

Deteksi Berita Hoaks Terkait Debat Capres Pemilu 2024 Pada Media Sosial Menggunakan Metode Bayesian Neural Network

Anandita Prakarsa Maulida
Fakultas Informatika
Universitas Telkom, Bandung
ananditaprakarsa@student.telkomuniversity.ac.id

Yuliant Sibaroni
Fakultas Informatika
Universitas Telkom, Bandung
yuliant@telkomuniversity.ac.id

Abstrak

Berita Hoaks di media sosial semakin mengkhawatirkan, terutama pada saat pemilu, di mana informasi ini dapat mempengaruhi opini publik dan mengganggu integritas pemilu. Penelitian ini bertujuan untuk mengembangkan sistem deteksi hoaks menggunakan metode *Bayesian Neural Network* (BNN) yang dioptimalkan dengan teknik *Term frequency-Inverse Document frequency* (TF-IDF). Hasil pengujian menunjukkan bahwa sistem ini berhasil mencapai akurasi tinggi dalam mengklasifikasikan berita hoaks dan non-hoaks. Dibandingkan dengan penelitian lain, seperti menggunakan metode *K-Nearest Neighbor* (K-NN) mencapai akurasi 85%, Naïve Bayes dengan akurasi 82,6%, dan penelitian data mining menggunakan TF-IDF mencapai akurasi rendah 57%. Dengan menggunakan metode ini mengklasifikasi berita hoaks secara otomatis dengan memanfaatkan distribusi probabilistik untuk meningkatkan akurasi deteksi. Pengujian ini berhasil mendapatkan akurasi dengan fitur TF-IDF mencapai 85,71%, fitur Word2Vec mencapai akurasi tinggi yaitu 90,24%, dan fitur BERT mendapatkan akurasi rendah 75,27%. Penelitian ini diharapkan dapat menjadi referensi bagi pengembangan lebih lanjut dalam sistem deteksi hoaks dan meningkatkan kesadaran masyarakat akan pentingnya verifikasi informasi. Kata kunci: hoaks, bayesian neural network (BNN), pemilu 2024, media sosial, TF-IDF.

Abstract

The spread of hoax news on social media is becoming increasingly concern, especially during elections, where such misinformation can influence public opinion and disrupt the integrity of the electoral process. This study aims to develop a hoax detection system using the Bayesian Neural Network (BNN) method optimized with Term Frequency-Inverse Document Frequency (TF-IDF) techniques. The results of the testing indicate that the system successfully achieves high accuracy in classifying hoax and non-hoax news. Compared to other studies, such as those using the K-Nearest Neighbor (K-NN) method which achieved an accuracy of 85%, and Naïve bayes an accuracy of 82,6%, as well as data mining research using TF-IDF that reached a lower accuracy of 57%. By leveraging probabilistic distributions, this method enables automatic classification of hoax news to improve hoax detection accuracy. The test results indicate that using TF-IDF features achieved 85,71% accuracy, Word2Vec features obtained a high accuracy of 90,24%, while BERT features yielded a lower accuracy of 75,27%. This research is expected to serve as a reference for further development of hoax detection systems and to raise public awareness of the importance of information verification.

Keywords: hoax, bayesian neural network (BNN), 2024 elections, social media, TF-IDF.

1. PENDAHULUAN

Indonesia menganut sistem pemerintahan presidensial, di mana presiden berfungsi sebagai kepala negara dan kepala pemerintahan. Berdasarkan UUD 1945, pasal 7, masa jabatan presiden di Indonesia dapat mencapai maksimal dua periode, atau 10 tahun[1]. Pemilihan presiden tahun 2024 menjadi momen politik yang signifikan dan menarik perhatian masyarakat luas. Dalam konteks ini, fenomena penyebaran berita hoaks akan semakin cepat menyebar tanpa verifikasi yang memadai dan dapat menjadi ancaman serius terhadap integritas debat capres yang berlangsung selama kampanye[2]. Berita hoaks juga dapat berpotensi mempengaruhi opini publik dan persepsi terhadap calon presiden serta hasil debat yang disampaikan.

Platform media sosial, seperti X, Facebook, dan Instagram, telah menjadi sumber utama informasi politik. Namun, berita hoaks dapat menyebar dengan cepat melalui platform ini, karena kurangnya kontrol dan verifikasi informasi[3][4]. Oleh karena itu, penting untuk mengembangkan sistem yang efektif untuk mendeteksi berita hoaks. Penelitian tentang deteksi hoaks menggunakan berbagai metode menunjukkan hasil yang bervariasi. Metode *K-Nearest Neighbor* (K-NN) menunjukkan akurasi tertinggi sebesar 85%[5], sementara metode Naïve Bayes menghasilkan akurasi rata-rata 82,6%[6], penelitian selanjutnya menggunakan metode *Feature Expansion Word2Vec* mendapatkan akurasi 89,53% pada algoritma *Random Forest*[7], dan penelitian selanjutnya menggunakan teknik Data Mining Scikit Learn mendapatkan akurasi beragam, seperti TF-IDF (57%), CountVectorizer (100%), SupportVector Classifier (57%)

dan Passive Aggressive Classifier (71%)[8]. Di sisi lain, *Bayesian Neural Network* (BNN) menawarkan pendekatan yang lebih kompleks dengan memanfaatkan distribusi probabilistik untuk klasifikasi otomatis berita hoaks berdasarkan konten media sosial.

Dengan menggunakan metode BNN serta pembobotan *Term Frequency Inverse Document Frequency* (TF-IDF), penulis berharap dapat mengembangkan model yang efektif dalam mengklasifikasikan berita hoaks dan berita non-hoaks terkait debat capres pada pemilu 2024. Berita hoaks berpotensi menimbulkan dampak negatif, seperti mempengaruhi opini publik, memicu konflik, atau mengganggu integritas pemilu[9]. Oleh karena itu, diperlukan metode yang efektif untuk mendeteksi berita hoaks guna memastikan masyarakat memperoleh informasi akurat dan membuat keputusan yang tepat dalam konteks pemilu.

Penelitian ini berfokus pada pengumpulan data berita dan konten terkait debat capres pemilu 2024 dari media sosial X, yang kemudian akan digunakan untuk melatih serta menguji model BNN dalam klasifikasi berita hoaks dan non-hoaks. Diharapkan hasil dari penelitian ini diharapkan memberikan wawasan berharga dalam mengatasi masalah berita hoaks dalam konteks debat capres pada pemilu 2024 di Indonesia. Selain itu, penelitian ini juga bertujuan untuk memberikan rekomendasi dalam meningkatkan kesadaran masyarakat terhadap pentingnya informasi yang valid selama pemilu. Bayesian Neural Network (BNN) merupakan metode pembelajaran mesin yang menggabungkan prinsip-prinsip jaringan saraf tiruan dengan pendekatan probabilistik Bayesian. Beberapa keunggulan BNN antara lain, kemampuannya dalam menangani overfitting, memberikan estimasi ketidakpastian dalam prediksi, serta menghasilkan distribusi posterior yang informatif untuk parameter model[10][11]. Dengan karakteristik ini, BNN dipilih sebagai metode utama dalam klasifikasi berita hoaks pada penelitian ini.

2. KAJIAN TEORI

2.1 Identifikasi Berita Hoaks

Penelitian ini mengacu pada beberapa studi terkait dalam

deteksi berita hoaks, salah satunya penelitian oleh Han et al. pada tahun 2007 dengan menerapkan metode Graph Neural Network (GNNs)[12]. Penelitian ini fokus pada pendekatan berbasis propagasi untuk deteksi berita hoaks dengan menggunakan pola penyebaran berita di media sosial sebagai acuan. Penelitian ini memberikan kontribusi dengan menyajikan metode yang menggunakan teknik dari pembelajaran kontinu untuk melatih GNNs secara bertahap sehingga dapat mencapai kinerja yang seimbang pada dataset. Pada penelitian ini tidak terdapat angka matriks akurasi, recall, precision dan F1-score. Namun, pada model GNNs dapat mencapai performa yang sebanding atau bahkan superior dibandingkan dengan metode state-of-the-art yang membutuhkan analisis sofistikasi konten X dan balasan pengguna.

Rahutomo et al. (2019) melakukan penelitian dengan menggunakan metode Naïve Bayes[6]. Penelitian ini dengan bertujuan mendeteksi hoaks dalam bahasa Indonesia menggunakan algoritma Naive Bayes dan fokus untuk menguji efektivitas algoritma ini adalah mengklasifikasikan berita sebagai valid atau hoaks berdasarkan fitur Term Frequency (TF).

Munawar dan Silitonga (2019) melakukan penelitian lanjutan dengan mengembangkan sistem deteksi hoaks di media sosial menggunakan teknik data mining[8]. Dalam penelitian tersebut, beberapa model klasifikasi diuji untuk mengukur akurasi dalam mendeteksi hoaks. Model pertama yang diuji adalah TF-IDF, yang menghasilkan akurasi sebesar 57%, selanjutnya CountVectorizer dengan akurasi awal 100%, namun pada pengujian berikutnya mengalami penurunan menjadi 71%. Model PassiveAggressive Classifier memiliki rata-rata akurasi 65,3%, sedangkan Support Vector Classifier mencatat akurasi rata-rata 60%. Hasil penelitian ini menegaskan pentingnya pengembangan sistem deteksi hoaks yang lebih efektif di media sosial guna melindungi masyarakat dari informasi yang menyesatkan.

Ibrahim et al. (2020) melakukan studi kasus kantor Tribun Medan dengan menggunakan algoritma Naïve Bayes Classifier untuk mendeteksi tingkat kredibilitas Hoaks pada sosial media di Indonesia[13]. Penelitian ini bertujuan

untuk mendeteksi hoaks di media sosial Indonesia menggunakan algoritma Naïve Bayes Classifier. Algoritma ini mampu mengklasifikasikan berita hoaks dan fakta dengan cepat dan akurat, memanfaatkan probabilitas prior, likelihood dan evidence dalam proses klasifikasinya.

Sriyano dan Setiawan (2021) melakukan penelitian menggunakan Metode Naïve Bayes Multinomial dengan pembobotan TF-IDF untuk mendeteksi hoaks di twitter[14]. Tujuan penelitian ini membangun sistem yang mendeteksi berita hoaks di *Twitter* menggunakan metode Naïve Bayes Multinomial dengan pembobotan TF-IDF dan menyelidiki bagaimana penggunaan N-Gram (Unigram, Bigram, Trigram) dalam meningkatkan akurasi klasifikasi. Friskandini dan Setiawan (2021) melakukan penelitian dengan mendeteksi hoaks berbahasa Indonesia di *Twitter* menggunakan fitur ekspansi Word2Vec[7].

Penelitian ini membahas implementasi metode feature expansion menggunakan Word2vec untuk mendeteksi berita hoaks berbahasa Indonesia di *Twitter*. Model Random Forest dengan ekspansi fitur menggunakan corpus IndonesiaNews dan kombinasi parameter (Top 5 Similarity) memberikan akurasi tertinggi sebesar 89,53%, meningkatkan sebesar 1,46% dibandingkan baseline.

Agustina et al. (2021) melakukan penelitian dengan menggunakan algoritma Naïve Bayes Classifier untuk menemukan berita hoaks di media sosial[15]. Penelitian ini membangun sistem berbasis web untuk mendeteksi hoaks dengan menerapkan Naïve Bayes Classifier. Hasil pada model Naïve Bayes mendapatkan akurasi 81%.

Roshinta et al. (2023) melakukan penelitian dengan sistem deteksi hoaks berbahasa Indonesia pada bidang kesehatan[16]. Dalam penelitian ini algoritma Naïve Bayes digunakan untuk merancang dan membangun sistem deteksi berita hoaks berbahasa Indonesia di bidang kesehatan dengan akurasi sebesar 90,9%. Ini menunjukkan bahwa algoritma ini memiliki tingkat keandalan yang sangat tinggi untuk mendeteksi berita hoaks.

Melalui studi terkait ini[12], [6], [8], [13], [14], [7], [15], [16] penelitian ini berupaya untuk mengisi celah dalam literatur yang ada dengan mengembangkan sistem deteksi berita hoaks menggunakan metode Bayesian Neural

Network (BNN) dan menganalisis sentimen masyarakat terhadap berita yang beredar di mediasosial, khususnya menjelang pemilu 2024.

2.2 Hoaks

Hoaks merujuk pada informasi yang salah atau menyesatkan yang disebar dengan tujuan untuk memanipulasi pembaca[17]. Berita hoaks sering kali disajikan dalam bentuk artikel, video, atau gambar yang tampak meyakinkan, sehingga dapat dengan mudah diterima oleh masyarakat sebagai kebenaran. Adapun tujuan dari penyebaran hoaks ini untuk mempengaruhi pandangan dan opini masyarakat terhadap isu tertentu, termasuk politik, kesehatan, atau sosial. Dengan menyebarkan informasi yang tidak akurat membuat persepsi publik mendukung agenda mereka. Adapun juga menyebarkan berita hoaks untuk mendukung ideologi atau agenda politik tertentu. Dengan menyebarkan informasi yang mendukung pandangan tertentu, penyebar hoaks berusaha memobilisasi dukungan atau menurunkan dukungan terhadap lawan politik.

2.3 Preprocessing

Untuk mempersiapkan data langkah awal yang dilakukan adalah *preprocessing*. Dalam penelitian ini, tahap *preprocessing* yang digunakan adalah *Case Folding*, *Normalization*, *Filtering*, dan *Stemming*.

2.4 Term Frequency-Inverse Document Frequency (TF-IDF)

Term Frequency (TF) dan *Inverse Document Frequency* (IDF) merupakan dua komponen utama yang digunakan dalam pengolahan teks dan analisis informasi untuk menilai seberapa penting sebuah kata dalam sebuah dokumen relatif terhadap kumpulan dokumen[19].

- *Term Frequency* (TF), mengukur seberapa sering sebuah kata muncul dalam dokumen tertentu. Semakin sering kata tersebut muncul, semakin tinggi nilai TF-nya.

$$TF(t, d) = \frac{f(t, d)}{\sum t' f(t', d)} \quad (1)$$

<i>Term Frequency</i>	Definisi Term Frequency
<i>t</i>	Kata tertentu
<i>d</i>	Dokumen tertentu
<i>F (t,d)</i>	Frekuensi kemunculan kata <i>t</i> dalam dokumen <i>d</i> Penyebut adalah total jumlah semua kata dalam dokumen <i>d</i>

- *Inverse Document Frequency* (IDF), digunakan untuk mengukur tingkat kepentingan suatu kata dalam keseluruhan kumpulan dokumen. Kata yang sering muncul di banyak dokumen akan memiliki nilai IDF yang rendah, sedangkan kata yang jarang muncul akan memiliki nilai IDF yang tinggi.

$$IDF(t) = \log \left(\frac{N}{1 + n(t)} \right) \quad (2)$$

<i>Inverse Document Frequency</i>	Definisi Inverse Document Frequency
<i>N</i>	Jumlah total dokumen
<i>n(t)</i>	Jumlah dokumen yang mengandung kata <i>t</i> Ditambahkan 1 pada penyebut untuk menghindari pembagian dengan nol.

- **TF-IDF Score**
 $TF\ IDF(t, d) = TF(t, d) \times IDF(t) \quad (3)$
 TF-IDF mempertimbangkan kata-kata yang sering muncul dalam suatu dokumen tetapi jarang muncul di dokumen lain dengan mengalikan kedua nilai ini. Teknik ini sangat bermanfaat untuk klasifikasi teks, termasuk untuk mendeteksi berita hoaks, karena memungkinkan model untuk lebih fokus pada kata-kata kunci yang lebih relevan dan informatif.

2.5 Word2Vec

Word2Vec adalah metode pembelajaran representasi kata dalam bentuk vektor berkelanjutan yang dikembangkan[20]. Model ini bertujuan untuk menangkap hubungan semantik dan sintaksis antar kata dengan menggunakan teknik pembelajaran berbasis jaringan saraf. Adapun dua pendekatan utama dalam Word2Vec adalah :

1. *Continuous Bag-of-Words* (CBOW) :

Memprediksi kata tengah berdasarkan hubungannya dengan kata lain.

2. *Skip-gram model* : memprediksi kata-kata di sekitar berdasarkan kata tengah.

Adapun rumus yang digunakan pada fitur ini adalah :

- Kompleksitas Pelatihan Word2Vec
 $O = E \times T \times Q \quad (1)$

Word2Vec	Definisi Kompleksitas Pelatihan Word2Vec
<i>E</i>	Jumlah epoch pelatihan
<i>T</i>	Jumlah kata dalam dataset
<i>Q</i>	Kompleksitas spesifik untuk setiap model

- *Continuous Bag-of-Words* (CBOW)
 $Q = N \times D + D \times \log 2(V) \quad (2)$

Word2Vec	Definisi CBOW
<i>N</i>	Jumlah kata dalam konteks
<i>D</i>	Dimensi vektor kata
<i>V</i>	Ukuran kosakata

- *Skip-gram Model*
 $Q = C \times (D + D \times \log 2(v)) \quad (3)$

Word2Vec	Definisi Skip-gram Model
<i>C</i>	Rentang konteks (jumlah kata sebelum dan sesudah kata target)
<i>D dan V</i>	Dimensi vektor kata dan ukuran kosakata

2.6 BERT

Bidirectional Encoder Representations from Transformers adalah model berbasis *Transformer* yang dikembangkan untuk pemahaman bahasa alami. Tidak seperti model sebelumnya yang hanya membaca teks dari satu arah (*unidirectional*), BERT mempelajari representasi kata dengan pendekatan dua arah (*bidirectional*). Hal ini memungkinkan model memahami konteks secara lebih

menyeluruh dalam sebuah kalimat[21]. BERT digunakan sebagai model pra-terlatih (*pre-trained*) yang dapat disesuaikan (*fine-tuned*) untuk berbagai tugas NLP, seperti klasifikasi teks, *question answering*, dan *named entity recognition* (NER). Namun, dalam tugas sentence embedding, BERT standar memiliki keterbatasan karena tidak secara langsung menghasilkan representasi kalimat yang optimal. BERT menggunakan dua tugas utama dalam pra-pelatihannya[22]:

1. *Masked Language Model* (MLM) : Token khusus yang digunakan untuk menggantikan beberapa kata dalam teks, dan model dilatih untuk menemukan kata yang hilang berdasarkan konteks. Fungsi loss yang digunakan :

$$P(w_i|w_1, \dots, w_i - 1, w_i + 1, \dots, w_n) = \frac{e^{s(w_i)}}{\sum_{w' \in V} e^{s(w')}} \quad (4)$$

Definisi Fungsi Loss pada MLM

Dimana $P(w_i)$ adalah probabilitas kata yang benar, $s(w_i)$ adalah skor logit, dan V adalah kosakata model. Skor logit adalah nilai sebelum diterapkannya fungsi aktivasi softmax dalam model klasifikasi berbasis jaringan saraf, termasuk dalam BERT. Skor ini menunjukkan seberapa besar kecenderungan suatu kelas atau token menjadi prediksi yang benar sebelum dikonversi menjadi probabilitas. Skor logit $s(w_i)$ diberikan oleh :

$$s(w_i) = Wh_i + b \quad (5)$$

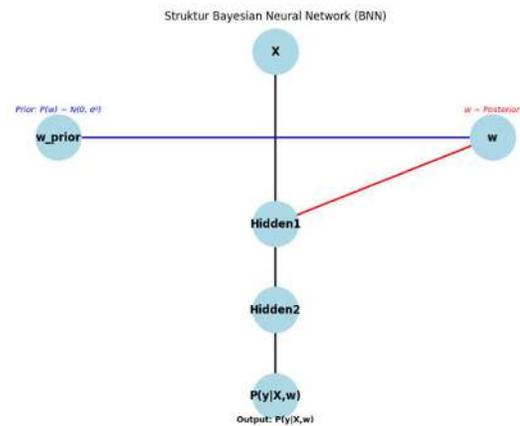
BERT	Definisi
W	Matriks bobot model
h_i	Vektor representasi fitur dari kata w_i ,
b	bias

2. *Translation Language Modeling* (TLM) : Menggunakan pasangan terjemahan dari dua bahasa yang digabungkan menjadi satu input, lalu sebagian kata dalam pasangan terjemahan tersebut dimasking. Model dilatih untuk memprediksi kata yang hilang menggunakan informasi dari kedua bahasa, sehingga meningkatkan pemahaman lintas bahasa.

Dengan pendekatan ini, fitur BERT berhasil mengungguli model NLP sebelumnya dalam memahami konteks bahasa secara lebih baik dan mendapatkan hasil terbaik dalam berbagai tugas pemrosesan bahasa alami.

2.7 Bayesian Neural Network

Bayesian neural network (BNN) adalah metode pendekatan dalam pembelajaran mesin yang menggabungkan jaringan saraf tiruan (*neural network*) dengan pemodelan stokastik[10]. BNN dirancang untuk mengatasi beberapa kelemahan dari jaringan saraf tradisional, terutama dalam hal yang tidak pasti dan overfitting. Dalam BNN, setiap parameter model dianggap sebagai variabel acak yang memiliki distribusi probabilitas tertentu[23]. Probabilitas keluaran dalam BNN dihitung menggunakan distribusi prediktif sebagai berikut[24][25]:



GAMBAR 1. Diagram Struktur Bayesian Neural Network

$$P(y|X, w) = f(X, w) \quad (6)$$

BNN	Definisi Distribusi Probabilitas
X	Input data (TF-IDF pada kasus)
w	Bobot jaringan saraf dengan distribusi probabilitas
$f(X, w)$	Jaringan saraf dengan bobot stokastik yang menghasilkan probabilitas output

Bobot w dianggap memiliki distribusi prior, biasanya Gaussian dengan mean 0 dan variansi kecil:

$$P(w) = N(0, \sigma^2) \quad (7)$$

BNN	Definisi Distribusi Posterior
$P(D w)$	Likelihood data yang dihasilkan dengan bobot tertentu
$P(w)$	Prior dari bobot jaringan saraf

$P(w|X)$ Evidence atau konstanta normalisasi

Likelihood mengukur seberapa baik model dengan bobot tertentu menjelaskan data:

$$P(y|X, w) = \prod_{i=1}^n B(y_i|\sigma(f(X_i, w))) \quad (8)$$

BNN	Definisi Likelihood
y_i	Label biner (Hoaks atau non-hoaks)
σ	Fungsi sigmoid yang digunakan di lapisan terakhir
B	Distribusi Bernoulli untuk klasifikasi biner

Prediksi probabilitas keluaran dihitung dengan mengintegrasikan bobot acak:

$$P(y^* | X^*, D) = \int P(y^* | X^*, w)P(w|D)dw \quad (9)$$

BNN	Definisi Hasil Prediksi BNN
$P(y^* X^*, D)$	Probabilitas prediksi output y^* untuk input baru X^* dengan mempertimbangkan data pelatihan D
$P(y^* X^*, w)$	Probabilitas prediksi output y^* dengan bobot parameter jaringan w
$P(w D)$	Posterior distribusi dari parameter w berdasarkan data pelatihan D , diperoleh dengan mengaplikasikan aturan bayes
Integral \int	Untuk menghitung probabilitas prediksi

Fungsi loss digunakan untuk menoptimasi model adalah Binary Cross Entropy:

$$BCE(y, y^{\wedge}) = -\frac{1}{N} \sum_{i=1}^N [y_i \log(y^{\wedge} i) + (1 - y_i) \log(1 - y^{\wedge} i)] \quad (10)$$

BNN	Definisi Binary Cross Entropy Loss
N	Jumlah sampel dalam dataset
y_i	Label sebenarnya untuk sampai ke-i(0 atau 1)
$y^{\wedge} i$	Probabilitas prediksi dari model untuk sampai ke-i
$\log(y^{\wedge} i)$	Logaritma dari probabilitas yang diprediksi, digunakan untuk menghitung loss prediksi yang benar
$1-y_i, 1-y^{\wedge} i$	Komponen untuk kesalahan prediksi kelas negatif
$-\frac{1}{N} \sum$	Rata-rata keseluruhan loss

2.5 Ukuran Performa

Metode evaluasi performa biasanya digunakan untuk

mengevaluasi akurasi mode, seperti akurasi, precision, recall, dan F1-score, serta dapat digunakan untuk membandingkan kinerja model yang berbeda. Confusion Matrix adalah alat evaluasi yang digunakan oleh penulis dalam penelitian ini. Metode evaluasi yang digunakan untuk mengevaluasi kinerja model klasifikasi dalam pembelajaran mesin adalah confusion matrix[26]. Matriks ini menunjukkan cara model mengklasifikasikan data uji dengan membandingkan hasil prediksi model dengan labelnya yang sebenarnya. Confusion Matrix dapat dilihat pada tabel berikut:

TABEL 1. Confusion Matrix

	Predicted Class Positive	Predicted Class Negative
Actual Class Positive	True Positive (TP)	False Negative (FN)
Actual Class Negative	False Positive (FP)	True Negative (TN)

True Positive (TP) merupakan jumlah kasus positif yang benar-benar terdeteksi sebagai positif oleh model, ini menunjukkan bahwa model berhasil mengidentifikasi kelas positif. False Positive (FP) merupakan jumlah kasus negatif yang salah terdeteksi sebagai positif oleh model, ini juga dikenal sebagai "Type I error" dimana model memberikan prediksi positif yang tidak akurat. True Negative (TN) merupakan jumlah kasus negatif yang benar-benar terdeteksi sebagai negatif oleh model, menunjukkan bahwa model berhasil mengidentifikasi kelas negatif. False negative (FN) merupakan jumlah kasus positif yang salah terdeteksi sebagai negatif oleh model, dikenal sebagai "Type II error" dimana model gagal mengidentifikasi kelas positif.

Pada tahap ini nilai performansi model akan diukur melalui perhitungan akurasi, precision, dan recall.

Akurasi adalah proporsi prediksi yang benar dari total prediksi.

$$Akurasi = \frac{TP + TN}{TP + FP + TN + FN} \quad (11)$$

Precision adalah persentase prediksi positif yang benar dari total prediksi positif.

$$Precision(P) = \frac{TP}{TP + FP} \quad (12)$$

Recall adalah persentase kasus positif yang benar-benar

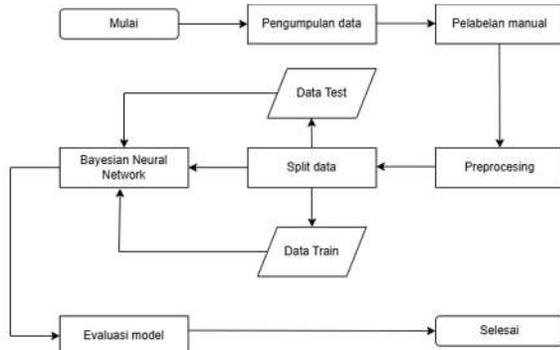
terdeteksi oleh model.

$$Recall (R) = \frac{TP}{TP + FN} \quad (13)$$

F1-Score adalah skor rata-rata harmonis untuk presisi dan recall, yang mengimbangi keduanya.

$$F1 = 2X \frac{Presisi \times Recall}{Presisi + Recall} \quad (14)$$

3. METODE



GAMBAR 2 Perancangan Sistem

3.1 Pengumpulan Data

Dalam penelitian ini langkah pertama dalam perancangan sistem adalah pengumpulan data, data yang diambil adalah berita terkait debat capres pada pemilu 2024 di media sosial terutama aplikasi X. Proses pengumpulan data akan melibatkan penggunaan API pada aplikasi X, yang berdasarkan keyword “Debat Capres 2024, Debat Capres Anies 2024, Debat Capres Prabowo 2024, dan Debat Capres Ganjar 2024, pada rentang waktu Desember 2023 - Oktober 2024. Total jumlah data yang didapatkan adalah 2.896 data.

3.2 Pelabelan Manual

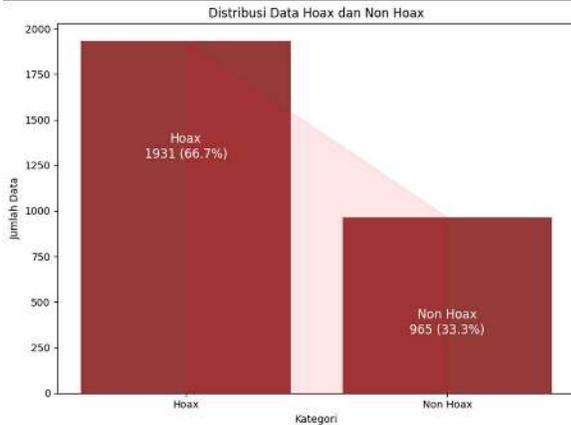
Pelabelan adalah proses menentukan kelas suatu data secara manual dengan memberi label hoaks dan non-hoaks. Dalam proses ini, setiap *tweet* dievaluasi oleh tiga orang berbeda, dan label yang paling banyak dipilih akan digunakan sebagai label akhir. Pelabelan dilakukan secara manual dan kriteria yang digunakan dalam pelabelan ini adalah dengan memperhatikan username yang digunakan mencerminkan identitas asli atau hanya nama samaran, mengecek apakah tweet menyertakan tautan ke sumber informasi terpercaya atau tidak, perbandingan jumlah

followers dan *following*, jika akun memiliki jumlah *followers* yang jauh lebih sedikit dibandingkan *following* cenderung merupakan akun baru yang digunakan untuk menyebarkan hoaks. Setelah tweet dilabeli berdasarkan kriteria, hasil pelabelan dari tiga orang yang dikumpulkan. Jika terdapat perbedaan pendapat, maka label yang paling banyak dipilih akan digunakan sebagai label akhir. Total data yang didapatkan setelah melakukan pelabelan manual adalah dengan 1.931 data label hoaks dan 965 data dengan label non-hoaks.

TABEL 2. Contoh Kalimat Hoax dan Non-Hoax

No.	Hoaks	Penjelasan
1.	“Jika Ingin Menang Pilpres, Anies Harus Geser dari 'Sayap Kanan' ke 'Tengah'”	Berita dilabeli hoaks dikarenakan tautan yang disertakan tidak dapat diakses (404 Not Found), yang menunjukkan bahwa berita tersebut telah dihapus atau tidak pernah ada.
2.	“Tidak Pamit Usai Lepas Jabatan Gubernur DKI, Anies Baswedan Dinilai Tak Berterima Kasih pada Prabowo”	Berita dilabeli hoaks dikarenakan sudah dikonfirmasi Geisz jikalau pak Anies berpamitan dengan sopan kepada pak Prabowo “Bahwa dia menyampaikan keinginannya untuk bertemu Pak Prabowo, dari jauh-jauh hari dia sudah sampaikan, dalam rangka pamitan ke mana pun dia pergi. Dia hadir berbagai pertemuan, bahkan termasuk dalam masyarakat miskin kota, bahkan kepada kepala dinas, bahkan kepada lurah Tanah Abang dia datangi, lurah loh, apalagi kepada Pak Prabowo yang memberikan jalan dia menjadi gubernur DKI,” sebut Geisz.

3.	“Prabowo Subianto Meninggal Dunia Karena Tenggelam”	Berita dilabeli hoaks dikarenakan berita ini sangat menyimpang dan tidak sesuai fakta bahwasanya bapak Prabowo masi hidup.
----	---	--



GAMBAR 3.
Distribusi Hoax dan Non-Hoax

3.3 Preprocessing

Setelah data diberi label, tahap *preprocessing* dilakukan untuk membersihkan dan mempersiapkan sebelum digunakan dalam proses selanjutnya. *Preprocessing* sangat penting dalam sistem berbasis teks karena berfungsi untuk menghitung bobot dengan lebih akurat serta mengatasi potensi kekurangan data. Dalam sistem yang dikembangkan, terdapat empat tahap preprocessing, yaitu:

- a. *Tokenisasi* memisahkan suatu kalimat menjadi kata-kata individu. Contohnya, kalimat “saya orang Indonesia” akan dipecah menjadi “saya”, “orang”, “Indonesia”.
- b. *Pembersihan data (Cleaning)* Menghapus elemen yang tidak diperlukan. seperti URL, mention, dan tagar, karena tidak memiliki arti penting.
- c. *Stemming* Menghilangkan imbuhan pada kata “me-“, “kan-“, dan lainnya, dengan tujuan mengurangi variasi kata dan menyederhanakan representasi teks.
- d. *Case Folding* mengubah semua huruf dalam teks menjadi huruf kecil agar konsistensi dalam pemrosesan data tetap terjaga.

Berikut contoh preprocessing yang pada Tabel 3.

TABEL 3
Contoh Preprocessing

Teks Awal	Wah, seru banget nih debat Capres 2024 bareng KADIN. Jadi penasaran sama visi-misi calon presiden yang lain!
Cleaning	Wah seru banget nih debat Capres bareng KADIN Jadi penasaran sama visi misi calon presiden yang lain
Stemming	['seru', 'banget', 'debat', 'capres', 'bareng', 'kadin', 'penasaran', 'visi', 'misi', 'calon', 'presiden']

3.4 Split Data

Setelah *preprocessing*, dataset akan dibagi menjadi subset data latih dan data uji, menggunakan metode *Term Frequency-Inverse Document Frequency (TF-IDF)*. Data latih ini akan digunakan untuk melatih model, sedangkan data uji digunakan untuk mengukur kinerja model setelah pelatihan. Pembagian ini dilakukan dengan menggunakan teknik stratified sampling untuk menjaga bahwa proporsi kelas yang seimbang dalam setiap subset data. Selain itu, proses evaluasi model dilakukan menggunakan teknik cross-validation dengan 10 fold. Dalam metode ini, dataset dibagi menjadi 10 bagian yang kurang lebih sama. Pada setiap iterasi, sembilan bagian digunakan sebagai data latih dan satu bagian digunakan sebagai data uji. Hasil evaluasi model kemudian diporelah dari rata-rata performa model di seluruh iterasi, sehingga memberikan penilaian yang stabil dan mengurangi risiko overfitting.

3.5 Pemodelan

Metode yang akan digunakan dalam sistem deteksi berita hoaks ini adalah *Bayesian Neural Network (BNN)*. *Bayesian Neural Network* adalah metode pembelajaran mesin yang menggabungkan prinsip-prinsip jaringan saraf tiruan dengan pendekatan probabilistik Bayesian[10]. Keunggulan pada model ini adalah dapat mengurangi overfitting, model ini dapat menangkap variasi dalam data dan memberikan hasil yang lebih akurat, dapat diterapkan pada berbagai jenis masalah terutama mendeteksi berita hoaks. Pada tahap pelatihan, model akan dilatih untuk mengoptimalkan parameter model agar dapat melakukan klasifikasi dengan akurat. Proses dimulai dengan mempersiapkan data, termasuk preprocessing yang meliputi pembersihan data dan penerapan TF-IDF untuk menghasilkan representasi numerik dari kata-kata. Setelah itu inisialisasi model dengan menentukan jumlah lapisan

dan neuron di setiap lapisan seperti ReLU atau sigmoid, setelah itu proses pelatihan melalui langkah forward pass, penghitungan loss, dan backpropagation. Hasil dari proses ini menunjukkan bahwa model BNN yang dilatih berhasil mencapai akurasi tinggi dalam mendeteksi hoaks.

3.6 Evaluasi Model

Setelah model dilatih, langkah selanjutnya adalah melakukan evaluasi terhadap performa model dalam mendeteksi berita hoaks. Pada tahap ini, kinerja model diukur untuk menentukan seberapa baik model dapat melakukan klasifikasi pada data yang tidak terlihat (data pengujian). Evaluasi dilakukan dengan menerapkan teknik cross-validation dan analisis confusion matrix guna memberikan gambaran yang lebih menyeluruh mengenai kemampuan model dalam mendeteksi hoaks. Cross validation merupakan teknik dalam pembelajaran mesin yang digunakan untuk menilai kinerja model dengan membagi dataset ke dalam beberapa subset. Tujuan dari metode ini adalah memastikan bahwa model dapat bekerja secara optimal pada data pelatihan serta tetap andal ketika diterapkan pada data yang belum pernah dilihat sebelumnya[27]. Cross validation juga dapat membantu menilai akurasi model deteksi berita hoaks yang dikembangkan. Dengan membagi data menjadi beberapa bagian, model dapat dilatih pada sebagian data dan diuji, sehingga memberikan gambaran tentang kemampuannya dalam melakukan generalisasi. Dalam penelitian ini, digunakan metode Stratified K-Fold Cross validation dengan 10 lipatan. Proses dimulai dengan membagi data menjadi bagian- bagian, dimana setiap fold model dilatih menggunakan sebagian data (data train) dan diuji pada sisa data (data test). Setelah semua fold dievaluasi, akurasi dari setiap fold dicatat dan rata-rata akurasi dihitung untuk memberikan gambaran umum tentang kinerja model. Selain akurasi, penelitian ini juga menghitung precision, recall, dan F1-score.

4. HASIL DAN PEMBAHASAN

4.1 Hasil Pengujian

Penulis mendapatkan pengumpulan data dengan *crawling* berjumlah 2.896 data, yang dimana data dibagi menjadi 1.931 data hoaks dan 965 data non-hoaks. Data tersebut telah dilakukan preprocessing dan dibagi menjadi data uji

dan data tes. Model yang digunakan untuk melakukan klasifikasi hoaksnya adalah Bayesian Neural Network (BNN) dengan arsitektur fully connected digunakan untuk melakukan klasifikasi hoaks. Hasil pengujian menunjukkan akurasi rata-rata yang dihasilkan 85,71%% dengan fold 10. Selain itu metrik tambahan seperti precision 0,87, recall 0,97, F1-score 0,92. Rincian dari hasil pengujian menggunakan model BNN bisa dilihat pada data dibawah :

TABEL 4.

Hasil Peformasi Keseluruhan Dari Model BNN dengan 10 Fold

K-Fold	Akurasi	F1-Score	Recall	Precision
Fold 1	0,83	0,91	1,00	0,83
Fold 2	0,84	0,91	1,00	0,83
Fold 3	0,83	0,90	0,98	0,83
Fold 4	0,83	0,91	1,00	0,83
Fold 5	0,83	0,91	0,99	0,83
Fold 6	0,90	0,93	0,92	0,93
Fold 7	0,87	0,91	1,00	0,83
Fold 8	0,90	0,94	0,89	0,99
Fold 9	0,91	0,94	0,93	0,96
Fold 10	0,85	0,91	1,00	0,83
Rata-rata akurasi seluruh fold				0,86

Pada pengujian selanjutnya, pengujian akan dibandingkan dengan fitur Word2Vec dan BERT. Perbandingan ini untuk melihat apakah model bekerja dengan baik apa tidak jika tidak memakai fitur TF-IDF.

TABEL 5.

Hasil Akurasi Dengan Menggunakan Fitur Word2vec dengan 10 Fold

K-Fold	Akurasi	F1-Score	Recall	Precision
Fold 1	0,89	0,93	0,88	0,99
Fold 2	0,90	0,93	0,88	1,00
Fold 3	0,89	0,93	0,87	1,00
Fold 4	0,90	0,93	0,88	1,00
Fold 5	0,90	0,93	0,88	0,99

5				
Fold 6	0,90	0,91	1,00	0,83
Fold 7	0,89	0,92	0,86	1,00
Fold 8	0,91	0,94	0,89	1,00
Fold 9	0,92	0,94	0,89	1,00
Fold 10	0,92	0,95	0,91	0,99
Rata-rata akurasi seluruh fold				0,92

TABEL 6.
Hasil Akurasi Dengan Menggunakan Fitur BERT dengan 5 Fold

	Akurasi pada BERT
Fold 1	0,84
Fold 2	0,81
Fold 3	0,82
Fold 4	0,46
Fold 5	0,85
Rata-rata akurasi seluruh fold	0,75

TABEL 7.
Hasil Rata-rata Akurasi dari Fitur TF-IDF, Word2Vec dan BERT

	Rata-rata Akurasi
TF-IDF	0,86
Word2Vec	0,92
BERT	0,75

Dari hasil pengujian, terlihat bahwa model klasifikasi dengan fitur Word2Vec menunjukkan hasil yang lebih baik dibandingkan dengan fitur BERT. Dengan akurasi rata-rata untuk TF-IDF adalah 85,71% menggunakan fitur TF-IDF, penggunaan fitur Word2vec menghasilkan akurasi yang lebih tinggi dibandingkan dengan TF-IDF dan BERT, dengan akurasi mencapai 90,24%. Hal ini menunjukkan bahwa representasi kata berbasis vektor yang dihasilkan oleh Word2Vec mampu menangkap hubungan semantik antar

kata dengan lebih baik, sehingga meningkatkan performa model klasifikasi. Sementara itu, fitur BERT menunjukkan akurasi terendah, yaitu 75,27%. Meskipun BERT merupakan model berbasis transformator yang mampu memahami konteks dalam kalimat, hasil ini mengindikasikan bahwa dalam penelitian yang dilakukan, representasi fitur BERT kurang optimal dibandingkan metode berbasis vektor seperti Word2Vec dan metode berbasis frekuensi seperti TF-IDF.

4.2 Analisis Hasil Pengujian

Secara keseluruhan, hasil analisis pengujian menunjukkan bahwa penggunaan Bayesian Neural Network (BNN) dalam penelitian ini berhasil mencapai akurasi 85,71% dengan menggunakan fitur TF-IDF, 90,24% dengan menggunakan fitur Word2Vec dan 75,27% menggunakan fitur BERT. Beberapa penelitian sebelumnya menggunakan berbagai metode untuk mendeteksi hoaks dan menghasilkan bervariasi akurasi. Penelitian menggunakan K- Nearest neighbor (KNN) mencapai akurasi 85%, setelah itu Naïve Bayes mencapai 82,6%, penelitian menggunakan data mining menunjukkan akurasi lebih rendah 57% dengan model TF-IDF[5] [6] [7] [8]. Perbandingan ini menunjukkan bahwa metode BNN yang digunakan dalam penelitian ini memberikan hasil yang baik daripada metode lain yang telah diuji, dengan penggabungan BNN dengan metode TF-IDF dan Word2Vec terbukti efektif dalam mendeteksi berita hoaks dengan lebih akurat. Sistem ini memberikan kontribusi yang signifikan dalam upaya melawan penyebaran informasi yang salah di media sosial, terutama selama periode pemilu 2024. Hasil dari deteksi hoaks dengan BNN ini menunjukkan potensi besar untuk meningkatkan akurasi dalam mengidentifikasi informasi yang salah, serta memberikan fondasi yang kuat untuk penelitian mengenai deteksi hoaks.

5. KESIMPULAN

Penelitian ini menunjukkan bahwa sistem deteksi hoaks yang dibangun menggunakan metode Bayesian Neural Network (BNN) yang dioptimalkan dengan teknik TF-IDF berhasil mencapai akurasi yang tinggi dalam mengklasifikasikan berita hoaks dan non-hoaks. Dengan

akurasi sebesar 85,71%, hasil ini membuat metode BNN efektif dalam mendeteksi berita hoaks. Saran untuk penelitian ini adalah melakukan pengujian lebih lanjut dengan menggunakan dataset yang lebih besar dan beragam untuk meningkatkan generalisasi model, dan diharapkan penelitian ini bisa dipraktikkan secara langsung kepada masyarakat. Selain itu, eksplorasi kombinasi metode lain dengan BNN juga dapat dipertimbangkan untuk meningkatkan performa sistem. Dengan ini, diharapkan sistem deteksi hoaks dapat terus diperbaiki dan lebih efektif dalam menghadapi tantangan informasi di era digital.

REFERENSI

- [1] “UNDANG-UNDANG DASAR NEGARA REPUBLIK INDONESIA 1945.”
- [2] M. Ravii Marwan Ahyad Jurusan Ilmu Komunikasi and F. Ilmu Komunikasi, “Analisis Penyebaran Berita Hoax di Indonesia.”
- [3] A. W. Fathurrahman *et al.*, “Penerapan Machine Learning untuk Pengklasifikasian Hoaks pada Platform Media Sosial: Studi Literatur.”
- [4] B. Imran, M. Nasirudin Karim, and N. Isna Ningsih, “Klasifikasi Berita Hoax Terkait Pemilihan Umum Presiden Republik Indonesia Tahun 2024 Menggunakan Naive Bayes dan SVM.” [Online]. Available: <http://jurnal.dinarek.unsoed.ac.id>
- [5] R. Darmawan, “Implementasi Metode K-Nearest Neighbor (K-NN) untuk Identifikasi Ujaran Kebencian Terhadap Tokoh Politik pada Twitter,” Pekanbaru, 2019.
- [6] F. Rahutomo, I. Yanuar Risca Pratiwi, D. Mayangsari Ramadhani, and P. Negeri Malang Jalan Soekarno Hatta No, “Eksperimen Naive Bayes pada Deteksi Berita Hoax Berbahasa Indonesia.”
- [7] F. Ismayanti and E. B. Setiawan, “Deteksi Konten Hoax Berbahasa Indonesia di Twitter Menggunakan Fitur Ekspansi dengan Word2Vec.”
- [8] Y. Riadi Silitonga, “Sistem Pendeteksi Berita Hoax di Media Sosial dengan Teknik Data Mining Scikit Learn,” 2019. [Online]. Available: www.beritasatu.com,
- [9] R. Manthovani Universitas Pancasila Jl Lenteng Agung Raya No, S. Sawah, K. Jagakarsa, K. Jakarta Selatan, and D. Khusus, “Dampak Berita Hoax Terhadap Keamanan Negara Dalam Persepektif Cyberlaw Bela Negara”, [Online]. Available: <https://www.kemhan.go.id/belanegara/opini/as>
- [10] E. Goan and C. Fookes, “Bayesian Neural Networks: An Introduction and Survey,” Jun. 2020, doi: 10.1007/978-3-030-42553-1_3.
- [11] D. Rosita, dan Syamsuddin Mallala, S. Informatika, S. Widya Cipta Dharma, T. Informatika, and P. Korespondensi, “Komparasi Data Mining Naive Bayes dan Neural Network Memprediksi Masa Studi Mahasiswa S1,” vol. 7, no. 3, pp. 443–452, 2020, doi: 10.25126/jtiik.202072093.
- [12] Y. Han, S. Karunasekera, and C. Leckie, “Graph Neural Networks with Continual Learning for Fake News Detection from Social Media,” Jul. 2020, [Online]. Available: <http://arxiv.org/abs/2007.03316>
- [13] M. Ibrahim, E. Bu, and I. Lubis, “RESOLUSI : Rekayasa Teknik Informatika dan Informasi Penerapan Algoritma Naive Bayes Classifier Untuk Mendeteksi Tingkat Kredibilitas Hoax News/ Fake News Pada Sosial Media Di Indonesia Berbasis Android (Studi Kasus : Kantor Tribun Medan),” *Media Online*, vol. 1, no. 1, 2020, [Online]. Available: <https://djournal.com/resolusi>
- [14] C. S. Sriyono and E. B. Setiawan, “Pendeteksian Berita Hoax Menggunakan Naive Bayes Multinomial Pada Twitter dengan Fitur Pembobotan TF-IDF.”
- [15] N. Agustina, A. Adrian, and M. Hermawati, “Implementasi Algoritma Naive Bayes Classifier untuk Mendeteksi Berita Palsu pada Sosial Media,” *Faktor Exacta*, vol. 14, no. 4, p. 206, Jan. 2022, doi: 10.30998/faktorexacta.v14i4.11259.
- [16] T. A. Roshinta, E. Kumala, and I. F. Dinata, “Sistem Deteksi Berita Hoax Berbahasa Indonesia Bidang Kesehatan,” *remik*, vol. 7, no. 2, pp. 1167–1173, Apr. 2023, doi: 10.33395/remik.v7i2.12369.
- [17] L. Asa Akhrani, I. Herani, I. Asqori Pohan, and M. Afif Alhad, “Kekacauan Pemilu 2019: Fenomena Firehose of falsehood dalam Relasi Sikap terhadap HOAX dan Kepercayaan Masyarakat Terhadap Komisi Pemilihan Umum,” *Jurnal Transformative*, vol. 6, no. 1, pp. 1–27, Mar. 2020, doi: 10.21776/ub.transformative.2020.006.01.1.
- [18] T. Ridwansyah, “KLIK: Kajian Ilmiah Informatika dan Komputer Implementasi Text Mining Terhadap Analisis Sentimen Masyarakat Dunia Di Twitter Terhadap Kota Medan Menggunakan K-Fold Cross Validation Dan Naive Bayes Classifier,” *Media Online*, vol.

- 2, no. 5, pp. 178–185, 2022, [Online]. Available: <https://djournals.com/klik>
- [19] L. Xiang, “Application of an Improved TF-IDF Method in Literary Text Classification,” *Advances in Multimedia*, vol. 2022, 2022, doi: 10.1155/2022/9285324.
- [20] T. Mikolov, K. Chen, G. Corrado, and J. Dean, “Efficient Estimation of Word Representations in Vector Space,” Jan. 2013, [Online]. Available: <http://arxiv.org/abs/1301.3781>
- [21] J. Devlin, M.-W. Chang, K. Lee, K. T. Google, and A. I. Language, “BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding.” [Online]. Available: <https://github.com/tensorflow/tensor2tensor>
- [22] F. Feng, Y. Yang, D. Cer, N. Arivazhagan, and W. Wang, “Language-agnostic BERT Sentence Embedding,” Jul. 2020, [Online]. Available: <http://arxiv.org/abs/2007.01852>
- [23] V. Mullachery, A. Khera, and A. Husain, “Bayesian Neural Networks.” [Online]. Available: <https://www.kaggle.com/dgaw>
- [24] R. M. Neal, “Bayesian Training of Backpropagation Networks by the Hybrid Monte Carlo Method,” 1992.
- [25] G. E. Hinton, N. Srivastava, A. Krizhevsky, I. Sutskever, and R. R. Salakhutdinov, “Improving neural networks by preventing co-adaptation of feature detectors,” Jul. 2012, [Online]. Available: <http://arxiv.org/abs/1207.0580>
- [26] M. Heydarian, T. E. Doyle, and R. Samavi, “MLCM: Multi-Label Confusion Matrix,” *IEEE Access*, vol. 10, pp. 19083–19095, 2022, doi: 10.1109/ACCESS.2022.3151048.
- [27] A. Vehtari, A. Gelman, and J. Gabry, “Practical Bayesian model evaluation using leave-one-out cross-validation and WAIC,” Jul. 2015, doi: 10.1007/s11222-016-9696-4.

