E-ISSN: 2715-2731

Perbandingan Sentimen Arus Mudik-Balik Pelaku Perjalanan Dalam Negeri (PPDN) Berbasis Algoritma Klasifikasi

Vion Age Tricahyo[1,*], Dessy Amry Raykhamna[2]

[1] Program Studi Ilmu Komputer, Universitas Nahdlatul Ulama Blitar, Indonesia
[2] Program Studi Teknik Informatika, Universitas Brawijaya, Indonesia

[1]vionage@unublitar.ac.id; [2]dessylie@gmail.com

* penulis korespondensi

INFO ARTIKEL

Sejarah Artikel
Diterima: 29 Juni 2022
Direvisi: 21 Agustus 2022
Diterbitkan: 24 Agustus 2022

Kata Kunci
Algoritma Klasifikasi
Covid-19
Mudik-Balik
Sentimen

ABSTRAK

Perjalanan mudik-balik sudah menjadi budaya dan kebiasaan setiap tahun bagi penduduk yang sedang merantau. Kebiasaan ini terhenti selama dua tahun dikarenakan pandemi Covid-19 menyebar ke seluruh negara termasuk Indonesia. Pada tahun ini melalui surat edaran satuan tugas no 16 tahun 2022, dinyatakan bahwa masyarakat diperbolehkan melakukan perjalanan dengan berbagai aturan dan syarat. Aturan dan syarat ini mengakibatkan banyak sentimen pro dan kontra yang terjadi di masyarakat terutama di media sosial twitter. Data berasal langsung dari *netizen* dan tanggapan dari media nasional menghasilkan variasi data. Data yang diolah total sebanyak 11.971 data dengan memanfaatkan *Word2Vec* sebagai ekstraksi fitur menggunakan kata kunci mudik dan balik. Hasil penelitian menggunakan *naïve bayes*, *decision tree* dan *support vector machine* menghasilkan perbandingan akurasi tertinggi yakni sebesar 97.5 % dan terendah 82.96 %.

PENDAHULUAN

Selama 2 tahun terakhir dunia menghadapi wabah virus yang terus menyebar ke seluruh negara di dunia tak terkecuali Indonesia. Wabah virus ini diidentifikasi bernama *Severe Acute Respiratory Syndrome Coronavirus 2* (SARS-COV2) dan lebih banyak dikenal dengan Covid-19. Persebaran Covid-19 ini lebih cepat dibandingkan dengan virus lain yang pernah menjadi wabah seperti SARS meski memiliki keidentikan sebesar 80% [1]. Wabah ini selain menelan banyak korban jiwa namun juga menyebabkan perkekonomian suatu negara menurun bahkan kacau. Salah satu yang dampak lain dan menimbulkan masalah baru adalah pembatasan sampai dilarangnya kegiatan sosial diluar ruangan. Masyarakat yang terbiasa dengan sosialisasi dibatasi dengan hanya berada didalam rumah tanpa mobilisasi.

Mobilisasi dalam dua tahun terakhir terutama antar negara bahkan antar kota-provinsi sangat dibatasi. Selain berbagai pembatasan, harga transportasi tiap perjalanan melambung tinggi. Hal ini belum ditambah dengan persyaratan tes antigen sampai PCR dan juga vaksinasi.Vaksinasi memang menjadi hal wajib bagi masyarakat yang melakukan perjalanan baik luar maupun dalam negeri, karena dengan vaksinasi membantu mengurangi penyebaran Covid-19 [2]. Namun keadaan dilapangan terutama masyarakat yang akan melalukan perjalanan menimbulkan pro dan kontra. Terlebih lagi perjalanan ini bukan perjalanan dinas atau liburan melainkan perjalanan mudik-balik dalam rangka hari raya Idul Fitri. Momen Idul Fitri memang sangat berbeda dan merupakan hal yang paling dinantikan bagi perantau. Pada momen tersebut dimanfaatkan bagi perantau untuk melakukan silaturahmi dan berkumpul dengan keluarga besar [3]. Penantian ini membuat masyarakat antusias ingin mudik dikarenakan ditambahnya libur Ramadan sampai Idul Fitri tiba. Ini belum ditambah dengan libur sekolah dan diterapkannya *Work Form Home* (WFH) selama akhir bulan April

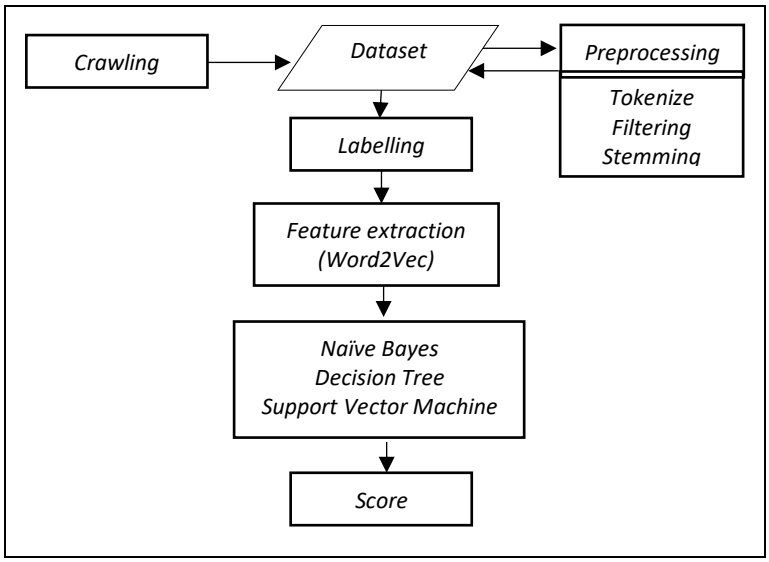
sampai Juni 2022. Melihat kebijakan yang diambil pemerintah dari awal pandemi hingga sekarang seperti dimulainya langkah 5M, Pembatasan Sosial Berskala Besar (PSBB), dan Pemberlakuan Pembatasan Kegiatan Masyarakat berbasis mikro (PPKM Mikro) sampai WFH kini telah berakhir. Hal ini telah disampaikan secara langsung dalam konferensi pers oleh Presiden Joko Widodo. Dengan kebijakan ini pada awal Juli 2022, ASN, swasta dan pekerja berbagai sektor bisa melakukan *Work From Office* (WFO) secara penuh. Hal ini juga di dukung dengan peraturan menteri dalam negeri bahwa PPKM level 1 WFO diperbolehkan 100% [4]. Dengan begitu sisa-sisa mudik-balik akan terjadi dan masyarakat yang masih melakukan WFH akan kembali ke domisili. Pembahasan kebijakan ini pada media sosial twitter masih menjadi topik yang cukup ramai diperbicangkan. Hal ini tidak mengherankan melihat Indonesia merupakan salah satu pengguna media sosial teraktif didunia. Data terbaru menunjukkan pengguna Twitter didunia mencapai 1,3 milliar akun dengan 500 juta lebih tweets yang terjadi setiap harinya [5]. Dan ini akan terus bertambah seiring waktu dan fenomena tertentu terjadi dan menjadi perbincangan publik.

Pada penelitian dengan topik "mudik" sebelumnya dilakukan oleh [6], menggunakan kata terbatas yaitu mudik. Selain itu metode yang digunakan adalah *Naïve Bayes* (NB) serta *Support Vector Machine* (SVM). Penelitian selanjutnya yaitu masih dengan penggunaan *Support Vector Machine* untuk sentimen analisis larangan mudik dilakukan oleh [7]. Dan yang terbaru adalah penggunaan *Naïve Bayes* dan *Decision Tree* dengan topik sama yang dilakukan oleh [8]. Dari seluruh penelitian terbauru diatas rentang akurasi yang dihasilkan adalah mulai dari 56,56 % sampai 87%. Dari semua penelitian diatas belum ada yang menggunakan *Word2Vec* sebagai metode untuk ekstraksi fitur serta hasil perbandingannya dalam dalam penerapan algoritma. Disamping dengan *Binary* TF, *Raw* TF dan TF.IDF yang sering digunakan dan telah diterapkan dipenelitian sebelumnya, penggunaan model *Word2Vec* dianggap lebih mumpuni karena *Word2Vec* membutuhkan contoh data dengan skala besar untuk merepresentasikan kata dan kata-kata mirip yang lebih dekat [9]. Berdasarkan hasil kajian penelitian diatas, peneliti akan menggabungkan ketiga metode yaitu *Naïve Bayes*, *Decision Tree* dan *Support Vector Machine* serta akan melakukan perbandingan didalamnya. Selain itu lingkup kata kunci akan ditambah sehingga akan menghasilkan data yang lebih bervariasi ketika diproses dalam menghasilkan nilai vektor. Fokus untuk mendapatkan akurasi yang tinggi juga sangat penting untuk mendukung keberlanjutan penelitian ini.

METODE

Pada penelitian yang dilakukan ini, pemrosesan dasar tetap dilakukan untuk mendapatkan nilai vektor setiap kata. Untuk menghasilkan setiap kata dibutuhkan pra-pemrosesan. Bagian paling awal adalah mendapatkan data dengan cara *crawling* di media sosial twitter. Setelah itu melakukan labeling untuk memisahkan kata positif dan negatif. Karena teknik penentuan label sangat berpengaruh dalam penelitian ini untuk mendukung *kesesuaian (ground truth data)*, bahkan sebelum proses *pre-processing* dimulai. Penentuan label dibantu dengan *Sastrawi library* sekaligus membantu dalam pengelompokan kata. Kata yang dikelompokkan akan menggunakan ekstraksi fitur *Word2Vec* untuk mendapatkan nilai vektor. Word2Vec adalah alat untuk menghitung representasi vektor kata yang dirilis oleh Google di 2013. Alat ini bisa mewakili kata input dengan vektor fitur fix-length dengan menyediakan dua model arsitektur utama untuk komputasi vektor secara berkelanjutan. Representasi vector tersebut adalah *Continuous Bag-Of-Words* (CBOW) model dan *Continuous Skip-Gram* (Skip-Gram) model [10].

Data yang dihasilkan dari proses crawling adalah sejumlah 11.971 tweets dengan kata kunci utama mudik dan balik. Data diambil dari tanggal 22 Mei sampai Juni tanggal 30. Pada penelitian ini difokuskan pada data Pelaku Perjalanan Dalam Negeri (PPDN) yaitu pemudik yang akan melakukan perjalanan "mudik" menuju kampung halamannya dan pemudik yang akan kembali "balik" ke tempat domisili tempat bekerja. Kurun waktu pengambilan data diatas adalah batas dari berahkhirnya PPKM (Pemberlakuan Pembatasan Kegiatan Masyarakat) Level 1 yang berdasarkan peraturan menteri dalam negeri berakhir pada tanggal 4 Juli 2022. Secara garis besar urutan yang dilakukan dalam penelitian ini ditampilkan pada Gambar 1.



Gambar 1. Alur Penelitian

Pada alur penelitian, setelah *crawling* selesai diakukan langkah selanjutnya adalah proses *case folding.* Proses *case folding* merupakan proses untuk mengubah kata ke dalam huruf kecil atau *lowercase* [11]. Setelah itu proses *tokenize* yaitu proses normalisasi untuk menghilangkan *usernamae,* URL dan RT pada *tweet*. Pada proses *tokenize* untuk mendapatkan kata unik maupun dasar sehingga diperoleh frekuensi dan nilai bobot disetiap kata. Kemudian proses dilanjutkan dengan *stopword removal* yaitu proses untuk menghilangkan kata yang tidak memiliki kaitan dengan analisis sentiment seperti kata penghubung dan berulang-ulang. Langkah terakhir adalah melakukan *stemming* yang digunakan untuk mendapatkan kata dasar murni tanpa tambahan *prefiks* maupun *sufiks.* Berikut ini adalah contoh tabel hasil crawling dan pengelompokannya berdasarkan kategori positif dan negatif:

Tabel 1. Contoh Dataset

Data	Label	Kata Kunci
1520108291245764609 2022-04-30 01:30:33 +0700 **Mudik**, Tol Jakarta-Cikampek di Bekasi Macet pollll https://t.co/YID8hobgGH	-	Mudik
1520070927228170246 2022-04-29 23:02:05 +0700 **Mudik** Gratis Sepeda Motor, Ribuan Orang Berdatangan Ke Pelabuhan Tanjung Priok https://t.co/wPlDH8Plwf	+	Mudik
1521716255786672129 2022-05-04 12:00:02 +0700 #SahabatKAI yg belum dpt tiket **balik**, bisa memanfaatkan program motor gratis (Motis). Daftar sekarang atau kunjungi stasiun yg melayani program #Motis2022. #KAI121 https://t.co/8m40VbfBww	+	Balik
1523328694671986689 2022-05-08 22:47:17 +0700 mengkhayal masih pgn dirumah smpe hari rabu tp skrg mau ga mau harus **balik** soalnya libur cuma 3 hari, so sadddd:(	-	Balik

Selanjutnya implementasi *stemming* dari *corpus* atau *library* dengan *stem_sastrawi* pada *library* yang ada pada *python*. Tujuan dari penggunaan dan penerapan *corpus* adalah untuk memperkuat data sentimen hasil label di atas. Sebelumnya berikut ini adalah contoh pengelompokan data dari kata positif dan negatif yang digunakan sebagai acuan *labeling*, ditunjukkan pada Tabel 2.

Tabel 2. List Label Setiap Kata

Daftar Kata	Label
Absurd, abnormal, acuh, belum pasti, bencana, cabul, cacat, dendam, egois, ejekan, fatal, fitnah, gebit, gila, hujat, ilegal, jahat, jelek, kebencian, kebiri, kritis, lemah, mabuk, mafia, murka, najis, nakal, omelan, otoriter, pecandu, polusi, racun, radikal, sabotase, sakit, tekanan, telanjang, terjelek, ugal-ugalan, virus, vulgar, zina	Negatif
Aman, amanah, berani, bergizi, bonus, cerdik, dinamis, ditegakkan, emas, empati, enak, favorit, fenomenal, gesit, giat, harum, hormat, ideal, idola, juara, jujur, kaya, keadilan, kuat, lancer, layak, manis, mulus, murah, nikmat, nyaman, optima, optimis, piala, pintar, rajin, riang, sempurna, senang, teladan, teliti, unik, yeah	Positif

Dengan daftar kata pada tabel diatas, protokol pelabelan semakin kuat dan dapat diuji kebenarannya. Sehingga proses *pre-processing* semakin cepat dan mudah. Proses *pre-processing* ini dimulai dari *case folding* hingga *stemming*. Dalam *stemming*, perhitungan dibentuk oleh bentuk kata dasar. Berikut ini adalah contoh stemming serta jumlah token per kata dalam setiap *tweet*, ditunjukkan pada Tabel 3.

Tabel 3. Stemming dengan token

Stemming List	Kunci
[('mudik', 4), ('tol', 2), ('jakarta', 2), ('cikampek', 2), ('bekasi', 1), ('macet', 1), ('gratis', 1), ('sepeda', 1), ('motor', 1), ('pelabuhan', 1), ('orang', 1)	Mudik
[('balik', 3), ('motor', 2), ('program', 2), ('stasiun', 2), ('rabu', 1), ('libur', 1), ('melayani', 1), ('hari', 1), ('indonesia', 1)	Balik

Dengan hasil *stemming* diatas menunjukkan bahwa setiap token memiliki nilai. Nilai ini didapatkan melalui proses selanjutnya yaitu ekstraksi fitur. *Feature Extraction* atau ekstraksi fitur adalah pengambilan ciri (*feature)* dari suatu bentuk kalimat atau objek yang nantinya nilai yang didapatkan akan dilakukan analisa pada proses selanjutnya. Dalam penelitian ini implementasi fitur yang digunakan adalah *Word2Vec*. Setelah selesai proses selanjutnya adalah klasifikasi sentimen menggunakan berbagai metode. Sebelum sampai pada proses tersebut perlu diperoleh nilai vektor dari setiap token. Token setiap kata mengikuti *list* yang ada pada *corpus sastrawi*. Penyajian data tetap terpisah untuk mengetahui perbandingan pro dan kontra dari setiap kata kunci.

Persamaan Matematika

Penerapan fitur Word2Vec akan menghitung vektor yang ada pada setiap token kata. Penggunaan Persamaan (1) dan (2) pada ekstraksi fitur ini untuk mendapatkan nilai vektor per token.

$$emb_i = \frac{\sum_{j}^{n}=i\ v_j^i}{n} \quad (1)$$

$$V_{doc} = [emb_1, emb_2, emb_3 \ldots emb_n] \quad (2)$$

Variabel emb_i merupakan representasi nilai vektor dari dokumen untuk dimensi ke-i. Variabel n bernilai total jumlah kata yang ada di dalam setiap dokumen. Sedangkan v_j^i merupakan isi elemen ke-i dari representasi nilai vektor kata ke-j. Dari rumus tersebut dihasilkan nilai persamaan vektor dokumen V_{doc}= $[emb_1, emb_2, emb_3 \ldots emb_n]$ sebagai model representasi dari sebuah kalimat di twitter. Pada implementasinya menggunakan *tools Rapid Miner*. Sebelum mendapatkan nilai vektor perlu *data training* untuk melatih data. Berikut ini training yang perlu dilakukan:

```
model = gensim.models.Word2Vec (document, vector_size = 100, window = 5, min_count = 1, workers = 8)
model.train (documents, total_examples = len (documents), epochs = 1)
```

Dari *code* dan data latih diatas akan didapatkan nilai *vector* sebagai berikut :

Tabel 4 Vektor Data 'Mudik'

Kata	emb_1	emb_1	emb_1	emb_n
mudik	0.10933	-0.0948	-0.0811	……
tol	0.08412	0.02427	-0.057	……
jakarta	0.07107	0.06424	0.04294	……
cikampek	-0.0525	-0.013	-0.1094	……
bekasi	-0.0586	-0.018	-0.119	……
macet	-0.0066	0.05306	-0.0124	……
Dst….	……	……	……	……

Tabel 5 Vektor Data 'Balik'

Kata	emb_1	emb_1	emb_1	emb_n
balik	0.084119	0.024265	-0.05702	……
motor	0.109329	-0.09475	-0.08109	……
program	0.071075	0.064239	0.042938	……
stasiun	-0.0525	-0.01297	-0.10939	……
rabu	-0.05858	-0.01803	-0.11897	……
libur	-0.00662	0.053064	-0.01236	……
Dst….	……	……	……	……

Setelah pengujian data latih selesai dan mendapat nilai vektor, bisa dilanjutkan dengan melakukan teknik *machine learning* yaitu klasifikasi menggunakan beberapa model seperti *Naïve Bayes, Decision Tree* dan *Support Vector Machine.*

Naïve Bayes

Teori *Bayesian* adalah bagian penting dari teori induktif *Bayesian* subjektif. *Naïve Bayes* merupakan pengambilan keputusan untuk memperkirakan probabilitas subjektif melalui perhitungan dari beberapa keadaan yang belum diketahui di bawah informasi yang tidak lengkap [12]. Untuk itu diperlukan cara yang optimal dengan cara memodifikasi nilai probabilitas. Salah satu ciri utama dari *naïve bayes* ini yaitu independensi yang kuat dari setiap suatu kondisi atau kejadian. Apabila terdapat dua kejadian berbeda yang terpisah (misalkan X dan H), maka *Teorema Bayes* dapat dirumuskan sesuai Persamaan (3).

$$P(H|X) = \frac{P(X|H)}{P(X)}.P(H) \quad (3)$$

Decision Tree

Algoritma *decision tree* digunakan untuk mengimplementasikan ekstraksi fitur menggunakan *information gain.* Information gain adalah ukuran perhitungan statistik untuk mengukur efektivitas dari sebuah atribut suatu sampel data kedalam sebuah klasifikasi [13]. Sampel yang belum diklasifikasi dimasukkan ke dalam kelas yang sudah tersedia. *Root node* merupakan jalur pertama dalam pengujian data dan memiliki nilai atribut terbesar. Dan *leaf node* adalah akhir dari pengujian yang dimana menghasilkan prediksi berdasarkan kelas data. Atribut yang memiliki nilai *continuous* perlu didiskritkan. Adapun perhitungan untuk mencari nilai tersebut dengan Persamaan (4).

$$I(S_{1.}\ S_{2.....}\ S_{n.}) = -\sum_{i=1}^{\infty} P_1\ log_2(P_i) \quad (4)$$

Support Vector Machine

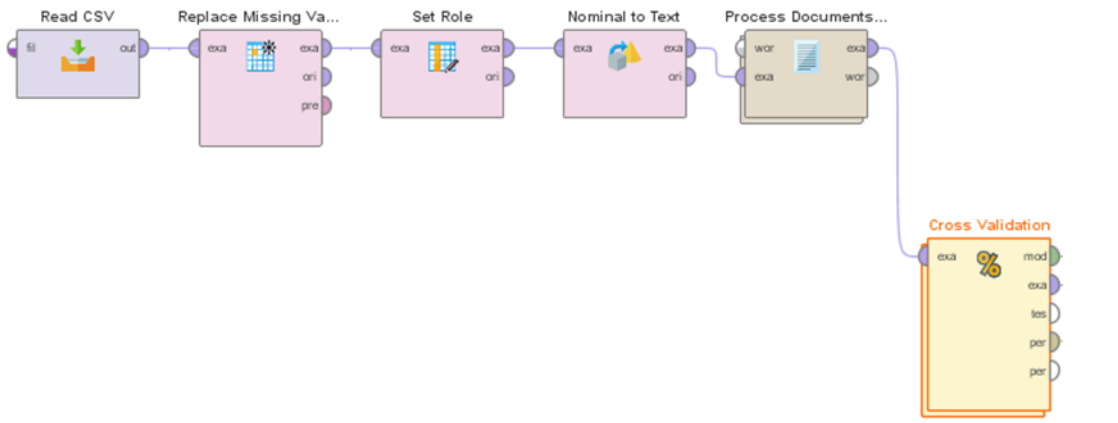
Support Vector Machine (SVM) merupakan salah satu metode handal dalam mengatasi permasalahan dalam klasifikasi data. SVM mengenali dan mempelajari pola yang nantinya akan membantu dalam klasifikasi dan analisis regresi. SVM merupakan kasus linier yang dipisahkan dan dibagi sesuai kelasnya [14]. Fungsi linier tersebut dapat didefinisikan menjadi Persamaan (5).

$$\text{K (x, xi)} = x \cdot x^T \quad (5)$$

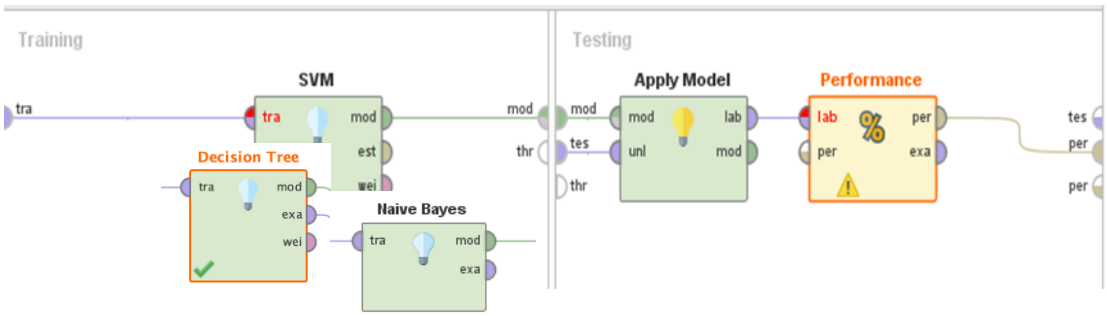
Karena sebuah SVM merupakan pengklasifikasi, maka berfungsi sebagai *hyperline* atau pemisah untuk menemukan fungsi terbaik melalui himpunan pelatihan. Algoritma SVM dalam data latih akan terbangun model yang dapat memprediksi apakah data yang baru masuk ke dalam suatu kategori atau jatuh dan tidak masuk ke dalam kategori.

HASIL DAN PEMBAHASAN

Setelah proses *pre-processing* dan melakukan training data selesai, langkah selanjutnya adalah implementasi klasifikasi dari algoritma yang dipilih. Pada tahap ini akan diperoleh nilai klasifikasi, perbedaan hasil klasifikasi dan menentukan hasil klasifikasi paling tinggi dan terbaik dalam kasus penelitian ini. Pengukuran kinerja dengan menggunakan persamaan *confusion matrix*, yang didalamnya terdapat 4 istilah. Keempat representasi klasifikasi tersebut adalah *True Positive* (TP), *True Negative* (TN), *False Positive* (FP) dan *False Negative* (FN). Dalam proses ini harus disiapkan data dalam bentuk Excel, CSV atau format lain yang didukung sebelum penentuan *replace missing values, set role, nominal to teks, process document* dan *cross-validation* yang saling terkait. Pada proses *cross-validation*, pada tahap ini dilakukan uji validasi hasil klasifikasi yang telah di *generate* menggunakan metode *K-Fold test data sharing* dengan k = 10. Berikut adalah gambar yang menunjukkan proses dan desain *cross-validation*. Diikuti oleh beberapa klasifikasi yang digunakan.



Gambar 2. Pengambilan-Pemrosesan Data



Gambar 3. *Training - Testing* Algoritma

Proses diatas adalah kelanjutan dari persiapan data yang dilakukan sebelumnya. Pada proses ini operator klasifikasi berada di data *training* yang mana seluruh data yang dipersiapkan sebelumnya dihubungkan oleh *cross validation* ke operator *training* klasifikasi. Setelah itu data *training* akan diimplementasikan di *apply model* dan dihitung nilai *performance*-nya. Pada tahap training diatas, operator klasifikasi yang digunakan adalah *Naïve Bayes, Decision Tree* dan *Support Vector Machine.*

Hasil Model

Dari hasil pengujian sentimen, seluruh data diuji untuk memperoleh nilai akurasi, *pressision, recall* dan jumlah nilai positif dan negatif dari masing-masing data. Pada tabel berikut diambil nilai tertinggi dari masing-masing proses, karena hal ini menunjukkan akurasi yang terbaik yang bisa dan telah dicapai oleh masing-masing data uji. Tabel 6 adalah hasil model klasifikasi dari kedua kata kunci.

Tabel 6 Hasil Perbandingan Klasifikasi.

Klasifikasi	Akurasi Mudik	Akurasi Balik	Cross Validation Mudik	Cross Validation Balik
Naïve Bayes	94.9 %	83.5 %	+/- 5.55 %	+/- 13.44 %
Decision Tree	97.5 %	93.33 %	+/- 6.91 %	+/- 12.05 %
SVM	83.67 %	82.96 %	+/- 12.32 %	+/- 8.39 %

Dari hasil yang diperoleh diatas, dapat diketahui dan diperoleh nilai dari berbagai klasifikasi dengan berbagai model. Data mudik mendominasi keunggulan akurasi dibandingan dengan data balik. Data mudik dengan akurasi tertinggi diperoleh dari metode

Decision Tree sebesar 97.5 % dan terendah menggunakan metode SVM sebesar 83.67 % . Sedangkan data balik memiliki akurasi tertinggi sebesar 93.33% dengan metode yang sama yaitu *Decision Tree*. Data dengan nilai terendah diperoleh melalui metode SVM dengan 82.96% dengan selisih diatasnya yaitu 83.5% dengan metode *Naïve Bayes.*

KESIMPULAN

Pada bagian ini peneliti dapat menyampaikan bahwa metode ekstraksi fitur *Word2Vec* dapat diimplementasikan dan terlihat cocok untuk mendapatkan nilai vektor. Selain itu fitur parameter yang ada pada *Word2Vec* sangat beragam dan memudahkan untuk mendapatkan hasil data yang bervariasi. Penelitian ini menghasilkan klasifikasi yang tinggi yang diperoleh dari kedua kata kunci yang terpisah untuk perbandingan. Nilai tertinggi klasifikasi diperoleh melalui klasifikasi *Decision Tree* dengan akurasi tertinggi mencapai 97,5% untuk kata kunci mudik. Sedangkan nilai akurasi terendah sebesar 81,96% menggunakan klasifikasi *Support Vector Machine* untuk kata kunci balik.

REFERENSI

[1] D. W. Nugroho, W. Indah, Alanish, N. Istiqomah, I. Cahyasari, M. Indrastuti, P. Sugondo and A. Isworo, "Literature Review : Transmisi Covid-19 dari Manusia ke Manusia Di Asia," *Journal of Bionursing*, pp. 101-112, 2020.

[2] P. Stepahnie, S. Enjelina, M. Angelica and I. Martinelli, "Aspek Hukum Pelaksanaan Vaksinasi Covid-19 di Indonesia," in *Prosiding Seminar Nasional Hasil Penelitian dan Pengabdian kepada Masyarakat (SENAPENMAS)*, Jakarta, 2021.

[3] A. .. Nugraha and D. Hidayat, "Persepsi Masyarakat Mengenai Peraturan Larangan Mudik Selama Covid-19," *Jurnal Ilmu Komunikasi*, pp. 84-97, 2021.

[4] G. Muhamad, "Instruksi Menteri Dalam Negeri Nomor 29 Tahun 2022," Satgas Covid-19, Jakarta, 2022.

[5] T. B. Lee, "Twitter Usage Statistics," Internet Live Stats, Boston, 2022.

[6] S. Sautomo, N. Hafidz, Y. .. Achyani and W. Gata, "Sentiment Analysis Due to "Mudik" Prohibited of Covid-19 Through Twitter," *Jurnal Ilmu Pengetahuan dan Teknologi Komputer*, pp. 7-12, 2020.

[7] R. Pebrianto, T. Rivanie, R. Nurfalah, W. Gata and M. .. Julianto, "Adopsi Algoritme Support Vector Machine untuk Analisis Sentimen Larangan Mudik Lebaran 2020 pada Twitter," *Jurnal Teknik Komputer AMIK BSI*, pp. 193-199, 2020.

[8] R. .. Artika and D. .. Pambudi, "Analisis Sentimen Terhadap Larangan Mudik Akibat Covid-19 Menggunakan Algoritma Naive Bayes dan Decision Tree," UISI, Gresik, 2021.

[9] A. M. Fauzi, "Word2Vec Model for Sentiment Analysis of Product Reviews in Indonesian Language," *International Journal of Electrical and Computer Engineering (IJECE)*, vol. 9, no. 1, pp. 525-530, 2019.

[10] V. .. Tricahyo and S. .. Isa, "Classification of Indonesian Presidential Campaign on Twitter Using Word2Vec," *International Journal of Advanced Trends in Computer Science and Engineering*, pp. 5501-5508, 2020.

[11] Rustiana and Rahayu, "Analisis Sentimen Pasar Otomotif Mobil : Tweet Twitter Menggunakan Naïve Bayes," *Jurnal Simetris Vol 8 No 1*, 2017.

[12] H. Chen, S. Hu and X. Zhao, "Improved Naive Bayes Classification Algorithm for Traffic Risk Management," *EURASIP Journal on Advances in Signal Processing*, pp. 1-12, 2021.

[13] M. Babaei, J. Kulshrestha, A. Chakraborty, E. .. Redmiles, M. Cha and K. .. Gummadi, "Analyzing Biases in Perception of Truth in News Stories and Their Implications for Fact Checking," *IEEE*, pp. 839-850, 2022.

[14] V. Nanda, A. Majumdar, C. Kolling, J. .. Dickerson, K. .. Gummadi, B. .. Love and A. Weller, "Exploring Representational Alignment with Human Perception Using Identically Represented Inputs," arXiv Labs, Cornell University, New York, 2022.