan & Claypool.
- [6] Turland, M. (2010). *php|architect’s Guide to Web Scaping With PHP*. In M. Turland, *php|architect’s Guide to Web Scaping With PHP* (p. 2). Marco Tabini.
- [7] *Twitter*. (2017). Retrieved from Wikipedia: <https://id.wikipedia.org/wiki/Twitter>
- [8] Sakar, D. (2016). *Text Analytics with Phyton*. Bangalore, Karnataka: apress.