

**SENTIMENT ANALYSIS ALIH FUNGSI
LAHAN PERTANIAN INDONESIA 2017 - 2021**

Alvina Maharani Hasibuan, Matthew Martianus Henry, M. Iqbal

**Anaraunga Team
14030011**



**IPB UNIVERSITY
2021**

ABSTRAK

Perkembangan infrastruktur merupakan tolak ukur kemajuan suatu negara di dunia. Tak terkecuali di Indonesia, perkembangan infrastruktur yang dilakukan pemerintah seringkali melibatkan alih fungsi lahan pertanian. Alih fungsi lahan pertanian kerap kali menimbulkan pro dan kontra di masyarakat. Tujuan penelitian ini adalah membuat suatu model analisis sentimen kebijakan alih fungsi lahan pertanian di Indonesia. Metode yang digunakan untuk analisis sentimen adalah model *logistic regression*, *support vector classifier* dan *long short term memory*. Hasil penelitian ini menunjukkan bahwa model *machine learning* terbaik untuk analisis sentimen alih fungsi lahan pertanian adalah *support vector classifier* dengan skor validasi 80% dan standar deviasi 4%. Skor validasi model *logistic regression* adalah 78% dengan standar deviasi 4%. Skor validasi model *long short term memory* adalah 76% dengan standar deviasi 5%. Model dengan skor validasi terbaik yaitu *support vector classifier* digunakan pada data uji dan menghasilkan akurasi sebesar 85%.

Kata Kunci : alih fungsi lahan pertanian, *logistic regression*, *long short term memory*, *support vector classifier*.

PENDAHULUAN

1. Latar Belakang

Pada pasal 1 nomor 15 UU No. 41 Tahun 2009 disebutkan bahwa alih fungsi lahan pertanian adalah perubahan fungsi lahan pertanian pangan berkelanjutan menjadi bukan lahan pertanian pangan berkelanjutan baik secara tetap maupun sementara. Dalam hal ini alih fungsi lahan sering kali dilakukan untuk diubah menjadi industri, perumahan, maupun bangunan lainnya. Beberapa implementasi dari kebijakan alih fungsi lahan pertanian tidak sesuai dengan UU No. 41 Tahun 2009 yang mengatur tentang perlindungan lahan pertanian pangan berkelanjutan. Sebagaimana dilansir Kompas (2019), Menteri Pertanian RI, Syahrul Yasin Limpo membenarkan alih fungsi lahan pertanian ke lahan non pertanian masih berlangsung bahkan cenderung meningkat. Hal ini jika dibiarkan akan mengancam ketahanan pangan dan memaksa pemerintah untuk membuka keran impor untuk memenuhi kebutuhan pangan masyarakat luas. Alih fungsi lahan pertanian yang masif ini secara tidak langsung bertentangan dengan Pasal 3b UU No. 41 Tahun 2009 yang berbunyi “Menjamin tersedianya lahan pertanian pangan secara berkelanjutan”. Potongan ayat tersebut menunjukkan kewajiban pemerintah dan juga masyarakat dalam menjaga ketersediaan lahan pertanian sehingga menghindari keadaan rawan pangan melalui pengendalian alih fungsi lahan pertanian. Pengembangan infrastruktur negara yang kerap kali diikuti dengan alih fungsi lahan pertanian menuai pro dan kontra di masyarakat.

Penggunaan media sosial di Indonesia semakin berkembang setiap tahunnya. Platform media sosial seperti Facebook, Twitter, maupun Instagram menjadi tempat mengekspresikan pendapat dan ekspresi seseorang dalam berbagai topik. Salah satu

media sosial yang diminati penduduk Indonesia adalah Twitter. Twitter sebagai salah satu platform media sosial memiliki fitur untuk memberikan cuitan (*tweet*) mengenai pendapat seseorang. *Tweet* tersebut dibatasi hanya 140 karakter. Pengguna dapat secara bebas mengemukakan pendapatnya mengenai sesuatu pada platform Twitter, serta memberikan tanggapan terhadap *tweet* orang lain. Jumlah pengguna Twitter di Indonesia sebanyak 15.1 juta penduduk dan menduduki peringkat lima dunia pada April 2021 (Statista 2021). Pengguna twitter di Indonesia sebagian besar adalah generasi muda Indonesia dengan jumlah terbanyak berada pada rentang usia 25 hingga 34 tahun.

Sentiment analysis cuitan pada twitter dapat digunakan untuk mengetahui penilaian masyarakat terhadap kebijakan yang ditetapkan pemerintah. Penelitian sebelumnya mengenai *sentiment analysis* kebijakan pemerintah pada platform Twitter dilakukan oleh Nandini *et al* (2018) mengenai kebijakan impor beras pemerintah menggunakan model *support vector machine* dan pembobotan jumlah *retweet*. Penelitian lainnya yang serupa dilakukan pula oleh Singh *et al* (2020) mengenai *sentiment analysis* kebijakan terbaru pemerintah India di bidang pertanian. Dalam penelitian tersebut, Singh *et al* (2020) mencoba membandingkan hasil dua buah pendekatan yaitu *fivegrams* dan *trigrams*.

2. Tujuan

Penelitian dilakukan dengan tujuan sebagai berikut:

- a. Membuat model *machine learning* untuk menganalisis *tweet* yang berhubungan dengan kebijakan alih fungsi lahan pertanian.
- b. Menganalisis pergerakan sentimen yang terjadi di masyarakat berkaitan dengan alih fungsi lahan pertanian.

3. Manfaat

Penelitian ini diharapkan dapat:

- a. Mempermudah analisis kebijakan pemerintah yang melibatkan alih fungsi lahan pertanian di Indonesia melalui model yang dibuat.
- b. Mengetahui penilaian masyarakat terhadap alih fungsi lahan pertanian yang dilakukan pemerintah.

4. Batasan Penelitian

Batasan penelitian ini meliputi:

- a. Data latih yang digunakan adalah *tweet* mengenai alih fungsi lahan pertanian di Twitter pada rentang waktu 2017 hingga Juni 2021.
- b. Data uji yang digunakan adalah *tweet* mengenai alih fungsi lahan pertanian di Twitter pada Juli 2021.
- c. Data yang diambil hanya berupa opini pengguna Twitter dan tidak meliputi *tweet* dari *news portal*.
- d. Pengolahan data dilakukan menggunakan bahasa pemrograman Python 3 dengan GUI Jupyter Notebook.

STUDI LITERATUR

Logistic Regression

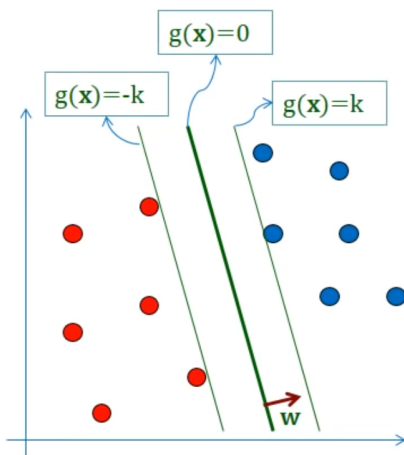
Logistic regression (LR) merupakan metode analisis yang digunakan untuk mendeskripsikan hubungan antara variabel terikat yang memiliki dua kategori atau lebih dengan satu atau lebih variabel bebas, berskala kategorik atau kontinu. Bentuk umum model LR adalah sebagai berikut

$$x = \frac{\exp(\beta_0 + \beta_1 x_{1i} + \beta_1 x_{2i} + \dots + \beta_p x_{pi})}{1 + \exp(\beta_0 + \beta_1 x_{1i} + \beta_1 x_{2i} + \dots + \beta_p x_{pi})}$$

Salah satu cara untuk mengelompokkan data adalah dengan menggunakan Regresi Logistik Biner. Metode Regresi Logistik Biner merupakan metode klasik yang digunakan untuk mengetahui pola hubungan antara variabel respon yang bersifat biner yakni terdiri dari 0 dan 1 dengan variabel prediktornya (Hosmer et al 2000). Hasil dari variabel respon yang terdiri dari 2 kategori yaitu sukses dan gagal yang dinotasikan dengan 1 (sukses) dan 0 (gagal).

Support Vector Classifier

Support vector classifier (SVC) merupakan salah satu dari algoritma *support vector machines* (SVMs) yang didesain untuk kasus klasifikasi data. SVC bekerja dengan melakukan konstruksi *hyperplane* untuk memisahkan data berdasarkan margin di dalam suatu ruang berdimensi N. Dari k *hyperplane* yang dibangun, akan dipilih suatu *hyperplane* yang mempunyai nilai margin maksimum yang disebut dengan *optimal hyperplane*.



$$g(x) = w^T x + b$$

Maximize k such that :

$$-w^T x + b \geq k \text{ for } d_i == 1$$

$$-w^T x + b \leq k \text{ for } d_i == -1$$

Value of $g(x)$ depends upon $\|w\|$:

1) Keep $\|w\| = 1$ and maximize $g(x)$ or,

2) $g(x) \geq 1$ and minimize $\|w\|$

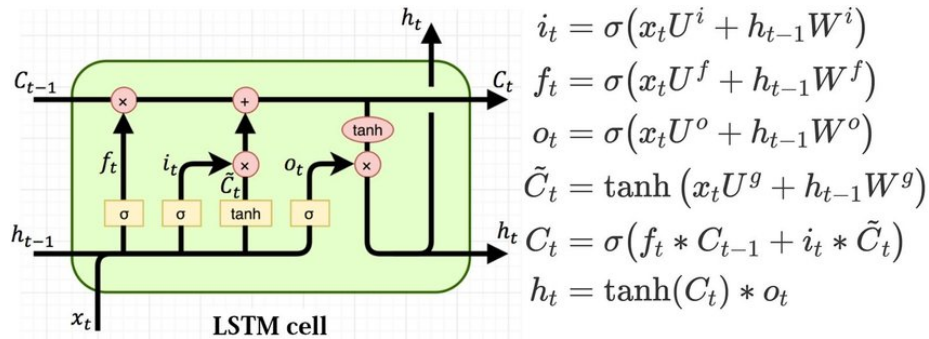
Untuk menemukan *optimal hyperplane* dibutuhkan parameter berupa regularisasi, *kernel*, γ dan margin. Parameter C (regularisasi) berfungsi untuk melakukan kontrol terhadap error yang dihasilkan saat klasifikasi. Parameter γ mempengaruhi besarnya kelengkungan dari suatu kurva yang berfungsi sebagai *decision boundary*. Parameter ini hanya berlaku jika

menggunakan *radial basis function (RBF) kernel*. Parameter *kernel* berperan dalam melakukan transformasi dari suatu masalah menggunakan beberapa konsep aljabar linear. Parameter margin mengukur seberapa baik suatu data dapat diklasifikasikan dengan benar.

Salah satu keunggulan algoritma ini terdapat di kemampuan belajarnya yang tidak bergantung terhadap dimensi ruang dari fitur tetapi menggunakan margin sebagai alat ukur dari kompleksitas hipotesis (Joachims 1998). Oleh karena itu, suatu data yang berdimensi tinggi dapat dilakukan proses generalisasi menggunakan *SVMs*.

Long Short Term Memory

Salah satu kelemahan utama dari *recurrent neural network (RNN)* adalah model RNN hanya bersifat *short term memory*. Penyebab utamanya terletak pada nilai gradien yang semakin mengecil atau membesar secara tidak terkendali ketika gradien dipropagasi kembali dalam model sehingga informasi yang tersimpan hilang atau tidak dapat diambil seluruhnya. Permasalahan ini dikenal dengan nama *vanishing gradient* dan dapat diatasi dengan suatu pengembangan dari model RNN yang disebut *long short term memory (LSTM)* (Sundermeyer et al 2012).



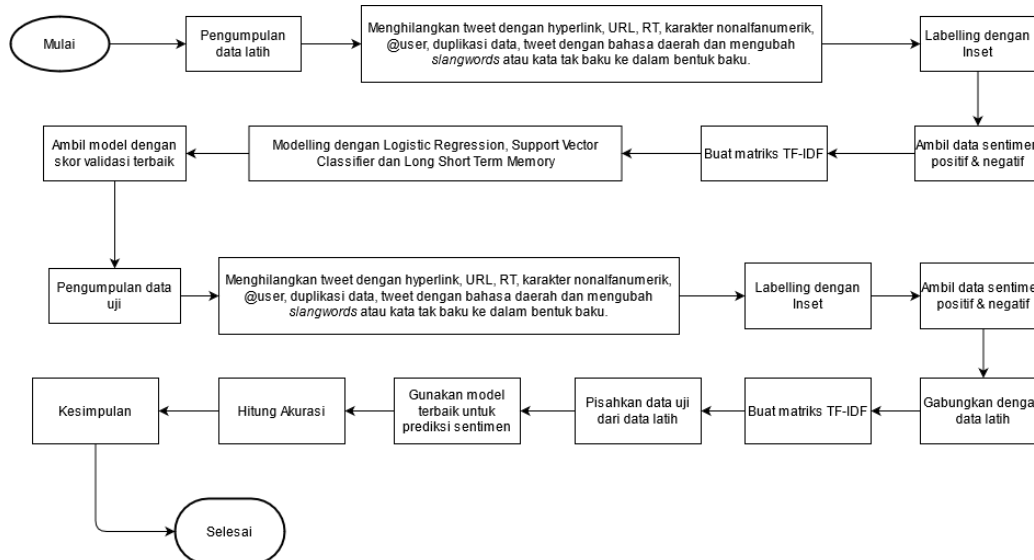
LSTM memiliki komponen utama yaitu *cell state*, *forget gate*, *input gate*, dan *output gate*. *Forget gate* berfungsi untuk menentukan informasi yang akan disimpan atau dibuang. Informasi yang diterima dari bagian sebelumnya dan informasi dari vektor input akan diproses dengan fungsi aktivasi sigmoid. Pada *input gate*, informasi dari vektor input dan *hidden state* sebelumnya akan dimasukkan ke fungsi aktivasi sigmoid dan tanh, lalu keduanya dikalikan. Hal ini bertujuan untuk mengontrol banyaknya informasi yang harus disimpan dalam sel. Hasil dari *input gate* dan *forget gate* akan menentukan nilai *cell state* yang baru. Nilai *cell state* yang baru akan dimasukkan ke fungsi aktivasi tanh dan hasilnya akan dikalikan dengan hasil masukan fungsi aktivasi sigmoid yang berisi informasi dari *hidden state* sebelumnya dan dari vektor input (Sundermeyer et al 2012).

Inset

InSet merupakan sebuah leksikon sentimen bahasa Indonesia yang memberikan bobot pada setiap kata positif dan negatif dalam rentang nilai -5 hingga 5. Bobot setiap kata akan dijumlahkan untuk menentukan sentimen dari sebuah kalimat. Apabila nilai keseluruhan bobot lebih besar dari 0, maka kalimat tersebut memiliki sentimen positif, begitu pula sebaliknya. Apabila total bobot adalah nol, maka kalimat tersebut adalah kalimat netral.

InSet dibuat berdasarkan algoritma AFINN dengan memilih 5000 kata yang paling sering muncul. Kata terpilih tersebut kemudian dilakukan pembobotan secara manual oleh dua orang Indonesia yang ditugaskan (Koto *et al* 2017). Kata yang dibobot tersebut kemudian dicari sinonimnya. Apabila sinonim kata tersebut memiliki perbedaan sentimen atau bobot yang signifikan, maka kata tersebut diberi bobot dan ditambahkan. Hasil pembobotan tersebut disimpan sebagai matriks dengan 2 kolom yaitu kata dan bobot.

METODE ANALISIS DATA



1. Pengumpulan Data

Data latih diambil dari platform twitter dengan berdasarkan hasil pencarian terkait topik peneliti, yaitu alih fungsi lahan pertanian. Peneliti menggunakan *API (Application Programming Interface)* twint dan tweepy untuk mendapatkan data dari Twitter. *Key* dan token API sudah disiapkan oleh twitter dan untuk menggunakannya peneliti dapat *login* ke akun Twitter. Data yang sudah didapatkan dari satu *keyword* akan digabungkan dengan data yang diambil dengan *keyword* yang lain. Jumlah data yang didapat adalah 2.843. Bentuk dataset terdiri atas ID pengguna, tanggal, dan *tweet*. Data yang diambil adalah data dalam periode Januari 2017 - Juni 2021. *Keyword* yang peneliti gunakan adalah alih fungsi lahan pertanian, kondisi lahan pertanian, pertanian digusur, lahan pertanian habis, konversi lahan sawah, sawah jadi jalan tol, sawah digusur, keadaan lahan pertanian, kebijakan ketersediaan lahan, penyusutan lahan pertanian, lahan gersang, ladang pertanian, lahan sawah berkurang, konversi lahan pertanian, dan alih fungsi lahan sawah.

2. Pra Proses Data

Setelah proses pengumpulan data dilakukan, data tersebut harus dilakukan pra proses data untuk menyiapkan data sebaik mungkin sebelum dilakukan *modelling*.

Tahapan awal dari pra proses data yaitu melakukan pembersihan *tweet*. Data tersebut dibersihkan dari link URL, RT, mentions, @user, hashtags, emoji, tag html (& dst), newline, angka dan *punctuations* (simbol). Setelah itu dilakukan *lowering text* dan tokenisasi untuk menghapus *whitespace* yang tersisa. Berikutnya kata-kata tidak formal (*slang words*) di dalam kolom *tweet* diubah dalam bentuk formalnya.

	tweet	clean_tweet
0	di gua udh ga ada sawah digusur sentraland sem...	di saya sudah tidak ada sawah digusur sentrala...
1	teriak² harga properti mahal akhirnya sawah di...	teriak harga properti mahal akhirnya sawah dig...
2	gunungnya udah dikeruk sawah udah jadi pabrik ...	gunungnya sudah dikeruk sawah sudah jadi pabri...
3	semua sawah akan digusur dinggo mbangun dalam ...	semua sawah akan digusur dinggo membangun dala...
4	terpikir jg klo fenomena skrg proyek bangunan ...	terpikir juga kalau fenomena sekarang proyek b...
...
2869	nggak punya lahan untuk tanaman makanya rumah ...	tidak punya lahan untuk tanaman makanya rumah ...
2870	memberi latihan dan memberi pengalaman secara ...	memberi latihan dan memberi pengalaman secara ...
2871	syoknya tggk pakaian cikguâ² ni hariâ² boleh pa...	syoknya tegak pakaian cikgu ini hari boleh pak...
2872	fungsi tanggul ini mampu melindungi kawasan la...	fungsi tanggul ini mampu melindungi kawasan la...
2873	tingkatkan proses dimana desa yg membutuhkan a...	tingkatkan proses dimana desa yang membutuhkan...

Tahap selanjutnya dilakukan proses *data labelling* yang bertujuan untuk mengklasifikasikan data berdasarkan sentimennya (positif, negatif, dan netral). Klasifikasi data dilakukan dengan menggunakan InSet. *Tweet* yang diklasifikasikan sebagai netral oleh InSet tidak akan digunakan dalam model.

3. Modelling

Tahapan selanjutnya yaitu menjadikan *tweet* yang sudah dibersihkan menjadi matriks *term frequency*. Proses pembuatan matriks *term frequency* menggunakan fungsi TF-IDF yang ada pada pustaka sklearn dan dapat diimplementasikan dengan Python. Hasilnya adalah sebuah matriks *sparse* dengan ukuran 2844×9571.

1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29
0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.200874
0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.000000
0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.000000
0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.000000
0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.000000

Matriks *sparse* yang terbentuk akan dimasukkan sebagai input ke dalam model LR, SVC dan LSTM. Untuk meningkatkan akurasi dari model, dilakukan *hyperparameter tuning* untuk menentukan parameter terbaik dari model. *Hyperparameter tuning* dilakukan dengan GridSearchCV dari pustaka sklearn dan pustaka Optuna. Parameter terbaik tiap model adalah sebagai berikut:

Model	Parameter
Logistic Regression (LR)	$C = 7.784550882354946$
Support Vector Classifier (SVC)	$C = 4.511114982618487$, kernel = linear
Long Short Term Memory (LSTM)	Optimizer = Adamax, Epoch = 20

Tabel 1. Parameter optimal dari masing-masing model

Parameter tersebut dimasukkan ke dalam model dan dilakukan *cross validation* sebanyak 10 *fold*. Untuk LSTM, dilakukan epoch sebanyak 20 kali dengan 1 *embedding layer*, 1 *layer* LSTM, dan 1 *output layer* dengan fungsi aktivasi softmax. Hasil rata rata akurasi skor validasi untuk LR, SVC, dan LSTM ditunjukkan pada tabel berikut.

Model	Skor Validasi
Logistic Regression (LR)	0.78 ± 0.04
Support Vector Classifier (SVC)	0.80 ± 0.04
Long Short Term Memory (LSTM)	0.76 ± 0.05

Tabel 2. Skor validasi dari masing-masing model

Dari skor validasi data, diketahui bahwa skor validasi tertinggi dihasilkan oleh model SVC. Model ini akan digunakan dalam tahap *model usage*.

4. Model Usage

Model SVC akan digunakan untuk memprediksi data baru, yaitu data *tweet* terkait alih fungsi lahan pertanian bulan Juli 2021 sebanyak 40 data. Data baru tersebut akan dilakukan pra proses terlebih dahulu, sebelum digabungkan dengan data latih untuk dibuat matriks *term frequency*. Data baru yang sudah berubah bentuk menjadi matriks *term frequency* tersebut kemudian dipisahkan kembali dan dilakukan *modelling* dengan SVC. Hasil prediksi model SVC dapat dilihat menggunakan *confusion matrix*. *Confusion matrix* yang dihasilkan adalah sebagai berikut

	Prediksi Positif	Prediksi Negatif
Sentimen Positif	20 (<i>True Positive</i>)	5 (<i>False Negative</i>)
Sentimen Negatif	1 (<i>False Positive</i>)	14 (<i>True Negative</i>)

Tabel 3. *Confusion matrix* data uji dengan model SVC

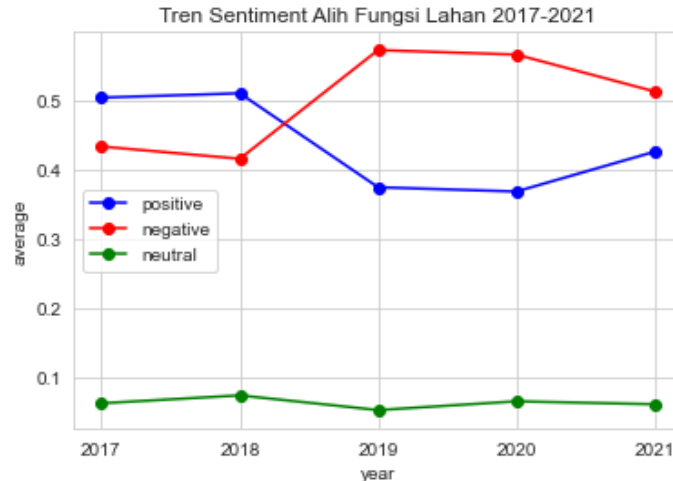
Hasil dari confusion matrix digunakan untuk menghitung akurasi.

$$\begin{aligned}
 Accuracy &= \frac{TP + TN}{TP + TN + FP + FN} \\
 &= \frac{20 + 14}{20 + 14 + 1 + 5} \\
 &= 0.85
 \end{aligned}$$

Akurasi yang didapat adalah 85%.

5. Analisis Hasil

Setelah dilakukan visualisasi tren sentimen alih fungsi lahan pertanian, diperoleh bahwa pada 2017-2018 sentimen positif cenderung lebih tinggi dibandingkan sentimen negatif. Akan tetapi pada periode 2019-2021, sentimen positif bergerak turun dan sentimen negatifnya lebih tinggi dibandingkan sentimen positif. Sentimen netral tiap tahunnya cenderung stagnan dan sangat kecil reratanya jika dibandingkan dengan sentimen positif dan negatif.



Berdasarkan perbandingan akurasi tiap model pada Tabel 2, SVC mempunyai akurasi model yang tertinggi sebesar 0.80 ± 0.04 jika dibandingkan dengan model LR (0.78 ± 0.04) dan LSTM (0.76 ± 0.05). Faktor utama yang mempengaruhi tingginya skor validasi model SVC adalah *kernel* yang digunakan yaitu *kernel* linear. *Kernel* linear cocok untuk data teks yang mempunyai banyak fitur sehingga pemetaan data ke ruang dimensi tinggi tidak dibutuhkan. Dengan kata lain, tidak terjadi peningkatan akurasi untuk pemetaan secara *nonlinear* (Chih-Wei *et al* 2016). Data teks memiliki sangat sedikit fitur yang tidak relevan karena data tersebut sudah dilakukan tahap pra proses. SVMs didukung dengan kemampuan untuk memasang sejumlah *kernel* sehingga fitur tidak relevan dapat dipisahkan dengan jarak yang maksimal. Data teks umumnya dapat dipisahkan secara linear, sehingga model SVMs cocok digunakan karena model tersebut selalu mencari pemisah linear dari data (Joachims 1998). Faktor lain yang mempengaruhi yaitu parameter C yang merupakan parameter regularisasi. Jika nilai parameter C

semakin besar, optimasi akan memilih *hyperplane* yang mempunyai nilai margin yang jauh lebih kecil dan begitu juga sebaliknya.

KESIMPULAN

Berdasarkan pengujian pada model LR, SVC dan LSTM, didapatkan bahwa model terbaik untuk memprediksi analisis sentimen alih fungsi lahan pertanian adalah model SVC dengan parameter C bernilai 4.511114982618487 dan *kernel* linear.

Hasil visualisasi menunjukkan bahwa pergerakan sentimen alih fungsi lahan pertanian di masyarakat cenderung mengarah ke sentimen negatif. Sentimen negatif mengalami peningkatan mengalahkan sentimen positif sejak pertengahan 2018.

DAFTAR PUSTAKA

- Chih-Wei H, Chih-Chung C, and Chih-Jen L. 2016. A Practical Guide to Support Vector Classification. Taipei (TW) : National Taiwan University
- Hosmer, David W. and Stanley Lemeshow. 2000. Applied Logistic Regression Second Edition. New York (US): John Wiley & Sons, Inc.
- Joachims T. 1998. Text Categorization with Support Vector Machines: Learning with Many Relevant Features. Dortmund (DE) : Technical University Dortmund.
- Nandini RA, Sari YA, Adikara PP. 2019. Analisis Sentimen Impor Beras 2018 Pada Twitter Menggunakan Metode *Support Vector Machine* dan Pembobotan Jumlah *Retweet*. Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer. 3(4): 3396 - 3406.
- Kompas. 2019. Mentan: Alih Fungsi dari Lahan Pertanian di 2019 Capai 150.000 Hektar. [internet].[diunduh 2021 Juni 30].
<https://money.kompas.com/read/2021/03/29/140755726/mentan-alih-fungsi-dari-lahan-pertanian-di-2019-capai-150000-hektar>
- Koto F, Rahmaningtyas G. 2017. InSet Lexicon: Evaluation of a Word List for Indonesian Sentiment Analysis in Microblogs. International Conference on Asian Language Processing (IALP). 10.1109/IALP.2017.8300625.
- Republik Indonesia. 2009. Undang Undang No.41 Tahun 2009, tentang Perlindungan Lahan Pertanian Pangan Berkelanjutan.
- Singh M, Goyal V, Raj S. 2021. Sentiment Analysis of Social Media Tweets on Farmer Bills 2020. Journal of Scientific Research. 65(3): 156 - 162.
- Sundermeyer M, Schlüter R, Ney H. 2012. LSTM Neural Network for Language Modelling. INTERSPEECH 2012. 194 - 197.
- Statista. 2021. Leading countries based on number of Twitter users as of April 2021 (in millions). [internet].[diunduh 2021 Juni 28].
<https://www.statista.com/statistics/242606/number-of-active-twitter-users-in-selected-countries/>