

Journal of Creativity Student

<http://journal.unnes.ac.id/journals/jcs>

Deteksi Sarcasme Pada Dataset News Headlines Menggunakan Artificial Neural Network Berbasis TF-IDF

Andhi Prasetyo*, Budi Harjo

Universitas Dian Nuswantoro, Indonesia

*Corresponding Author: p31202502655@mhs.dinus.ac.id

Abstract

Sarcasm is a style of language frequently used in news headlines, especially in satirical media, thus posing a challenge for natural language processing systems in understanding the true meaning. The system's inability to recognize sarcasm can lead to misinterpretation in various tasks such as sentiment analysis and text classification. This study aims to detect sarcasm in news headlines using an Artificial Neural Network (ANN) with Term Frequency–Inverse Document Frequency (TF-IDF) feature representation. The dataset used is the News Headlines Dataset for Sarcasm Detection, which consists of approximately 28,000 news headlines from two sources, namely The Onion as sarcastic news and HuffPost as non-sarcastic news. The data is processed through a text pre-processing stage, then represented using TF-IDF with unigram and bigram schemes. The ANN model is trained using an 80:20 data split scheme and evaluated using accuracy, precision, recall, and F1-score metrics. Experimental results show that the TF-IDF-based ANN model achieved an accuracy of 85.02% and an F1-score of 84.44% on the test data. These results demonstrate that the ANN approach with TF-IDF representation remains effective and efficient as a baseline method for detecting sarcasm in short texts such as news headlines.

Keywords: sarcasm detection, news headlines, artificial neural network, TF-IDF, text classification

INTRODUCTION

Perkembangan platform berita digital dan media sosial membuat judul berita (headline) menjadi elemen kunci untuk menarik perhatian pembaca. Dalam praktiknya, *headline* tidak selalu disusun secara literal; banyak *headline* memanfaatkan gaya bahasa figuratif seperti ironi dan sarkasme untuk menghasilkan efek humor, kritik, atau sindiran. Keberadaan sarkasme ini penting bagi sistem pemrosesan bahasa alami (NLP), karena dapat menyebabkan interpretasi yang keliru pada tugas-tugas seperti analisis sentimen, opini publik, moderasi konten, dan peringkasan berita (Jurafsky & Martin, 2025; Zhang et al., 2025).

Deteksi sarkasme merupakan tugas yang menantang karena maknanya sering “berlawanan” dengan kata-kata yang digunakan, serta dapat bergantung pada konteks dan pengetahuan dunia. Tantangan semakin besar pada teks pendek seperti headline karena minimnya konteks, struktur kalimat yang padat, serta pemilihan diksi yang sengaja dibuat “menggiring” pembaca (Sharma et al., 2025). Oleh karena itu, penelitian deteksi sarkasme pada headline relevan baik secara ilmiah maupun aplikatif, khususnya untuk meningkatkan reliabilitas sistem klasifikasi teks di ranah berita.

Salah satu dataset yang banyak digunakan untuk penelitian ini adalah *News Headlines Dataset For Sarcasm Detection*, yang berisi sekitar 28 ribu headline dari dua sumber: *The Onion* (satir/sarkasme) dan *HuffPost* (berita non-sarkasme). Dataset ini dikurasi untuk mengatasi kelemahan dataset sarkasme berbasis Twitter yang cenderung noisy, banyak slang/salah eja, dan sering membutuhkan konteks percakapan (*reply*) (Misra, 2020a, 2020b). Dengan demikian, dataset headline *The Onion vs HuffPost* menawarkan data yang relatif lebih “rapi” dan label yang lebih kuat untuk studi klasifikasi sarkasme. Dataset ini juga telah dimanfaatkan dalam beberapa penelitian terdahulu sebagai dasar pengembangan model deteksi sarkasme pada judul berita, baik sebagai baseline maupun pembanding metode (Singh & Mehta, 2023).

Secara historis, pendekatan deteksi sarkasme banyak memanfaatkan fitur permukaan (*surface*

features) dan representasi berbasis frekuensi kata seperti *Bag-of-Words* atau TF-IDF, yang kemudian diklasifikasikan memakai model klasik (misalnya *Logistic Regression* atau SVM). Namun, beberapa studi melaporkan bahwa model *deep learning* seperti BiLSTM atau CNN dapat memberi peningkatan performa karena mampu menangkap pola sekuens dan interaksi kata yang lebih kompleks (Al-Ayyoub et al., 2023; Reddy & Kumar, 2023). Meski begitu, hasil penelitian tertentu juga menunjukkan bahwa pengaruh BoW vs TF-IDF bisa relatif kecil pada beberapa pengaturan dan model, sehingga evaluasi yang terstruktur tetap diperlukan (Verma & Mishra, 2024).

Dalam beberapa tahun terakhir, pendekatan Transformer (misalnya BERT/RoBERTa) menjadi sorotan karena kemampuan representasi kontekstual yang kuat dan sering mencapai hasil terbaik pada berbagai tugas NLP (Devlin et al., 2019; Liu et al., 2019; Vaswani et al., 2017). Selain model transformer, beberapa penelitian terbaru juga menerapkan pendekatan *deep learning* untuk mendeteksi sarkasme pada judul berita, termasuk pada konteks berita aktual atau breaking news (Gupta & Kumar, 2025). Pada domain sarcasm detection headline, studi-studi terbaru menunjukkan bahwa model transformer dan variasi hybrid (misalnya transformer + BiLSTM/attention) dapat meningkatkan performa, terutama ketika dirancang untuk menangkap nuansa bahasa dan ambiguitas sarkasme pada teks pendek (Patel & Shah, 2025; Sharma et al., 2025). Pendekatan *ensemble deep learning* juga telah diusulkan untuk meningkatkan kinerja deteksi sarkasme pada *News Headlines Dataset* melalui kombinasi beberapa model pembelajaran (Kaur & Gupta, 2023). Meski demikian, pendekatan transformer sering memerlukan sumber daya komputasi lebih besar dan proses tuning yang lebih kompleks, sehingga di banyak konteks riset terapan (kampus/industri kecil) masih dibutuhkan baseline yang efisien namun kompetitif.

Berdasarkan latar belakang tersebut, penelitian ini berfokus pada pembangunan model *Artificial Neural Network* (ANN) yang efisien untuk deteksi sarkasme headline dengan representasi TF-IDF. ANN berbasis TF-IDF dipilih karena: (a) mudah direplikasi, (b) biaya komputasi relatif rendah, (c) kinerja sering kuat sebagai baseline untuk klasifikasi teks, dan (d) memungkinkan analisis interpretabilitas berbasis fitur kata (misal bobot TF-IDF) sebagai bahan pembahasan. Penelitian ini diharapkan menghasilkan pipeline yang terstruktur, evaluasi yang ketat (*accuracy, precision, recall, F1*), serta analisis hasil yang dapat dibandingkan dengan kecenderungan penelitian terdahulu (termasuk arah SOTA transformer) sebagai konteks ilmiah (Jurafsky & Martin, 2025; Sharma et al., 2025; Sokolova & Lapalme, 2009; Zhang et al., 2025).

Landasan Teori

Sarkasme dalam Pemrosesan Bahasa Alami

Sarkasme adalah gaya bahasa yang menyampaikan makna implisit yang sering bertentangan dengan makna literalnya. Dalam NLP, sarkasme dianggap “pengganggu” (*confounder*) utama karena dapat membalik polaritas makna, misalnya kalimat bernada positif secara literal namun berniat negatif secara pragmatik (Zhang et al., 2025). Kesulitan utama deteksi sarkasme meliputi ambiguitas, kebutuhan konteks, serta adanya isyarat halus (misalnya ketidaksesuaian sentimen, hiperbola, atau framing satir) yang tidak selalu hadir secara eksplisit pada teks pendek (Sharma et al., 2025; Zhang et al., 2025).

News Headlines Dataset (The Onion vs HuffPost)

Dataset yang digunakan pada penelitian ini adalah *News Headlines Dataset For Sarcasm Detection* yang dikurasi dari dua situs: *The Onion* (satir) dan *HuffPost* (berita non-sarkasme). Kelebihan dataset ini dibanding dataset Twitter adalah kualitas bahasa yang lebih formal, minim salah eja/informal slang, serta label yang cenderung lebih stabil karena sumber satir dan non-satirnya jelas (Misra, 2020b; Sokolova & Lapalme, 2009). Selain itu, ukuran dataset sekitar 28K headline memberikan basis yang cukup untuk melatih model klasifikasi dan melakukan evaluasi yang reliabel (Misra, 2020a).

Pra-pemrosesan Teks (*Text Preprocessing*)

Pra-pemrosesan bertujuan mengurangi noise, menyamakan format teks, dan menyiapkan input yang konsisten bagi ekstraksi fitur. Pada teks *headline* berbahasa Inggris (umumnya pada dataset ini), langkah yang lazim meliputi: lowercasing, penghapusan tanda baca tertentu (opsional), normalisasi spasi, tokenisasi, serta penghapusan *stopwords* atau lemmatisasi/*stemming* (opsional tergantung eksperimen) (Jurafsky & Martin, 2025). Penting dicatat bahwa pada deteksi sarkasme, beberapa tanda

baca atau bentuk penekanan kata kadang mengandung sinyal pragmatik; karena itu, keputusan menghapus atau mempertahankan fitur seperti tanda seru, tanda tanya, atau huruf kapital sebaiknya konsisten dan dijelaskan sebagai bagian dari desain eksperimen (Sharma et al., 2025; Zhang et al., 2025).

Representasi Fitur TF-IDF

TF-IDF (*Term Frequency–Inverse Document Frequency*) adalah metode pembobotan kata untuk merepresentasikan dokumen sebagai vektor numerik. Intuisinya: kata yang sering muncul dalam satu dokumen namun jarang muncul di seluruh korpus dianggap lebih informatif. Konsep pembobotan term untuk temu-kembali informasi dan representasi berbasis *term-weighting* dibahas klasik oleh Salton & Buckley, dan TF-IDF tetap luas dipakai pada klasifikasi teks karena sederhana, kuat sebagai *baseline*, dan efisien (Salton & Buckley, 1988). Dalam praktik modern, TF-IDF sering dipadukan dengan n-gram (mis. unigram–bigram) agar model dapat menangkap frasa khas yang muncul pada *headline satir* (Jurafsky & Martin, 2025). Beberapa studi juga membahas modifikasi TF-IDF atau variasi input untuk meningkatkan performa model sekuens tertentu (Li et al., 2022).

Artificial Neural Network (ANN) untuk Klasifikasi Teks

ANN pada konteks penelitian ini merujuk pada *feedforward neural network / Multi-Layer Perceptron* (MLP) yang menerima input vektor TF-IDF dan memprediksi label kelas (sarkasme vs non-sarkasme). MLP umumnya terdiri dari: lapisan input (fitur TF-IDF), satu atau beberapa hidden layer (aktivasi ReLU/tanh), dan lapisan output (sigmoid untuk biner). Keunggulan MLP adalah kemampuannya memodelkan non-linearitas dan interaksi antar fitur, relatif mudah *ditraining*, dan cocok sebagai baseline neural untuk data tabular/vektor (Jurafsky & Martin, 2025). Sebagai konteks, model neural lain yang sering digunakan pada teks adalah CNN untuk klasifikasi kalimat (Kim, 2014) dan RNN/LSTM untuk memodelkan urutan kata (Hochreiter & Schmidhuber, 1997); bahkan transformer berbasis attention kini menjadi SOTA di banyak tugas NLP (Devlin et al., 2019; Vaswani et al., 2017). Namun, MLP+TF-IDF tetap relevan untuk penelitian yang menekankan efisiensi dan replikasi, atau sebagai pembanding terhadap model yang lebih kompleks (Reddy & Kumar, 2023; Verma & Mishra, 2024).

Evaluasi Kinerja Klasifikasi

Evaluasi model klasifikasi biner umumnya meliputi *accuracy, precision, recall, dan F1-score*. *Accuracy* mengukur proporsi prediksi benar, namun pada distribusi kelas yang tidak seimbang, *precision/recall/F1* lebih informatif untuk memahami kesalahan tipe tertentu. Kajian sistematis mengenai ukuran evaluasi pada tugas klasifikasi (termasuk interpretasi dan implikasi perubahan distribusi label) dibahas oleh Sokolova & Lapalme (Sokolova & Lapalme, 2009). Selain *metrik, confusion matrix* penting untuk melihat pola salah-klasifikasi (*false positive vs false negative*), terutama karena konsekuensi salah mengidentifikasi sarkasme dapat berbeda pada aplikasi tertentu (misalnya analisis opini vs moderasi konten) (Sokolova & Lapalme, 2009; Zhang et al., 2025).

METODE PENELITIAN

Desain Penelitian

Penelitian ini menggunakan desain eksperimen kuantitatif dengan pendekatan supervised learning untuk klasifikasi biner: *sarcastic* (1) dan *non-sarcastic* (0). Sistem dibangun melalui pipeline: pengumpulan data → pra-pemrosesan → ekstraksi fitur TF-IDF → pelatihan ANN → evaluasi kinerja → analisis hasil. Dataset dan tugas klasifikasi mengikuti definisi yang disediakan oleh penyusun dataset (*The Onion vs HuffPost*) (Misra, 2020a, 2020b).

Dataset dan Sumber Data

- a) Nama dataset: *News Headlines Dataset For Sarcasm Detection*
- b) Sumber: headline dari *The Onion* (satir) dan *HuffPost* (non-satir)
- c) Skala: sekitar 28K headline, dengan proporsi kelas yang dilaporkan sekitar 13K headline sarkastik (Misra, 2020a).
- d) Unit analisis: satu headline = satu sampel data (dokumen pendek).

Dataset dipilih karena kualitas bahasa relatif formal dan label sarkasme yang lebih “bersih” dibanding data media sosial berbasis hashtag (Misra, 2020a, 2020b).

Alur Kerja Penelitian

Alur kerja metodologi dirancang sebagai berikut:

- a) *Data acquisition & inspection*
- b) *Data cleaning & preprocessing*
- c) *Feature extraction (TF-IDF)*
- d) *Data splitting (train/validation/test) atau cross-validation*
- e) *Training ANN (MLP)*
- f) Hyperparameter tuning (terbatas dan terukur)
- g) *Evaluation & error analysis*
- h) Pelaporan hasil (tabel metrik, *confusion matrix*, analisis kata/fitur dominan)

Pra-pemrosesan Data

Langkah pra-pemrosesan dijalankan konsisten untuk seluruh data:

- a) *Lowercasing*: menyamakan bentuk huruf untuk mengurangi sparsity fitur (Jurafsky & Martin, 2025).
- b) *Cleaning karakter*: normalisasi spasi; penghapusan karakter non-alfanumerik tertentu dapat dilakukan secara hati-hati (opsional) karena tanda baca kadang menjadi sinyal sarkasme (Sharma et al., 2025; Zhang et al., 2025).
- c) Tokenisasi: memecah headline menjadi token kata/fragmen untuk keperluan vektorisasi (Jurafsky & Martin, 2025).
- d) *Stopword removal & lemmatization* (opsional): diuji terbatas; pada sebagian kasus, stopword dapat membantu konteks frasa, sehingga keputusan akhir didasarkan pada evaluasi validasi. Semua opsi yang dipilih harus dilaporkan agar replikasi terjaga (Jurafsky & Martin, 2025).

Output tahap ini adalah korpus headline yang siap diekstraksi fiturnya.

Ekstraksi Fitur dengan TF-IDF

Fitur dibentuk menggunakan TF-IDF *vectorizer*:

- a) Skema n-gram: minimal unigram; dapat diperluas ke bigram untuk menangkap frasa satir (mis. “area man”, “report says”, dll.) yang sering muncul sebagai pola gaya headline (Jurafsky & Martin, 2025).
- b) Parameter penting (ditetapkan dan dilaporkan):
 - `max_features` (mis. 5.000–30.000) untuk mengontrol dimensi,
 - `min_df` dan `max_df` untuk menyaring term terlalu langka/terlalu umum,
 - pilihan `sublinear_tf` (opsional) dan normalisasi vektor.

Secara konseptual, TF-IDF menguatkan term yang diskriminatif terhadap kelas, dan telah lama menjadi pendekatan representasi standar untuk *retrieval* dan klasifikasi teks (Salton & Buckley, 1988). Variasi/peningkatan TF-IDF juga pernah diusulkan untuk meningkatkan performa pada skenario tertentu (Li et al., 2022).

Perancangan Model ANN (MLP)

Model utama adalah MLP untuk klasifikasi biner:

- a) *Input layer*: vektor TF-IDF berdimensi d (sesuai `max_features` dan n-gram).
- b) *Hidden layer*: 1–3 lapisan dense (mis. 128–512 unit) dengan aktivasi ReLU.
- c) Regularisasi: dropout dan/atau L2 weight decay untuk mengurangi *overfitting* pada fitur berdimensi tinggi.
- d) *Output layer*: 1 neuron dengan aktivasi sigmoid (probabilitas sarkasme).
- e) *Loss function*: *binary cross-entropy*.
- f) *Optimizer*: Adam (umum untuk stabilitas *training*).

Desain ini dipilih karena ANN *feedforward* efektif untuk input berbentuk vektor (TF-IDF) dan memodelkan hubungan non-linear antar fitur (Jurafsky & Martin, 2025). Sebagai pembanding konseptual, studi lain juga mengevaluasi arsitektur yang lebih kompleks seperti CNN/LSTM atau transformer pada task serupa (Al-Ayyoub et al., 2023; Hochreiter & Schmidhuber, 1997; Patel & Shah, 2025; Sharma et al., 2025).

Strategi Pelatihan dan Validasi

Agar evaluasi kuat, gunakan salah satu (atau gabungan) strategi berikut:

- a) *Hold-out split*: misal 80:20 atau 70:30 dengan stratifikasi label agar distribusi kelas seimbang di train/test.
- b) *K-Fold Cross-Validation* (opsional tapi direkomendasikan): mis. 5-fold, untuk memperoleh estimasi performa yang lebih stabil pada dataset ukuran menengah (Sokolova & Lapalme, 2009).

Selama training:

- a) Terapkan early stopping berdasarkan performa validasi untuk mencegah *overfitting*.
- b) Catat konfigurasi training (epoch maksimum, *batch size*, *learning rate*) agar eksperimen dapat direplikasi.

Metode Evaluasi dan Analisis Hasil

Evaluasi dilakukan pada test set (atau rata-rata CV) menggunakan:

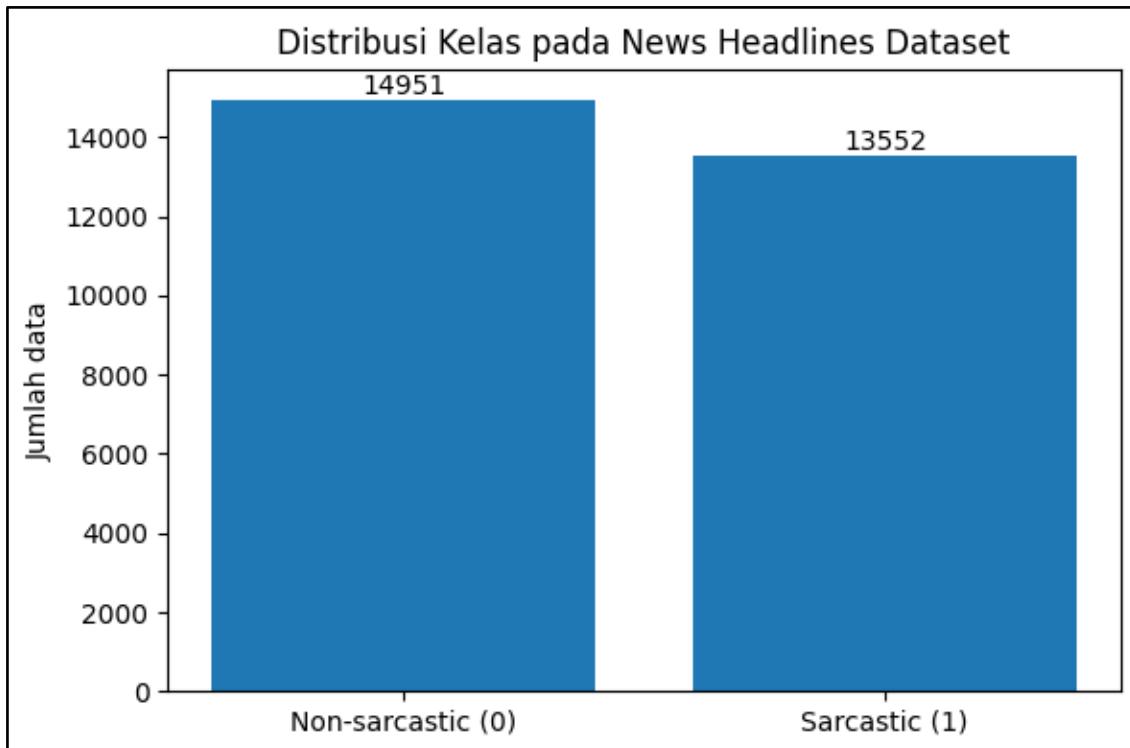
- a) *Accuracy*, *Precision*, *Recall*, *F1-score* (Sokolova & Lapalme, 2009).
- b) *Confusion matrix* untuk melihat pola error (FP/FN).
- c) Analisis error (kualitatif): ambil contoh headline yang sering salah diprediksi untuk mengidentifikasi pola seperti permainan kata, referensi budaya, atau ambiguitas yang sulit ditangkap TF-IDF (Sharma et al., 2025; Zhang et al., 2025).
- d) Analisis fitur (opsional): inspeksi term/bi-gram dengan bobot TF-IDF tinggi atau kontribusi besar terhadap prediksi (mis. melalui pendekatan interpretasi model linear sebagai pembanding) untuk memperkaya pembahasan.

HASIL DAN PEMBAHASAN

Deskripsi Dataset dan Skema Eksperimen

Eksperimen pada penelitian ini menggunakan *News Headlines Dataset for Sarcasm Detection*, yang terdiri dari judul berita sarkastik dan non-sarkastik. Setelah tahap pra-pemrosesan dan seleksi fitur, dataset dibagi menggunakan skema split validation 80:20 dengan stratifikasi label, sehingga distribusi kelas pada data latih dan data uji tetap proporsional. Distribusi kelas pada News Headlines Dataset yang menunjukkan jumlah data non-sarkastik (0) dan sarkastik (1) dapat dilihat pada Gambar 1, sedangkan contoh data dapat dilihat pada Tabel 1.

Data uji yang digunakan pada tahap evaluasi berjumlah 5.701 headline, terdiri dari 2.990 data non-sarkastik (label 0) dan 2.711 data sarkastik (label 1). Distribusi ini menunjukkan bahwa dataset relatif seimbang, sehingga metrik evaluasi seperti *accuracy* dan *F1-score* dapat digunakan secara representatif untuk menilai kinerja model.



Gambar 1. Distribusi kelas pada News Headlines Dataset

Tabel 1. Contoh beberapa data pada dataset

index	headline	is_sarcastic
0	<i>thirtysomething scientists unveil doomsday clock of hair loss</i>	1
1	<i>dem rep. totally nails why congress is falling short on gender, racial equality</i>	0
2	<i>eat your veggies: 9 deliciously different recipes</i>	0
3	<i>inclement weather prevents liar from getting to work</i>	1
4	<i>mother comes pretty close to using word 'streaming' correctly</i>	1

Hasil Pelatihan Model ANN Berbasis TF-IDF

Model klasifikasi yang digunakan adalah *Artificial Neural Network* (ANN) dengan arsitektur *Multi-Layer Perceptron*, yang menerima input berupa vektor TF-IDF (unigram dan bigram). Proses pelatihan dilakukan menggunakan optimizer Adam dan fungsi *loss binary cross-entropy*. Berdasarkan log pelatihan, model mencapai konvergensi secara stabil dalam waktu pelatihan yang relatif singkat, tanpa indikasi overfitting yang signifikan.

Hasil evaluasi pada data uji menunjukkan bahwa model ANN berbasis TF-IDF mampu mencapai:

- a) *Accuracy : 85,02%*
- b) *F1-score : 84,44%*

Nilai ini menunjukkan bahwa model memiliki kemampuan klasifikasi yang baik dalam membedakan headline sarkastik dan non-sarkastik. Gambar 2 menampilkan detail hasil evaluasi pada data uji.

179/179	2s 11ms/step																														
Accuracy:	0.8502017189966673																														
F1-score:	0.8444444444444444																														
Classification Report:																															
<table> <thead> <tr> <th></th><th>precision</th><th>recall</th><th>f1-score</th><th>support</th></tr> </thead> <tbody> <tr> <td>0</td><td>0.87</td><td>0.85</td><td>0.86</td><td>2990</td></tr> <tr> <td>1</td><td>0.83</td><td>0.86</td><td>0.84</td><td>2711</td></tr> <tr> <td>accuracy</td><td></td><td></td><td>0.85</td><td>5701</td></tr> <tr> <td>macro avg</td><td>0.85</td><td>0.85</td><td>0.85</td><td>5701</td></tr> <tr> <td>weighted avg</td><td>0.85</td><td>0.85</td><td>0.85</td><td>5701</td></tr> </tbody> </table>			precision	recall	f1-score	support	0	0.87	0.85	0.86	2990	1	0.83	0.86	0.84	2711	accuracy			0.85	5701	macro avg	0.85	0.85	0.85	5701	weighted avg	0.85	0.85	0.85	5701
	precision	recall	f1-score	support																											
0	0.87	0.85	0.86	2990																											
1	0.83	0.86	0.84	2711																											
accuracy			0.85	5701																											
macro avg	0.85	0.85	0.85	5701																											
weighted avg	0.85	0.85	0.85	5701																											

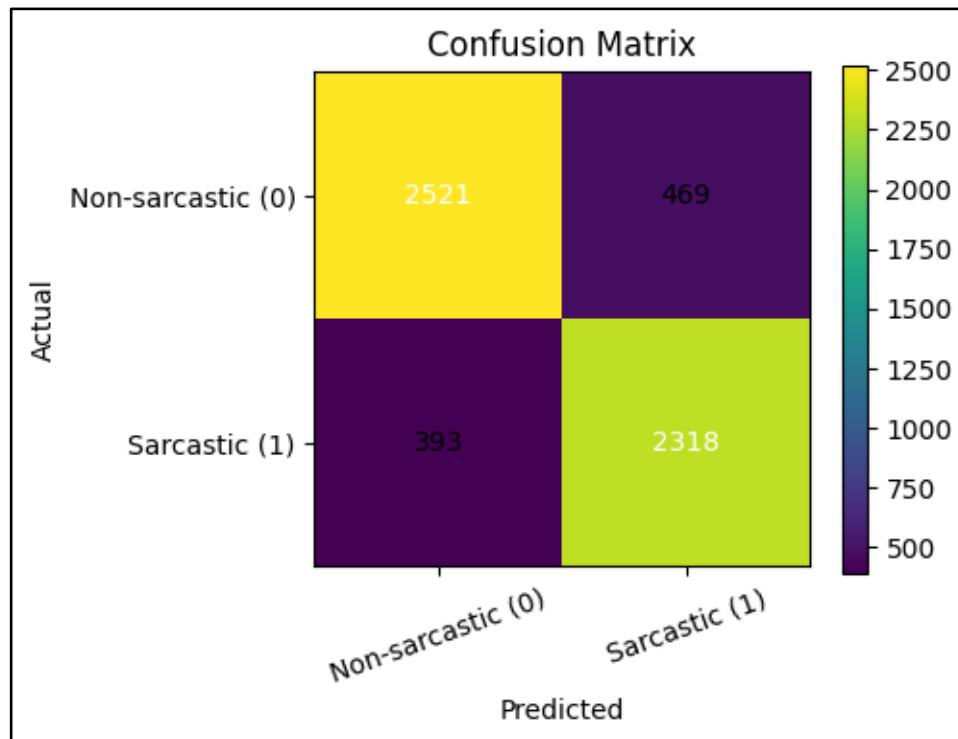
Gambar 2. Hasil Evaluasi pada Data Uji

Analisis Classification Report

Hasil evaluasi lebih rinci ditunjukkan melalui *classification report* yang mencakup nilai *precision*, *recall*, dan *F1-score* untuk masing-masing kelas. Untuk kelas non-sarkastik (label 0), model memperoleh *precision* sebesar 0,87 dan *recall* sebesar 0,85, yang menunjukkan bahwa sebagian besar *headline* yang diprediksi sebagai non-sarkastik memang benar, serta mayoritas *headline* non-sarkastik berhasil dikenali dengan baik.

Sementara itu, untuk kelas sarkastik (label 1), model mencapai *precision* sebesar 0,83 dan *recall* sebesar 0,86. Nilai *recall* yang relatif tinggi pada kelas sarkastik menunjukkan bahwa model cukup sensitif dalam mendeteksi *headline* sarkastik, meskipun masih terdapat sebagian kecil kesalahan prediksi yang disebabkan oleh sifat sarkasme yang implisit dan bergantung pada konteks pragmatik.

Nilai *macro average F1-score* sebesar 0,85 menunjukkan bahwa performa model relatif seimbang pada kedua kelas, tanpa bias yang signifikan terhadap salah satu label. Hal ini memperkuat validitas pendekatan ANN berbasis TF-IDF yang digunakan dalam penelitian ini. Gambar 3 menampilkan confusion matrix hasil klasifikasi pada data uji.



Gambar 3. Confusion matrix hasil klasifikasi ANN berbasis TF-IDF pada data uji

Pembahasan Hasil

Hasil eksperimen menunjukkan bahwa kombinasi TF-IDF sebagai representasi fitur dan *Artificial Neural Network* sebagai *classifier* mampu memberikan performa yang kompetitif untuk tugas deteksi sarkasme pada headline berita. Akurasi di atas 85% mengindikasikan bahwa model berhasil mempelajari pola-pola leksikal dan frasa khas yang membedakan berita satir dari berita non-satir.

Jika dibandingkan secara konseptual dengan pendekatan berbasis transformer yang dilaporkan dalam penelitian-penelitian terbaru, performa ANN berbasis TF-IDF memang cenderung lebih rendah. Namun demikian, pendekatan ini memiliki keunggulan dari sisi efisiensi komputasi, kesederhanaan arsitektur, dan kemudahan replikasi, sehingga tetap relevan sebagai baseline yang kuat, khususnya dalam konteks penelitian akademik dengan keterbatasan sumber daya.

Selain itu, hasil yang diperoleh menunjukkan bahwa peningkatan performa tidak selalu diperoleh dengan menambah kompleksitas model secara berlebihan. Temuan ini sejalan dengan literatur yang menyatakan bahwa untuk representasi fitur berbasis TF-IDF yang bersifat sparse dan hampir linear separable, model ANN dengan kedalaman moderat sudah cukup efektif tanpa perlu arsitektur yang terlalu dalam.

Analisis Kesalahan dan Keterbatasan

Meskipun menunjukkan performa yang baik, model masih mengalami kesalahan klasifikasi pada sejumlah headline, khususnya yang memiliki bentuk sarkasme implisit atau mengandalkan referensi budaya dan pengetahuan dunia. Hal ini merupakan keterbatasan umum dari pendekatan TF-IDF yang tidak mampu menangkap konteks semantik mendalam dan relasi antar kata secara kontekstual.

Selain itu, penelitian ini hanya mengevaluasi model pada satu dataset, sehingga generalisasi ke domain lain (misalnya berita politik atau media sosial) belum dapat dipastikan. Oleh karena itu, hasil yang diperoleh perlu diinterpretasikan dalam konteks dataset yang digunakan.

Ringkasan Bab

Secara keseluruhan, hasil eksperimen membuktikan bahwa *Artificial Neural Network* berbasis TF-IDF mampu mencapai kinerja yang baik dalam mendeteksi sarkasme pada headline berita, dengan akurasi 85,02% dan *F1-score* 84,44%. Hasil ini menunjukkan bahwa pendekatan yang relatif sederhana masih efektif untuk tugas klasifikasi sarkasme pada teks pendek dan terstruktur.

KESIMPULAN

Penelitian ini membahas deteksi sarkasme pada judul berita menggunakan pendekatan *Artificial Neural Network* (ANN) dengan representasi fitur *Term Frequency–Inverse Document Frequency* (TF-IDF). Dataset yang digunakan adalah *News Headlines Dataset for Sarcasm Detection* yang bersumber dari *The Onion* dan *HuffPost*. Melalui tahapan pra-pemrosesan teks, ekstraksi fitur TF-IDF, serta pelatihan model ANN, sistem klasifikasi sarkasme berhasil dibangun secara terstruktur dan dapat direplikasi.

Hasil eksperimen menunjukkan bahwa model ANN berbasis TF-IDF mampu mencapai akurasi sebesar 85,02% dan *F1-score* sebesar 84,44% pada data uji. Nilai ini mengindikasikan bahwa model memiliki kinerja yang baik dan seimbang dalam membedakan headline sarkastik dan non-sarkastik. Temuan ini menegaskan bahwa pendekatan ANN dengan representasi TF-IDF masih efektif sebagai metode *baseline* untuk deteksi sarkasme pada teks pendek seperti judul berita, dengan keunggulan pada efisiensi komputasi dan kesederhanaan arsitektur.

Meskipun demikian, pendekatan yang digunakan masih memiliki keterbatasan dalam menangkap sarkasme implisit yang bergantung pada konteks semantik dan pengetahuan dunia. Oleh karena itu, penelitian selanjutnya dapat mengembangkan pendekatan ini dengan memanfaatkan representasi fitur yang lebih kontekstual, seperti word embedding atau model transformer, serta melakukan evaluasi pada dataset yang lebih beragam untuk meningkatkan kemampuan generalisasi sistem.

DAFTAR PUSTAKA

- Al-Ayyoub, M., Alhindawi, N., & Jararweh, Y. (2023). Deep learning for sarcasm identification in

- news headlines. *Applied Sciences*.
- Devlin, J., Chang, M.-W., Lee, K., & Toutanova, K. (2019). BERT: Pre-training of deep bidirectional transformers for language understanding. arXiv Preprint, arXiv:1810.04805.
- Gupta, S., & Kumar, A. (2025). Sarcasm detection in breaking news headlines using deep learning. In Intelligent Computing and Communication (pp. 115–127). Springer.
- Hochreiter, S., & Schmidhuber, J. (1997). Long short-term memory. *Neural Computation*, 9(8), 1735–1780.
- Jurafsky, D., & Martin, J. H. (2025). Speech and language processing (3rd ed.). Stanford University.
- Kaur, N., & Gupta, V. K. (2023). Sarcasm detection in news headline dataset using ensemble deep learning models. *Journal of Information and Computing Science*, 18(2), 45–56.
- Kim, Y. (2014). Convolutional neural networks for sentence classification. arXiv Preprint, arXiv:1408.5882.
- Li, Y., Zhou, X., & Wang, J. (2022). Research on text classification techniques based on improved TF-IDF. *Procedia Computer Science*, 199, 143–150.
- Liu, Y., Ott, M., & Goyal, N. (2019). RoBERTa: A robustly optimized BERT pretraining approach. arXiv Preprint, arXiv:1907.11692.
- Misra, R. (2020a). News headlines dataset for sarcasm detection. arXiv Preprint, arXiv:2009.05446.
- Misra, R. (2020b). News headlines dataset for sarcasm detection [Dataset]. GitHub. <https://github.com/rishabhmisra/News-Headlines-Dataset-For-Sarcasm-Detection>
- Patel, R., & Shah, S. (2025). A hybrid transformer-based model for sarcasm detection from news headlines. *Neural Computing and Applications*.
- Reddy, S. R., & Kumar, P. R. (2023). Comparison of artificial decision techniques for detection of sarcastic news headlines. *Expert Systems with Applications*.
- Salton, G., & Buckley, C. (1988). Term-weighting approaches in automatic text retrieval. *Information Processing & Management*, 24(5), 513–523.
- Sharma, A., Gupta, R., & Sahay, S. K. (2025). Sarcasm detection on news headlines using transformer-based models. *The Journal of Supercomputing*.
- Singh, A., & Mehta, R. (2023). Sarcasm detection using news headlines dataset. *Procedia Computer Science*, 218, 223–230.
- Sokolova, M., & Lapalme, G. (2009). A systematic analysis of performance measures for classification tasks. *Information Processing & Management*, 45(4), 427–437. <https://doi.org/10.1016/j.ipm.2009.03.002>
- Vaswani, A., Shazeer, N., & Parmar, N. (2017). Attention is all you need. *Advances in Neural Information Processing Systems*, 5998–6008.
- Verma, P., & Mishra, S. (2024). Sarcasm detection in newspaper headlines. In Advances in Artificial Intelligence and Data Science (pp. 89–101). Springer.
- Zhang, L., Wang, H., & Liu, J. (2025). Unpacking sarcasm: A contextual and transformer-based approach for news headlines. *Computers*.