

Perbandingan SVM dan LSTM Untuk Memprediksi Gangguan Kecemasan (*Anxiety Disorder*) Berdasarkan Cuitan di Platform Aplikasi X (Twitter)

Nurochman¹, Luthfia Ashilah^{2*}

^{1,2}UIN Sunan Kalijaga Yogyakarta

¹nurochman@uin-suka.ac.id

^{2*}19106050021@student.uin-suka.ac.id

Diterima	Direvisi	Disetujui
05-08-2024	08-08-2024	14-08-2024.

Abstrak - *Anxiety Disorder* adalah gangguan kecemasan yang menyerang kesehatan mental seseorang sehingga bisa mengganggu aktivitas. Pengaruh media sosial menjadi salah satu adanya pemicu gangguan kecemasan, dimana media sosial dijadikan objek untuk meluapkan perasaan yang dialami pengguna, contohnya Twitter atau aplikasi X. Penelitian ini dilakukan untuk memprediksi adanya *anxiety disorder* pada pengguna Twitter dilihat berdasarkan cuitan yang ada dengan menggunakan dua model, yaitu Support Vector Machine (SVM) dan Long Short-Term Memory (LSTM) sehingga bisa dibandingkan mana yang lebih baik diantara dua metode tersebut. Data yang digunakan merupakan data hasil dari *crawling* yang kemudian dilakukan beberapa pemrosesan sehingga bisa diolah sesuai dengan medel yang ada. Hasil dari penelitian ini menunjukkan bahwa untuk memprediksi *anxiety disorder* menggunakan Long Short-Term Memory (LSTM) lebih unggul daripada Support Vector Machine (SVM) dari 3 matriks perhitungan yang ada, yaitu *precision*, *recall*, dan *f1-score* dengan nilai 75%. Sedangkan, untuk metode Support Vector Machine (SVM) hanya unggul dalam perhitungan nilai akurasi, yaitu 81%.

Kata Kunci: *Anxiety Disorder*, Support Vector Machine, Long Short-Term Memory

Abstract - *Anxiety disorder* is a disorder that attacks a person's mental health so that it can interfere with activities. The influence of social media is one of the triggers for anxiety disorders, where social media is used as an object to express the feelings experienced by users, for example Twitter or the X application. This research was conducted to predict the presence of anxiety disorder in Twitter users seen based on existing tweets using two models, namely Support Vector Machine (SVM) and Long Short-Term Memory (LSTM) so that it can be compared which is better between the two methods. The data used is the result of crawling data which is then subjected to some processing so that it can be processed in accordance with existing models. The results of this study show that to predict anxiety disorder using Long Short-Term Memory (LSTM) is superior to Support Vector Machine (SVM) from 3 existing calculation matrices, namely *precision*, *recall*, and *f1-score* with a value of 75%. Meanwhile, the Support Vector Machine (SVM) method is only superior in calculating the accuracy value, which is 81%.

Keywords: *Anxiety Disorder*, Support Vector Machine, Long Short-Term Memory

PENDAHULUAN

Kecemasan merupakan bagian dari kesehatan mental yang mempengaruhi perasaan seseorang sehingga membuat orang tersebut selalu merasa waswas dan khawatir seolah-olah akan terjadi sesuatu hal buruk dan merasa tidak nyaman seakan ada sebuah ancaman yang disertai dengan beberapa gejala fisik, seperti jantung berdebar-debar, mengeluarkan keringat dingin, dan tangan gemetar (Keliat dkk, 2011). Kaplan (2007) mengungkapkan bahwa kecemasan atau anxiety disorder merupakan kelompok gangguan psikiatri yang paling sering ditemukan. National study melaporkan bahwa satu di antara empat orang memenuhi kriteria untuk bisa dikatakan mengidap *anxiety disorder*.

Menurut hasil Riset Kesehatan Dasar (Riskesdas) pada tahun 2018, jumlah prevalensi terkait anxiety di Indonesia, yaitu sebesar 6 untuk usia 15 tahun keatas atau sekitar 14 juta penduduk Indonesia mengalami gangguan Kesehatan mental yang ditunjukkan dari gejala anxiety dan depresi. Selain itu, hasil penelitian yang dilakukan oleh Fakultas Kesehatan Universitas Indonesia pada tahun 2021 menemukan bahwa mayoritas remaja dan dewasa muda berusia 16-24 tahun memasuki periode kritis kesehatan mental, dengan persentase bahwa hampir 96% remaja dan dewasa muda 17 mengalami gejala kecemasan (*anxiety*) dan 88% diantaranya mengalami gejala depresi. Menurut Elavarasan et al., pada penelitiannya



menyebutkan bahwa kecanduan media sosial memiliki resiko tiga kali lebih tinggi untuk mengalami *anxiety* dibanding orang-orang yang tidak kecanduan sosial media.

Di era globalisasi saat ini teknologi berkembang dengan sangat pesat dalam segala bidang. Adanya internet membuat kemajuan teknologi lebih mendominasi lagi di dunia. Secara global ada sekitar 4,66 miliar atau sekitar 59,5% dari total seluruh populasi manusia di seluruh dunia yang telah menggunakan internet. Dan di Indonesia jumlah pengguna internet menurut Hootsuite (We Are Social) telah mencapai 202,6 juta jiwa atau sekitar 73,7% dari total populasi yang ada (We Are Social, & HootSuite. 2021). Media sosial merupakan saluran komunikasi digital yang bisa memberikan ruang bagi penggunanya untuk menciptakan tempat berbagi opini dan informasi dalam segala aspek pembahasan yang ada. Salah satu contoh media sosial yang bisa dikategorikan sebagai media komunikasi, yaitu aplikasi Twitter atau yang sekarang dikenal sebagai aplikasi X.

Twitter merupakan media sosial daring dan layanan jejaring sosial yang diciptakan oleh Jack Dorsey, Noah Glass, Biz Stone, dan Evan Williams pada bulan maret 2006 dan sekarang dioperasikan oleh X Corp yang merupakan sebuah perusahaan yang ada di Amerika Serikat. Menurut laporan We Are Social yang diunggah oleh databoks.katadata.co.id ada sekitar 666,2 juta pengguna Twitter di dunia dan ada sekitar 27,5 juta pengguna Twitter yang ada di Indonesia per bulan Oktober 2023 serta Indonesia menduduki peringkat keempat global untuk jumlah pengguna Twitter.

Besarnya pasar Twitter di Indonesia membuat banyak orang menjadikannya sebagai media untuk menyalurkan isi hati mereka bahkan mereka juga menuliskan cuitan mengenai gejala *anxiety disorder* yang dialaminya. Hal tersebut bisa dijadikan sebagai bahan penelitian dengan menggunakan beberapa teknik dari machine learning, yaitu Support Vector Machine (SVM) dan Long Short-Term Memory (LSTM). Oleh karena itu, hipotesis disini bertujuan untuk membandingkan kedua metode tersebut untuk memprediksi adanya gangguan kecemasan atau *anxiety disorder* para pengguna Twitter berdasarkan cuitan yang diunggah. Machine learning merupakan sebuah metode komputasi berdasarkan informasi yang telah tersedia sebelumnya untuk meningkatkan performa atau membuat prediksi yang akurat (Mohri et.al, 2012).

Pengembangan machine learning untuk melakukan prediksi terhadap kasus tentang kesehatan mental pernah dilakukan oleh Sofianita et al. (2021). Penelitian ini dilakukan untuk mendapatkan hasil akurasi terbaik dari beberapa algoritma yang ada pada machine learning yang digunakan dalam pendeteksian faktor-faktor yang mempengaruhi kesehatan mental pada suatu Lembaga Pendidikan di Malaysia yang setara dengan Sekolah Menengah Atas di Indonesia. Output yang dihasilkan, yaitu berupa algoritma yang

memiliki nilai akurasi prediksi yang terbaik untuk mendeteksi stress adalah Decision Tree. Selanjutnya, metode terbaik yang digunakan untuk mendeteksi depresi yaitu Support Vector Machine. Dan yang terakhir, yaitu metode terbaik yang digunakan untuk mendeteksi kecemasan menurut penelitian tersebut adalah algoritma Linear Regression dan Neural Network yang menghasilkan akurasi sekitar 68%-88%.

Dalam penelitian “Implementasi Algoritma Naïve Bayes Untuk Analisis Klasifikasi Survei Kesehatan Mental (Studi Kasus: Open Sourcing Mental Illness)” yang dilakukan oleh Reza et al. (2022) merupakan sebuah penelitian yang dilakukan untuk klasifikasi dan prediksi dari sebuah data survei mengenai pengaruh kesehatan mental dengan objek peneliti para pekerja yang ada di dalam perusahaan industri teknologi. Penelitian tersebut menggunakan algoritma Naïve Bayes karena tingkat akurasi yang cukup tinggi dan mudah diaplikasikan pada studi kasus yang akan diteliti. Data yang digunakan diambil dari sebuah website official Open Source Mental Illness yang berkaitan dengan topik yang akan digunakan. Output yang didapatkan dari penelitian tersebut adalah 72% pada ukuran data uji 30% dan data latih 70% yang hal tersebut merupakan hasil akurasi tertinggi yang bisa didapatkan oleh peneliti menggunakan algoritma Naïve Bayes.

Penelitian yang dilakukan oleh Satrio Yudho et al. (2019) merupakan penelitian yang dilakukan untuk memprediksi hasil akurasi dari sebuah data tentang sikap politik masyarakat terhadap partai politik yang ada di Indonesia. Data diambil dari animo demokrasi yang ada di aplikasi Twitter pada saat momen pemilu tahun 2019 dengan menggunakan algoritma Support Vector Machine. Data yang diambil oleh peneliti dipilah menjadi 2 kategori, yaitu positif dan negatif yang kemudian akan dilakukan preprocessing, pembobotan kata dengan TF-IDF, dan yang terakhir baru dilakukan pembuatan model klasifikasi dengan pendekatan Support Vector Machine dan K-Fold Cross Validation untuk menguji tingkat akurasi dari datanya. Output yang didapatkan dengan total data 900 yaitu diperoleh rata – rata akurasi sebesar 71%, error rate sebesar 29%, precision sebesar 73%, recall sebesar 70%, dan f-score sebesar 71%.

Penelitian selanjutnya, yaitu “Analisis dan Implementasi Long Short-Term Memory Neural Network untuk Prediksi Harga Bitcoin” yang dilakukan oleh Muhammad Wildan et al. (2018) untuk memprediksi pergerakan kurs nilai tukar pada Bitcoin sehingga bisa mengantisipasi resiko yang mungkin saja bisa terjadi saat berinvestasi menggunakan Bitcoin dengan bantuan algoritma Long Short-Term Memory Neural Network. Data yang diambil berdasarkan blockchain.info yang merupakan data jual beli indeks harga Bitcoin terhadap US Dollar. Output yang didapatkan adalah hasil prediksi harga Bitcoin dengan komposisi data latih 70% dan data uji 30%, parameter 1 pola time series, jumlah 25 neuron hidden, dan max

epoch adalah 100 dengan akurasi rata-rata pada data latih 95.36% dan data testing 93.5%.

Berdasarkan dari penelitian di atas, dapat disimpulkan bahwa banyak penelitian sebelumnya yang telah membahas mengenai perbandingan beberapa metode untuk melakukan sebuah prediksi terhadap suatu kasus. Maka, penelitian ini sangat memungkinkan untuk dilakukan oleh karena itu, dalam penelitian ini penulis memutuskan untuk menggunakan algoritma Support Vector Machine (SVM) dan Long Short-Term Memory (LSTM) untuk memprediksi adanya gangguan kecemasan pada pengguna Twitter berdasarkan cuitan yang diunggah sehingga bisa menghasilkan output yang nantinya sesuai dengan apa yang diinginkan penulis. Dengan dibuatnya penelitian ini, diharapkan nantinya penelitian ini bisa bermanfaat bagi penelitian-penelitian lain dan dapat dikembangkan menjadi sebuah sistem yang kompleks sehingga bisa dimanfaatkan lagi secara lebih luas.

METODE PENELITIAN

1. Pengambilan Data

Pengambilan data atau *crawling* data dilakukan dengan bantuan script Chromium Browser yang dieksekusi menggunakan Google Colab menggunakan bantuan dari auth token yang dimiliki di Twitter peneliti. Pengambilan data dilakukan berdasarkan rentang waktu dari tanggal 1 Januari 2023 sampai 13 Desember 2023. Data diambil berdasarkan kata kunci atau keyword yang sesuai dengan topik dari *anxiety disorder*. Keyword yang digunakan dalam penelitian ini adalah kata-kata yang mengarah pada adanya indikasi *anxiety disorder*, misalnya gelisah, takut, cemas, khawatir, panik, resah, stress, dan *anxiety*. Data yang diambil di dalam penelitian ini, yaitu 5.162 data yang disimpan dalam format .csv dan dapat dibuka secara manual menggunakan Microsoft Excel.

2. Pra-Pemrosesan Data

2.1 Cleansing

Hal pertama yang dilakukan di dalam tahap preprocessing data yaitu melakukan *cleansing* dimana kalimat yang ada di dalam data tersebut akan dihilangkan URL, hashtag, karakter HTML, tanda baca, dan emoticon sticker. Hasil dari proses ini bisa dilihat pada Tabel 1. Dan kemudian hasil pada proses ini akan dilanjutkan untuk proses preprocessing data selanjutnya.

Tabel 1. Hasil Proses *Cleansing*

No	Tweet	Hasil
1	Lagi berusaha mengurangi cemas, dan meredam gejala panik dengan melakukan apapun biar bisa berkurang.. Esok sore kalau dah	Lagi berusaha mengurangi cemas dan meredam gejala panik dengan melakukan apapun biar bisa berkurang Esok sore

	sampai jkt, reda nii ðŸ˜ˆ-ðŸ˜ˆ-	kalau dah sampai jkt reda nii
2	@imanlagi Bisa yuk dikit lagi, udah mulai emosi nih wkwk	Bisa yuk dikit lagi udah mulai emosi nih wkwk
3	Hari ini gw cemburu banget liat cewek gw Deket cowok wajarkan gw cemburu meskipun dia cuma sekedar ngobrol dan Krna cemburu itu gw ngalamin kejadian gelisah gemetar pikiran kosong rasa takut, sampai" GK ke kontrol alhasil gw susah nafas GK bisa nafas kayak mau mati jadinya"	Hari ini gw cemburu banget liat cewek gw Deket cowok wajarkan gw cemburu meskipun dia cuma sekedar ngobrol dan Krna cemburu itu gw ngalamin kejadian gelisah gemetar pikiran kosong rasa takut sampai GK ke kontrol alhasil gw susah nafas GK bisa nafas kayak mau mati jadinya
4	Social culture masyarakat yg bikin hidup jadi ga woless. Padahal ga ada kewajiban buat kita ngikutin standar hidup kebanyakan orang. Kita ngerasa gelisah krn tanpa sadar masih pay attention ke omongan org yg nyuruh kita begini, begitu. Pdhal ga didengerin jg gpp	Social culture masyarakat yg bikin hidup jadi ga woless. Padahal ga ada kewajiban buat kita ngikutin standar hidup kebanyakan orang Kita ngerasa gelisah krn tanpa sadar masih pay attention ke omongan org yg nyuruh kita begini begitu Pdhal ga didengerin jg gpp

Sumber: Cuitan aplikasi X (Twitter)

2.2 Case Folding

Setelah proses *cleansing* dilakukan, data akan lanjut pada proses *case folding*. Proses ini dilakukan dengan tujuan untuk mengubah seluruh huruf pada teks menjadi huruf kecil (*lower case*) sehingga hasil dari kalimat yang ada pada data sama rata dilihat dari komposisi hurufnya. Hasil dari proses ini bisa dilihat pada Tabel 2. Dan kemudian hasil pada proses ini akan dilanjutkan untuk proses *preprocessing* data selanjutnya.

Tabel 2. Hasil Proses *Case Folding*

No	Data Hasil <i>Cleansing</i>	Hasil
1	Lagi berusaha mengurangi cemas dan meredam gejala panik dengan melakukan apapun	lagi berusaha mengurangi cemas dan meredam gejala panik dengan melakukan apapun biar bisa berkurang esok sore

	biar bisa berkurang Esok sore kalau dah sampai jkt reda nii	kalau dah sampai jkt reda nii
2	Bisa yuk dikit lagi udah mulai emosi nih wkwk	bisa yuk dikit lagi udah mulai emosi nih wkwk
3	Hari ini gw cemburu banget liat cewek gw Deket cowok wajarkan gw cemburu meskipun dia cuma sekedar ngobrol dan Krna cemburu itu gw ngalamin kejadian gelisah gemetar pikiran kosong rasa takut sampai GK ke kontrol alhasil gw susah nafas GK bisa nafas kayak mau mati jadinya	hari ini gw cemburu bangett liat cewek gw deket cowok wajarkan gw cemburu meskipun dia cuma sekedar ngobrol dan krna cemburu itu gw ngalamin kejadian gelisah gemetar pikiran kosong rasa takut sampai gk ke kontrol alhasil gw susah nafas gk bisa nafas kayak mau mati jadinya
4	Social culture masyarakat yg bikin hidup jadi ga woless Padahal ga ada kewajiban buat kita ngikutin standar hidup kebanyakan orang Kita ngerasa gelisah krn tanpa sadar masih pay attention ke omongan org yg nyuruh kita begini begitu Pdhal ga didengerin jg gpp	social culture masyarakat yg bikin hidup jadi ga woless padahal ga ada kewajiban buat kita ngikutin standar hidup kebanyakan orang kita ngerasa gelisah krn tanpa sadar masih pay attention ke omongan org yg nyuruh kita begini begitu pdhal ga didengerin jg gpp

Sumber: Cuitan aplikasi X (Twitter)

2.3 Tokenization

Jika proses case folding telah dilakukan, maka data akan lanjut pada proses tokenization. Proses ini dilakukan dengan tujuan untuk memecah kalimat yang ada pada data menjadi bagian yang lebih kecil berupa kata-kata penyusunnya sehingga mempermudah pendeteksian kata yang nantinya akan dihilangkan pada proses filtering/stopword removal. Hasil dari proses ini bisa dilihat pada Tabel 3. Dan kemudian hasil pada proses ini akan dilanjutkan untuk proses preprocessing data selanjutnya.

Tabel 3. Hasil Proses *Tokenization*

No	Data Hasil <i>Case Folding</i>	Hasil
1	lagi berusaha mengurangi cemas dan meredam gejala panik dengan melakukan apapun biar bisa berkurang esok sore kalau dah sampai jkt reda nii	"['lagi', 'berusaha', 'mengurangi', 'cemas', 'dan', 'meredam', 'gejala', 'panik', 'dengan', 'melakukan', 'apapun', 'biar', 'bisa', 'berkurang', 'esok', 'sore', 'kalau', 'dah', 'sampai', 'jkt', 'reda', 'nii']"
2	bisa yuk dikit lagi udah mulai emosi nih wkwk	"['bisa', 'yuk', 'dikit', 'lagi', 'udah', 'mulai', 'emosi', 'nih', 'wkwk']"
3	hari ini gw cemburu banget liat cewek gw deket cowok wajarkan gw cemburu meskipun dia cuma sekedar ngobrol dan krna cemburu itu gw ngalamin kejadian gelisah gemetar pikiran kosong rasa takut sampai gk ke kontrol alhasil gw susah nafas gk bisa nafas kayak mau mati jadinya	"['hari', 'ini', 'gw', 'cemburu', 'bangett', 'liat', 'cewek', 'gw', 'deket', 'cowok', 'wajarkan', 'gw', 'cemburu', 'meskipun', 'dia', 'cuma', 'sekedar', 'ngobrol', 'dan', 'krna', 'cemburu', 'itu', 'gw', 'ngalamin', 'kejadian', 'gelisah', 'gemetar', 'pikiran', 'kosong', 'rasa', 'takut', 'sampai', 'gk', 'ke', 'kontrol', 'alhasil', 'gw', 'susah', 'nafas', 'gk', 'bisa', 'nafas', 'kayak', 'mau', 'mati', 'jadinya']"
4	social culture masyarakat yg bikin hidup jadi ga woless Padahal ga ada kewajiban buat kita ngikutin standar hidup kebanyakan orang Kita ngerasa gelisah krn tanpa sadar masih pay attention ke omongan org yg nyuruh kita begini begitu Pdhal ga didengerin jg gpp	"['social', 'culture', 'masyarakat', 'yg', 'bikin', 'hidup', 'jadi', 'ga', 'woless', 'padahal', 'ga', 'ada', 'kewajiban', 'buat', 'kita', 'ngikutin', 'standar', 'hidup', 'kebanyakan', 'orang', 'kita', 'ngerasa', 'gelisah', 'krn', 'tanpa', 'sadar', 'masih', 'pay', 'attention', 'ke', 'omongan', 'org', 'yg', 'nyuruh', 'kita', 'begini', 'begitu', 'pdhal', 'ga', 'didengerin', 'jg', 'gpp']"

Sumber: Cuitan aplikasi X (Twitter)

2.4 Filtering/Stopword Removal

Proses selanjutnya yang akan dilakukan yaitu proses filtering/stopword removal. Proses ini dilakukan dengan tujuan untuk menghapus kata-kata yang tidak memiliki makna dan tidak berpengaruh pada proses selanjutnya. Selain itu, pada proses ini digunakan juga untuk menghapus kata stopwords yang juga biasa disebut dengan kata penghubung dari suatu kalimat. Contoh kata penghubung yang biasa digunakan, yaitu yang, dan, tetapi, atau, dan sebagainya. Hasil dari proses ini bisa dilihat pada Tabel 4. Dan kemudian hasil pada proses ini akan dilanjutkan untuk proses preprocessing data selanjutnya.

Tabel 4. Hasil Proses *Filtering/Stopwords Removal*

No	Data Hasil <i>Tokenization</i>	Hasil
1	"['lagi', 'berusaha', 'mengurangi', 'cemas', 'dan', 'meredam', 'gejala', 'panik', 'dengan', 'melakukan', 'apapun', 'biar', 'bisa', 'berkurang', 'esok', 'sore', 'kalau', 'dah', 'sampai', 'jkt', 'reda', 'nii']"	"['berusaha', 'mengurangi', 'cemas', 'meredam', 'gejala', 'panik', 'apapun', 'biar', 'berkurang', 'esok', 'sore', 'dah', 'jkt', 'reda', 'nii']"
2	"['bisa', 'yuk', 'dikit', 'lagi', 'udah', 'mulai', 'emosi', 'nih', 'wkwk']"	"['yuk', 'dikit', 'udah', 'emosi', 'nih', 'wkwk']"
3	"['hari', 'ini', 'gw', 'cemburu', 'bangett', 'liat', 'cewek', 'gw', 'deket', 'cowok', 'wajarkan', 'gw', 'cemburu', 'sekedar', 'deket', 'cowok', 'gw', 'wajarkan', 'gw', 'cemburu', 'gelisah', 'gemetar', 'pikiran', 'meskipun', 'dia', 'cuma', 'sekedar', 'kontrol', 'alhasil', 'gw', 'ngobrol', 'dan', 'krna', 'cemburu', 'itu', 'gw', 'ngalamin', 'kejadian', 'gelisah', 'gemetar', 'pikiran', 'kosong', 'rasa', 'takut', 'sampai', 'gk', 'ke', 'kontrol', 'alhasil', 'gw', 'susah', 'nafas', 'gk', 'nafas', 'nafas', 'kayak', 'mati']"	"['gw', 'cemburu', 'bangett', 'liat', 'cewek', 'gw', 'deket', 'cowok', 'wajarkan', 'gw', 'cemburu', 'sekedar', 'deket', 'cowok', 'gw', 'wajarkan', 'gw', 'cemburu', 'gelisah', 'gemetar', 'pikiran', 'kosong', 'takut', 'gk', 'kontrol', 'alhasil', 'gw', 'susah', 'nafas', 'gk', 'nafas', 'kayak', 'mati']"

'gw', 'susah',
'nafas', 'gk', 'bisa',
'nafas', 'kayak',
'mau', 'mati',
'jadinya']"

4	"['social', 'culture', 'masyarakat', 'yg', 'bikin', 'hidup', 'jadi', 'ga', 'woles', 'padahal', 'ga', 'ada', 'kewajiban', 'buat', 'kita', 'ngikutin', 'standar', 'hidup', 'keanyakan', 'orang', 'ngerasa', 'gelisah', 'krn', 'ngikutin', 'sadar', 'pay', 'attention', 'standar', 'hidup', 'omongan', 'org', 'yg', 'keanyakan', 'nyuruh', 'pdhal', 'ga', 'orang', 'kita', 'didengerin', 'jg', 'gpp']"	"['social', 'culture', 'masyarakat', 'yg', 'bikin', 'hidup', 'ga', 'woles', 'ga', 'kewajiban', 'ngikutin', 'standar', 'hidup', 'keanyakan', 'orang', 'ngerasa', 'gelisah', 'krn', 'ngikutin', 'sadar', 'pay', 'attention', 'omongan', 'org', 'yg', 'nyuruh', 'pdhal', 'ga', 'didengerin', 'jg', 'gpp']"
---	---	---

Sumber: Cuitan aplikasi X (Twitter)

2.5 Stemming

Proses terakhir yang dilakukan pada tahap preprocessing data, yaitu proses stemming. Proses ini dilakukan dengan tujuan untuk menghilangkan imbuhan pada kata yang ada sehingga mengurangi ambiguitas yang bisa mengganggu proses identifikasi dan juga bisa membantu meningkatkan akurasi dalam proses penelitian ini karena stemming juga mengurangi keberagaman kata dalam suatu kalimat. Hasil dari proses ini bisa dilihat pada Tabel 5. Dan kemudian hasil pada proses ini akan dilanjutkan untuk tahap selanjutnya, yaitu tahap labeling data.

Tabel 5. Hasil Proses *Stemming*

No	Data Hasil <i>Filtering/Stopword Removal</i>	Hasil
1	"['berusaha', 'mengurangi', 'cemas', 'meredam', 'gejala', 'panik', 'apapun', 'biar', 'berkurang', 'esok', 'sore', 'dah', 'jkt', 'reda', 'nii']"	usaha kurang cemas redam gejala panik apa biar kurang esok sore dah jkt reda nii

2	"['yuk', 'dikit', 'udah', 'emosi', 'nih', 'wkwk']"	yuk dikit udah emosi nih wkwk
3	"['gw', 'cemburu', 'bangett', 'liat', 'cewek', 'gw', 'deket', 'cowok', 'wajarkan', 'gw', 'cemburu', 'sekedar', 'ngobrol', 'krna', 'cemburu', 'gw', 'ngalamin', 'kejadian', 'gelisah', 'gemetar', 'pikiran', 'kosong', 'takut', 'gk', 'kontrol', 'alhasil', 'gw', 'susah', 'nafas', 'gk', 'nafas', 'kayak', 'mati']"	gw cemburu bangett liat cewek gw deket cowok wajar gw cemburu dar ngobrol krna cemburu gw ngalamin jadi gelisah gemetar pikir kosong takut gk kontrol alhasil gw susah nafas gk nafas kayak mati
4	"['social', 'culture', 'masyarakat', 'yg', 'bikin', 'hidup', 'ga', 'woles', 'ga', 'kewajiban', 'ngikutin', 'standar', 'hidup', 'kebanyakan', 'orang', 'ngerasa', 'gelisah', 'krn', 'sadar', 'pay', 'attention', 'omongan', 'org', 'yg', 'nyuruh', 'pdhal', 'ga', 'didengerin', 'jg', 'gpp']"	social culture masyarakat yg bikin hidup ga woles ga wajib ngikutin standar hidup banyak orang ngerasa gelisah krn sadar pay attention omong org yg nyuruh pdhal ga didengerin jg gpp

Sumber: Cuitan aplikasi X (Twitter)

3 Labeling Data

Tahap selanjutnya setelah melakukan preprocessing data, yaitu tahap labeling data. Pada tahap ini akan dilakukan pelebelan terhadap data yang tersedia sehingga nantinya data bisa digunakan pada tahap selanjutnya. Pelebelan dari setiap data akan menggunakan algoritma Transformers dimana algoritma ini memiliki mekanisme otomatis dalam melakukan suatu pelabelan data dan juga merupakan salah satu model dari Natural Language Processing yang menyediakan arsitektur lengkap untuk dilakukan proses representasi data yang kemudian akan diolah lagi sehingga menghasilkan data yang telah mendapatkan label, baik itu positif atau negatif.

Pada tahap labeling data ini ada sekitar 5.033 data yang dilakukan pelabelan dan jumlah tersebut berkurang jika dilihat dari data hasil *crawling*.

Pengurangan data itu terjadi pada saat tahap preprocessing. Pada tahap itu data yang sama akan di take out dari dataset sehingga semua data yang ada di dalamnya merupakan data tunggal. Hasil dari labeling ini, yaitu ada sekitar 3886 data berlabel negatif dan 1147 data berlabel positif. Hasil dari tahap ini bisa dilihat pada Tabel 6.

Tabel 6. Hasil Labeling Data

No	Hasil <i>Preprocessing Data</i>	Label
1	usaha kurang cemas redam gejala panik apa biar kurang esok sore dah jkt reda nii	Negatif
2	yuk dikit udah emosi nih wkwk	Positif
3	gw cemburu bangett liat cewek gw deket cowok wajar gw cemburu dar ngobrol krna cemburu gw ngalamin jadi gelisah gemetar pikir kosong takut gk kontrol alhasil gw susah nafas gk nafas kayak mati	Negatif
4	social culture masyarakat yg bikin hidup ga woles ga wajib ngikutin standar hidup banyak orang ngerasa gelisah krn sadar pay attention omong org yg nyuruh pdhal ga didengerin jg gpp	Positive

Sumber: Cuitan aplikasi X (Twitter)

4 Modeling Support Vector Machine (SVM)

Pada tahap pemodelan topik untuk metode Support Vector Machine (SVM) dilakukan menggunakan bantuan *library* Scikit-Learn yang merupakan package yang dimiliki oleh bahasa pemrograman Python sehingga memudahkan peneliti saat mengimplementasikan kode. Pada metode ini data akan dibagi menjadi dua bagian, yaitu data training dan data testing dengan melakukan empat jenis percobaan yang berbeda dilihat berdasarkan perbedaan perbandingan jumlah data training dan data testing. Setelah itu, metode Support Vector Machine (SVM) akan mengimplementasi TF-IDF (Term Frequency-Inverse Document Frequency) untuk melengkapi kode yang akan digunakan sehingga saat dilakukan tahap uji coba mendapatkan hasil yang sesuai dengan napa yang diinginkan oleh penulis. Berikut merupakan parameter yang digunakan di dalam metode Support Vector Machine (SVM) yang bisa dilihat pada Gambar 1.

```
Parameter Model SVM:
Kernel: linear
C: 1.0
Degree: 3
Intercept: [-0.64405032]
Support Vectors: (0, 940) 0.3680539278326171
(0, 2697) 0.11120384708026344
(0, 3162) 0.2480982923495595
(0, 5032) 0.37735507963339865
(0, 7341) 0.29950317015574973
(0, 7993) 0.10356164040089362
(0, 8203) 0.3533777991658276
(0, 8483) 0.49300788165814274
(0, 8992) 0.4240997396762165
(1, 1202) 0.7669037960489403
(1, 4804) 0.6108185353136503
(1, 7993) 0.1968737781498908
(2, 623) 0.22596214246247762
(2, 672) 0.24318037005373225
(2, 712) 0.24932582148298216
(2, 1231) 0.24318037005373225
(2, 1501) 0.3957751092598808
(2, 2681) 0.09802886417953267
(2, 3832) 0.17923478442146853
(2, 3924) 0.21462078588453015
...
(2107, 5435) 0.4635061284934951
(2107, 6482) 0.5981956878430008
(2107, 7526) 0.3455148604542035
(2107, 8574) 0.33836430795447925
```

Gambar 1. Parameter SVM

5 Modeling Long Short-Term Memory (LSTM)

Pada tahap pemodelan topik untuk metode Long Short-Term Memory (LSTM) dilakukan menggunakan bantuan dari library TensorFlow yang merupakan package yang dimiliki oleh bahasa pemrograman Python sehingga memudahkan peneliti saat mengimplementasikan kode. Pada metode ini dibutuhkan tokenizer untuk memecah kalimat yang ada pada dataset menjadi bagian yang lebih kecil dimana setiap kata dari sebuah kalimat akan mendapatkan penomorannya masing-masing dengan ketentuan kata yang memiliki frekuensi paling banyak akan menjadi kata awal yang mendapatkan penomoran dan seterusnya sampai semua kata tersebut habis teridentifikasi sehingga memudahkan proses yang akan dilalui selanjutnya.

Setelah pengidentifikasian menggunakan tokenizer selesai, akan dilakukan proses sequences, dimana sequences tersebut akan diterapkan pada dataset untuk mengubah kalimat dari satu dataset menjadi bagian-bagian bernomor sesuai dengan nomor yang sudah dijabarkan oleh tokenizer sebelumnya. Selanjutnya akan dilakukan pelapisan kalimat supaya setiap dataset yang ada memiliki panjang kalimat yang sama, yaitu maksimal 250 kata yang telah diimplementasikan ke dalam setiap nomor dari setiap kata yang ada. Sebelum dilakukan penentuan pada parameter metode ini, dibuatlah terlebih dahulu variable yang akan merepresentasikan data training dan data testing yang nantinya akan diujikan. Dan proses terakhir sebelum melakukan pengujian data, yaitu melakukan klasifikasi terhadap Long Short-Term Memory (LSTM) itu sendiri dengan menentukan atribut parameter yang akan digunakan di dalam penelitian ini. Atribut yang digunakan dapat dilihat pada Gambar 2.

2.9 Pengujian Data

Pada tahap ini akan dilakukan pengujian data terhadap dataset yang sebelumnya telah dilakukan

Layer (type)	Output Shape	Param #
embedding (Embedding)	(None, 250, 100)	5000000
spatial_dropout2d (Spatial DropoutID)	(None, 250, 100)	0
lstm (LSTM)	(None, 100)	80400
dense (Dense)	(None, 2)	202
Total params: 5080602 (19.38 MB)		
Trainable params: 5080602 (19.38 MB)		
Non-trainable params: 0 (0.00 Byte)		
None		

Gambar 2. Parameter LSTM

preprocessing dan labeling data. Dimana di dalam masing-masing metode sudah diimplementasikan parameter-parameter yang dibutuhkan dan beberapa atribut lainnya sehingga tahap pengujian data bisa dilakukan. Untuk metode Support Vector Machine (SVM), data siap uji akan dilakukan penghitungan menggunakan confusion matrix dimana hasil yang didapatkan nanti bisa digunakan untuk melakukan evaluasi terhadap performa dari metode Support Vector Machine (SVM). Dan untuk metode Long Short-Term Memory (LSTM) akan memanfaatkan atribut parameter yang telah ditentukan sebelumnya untuk mengeksekusi dataset berdasarkan beberapa jumlah epoch. Masing-masing jumlah epoch yang dilakukan tahap pengujian data ini akan menghasilkan nilai accuracy dan nilai loss yang berbeda-beda sehingga bisa dilakukan pengamatan sehingga mendapatkan hasil yang digunakan untuk melakukan evaluasi terhadap performa dari metode Long Short-Term Memory (LSTM).

HASIL DAN PEMBAHASAN

1 Support Vector Machine (SVM)

Hasil pengujian data pada metode Support Vector Machine (SVM) untuk mendeteksi adanya gejala *anxiety disorder* berdasarkan kicauan di Twitter dilakukan menggunakan empat kali percobaan dilihat dari perbandingan rasio data training dan data testing. Empat percobaan tersebut dilakukan dengan perbandingan data training dan data testing, yaitu 50:50, 60:40, 70:30, dan 80:20. Pada Tabel 7 berikut merupakan hasil performa dari masing-masing percobaan dilihat berdasarkan confusion matrix yang telah diujikan terhadap dataset.

Tabel 7. Perbandingan Hasil *Confusion Matrix* Metode SVM

No	Data Latih	Data Uji	Accuracy	Pre cisi	Recall	f1-Score

1	50	50	0,80	0,71	0,56	0,63
2	60	40	0,79	0,70	0,56	0,62
3	70	30	0,80	0,74	0,58	0,65
4	80	20	0,81	0,74	0,57	0,65

Tabel 8.b Perbandingan Hasil Confusion Matrix Metode LSTM

Rasio Data Latih-Uji	Epoch	Val_Acc	Val_Prec	Val_Recall	Val_f1-Score
80-20	60	0,75	0,75	0,75	0,75
70-30	90	0,7252	0,7252	0,7252	0,7252
60-40	40	0,7144	0,7144	0,7144	0,7144
60-40	80	0,7343	0,7343	0,7343	0,7343
50-50	30	0,7477	0,7477	0,7477	0,7477
50-50	40	0,7457	0,7457	0,7457	0,7457

Sumber: Hasil eksperimentasi SVM

Berdasarkan Tabel 7 di atas yang memiliki hasil pengujian paling optimal untuk akurasi, precision, dan f1-score diantara empat kali percobaan, yaitu pada percobaan ke 4. Nilai accuracy yang dihasilkan terhadap percobaan tersebut, yaitu 81% dengan nilai precision 74%, nilai recall 57%, dan nilai dari F1-Score 65%. Dan untuk percobaan ke-3 memiliki hasil pengujian paling optimal untuk precision, recall, dan f1-score. Nilai accuracy yang dihasilkan terhadap percobaan tersebut, yaitu 80% dengan nilai precision 74%, nilai recall 58%, dan nilai dari F1-Score 65%.

2 Long Short-Term Memory (LSTM)

Hasil pengujian ini dilakukan untuk menentukan nilai akurasi terbaik dari Long Short-Term Memory (LSTM) untuk mendeteksi adanya gejala *anxiety disorder* berdasarkan kicauan di Twitter dengan cara menggunakan rasio dari data training dan data testing yang didapatkan dari dataset yang ada. Rasio data training dan data testing yang akan dilakukan empat kali percobaan pada sistem yang telah dibuat akan ditentukan berdasarkan dengan perbandingan 50:50, 60:40, 70:30 dan 80:20. Dilihat dari Tabel 4.8, dapat disimpulkan bahwa untuk pengujian data paling optimal pada metode Long Short-Term Memory (LSTM), yaitu terjadi pada perbandingan data 80:20 dengan epoch 60 yang menghasilkan validation accuracy, validation precision, validation recall, dan validation f1-score berjumlah sama, yaitu 0,75 atau 75%.

Tabel 8.a Perbandingan Hasil Confusion Matrix Metode LSTM

Rasio Data Latih-Uji	Epoch	Acc	Prec	Recall	f1-Score
80-20	60	1	1	1	1
70-30	90	1	1	1	1
60-40	40	1	1	1	1
60-40	80	0,9994	0,9994	0,9994	0,9994
50-50	30	0,998	0,998	0,998	0,998
50-50	40	1	1	1	1

Sumber: Hasil eksperimentasi LSTM

Sumber: Hasil eksperimentasi LSTM

KESIMPULAN

Berdasarkan penelitian yang telah dilakukan untuk menyelesaikan tugas akhir ini, dapat disimpulkan bahwa model Support Vector Machine (SVM) dan Long Short-Term Memory (LSTM) yang dibangun untuk melakukan pengujian terhadap data kicauan dari Twitter untuk memprediksi gangguan kecemasan (*anxiety disorder*) berhasil dilakukan. Hasil yang di dapat dari penelitian ini, yaitu bahwa untuk *anxiety disorder* penggunaan metode Long Short-Term Memory (LSTM) lebih unggul daripada Support Vector Machine (SVM) dari 3 matriks perhitungan yang ada, yaitu precision, recall, dan f1-score dengan nilai 75%. Sedangkan, untuk metode Support Vector Machine (SVM) hanya unggul dalam perhitungan nilai akurasi, yaitu 81%. Saran untuk penelitian selanjutnya, yaitu penelitian bisa dilakukan lebih mendetail berdasar aspek lain, penelitian selanjutnya juga bisa menggunakan dataset dengan jumlah yang lebih banyak lagi sehingga kualitas penelitian meningkat, dan parameter yang digunakan juga bisa lebih bervariasi sehingga mendapatkan hasil yang lebih baik.

REFERENSI

- Aldi, Muhammad Wildan Putra., Jondri., & Annisa Aditsania. "Analisis dan Implementasi Long Short Term Memory Neural Network untuk Prediksi Harga Bitcoin", e-Proceeding of Engineering, Volume 5, No.2, pp.3548, 2018.
- Annisa, Dona Fitri., & Ifdil. "Konsep Kecemasan (Anxiety) pada Lanjut Usia (Lansia)", Konselor, Volume 5, No.2, pp.93-99, 2016.
- Azid, Dian Faruqi., Budhi Irawan., & Casi Setianingsih. "Translation Russian Cyrillic To Latin Alphabet Using SVM (Support Vector Machine)", e-Proceeding of Engineering,

- Volume 4, No.3, pp.4007, 2017.
- Chaerudin, Reza Alfaresy., Ermatita., & Ruth Mariana Bunga Wadu. "Implementasi Algoritma Naïve Bayes Untuk Analisis Klasifikasi Survei Kesehatan Mental (Studi Kasus: Open Sourcing Mental Illness)", Seminar Nasional Mahasiswa Ilmu Komputer dan Aplikasinya (SENAMIKA), Volume 3, No.2, pp.623-632, 2022.
- Dinata, Rozzi Kesuma., & Novia Hasdyna. 2020. MACHINE LEARNING. Lhokseumawe: Unimal Press.
- Deng, Li., & Dong Yu "Deep Learning: Methods and Applications", Foundations and Trends in Signal Processing, Volume 7, No.3-4, pp.197-387, 2014.
- Fakhriyani, Diana Vidya. 2019. KESEHATAN MENTAL. Pamekasan: Duta Media Publishing.
- Hochreiter, S., & J. Schmidhuber. "Long Short-Term Memory", Neural Comput, Volume 9, No.8, pp.1735-1780, 1997.
- Heaton, Jeff. 2015. AIFH, Volume 3: Deep Learning and Neural Networks. Washington: Heaton Research, Inc.
- Li, Dan-Yang., & Jiang Qian. "Text sentiment analysis based on long short-term memory", First IEEE International Conference on Computer Communication and the Internet (ICCCI), pp.471-475, 2016
- Mutalib, Sofianita., Nor Safika Mohd Shafiee., & Shuzlina Abdul-Rahman. "Mental Health Prediction Models Using Machine Learning in Higher Education Institution", Turkish Journal of Computer and Mathematics Education, Volume 12, No.5, pp.1782-1792, 2021.
- Nugraini, Yunis Familia., Rd. Rohmat Saedudin., & Rachmadita Andreswari. "Implementasi Data Mining Dalam Kasus Mental Health Pada Sosial Media Twitter Menggunakan Metode Naive Bayes", e-Proceeding of Engineering, Volume 8, No.5, pp.9260, 2021.
- Nurrochman, Miftah Meinaldi., Anggunmeka Luhur Prasasti., & Ratna Astuti Nugrahaeni. "Implementasi Machine Learning Untuk Mendeteksi Unsur Depresi Pada Tweet Menggunakan Metode Naïve Bayes (Machine Learning Implementation For Depression Detection In Tweet Using Naïve Bayes Method)", e-Proceeding of Engineering, Volume 8, No.5, pp.6250, 2021.
- Octaviani, Pusphita Anna., Yuciana Wilandari., & Dwi Ispriyanti. "Penerapan Metode Klasifikasi Support Vector Machine (SVM) Pada Data Akreditasi Sekolah Dasar (SD) Di Kabupaten Magelang", Jurnal Gaussian, Volume 3, No.4, pp.811-820, 2014.
- O'Reilly, Tim., & Sarah Milstein. 2011. The Twitter Book, 2nd Edition. 1055 Gravenstein Highway North: O'Reilly Media.
- Pangestu, Satrio Yudho., Yuli Astuti., & Lilis Dwi Farida. "Algoritma Support Vector Machine Untuk Klasifikasi Sikap Politik Terhadap Partai Politik Indonesia", Jurnal Mantik Penusa, Volume 3, No.1, pp.236-241, 2019.
- Pragholapati, Andria., Rinanda Rizky Amalia Shaleha., Lina Anisa Nasution., & Irma Darmawati. "Reduksi Stigma terhadap Orang Dengan Gangguan Mental", Jurnal Paradigma, Volume 3, No.2, pp.1-6, 2021.
- Raup, Abdul., Wawan Ridwan., Yayah Khoeriyah., Supiana., & Qiqi Yuliati Zaqiah. "Deep Learning dan Penerapannya dalam Pembelajaran", JIIP (Jurnal Ilmiah Ilmu Pendidikan), Volume 5, No.9, pp.3258-3267, 2022.
- Rezeki, Syailendra Reza Irwansyah., Yuliana Restiviani., & Rita Zahara. "PENGUNAAN SOSIAL MEDIA TWITTER DALAM KOMUNIKASI ORGANISASI (Studi Kasus Pemerintah Provinsi DKI Jakarta Dalam Penanganan Covid-19)", JOURNAL OF ISLAMIC AND LAW STUDIES, Volume 4, No.2, pp.63-78, 2020.
- Romzi, Muhammad., & Budi Kurniawan. "PEMBELAJARAN PEMROGRAMAN PYTHON DENGAN PENDEKATAN LOGIKA ALGORITMA", Jurnal Teknik Informatika Mahakarya, Volume 3, No.2, pp.37-44, 2020.
- Setyani, Nomorvia Ika. "Pengguna Media Sosial Sebagai Sarana Komunikasi Bagi Komunitas", Jurnal Komunikasi Surakarta, h.8, 2013.
- Smagulova, Kamilya., & Alex Pappachen James. "A survey on LSTM memristive neural network architectures and applications", The European Physical Journal Special Topics, 2016.
- Tarkus, Exel Defrisco., Sherwin R.U.A. Sompie., & Agustinus Jacobus. "Implementasi Metode Recurrent Neural Network pada Pengklasifikasian Kualitas Telur Puyuh", Jurnal Teknik Informatika, Volume 15, No.2, pp.137-144, 2020.
- Tim Wikipedia.com, Twitter, dikutip dari <https://id.wikipedia.org/wiki/Twitter> diakses pada 16 Januari 2023 pukul 09.15 WIB.