

## DETEKSI EMOSI PADA TEKS BERBAHASA INDONESIA MENGGUNAKAN PENDEKATAN *ENSEMBLE*

Faisal Abdullah<sup>1</sup>, Syafrial Fachri Pane<sup>2</sup>, Roni Habibi<sup>3</sup>

<sup>1</sup>Teknik Informatika, Universitas Logistik dan Bisnis Internasional  
Email: <sup>1</sup>faisalabdullah168@gmail.com, <sup>2</sup>syafrial.fachri@ulbi.ac.id, <sup>3</sup>roni.habibi@ulbi.ac.id

### Abstrak

*Abstrak--*. Emosi dalam tulisan sering kali sulit dikenali karena tidak adanya petunjuk visual seperti ekspresi wajah atau intonasi suara yang biasanya membantu dalam memahami perasaan seseorang. Penelitian ini bertujuan untuk mengatasi tantangan tersebut dengan membangun model deteksi emosi pada teks berbahasa Indonesia. Pendekatan yang digunakan adalah *Ensemble Learning*, yang menggabungkan tiga model *Machine Learning*, yaitu SVM, KNN, dan XGBoost, untuk mengoptimalkan hasil deteksi emosi. Kontribusi utama penelitian ini adalah penerapan metode *Ensemble* dalam mendeteksi emosi pada teks berbahasa Indonesia dan evaluasi kinerja model dengan metrik akurasi, presisi, recall, F1 score, dan ROC AUC. Hasil evaluasi menunjukkan bahwa model *Ensemble* yang dibangun mampu mengungguli model sebelumnya, dengan akurasi, presisi, recall, dan F1 score masing-masing sebesar 87,14%, serta nilai ROC AUC mencapai 97,90%. Untuk meningkatkan performa, penelitian ini juga menggunakan *GridSearchCV* untuk *hyperparameter tuning* pada model SVM dan XGBoost, serta teknik *Automated Machine Learning* (AutoML) TPOT untuk menghasilkan model KNN.

**Kata Kunci:** Deteksi Emosi, Emosi pada teks, *Ensemble*, *Machine Learning*

### Abstract

*Abstract--*. Emotion in written text is often difficult to recognize due to the absence of visual cues such as facial expressions or vocal intonation, which typically aid in understanding a person's feelings. This research aims to address this challenge by developing an emotion detection model for Indonesian text. The approach used is *Ensemble Learning*, combining three *Machine Learning* models: SVM, KNN, and XGBoost, to optimize emotion detection results. The main contribution of this research is the implementation of the *Ensemble* method for detecting emotions in Indonesian text, with performance evaluated using metrics such as accuracy, precision, recall, F1 score, and ROC AUC. The evaluation results show that the *Ensemble* model outperforms previous models, achieving an accuracy, precision, recall, and F1 score of 87.14%, and a ROC AUC score of 97.90%. To further enhance performance, this study utilizes *GridSearchCV* for *hyperparameter tuning* of the SVM and XGBoost models and employs the *Automated Machine Learning* (AutoML) tool TPOT to generate the KNN model.

**Keywords:** Emotion Detection, Emotion on text, *Ensemble*, *Machine Learning*

## I. PENDAHULUAN

Menurut (Chowanda et al., 2021) emosi merupakan hal dasar dalam berkomunikasi yang mempengaruhi pandangan seseorang terhadap lawan bicaranya. Emosi juga adalah hasil tanggapan seseorang terhadap apa yang dialami dan bahkan diketahui. Terdapat beberapa jenis emosi yang umum seperti senang, sedih, marah, takut dan terkejut (Ranganathan & Tzacheva, 2019). Raut wajah, intonasi suara, gerakan tubuh, dan bahkan tulisan adalah cara umum untuk menyampaikan

emosi. Selama bertahun-tahun, orang telah menggunakan media tulis, seperti surat, puisi, lirik lagu, dan berbagai bentuk tulisan lainnya, sebagai sarana untuk mengungkapkan perasaan mereka.

Perkembangan teknologi telah mengdigitalisasi hampir semua aspek kehidupan, termasuk media tulis. Banyak orang beralih dari penggunaan media tulis tradisional ke media digital seperti Facebook, Twitter, Instagram, dan platform lainnya (Parvin & Hoque, 2021). Media sosial memungkinkan mereka untuk berbagi konten berupa audio, video, visual, dan teks kepada publik. Emosi yang terungkap

dalam audio, video, dan visual lebih mudah dikenali daripada tulisan, karena dalam tulisan, emosi disampaikan melalui gaya penulisan, pilihan kata-kata, dan struktur kalimat yang mencerminkan kesan emosional (Slavova & Andonov, 2022). Memahami hal ini sering memerlukan bantuan dari seorang pakar bahasa. Namun, jika teks yang harus dianalisis cukup besar, prosesnya akan memakan waktu yang cukup lama. Oleh karena itu, diperlukan pendekatan cerdas seperti NLP (Natural Language Processing) untuk mendeteksi emosi berdasarkan teks dengan lebih efisien.

NLP merupakan cabang kecerdasan buatan yang berfokus pada interaksi antara manusia dan bahasa alami (N. George et al., 2021). Di dalam NLP terdapat berbagai sub-area, seperti klasifikasi, ekstraksi informasi, pencocokan kemiripan dokumen, ringkasan dokumen, dan generasi bahasa alami (Zadgaonkar & Agrawal, 2021). Pada kasus ini, sub-area yang digunakan adalah klasifikasi. Klasifikasi pada teks terdiri dari analisis sentiment, deteksi berita palsu, deteksi emosi, deteksi *bot*, deteksi ujaran kebencian dan lain-lain (Sultana & Nishino, 2023). Pada penelitian ini, klasifikasi yang digunakan adalah deteksi emosi atau disebut dengan *Text-Based-Emotion Detection* (Kumar & Raman, 2022).

Terdapat banyak penelitian mengenai deteksi emosi pada teks, tetapi sebagian besar menggunakan data berbahasa Inggris yang telah tersedia di internet, seperti SemEval-2018 dan NLPCC 2018, yang telah dianotasi dengan label emosi untuk mengevaluasi model deteksi emosi. Dalam upaya ini, berbagai model deteksi telah digunakan, termasuk pendekatan *Machine Learning* dan *Deep Learning*, seperti ANN (*Artificial Neural Network*), SVM (*Support Vector Machine*), KNN (*K-Nearest Neighbor*), CNN (*Convolutional Neural Network*), RNN (*Recurrent Neural Network*), *Random Forest*, dan model klasifikasi lainnya (Ibrahim & Abdulazeez, 2021).

*Machine Learning* merupakan bagian dari *Artificial Intelligence* yang memungkinkan tugas-tugas tertentu dilakukan oleh algoritma berdasarkan pola data dan pengalaman (Le Glaz et al., 2021). Sedangkan *Deep Learning* adalah cabang dari *Machine Learning* yang berfokus pada penerapan algoritma yang terinspirasi oleh jaringan saraf otak manusia (Sorin et al., 2020).

Beberapa penelitian sebelumnya telah dilakukan mengenai deteksi emosi pada teks, salah satunya adalah penelitian (Plaza-del-Arco et al., 2020). Penelitian ini fokus pada deteksi emosi dalam teks media sosial di *Spanyol* menggunakan pendekatan *Machine Learning*. Hasil penelitian ini

menghasilkan model SVM dengan presisi sebesar 76%, *recall* sebesar 75%, dan nilai F1 sebesar 75%.

Kemudian penelitian (Ma et al., 2021) menggunakan pendekatan *Deep Learning* untuk mendeteksi emosi pada teks dialog (percakapan). Model yang digunakan pada penelitian ini adalah HAN-ReGRU (*Hierarchical Attention Network-Residual Gated Recurrent*). Dengan *pre-trained* BERT-Large untuk mendapatkan representasi kalimat, kemudian menggunakan HAN untuk menghasilkan representasi dokumen. Selanjutnya ReGRU digunakan untuk pemodelan urutan waktu dari dokumen. Penelitian ini menggunakan data berbahasa Inggris yang berasal dari Friends TV dan EmotionsPush dataset. Kinerja yang dihasilkan model tersebut adalah akurasi sebesar 69,9%.

Kemudian penelitian oleh (Chatterjee et al., 2019) menerapkan pendekatan *Deep Learning* untuk mengklasifikasi emosi pada data sosial media yaitu Twitter. Model yang digunakan pada penelitian tersebut adalah SS-BED (*Sentiment and Semantic based Emotion Detector*). Hasil dari penelitian tersebut menunjukkan bahwa model yang dikembangkan memiliki presisi sebesar 80,87%, *recall* sebesar 64,08%, dan nilai F1 sebesar 71,34%.

Selanjutnya adalah penelitian oleh (Nimmi et al., 2022) yang bertujuan untuk mendeteksi emosi ketika pandemi Covid-19 berdasarkan data berbahasa Inggris yang telah diberi label. Data yang digunakan merupakan gabungan antara Twitter dan ERSS. Model yang digunakan adalah model *Ensemble* yang terdiri dari beberapa model yang sudah dilatih sebelumnya yaitu DistilBERT, RoBERT dan BERT, model tersebut diberi nama AVELD. Hasil penelitian menunjukan akurasi 84,46%, presisi 85,49%, *recall* 85,07% dan nilai f1 85,20%.

Berdasarkan beberapa penelitian sebelumnya, dapat disimpulkan bahwa model *Ensemble* menunjukkan kinerja yang lebih baik dibandingkan dengan single model dalam mendeteksi emosi. Namun, perlu diperhatikan bahwa sebagian besar penelitian tersebut menggunakan dataset berbahasa Inggris. Oleh karena itu, penelitian ini bertujuan untuk mengusulkan sebuah model *Ensemble* yang difokuskan pada deteksi emosi dalam dataset berbahasa Indonesia. Evaluasi kinerja model akan dilakukan menggunakan metrik akurasi, presisi, *recall*, nilai F1, dan ROC AUC. Dengan demikian, diharapkan penelitian ini dapat memberikan kontribusi dalam pengembangan deteksi emosi untuk bahasa Indonesia.

Penelitian ini memberikan kontribusi pada bidang NLP dengan dua hal utama. Pertama, penelitian ini menerapkan pendekatan *Ensemble* untuk mendeteksi emosi pada teks berbahasa

Indonesia. Model *Ensemble* ini menggunakan *estimator* berupa SVM, KNN, dan XGBoost untuk mengoptimalkan hasil deteksi emosi. Pendekatan *Ensemble* ini bertujuan untuk meningkatkan kinerja dan akurasi deteksi emosi secara keseluruhan.

Kedua, model yang dibangun dalam penelitian ini dievaluasi menggunakan metrik yang relevan seperti akurasi, presisi, *recall*, nilai F1, dan ROC AUC. Dengan menghitung metrik-metrik ini, penelitian ini akan memberikan gambaran tentang seberapa baik model dalam mengklasifikasi emosi pada teks bahasa Indonesia.

Kemudian artikel ini dibagi menjadi beberapa bagian yaitu: Bagian II berisi penjelasan metodologi penelitian seperti algoritma atau teori-teori yang dipakai pada penelitian. Bagian III berisi hasil dan pembahasan mengenai proses implementasi dan pengujian penelitian. Bagian IV merupakan kesimpulan dari hasil penelitian dan saran. Bagian V merupakan daftar pustaka yang digunakan pada penelitian.

## II. METODE

Bagian ini menjelaskan setiap langkah yang dilakukan ketika penelitian dengan penjelasannya digambarkan pada Gambar 1.

Gambar 1 menjelaskan alur penelitian yang dilakukan, mulai dari mengumpulkan data, *pre-processing*, *feature extraction*, *modeling* dan *evaluation*. Berikut merupakan penjelasan langkah-langkah yang dilakukan dalam penelitian:

### A. Data *Pre-processing*

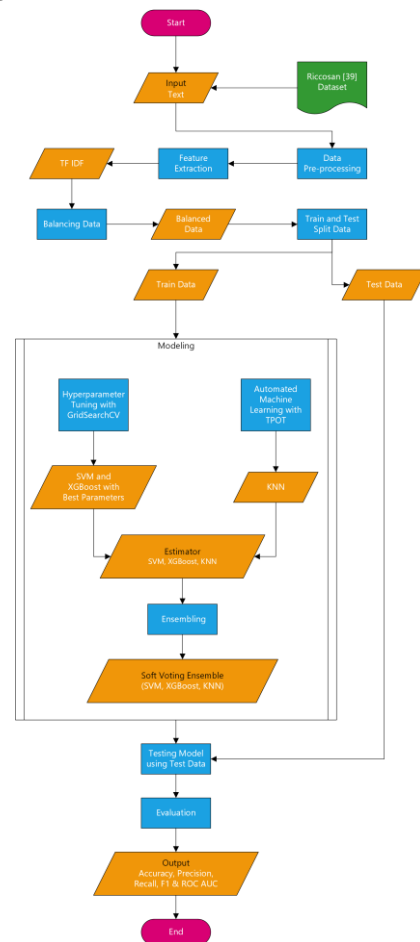
*Pre-processing* adalah tahapan penting dalam penelitian yang bertujuan untuk menyiapkan data dengan menghilangkan elemen-elemen yang tidak dibutuhkan (Pane et al., 2023). Tahapan ini meliputi membersihkan teks, tokenisasi, menghapus *stopwords*, *stemming*, memperbaiki *slang word* dan menghapus *stopword* lagi jika ada kata yang masih tidak dibutuhkan setelah proses perbaikan *slang word*.

Berikut merupakan penjelasan mengenai setiap langkah pada tahap *pre-processing* berdasarkan Gambar 2. Pertama adalah *cleaning text* atau membersihkan teks. Setiap teks dibersihkan dari *link*, tagar, username dan karakter yang tidak termasuk ke dalam huruf.

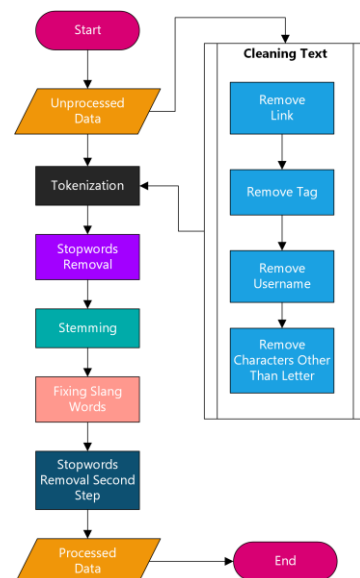
Langkah selanjutnya adalah tokenisasi, di mana hasil dari tahap sebelumnya akan dipisahkan ke dalam bentuk yang lebih sederhana, yang disebut dengan token (Mielke et al., 2021).

Setelah teks ditokenisasi, langkah selanjutnya adalah *Stopword Removal* atau menghapus *stopwords*. *Stopwords* adalah kata-kata yang sering

muncul namun tidak memberikan kontribusi penting (Cunningham-Nelson et al., 2019).



Gambar 1. Metodologi Penelitian



Gambar 2. *Pre-processing* Data

Langkah selanjutnya adalah proses *stemming*, dimana kata-kata akan diubah menjadi bentuk dasarnya (S et al., 2021). Pada tahapan ini, *library* Sastrawi digunakan. Sastrawi merupakan library NLP yang secara khusus digunakan untuk bahasa Indonesia.

Setelah *stemming* dilakukan, tahapan selanjutnya adalah *fixing slang words* atau disebut sebagai memperbaiki kata-kata gaul atau tidak baku. *Slang Word* merupakan kosa kata yang tidak sesuai dengan ketentuan atau tidak baku (Pei et al., 2019).

Dengan dilakukannya tahapan ini, kata-kata sebelumnya tidak terdeteksi ke dalam stopwords dapat dihilangkan dengan melakukan tahapan menghapus *stopword* kembali atau *stopwords removal second step*.

Setelah tahapan *pre-processing* dilakukan, tahap selanjutnya adalah tahapan *feature extraction* yaitu sebagai berikut.

**B. Feature Extraction**

Dalam mengenali dan memahami teks dalam bentuk numerik atau angka, diperlukan proses vektorisasi. Beberapa metode yang umum digunakan adalah *Bag of Words*, *Ngram*, TF IDF dan *Word Embedding* (Nandwani & Verma, 2021). Pada penelitian ini, yang digunakan untuk vektorisasi adalah TF IDF.

TF IDF merupakan singkatan dari *Term Frequency-Inverse Document Frequency*. TF IDF adalah sebuah metode untuk merepresentasikan teks ke dalam bentuk matriks (Artama et al., 2020). Cara kerja dari metode ini adalah dengan menentukan tingkat pentingnya suatu kata dalam dokumen (Lan, 2022). Berikut merupakan persamaan dari TF IDF.

$$TF = \frac{\text{Frekuensi kata "X" muncul dalam dokumen}}{\text{Jumlah kata yang ada dalam dokumen}}$$

$$IDF = \frac{\text{Jumlah dokumen yang ada dalam korpus}}{\text{Jumlah dokumen tempat kata "X" muncul}}$$

$$TF\ IDF = TF * IDF \tag{1}$$

Berdasarkan persamaan (1), untuk menentukan skor dari TF IDF adalah mengkali nilai dari TF dengan nilai IDF.

**C. Balancing Data**

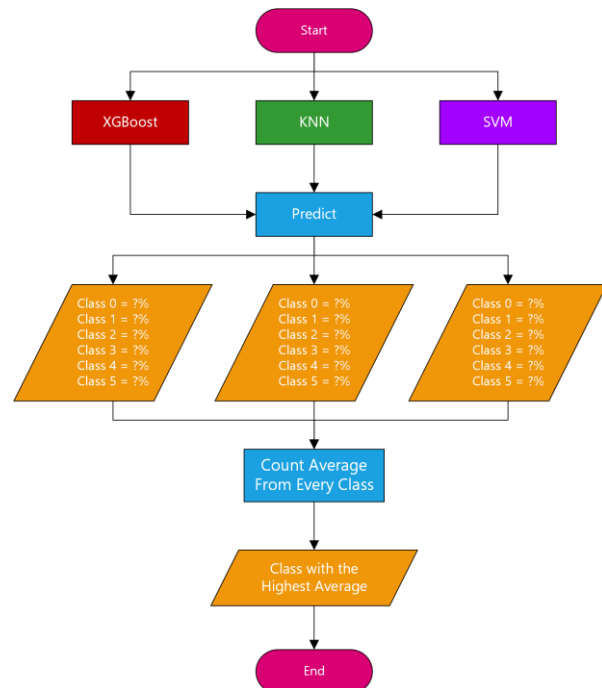
Setelah tahap *feature extraction*, tahapan berikutnya adalah tahapan *balancing* data. *Balancing* merupakan teknik untuk mengatasi data yang tidak seimbang (*imbalance*) antara kelas

menjadi seimbang (*balance*) (S. George & Srividhya, 2022). Terdapat beberapa cara *balancing* seperti *sampling* dan metode *ensemble*. Pada penelitian ini, peneliti menggunakan *sampling* untuk mengatasi *imbalance* data. Terdapat 2 (dua) teknik *sampling* yang digunakan oleh peneliti yaitu *Random Oversampling* dan *Random Undersampling*. Teknik *Random Oversampling* bekerja dengan cara meningkatkan jumlah sampel kelas yang sedikit (*minoritas*). Sedangkan *Random Undersampling* bekerja dengan cara mengurangi sampel kelas yang banyak (*mayoritas*).

**D. Modeling**

Dalam penelitian ini, model yang digunakan adalah model *ensemble*. *Ensemble* merupakan pendekatan pada *Machine Learning* dengan memanfaatkan beberapa model secara bersamaan untuk memprediksi hasil terbaik (Aslam et al., 2020). Terdapat beberapa macam model *Ensemble* yaitu *bagging*, *boosting*, *stacking*, *voting (hard voting, weighted voting, soft voting)* dan *cascade*. Pada penelitian ini, model *Ensemble* yang digunakan adalah *Soft Voting Ensemble*.

*Soft Voting* adalah *voting* dengan hasil prediksi berdasarkan probabilitas tertinggi dari beberapa model yang diterapkan (Xu et al., 2021). Berikut merupakan gambaran dari *soft voting*.



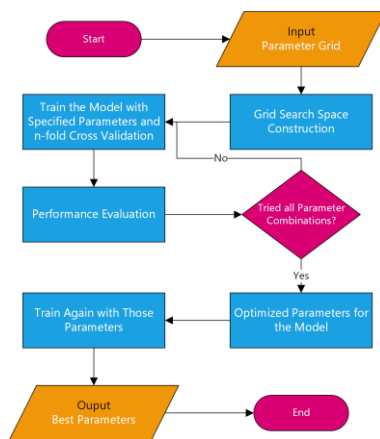
Gambar 3. *Soft Voting Ensemble*

Berdasarkan Gambar 3, cara kerja *soft voting ensemble* pada penelitian ini adalah melibatkan tiga model yang terdiri dari XGBoost, KNN, dan SVM

menghasilkan prediksi dengan persentase yang berbeda pada setiap kelas, kemudian persentase setiap kelas dihitung rata-ratanya, lalu kelas yang memiliki nilai rata-rata paling tinggi akan dipilih (*soft voting*) menjadi prediksi final. Model *Ensemble* ini terdiri dari 3 model yang digunakan sebagai estimator yaitu SVM, KNN dan XGBoost.

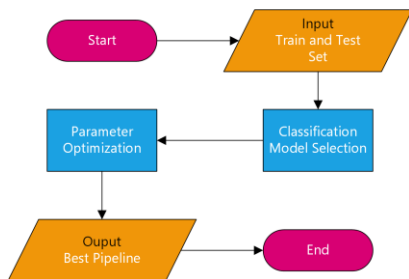
Pada penelitian ini, peneliti melakukan teknik *hyperparameter tuning* pada setiap *estimator* sehingga model memiliki kinerja yang baik.

*Hyperparameter tuning* merupakan suatu cara untuk menemukan kombinasi *hyperparameter* yang paling baik atau optimal pada model (Radzi et al., 2021). Pada penelitian ini, *hyperparameter tuning* dilakukan menggunakan *GridSearchCV*. Berikut merupakan *flow* dari *GridSearchCV*.



Gambar 4. GridSearchCV

Berdasarkan Gambar 4, dapat disimpulkan bahwa *GridSearchCV* bekerja dengan mencoba semua kombinasi parameter yang ditentukan pada model untuk menemukan parameter terbaik. Selanjutnya, peneliti menggunakan *tool automated machine leaning* yaitu TPOT.



Gambar 5. TPOT

Gambar 5 menunjukkan cara kerja dari TPOT (*Tree-based Pipeline Optimization Model*) yaitu mengeksplorasi berbagai kemungkinan, memilih *parameter*, dan algoritma *Machine Learning* untuk menentukan *pipeline* terbaik (Romano et al., 2021).

Selanjutnya adalah penjelasan singkat tentang setiap model yang digunakan sebagai *estimator*. SVM atau singkatan dari *Support Vector Machine* adalah *Machine Learning* yang dapat digunakan pada tugas klasifikasi maupun regresi. Cara kerja dari SVM adalah dengan mencari *hyperlane* optimal untuk membagi kelas. *Hyperlane* ini berfungsi untuk batas keputusan antar kelas (Rizwan et al., 2021).

Kemudian estimator selanjutnya adalah KNN. KNN merupakan singkatan dari *K-Nearest Neighbors* yang bekerja dengan menentukan prediksi ataupun klasifikasi berdasarkan objek terdekatnya (Yu et al., 2019).

Selanjutnya estimator terakhir yaitu XGBoost. XGBoost (*eXtrem Gradient Boosting*) merupakan model *boosting* yang bekerja dengan cara membangun banyak *decision tree* (pohon keputusan) untuk memperbaiki setiap kesalahan oleh *tree* sebelumnya (Kiangala & Wang, 2021). Setelah model dibangun, kemudian model akan dievaluasi menggunakan akurasi (*accuracy*), presisi (*precision*), *recall*, F1 dan ROC AUC. Berikut penjelasannya.

### E. Evaluation

Pada penelitian ini, model yang digunakan dievaluasi untuk mengukur kinerjanya dengan metrik akurasi, presisi, *recall*, dan F1. Akurasi merupakan tingkat ketepatan suatu model dalam memprediksi hasil. (Ahmad et al., 2021). Berikut persamaan akurasi:

$$Akurasi = \frac{BP + BN}{BP + SP + BN + SN} \tag{2}$$

Berdasarkan persamaan (2), BP adalah jumlah dari *True Positive*, BN adalah jumlah dari *True Negative*, SP adalah jumlah dari *False Positive* dan SN adalah jumlah *False Negative*.

Kemudian adalah presisi. Presisi merupakan persentase data yang benar dengan jumlah total data yang benar (Chen et al., 2020). Berikut persamaan presisi.

$$Presisi = \frac{BP}{BP + SP} \tag{3}$$

Berdasarkan (3), untuk menghitung presisi adalah membagi jumlah BP dengan BP ditambah SP.

Setelah itu adalah *recall*, *recall* merupakan persentase perbandingan antara nilai *True Positive* dengan nilai yang sebenarnya positif (Chen et al., 2020). Berikut persamaan *recall*.

$$Recall = \frac{BP}{BP + SN} \tag{4}$$

Berdasarkan persamaan (4), untuk mendapatkan nilai *recall* adalah membagi nilai BP dengan BP ditambah SN.

Kemudian ada nilai F1. F1 merupakan nilai yang memberikan bobot yang sama untuk presisi dan *recall* (Hasan et al., 2019). Berikut persamaannya.

$$F1 = \frac{2 * (P * R)}{P + R} \tag{5}$$

Berdasarkan persamaan (5), nilai P adalah presisi dan nilai R adalah *recall*.

Dan yang terakhir adalah ROC AUC (*Receiver Operating Characteristic - Area Under the Curve*). ROC AUC merupakan pengukuran kinerja untuk model klasifikasi, dimana semakin besar nilainya atau mendekati 1, maka semakin baik kinerja suatu model (Hu et al., 2019). ROC dihasilkan berdasarkan TPR terhadap FPR dimana TPR berada di sumbu y dan FPR berada di sumbu x. Berikut persamaan dari ROC AUC.

$$TPR = \frac{BP}{BP + SN} \tag{6}$$

$$FPR = \frac{SP}{SP + BN} \tag{7}$$

$$Specificity = \frac{BN}{BN + SN} \tag{8}$$

$$FPR = 1 - Specificity \tag{9}$$

Pada persamaan (6,7,8,9), TPR merupakan *True Positive Rate* dan FPR merupakan *False Positive Rate*. Setelah semua teori dijelaskan, bagian artikel selanjutnya adalah menampilkan hasil dari penelitian yaitu pada bagian hasil dan pembahasan.

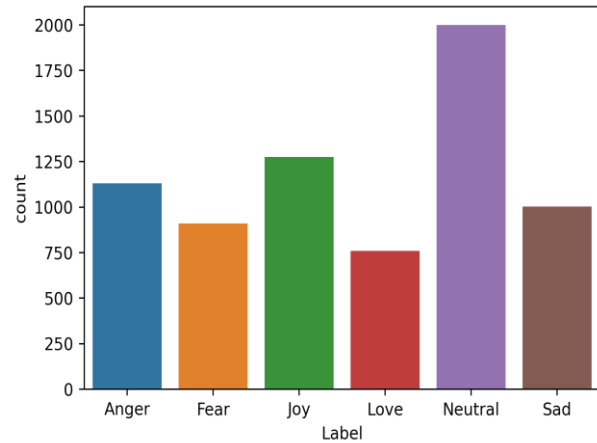
### III. HASIL DAN PEMBAHASAN

Pada bagian ini, eksperimen yang telah dilakukan kemudian dijelaskan dan dibahas hasilnya. Berikut merupakan hasil dan pembahasan mengenai eksperimen yang telah dilakukan.

#### A. Data

Data yang digunakan merupakan data dari hasil penelitian (Riccosan et al., 2022). Data tersebut diperoleh dari Twitter berbahasa Indonesia yang diambil menggunakan Twitter API. Data tersebut mengandung nilai-nilai emosi pada opini publik di berbagai Topik yang semua teks nya sudah huruf

kecil (*lowercase*). Data telah dianotasikan dengan enam label emosi yaitu marah (*anger*), takut (*fear*), senang (*joy*), cinta (*love*), sedih (*sad*) dan netral (*neutral*). Jumlah keseluruhan data adalah 7080 dengan jumlah data berlabel *anger* 1130, 911 pada label *fear*, 1275 data pada label *joy*, 760 data untuk label *love*, 1003 data pada label *sad* dan 2001 data pada label *neutral*.



Gambar 6. Sebaran Data

Gambar 6 menunjukkan perbandingan jumlah data pada masing-masing label dimana kelas netral memiliki jumlah terbanyak.

Tabel 1. Contoh Kalimat pada Setiap Emosi

Tweet	Label
bacoot kadrin emang bangsat	<i>anger</i>
plis nilai akm keluarnya lama banget, takut	<i>fear</i>
aw ikut senang semangat ya kuliahnya!	<i>joy</i>
ayang kenapa?? yuk semangat yuk	<i>love</i>
sarapan pake nasgor enak gk yh	<i>neutral</i>
gimana damai dengan rasa kecewa. gimana.	<i>sad</i>

Tabel 1 menunjukkan setiap contoh kalimat yang ada pada kelasnya masing-masing. Kemudian, langkah selanjutnya adalah tahapan *pre-processing*. Berikut merupakan hasil dari tahapan *pre-processing*.

#### B. Pre-processing Data

Setelah data didapatkan tahapan selanjutnya adalah melakukan *pre-processing* data. Berikut merupakan hasil dari *pre-processing* data:



Tabel 2. Hasil Tahapan *Pre-processing*

Sebelum	Sesudah
kok stabilitas negara, memange 10 thn negara tdk aman, bahkan sby menyuburkan ormas2 radikal, intoleran, teroris, yg berafiliasi ke partai tertentu..narasi klhntn intelektual tp bodoh..	stabilitas negara memange negara aman sby subur ormas radikal intoleran teroris afiliasi partai tertentunaras klhntn intelektual bodoh

Setelah membersihkan kata-kata, angka, dan karakter yang tidak diperlukan seperti yang ditunjukkan dalam Tabel 2, tahapan selanjutnya adalah tahapan *feature extraction* menggunakan TF IDF.

### C. Feature Extraction

Pada tahapan ini, data yang sudah bersih diubah kedalam bentuk vektor menggunakan metode TF IDF, dan berikut adalah hasilnya:

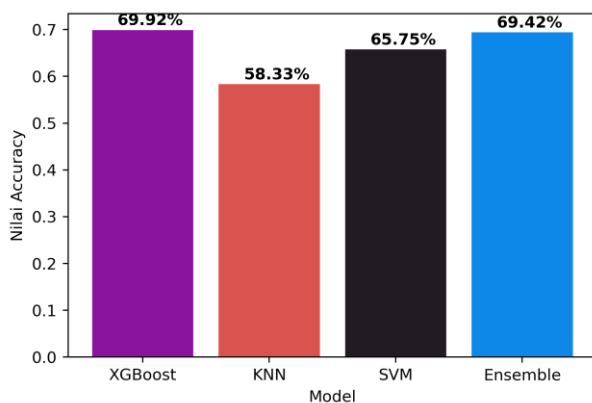
Tabel 3. Hasil TF IDF

Sebelum	Sesudah
['emosi', (0, 3338) 'pagi'] (0, 9343)	0.7498905915162152 0.6615618646471856

Tabel 3 menunjukan bahwa teknik TF IDF telah mengubah kata-kata kedalam bentuk vektornya masing-masing.

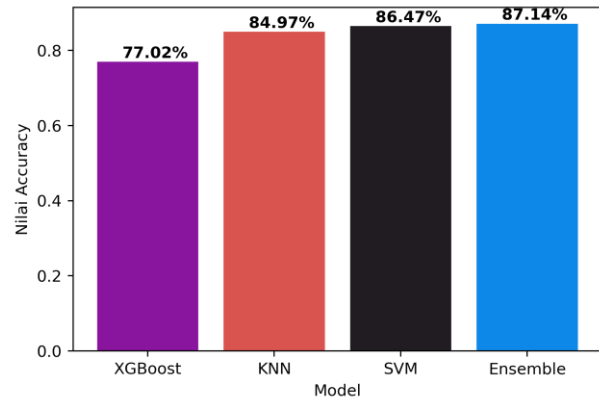
### D. Balancing Data

Pada tahapan ini, data yang sebelumnya tidak seimbang seperti yang ditunjukkan pada Gambar 6, telah diubah menjadi seimbang dengan jumlah yang sama pada setiap kelas. Peneliti menggunakan melakukan *balancing* data karena dapat meningkatkan performa dari model, berikut perbandingan akurasi antara menggunakan *balancing* data dengan yang tidak:



Gambar 7. Akurasi Model pada *Unbalance* Data

Pada Gambar 7 ditunjukkan bahwa model *ensemble* memiliki akurasi 69,42%, bahkan lebih kecil dari salah satu model *estimator*-nya yaitu XGBoost dengan akurasi model *Ensemble* hanya 69,92%.



Gambar 8. Akurasi Model pada *Balanced* Data

Dengan menggunakan teknik *balancing* data, akurasi model dapat meningkat seperti yang ditunjukkan pada Gambar 8.

### E. Train Test Split Data

Pada tahapan ini, peneliti memilih untuk membagi data menjadi *train* dan *test* dengan persentase 80:20. Keputusan ini diambil karena persentase tersebut memiliki akurasi yang terbaik dibandingkan dengan persentase lainnya.

Tabel 4. Hasil *Train Test Split*

Persentase	Akurasi
80:20	87,14%
70:30	85,34%
60:40	83,57%

Tabel 4 menunjukan bahwa persentase 80:20 untuk data *train* dan *test* memiliki akurasi terbaik yaitu 87,14%.

### F. Modeling

Peneliti pertama mendefinisikan probabilitas untuk setiap nilai *hyperparameter* dalam setiap model.

Tabel 5. Konfigurasi *GridSearchCV*

Model	Hyperparameter	Nilai
SVM	<i>C</i>	0.1, 1, 10, 100, 1000
	<i>kernel</i>	1, 0.1, 0.01, 0,001, 0,0001
	<i>gamma</i>	<i>rbf</i> , <i>linear</i> , <i>sigmoid</i> , <i>poly</i>
XGBoost	<i>max_depth</i>	3, 4, 5
	<i>learning_rate</i>	0.1, 0.01, 0.05
	<i>n_estimator</i>	50, 100, 150

Tabel 5 menunjukkan nilai-nilai probabilitas *hyperparameter* seperti pada model SVM yang memiliki *hyperparameter* C, kernel dan gamma. Kemudian hasil dari *GridSearchCV* yaitu:

Tabel 6. Hyperparameter tiap Estimator

Model	Hyperparameter Terbaik
SVM	C = 100 Kernel = rbf Gamma = 1
XGBoost	Probability = true max_depth = 5 n_estimator = 150 learning_rate = 0,1

Tabel 6 menunjukkan model SVM dan XGBoost beserta *hyperparameter* terbaik yang terpilih. Untuk hasil dari TPOT adalah:

Tabel 7. Hasil TPOT

Model	Hyperparameter
KNN	n_neighbors = 91 p = 2 weights = distance

Tabel 7 menunjukkan KNN merupakan model yang terpilih dengan *n\_neighbors* sebanyak 91, p (*power*) sama dengan 2 (*arbitrary p*) dan *weights* yang digunakan adalah tipe *distance*.

Tabel 8. Ensemble

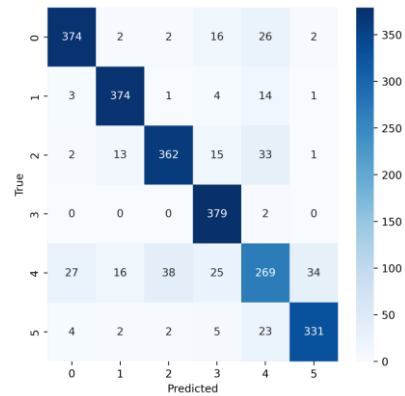
Parameter	Nilai
estimator	SVM, KNN, XGBoost
voting	soft
n_jobs	-1
flatten_transform	True
verbose	1

Setelah semua model siap, peneliti menggabungkannya ke dalam model *ensemble* menjadi *estimator*, seperti pada Tabel 8.

### G. Evaluation

Model kemudian dievaluasi kinerjanya dengan akurasi, presisi, *recall*, f1 dan ROC AUC. Berikut hasil *confusion matrix* dari model *Ensemble* yang dibangun.

Gambar 9 menunjukkan jumlah citra yang berhasil diklasifikasi secara tepat dimana kelas 3 (*love*) memiliki jumlah terbanyak yaitu 379. Kemudian hasil evaluasi dari model *ensemble* dan *estimator*-nya dapat dilihat pada Tabel 9.



Gambar 9. Confusion Matrix pada Ensemble

Tabel 9. Evaluasi Model

Model	Evaluasi			
	Akurasi	Presisi	Recall	F1
XGBoost	77,02	77,02	77,02	77,02
KNN	84,97	84,97	84,97	84,97
SVM	86,47	86,47	86,47	86,47
<b>Proposed Method (Ensemble)</b>	<b>87,14</b>	<b>87,14</b>	<b>87,14</b>	<b>87,14</b>

Tabel 9 menunjukkan model *Ensemble* yang dibangun memiliki kinerja lebih baik. Kemudian, langkah evaluasi model yang terakhir adalah dengan menghitung ROC AUC seperti yang ditunjukkan pada Tabel 10.

Tabel 10. Hasil ROC AUC

Model	ROC AUC (Multiclass)
XGBoost	94,10%
KNN	97,50%
SVM	97,70%
<b>Ensemble</b>	<b>97,90%</b>

Dapat diamati pada Tabel 10, bahwa nilai ROC AUC pada model *ensemble* secara signifikan lebih tinggi daripada model *estimator* individunya, mengindikasikan bahwa model *ensemble* menunjukkan kinerja yang lebih unggul. Selanjutnya, peneliti membandingkan model yang dibangun dengan model-model penelitian terkait.

Tabel 11. Author

No	Peneliti	Model
1	(Plaza-del-Arco et al., 2020)	SVM
2	(Ma et al., 2021)	Han- ReGRU
3	(Chatterjee et al., 2019)	SS-BED
4	(Nimmi et al., 2022)	Ensemble (AVEDL)
5	<b>PM (Proposed Method)</b>	<b>Soft Voting Ensemble</b>



Tabel 11, menunjukkan anotasi nomor untuk nama peneliti yang digunakan pada Tabel 13 dan Tabel 14.

Tabel 12. Perbandingan dengan Penelitian Terdahulu

Model	Acc	P	R	F1	ROC AUC
SVM (Ma et al., 2021)	-	76,0	75,0	75,0	-
Han-ReGRU (Ma et al., 2021)	69,9	-	-	-	-
SS-BED (Chatterjee et al., 2019)	-	80,87	64,08	71,34	-
Ensemble (AVELD) (Nimmi et al., 2022)	86,4 6	85,49	85,07	85,20	-
<b>PM (Soft Voting Ensemble)</b>	<b>87,1 4</b>	<b>87,14</b>	<b>87,14</b>	<b>87,14</b>	<b>97,90</b>

Tabel 12 menunjukkan bahwa model *Ensemble* yang dibangun dalam penelitian ini memiliki kinerja yang lebih baik dibandingkan dengan model-model pada penelitian terkait.

## IV. PENUTUP

### A. Kesimpulan

Hasil penelitian menunjukkan bahwa model *ensemble* yang dibangun untuk mendeteksi emosi berdasarkan teks berbahasa Indonesia memiliki kinerja yang baik dan dapat mengungguli model-model pada penelitian yang sudah ada dengan memiliki akurasi, presisi, *recall*, f1 sebesar 87,14% dan ROC AUC 97,90% seperti yang ditunjukkan pada Tabel 12. Dengan menggunakan persentase data *train* dan *test* yang tepat, model yang dibangun dapat mencapai kinerja terbaiknya. Ditunjukkan bahwa persentase 80:20 memiliki kinerja paling baik dengan akurasi sebesar 87,14%, kemudian diikuti oleh persentase 70:30 dengan akurasi 85,34% dan persentase 60:40 dengan akurasi 83,57%. Selain itu, teknik *balancing* data juga berdampak signifikan terhadap kinerja model. Hasil penelitian menunjukkan bahwa data yang seimbang (*balance*) memberikan kinerja yang jauh lebih baik dengan akurasi mencapai 87,14%, sementara data yang tidak seimbang (*unbalanced*) hanya mencapai akurasi 69,42%, menunjukkan perbedaan yang cukup besar dalam kualitas prediksi model. Kemudian, penelitian ini memiliki keunggulan dibandingkan dengan

penelitian-penelitian sebelumnya karena melibatkan tahapan yang lebih lengkap setelah proses pembersihan teks (*cleaning text*), sebagaimana terlihat pada Tabel 13.

Tabel 13. Pre-processing Proposed Method (PM)

NO	Preprocessing			
	Tokenisasi	Stopword	Stemming	Slang-word
1	✓	-	-	-
2	✓	-	-	✓
3	✓	-	-	✓
4	✓	-	-	-
<b>PM</b>	<b>✓</b>	<b>✓</b>	<b>✓</b>	<b>✓</b>

Selain itu, penelitian ini juga memiliki keunggulan lain dibandingkan dengan penelitian-penelitian sebelumnya, yaitu menyajikan tahapan-tahapan evaluasi yang lebih komprehensif, sebagaimana terlihat dalam Tabel 14.

Tabel 14. Evaluasi Proposed Method (PM)

NO	Evaluasi Model				
	Akurasi	Presisi	Recall	F1	ROC AUC
1	-	✓	✓	✓	-
2	✓	-	-	-	-
3	-	✓	✓	✓	-
4	✓	✓	✓	✓	-
<b>PM</b>	<b>✓</b>	<b>✓</b>	<b>✓</b>	<b>✓</b>	<b>✓</b>

Tabel 14 menunjukkan bahwa penelitian ini menyajikan semua tahapan evaluasi untuk model klasifikasi dari mulai akurasi, presisi, *recall*, nilai f1 dan ROC AUC.

### B. Saran

Walaupun model *Ensemble* ini memberikan kinerja yang baik, pertama peneliti menyarankan untuk mencoba berbagai cara untuk meningkatkan kinerja model seperti, vektorisasi menggunakan *word2vec* ataupun *doc2vec*. Kedua, penanganan *emoticon* ketika *pre-processing* data. Ketiga, menggunakan data yang berisi campuran antara bahasa Indonesia dan bahasa Inggris contohnya bahasa khas yang sering digunakan oleh masyarakat Jakarta Selatan sebagai tantangan untuk penelitian selanjutnya.

**V. DAFTAR PUSTAKA**

- Ahmad, A., Saraswat, D., Aggarwal, V., Etienne, A., & Hancock, B. (2021). Performance of deep learning models for classifying and detecting common weeds in corn and soybean production systems. *Computers and Electronics in Agriculture*, 184. <https://doi.org/10.1016/j.compag.2021.106081>
- Artama, M., Sukajaya, I. N., & Indrawan, G. (2020). Classification of official letters using TF-IDF method. *Journal of Physics: Conference Series*, 1516(1). <https://doi.org/10.1088/1742-6596/1516/1/012001>
- Aslam, Z., Javid, N., Ahmad, A., Ahmed, A., & Gulfam, S. M. (2020). A combined deep learning and ensemble learning methodology to avoid electricity theft in smart grids. *Energies*, 13(21). <https://doi.org/10.3390/en13215599>
- Chatterjee, A., Gupta, U., Chinnakotla, M. K., Srikanth, R., Galley, M., & Agrawal, P. (2019). Understanding Emotions in Text Using Deep Learning and Big Data. *Computers in Human Behavior*, 93, 309–317. <https://doi.org/10.1016/j.chb.2018.12.029>
- Chen, R. C., Dewi, C., Huang, S. W., & Caraka, R. E. (2020). Selecting critical features for data classification based on machine learning methods. *Journal of Big Data*, 7(1). <https://doi.org/10.1186/s40537-020-00327-4>
- Chowanda, A., Sutoyo, R., Meiliana, & Tanachutiwat, S. (2021). Exploring Text-based Emotions Recognition Machine Learning Techniques on Social Media Conversation. *Procedia Computer Science*, 179, 821–828. <https://doi.org/10.1016/j.procs.2021.01.099>
- Cunningham-Nelson, S., Baktashmotlagh, M., & Boles, W. (2019). Visualizing Student Opinion through Text Analysis. *IEEE Transactions on Education*, 62(4), 305–311. <https://doi.org/10.1109/TE.2019.2924385>
- George, N., Khan, M., Velu, A., & Whig, P. (2021). *Framework of Perceptive Artificial Intelligence using Natural Language Processing (P.A.I.N)*. [https://acors.org/ijacoi/VOL2\\_ISSUE2\\_3.pdf](https://acors.org/ijacoi/VOL2_ISSUE2_3.pdf)
- George, S., & Srividhya, V. (2022). Performance Evaluation of Sentiment Analysis on Balanced and Imbalanced Dataset Using Ensemble Approach. *Indian Journal of Science and Technology*, 15(17), 790–797. <https://doi.org/10.17485/IJST/v15i17.2339>
- Hasan, M., Milon Islam, M., Ishrak Islam Zarif, M., & Hashem, M. (2019). *Attack and anomaly detection in IoT sensors in IoT sites using machine learning approaches*. <https://doi.org/10.1016/j.iot.2019.10>
- Hu, J., Wang, J., Lin, J., Liu, T., Zhong, Y., Liu, J., Zheng, Y., Gao, Y., He, J., & Shang, X. (2019). MD-SVM: A novel SVM-based algorithm for the motif discovery of transcription factor binding sites. *BMC Bioinformatics*, 20. <https://doi.org/10.1186/s12859-019-2735-3>
- Ibrahim, I., & Abdulazeez, A. (2021). The Role of Machine Learning Algorithms for Diagnosing Diseases. *Journal of Applied Science and Technology Trends*, 2(01), 10–19. <https://doi.org/10.38094/jastt20179>
- Kiangala, S. K., & Wang, Z. (2021). An effective adaptive customization framework for small manufacturing plants using extreme gradient boosting-XGBoost and random forest ensemble learning algorithms in an Industry 4.0 environment. *Machine Learning with Applications*, 4, 100024. <https://doi.org/10.1016/j.mlwa.2021.100024>
- Kumar, P., & Raman, B. (2022). A BERT based dual-channel explainable text emotion recognition system. *Neural Networks*, 150, 392–407. <https://doi.org/10.1016/j.neunet.2022.03.017>
- Lan, F. (2022). Research on Text Similarity Measurement Hybrid Algorithm with Term Semantic Information and TF-IDF Method. *Advances in Multimedia*, 2022. <https://doi.org/10.1155/2022/7923262>
- Le Glaz, A., Haralambous, Y., Kim-Dufor, D. H., Lenca, P., Billot, R., Ryan, T. C., Marsh, J., DeVylder, J., Walter, M., Berrouiguet, S., & Lemey, C. (2021). Machine learning and natural language processing in mental health: Systematic review. In *Journal of Medical Internet Research* (Vol. 23, Issue 5). JMIR Publications Inc. <https://doi.org/10.2196/15708>
- Ma, H., Wang, J., Qian, L., & Lin, H. (2021). HAN-ReGRU: hierarchical attention network with residual gated recurrent unit for emotion recognition in conversation. *Neural Computing and Applications*, 33(7), 2685–2703. <https://doi.org/10.1007/s00521-020-05063-7>
- Mielke, S. J., Alyafeai, Z., Salesky, E., Raffel, C., Dey, M., Gallé, M., Raja, A., Si, C., Lee, W. Y., Sagot, B., & Tan, S. (2021). *Between words and characters: A Brief History of*

- Open-Vocabulary Modeling and Tokenization in NLP*. <http://arxiv.org/abs/2112.10508>
- Nandwani, P., & Verma, R. (2021). A review on sentiment analysis and emotion detection from text. In *Social Network Analysis and Mining* (Vol. 11, Issue 1). Springer. <https://doi.org/10.1007/s13278-021-00776-6>
- Nimmi, K., Janet, B., Selvan, A. K., & Sivakumaran, N. (2022). Pre-trained ensemble model for identification of emotion during COVID-19 based on emergency response support system dataset. *Applied Soft Computing*, 122. <https://doi.org/10.1016/j.asoc.2022.108842>
- Pane, S. F., Ramdan, J., Putrada, A. G., Fauzan, M. N., Awangga, R. M., & Alamsyah, N. (2023). A Hybrid CNN-LSTM Model With Word-Emoji Embedding For Improving The Twitter Sentiment Analysis on Indonesia's PPKM Policy. 51–56. <https://doi.org/10.1109/icitisee57756.2022.10057720>
- Parvin, T., & Hoque, M. M. (2021). An Ensemble Technique to Classify Multi-Class Textual Emotion. *Procedia Computer Science*, 193, 72–81. <https://doi.org/10.1016/j.procs.2021.10.008>
- Pei, Z., Sun, Z., & Xu, Y. (2019). *Slang detection and identification*.
- Plaza-del-Arco, F. M., Martín-Valdivia, M. T., Ureña-López, L. A., & Mitkov, R. (2020). Improved emotion recognition in Spanish social media through incorporation of lexical knowledge. *Future Generation Computer Systems*, 110, 1000–1008. <https://doi.org/10.1016/j.future.2019.09.034>
- Radzi, S. F. M., Karim, M. K. A., Saripan, M. I., Rahman, M. A. A., Isa, I. N. C., & Ibahim, M. J. (2021). Hyperparameter tuning and pipeline optimization via grid search method and tree-based autoML in breast cancer prediction. *Journal of Personalized Medicine*, 11(10). <https://doi.org/10.3390/jpm11100978>
- Ranganathan, J., & Tzacheva, A. (2019). Emotion mining in social media data. *Procedia Computer Science*, 159, 58–66. <https://doi.org/10.1016/j.procs.2019.09.160>
- Riccosan, Saputra, K. E., Pratama, G. D., & Chowanda, A. (2022). Emotion dataset from Indonesian public opinion. *Data in Brief*, 43. <https://doi.org/10.1016/j.dib.2022.108465>
- Rizwan, A., Iqbal, N., Ahmad, R., & Kim, D. H. (2021). Wr-svm model based on the margin radius approach for solving the minimum enclosing ball problem in support vector machine classification. *Applied Sciences (Switzerland)*, 11(10). <https://doi.org/10.3390/app11104657>
- Romano, J. D., Le, T. T., Fu, W., & Moore, J. H. (2021). TPOT-NN: augmenting tree-based automated machine learning with neural network estimators. *Genetic Programming and Evolvable Machines*, 22(2), 207–227. <https://doi.org/10.1007/s10710-021-09401-z>
- S, S. B., Khyani, D., M, N. N., & M, D. B. (2021). *An Interpretation of Lemmatization and Stemming in Natural Language Processing*. <https://www.researchgate.net/publication/348306833>
- Slavova, V., & Andonov, F. (2022). Bad news or good news when recognizing emotional valence using phonemic content. 2022 21st International Symposium INFOTEH-JAHORINA, INFOTEH 2022 - Proceedings. <https://doi.org/10.1109/INFOTEH53737.2022.9751339>
- Sorin, V., Barash, Y., Konen, E., & Klang, E. (2020). Deep Learning for Natural Language Processing in Radiology—Fundamentals and a Systematic Review. *Journal of the American College of Radiology*, 17(5), 639–648. <https://doi.org/10.1016/j.jacr.2019.12.026>
- Sultana, R., & Nishino, T. (2023). *EPiC Series in Computing Fake News Detection System: An implementation of BERT and Boosting Algorithm* (Vol. 91).
- Xu, T., Ma, Y., & Kim, K. (2021). Telecom churn prediction system based on ensemble learning using feature grouping. *Applied Sciences (Switzerland)*, 11(11). <https://doi.org/10.3390/app11114742>
- Yu, H., Ji, N., Ren, Y., & Yang, C. (2019). A special event-based K-nearest neighbor model for short-term traffic state prediction. *IEEE Access*, 7, 81717–81729. <https://doi.org/10.1109/ACCESS.2019.2923663>
- Zadgaonkar, A. V., & Agrawal, A. J. (2021). An overview of information extraction techniques for legal document analysis and processing. *International Journal of Electrical and Computer Engineering*, 11(6), 5450–5457. <https://doi.org/10.11591/ijece.v11i6.pp5450-5457>