

**Fitur N-Gram Untuk Perbandingan Metode Machine Learning Pada
Sentimen Judul Berita Keuangan**



TIM PENGUSUL

Arif Mudi Priyatno, S.T., M.Kom
Fahmi Iqbal Firmananda, S.Kom., M.M

UNIVERSITAS PAHLAWAN
JULI 2022

HALAMAN PENGESAHAN

Judul Penelitian : Fitur N-Gram Untuk Perbandingan Metode Machine Learning Pada Sentimen Judul Berita Keuangan

Kode>Nama Rumpun Ilmu : 458 / Teknik Informatika

Jenis Penelitian : Penelitian Dasar

Bidang Fokus : Data Mining

Ketua :

a. Nama Lengkap : Arif Mudi Priyatno

b. NIDN : 1023059501

c. Jabatan Fungsional : -

d. Program Studi : S1 Bisnis Digital

e. Mata Kuliah yang diampu : Dasar Pemrograman, Database

f. Nomor HP dan email : 0823-9044-9323 / arifmudi@universitaspahlawan.ac.id

Anggota Peneliti : 1. Fahmi Iqbal Firmananda / 1014119401 / S1 Bisnis Digital
: 2. -

Peneliti (MITRA)

a. Nama Lengkap : -

b. NIDN : -

c. Instansi : -

d. Jabatan : -

Biaya Penelitian Keseluruhan : Rp 10.421.000,-

Biaya Penelitian :

- dana mandiri PT : Rp 10.421.000,-

- dana Internal PT : Rp -

- dana institusi mitra : Rp -

Mengetahui,
Ketua Prodi



(Rizqon Jamil Farhas, S.E, M.Si)
NIDN: 1015049203

Bangkinang, 31 Juli 2022
Ketua Pelaksana



(Arif Mudi Priyatno, S.T., M.Kom)
NIDN: 1023059501

Menyetujui,
Ketua LPPM



(Dr. Musnar Indra Daulay, M.Pd)

Nip-TT: 96542108

IDENTITAS DAN URAIAN UMUM

1. Judul Penelitian : Fitur N-Gram Untuk Perbandingan Metode Machine Learning Pada Sentimen Judul Berita Keuangan

2. Tim Peneliti

No	Nama	Jabatan	Bidang Keahlian	Instansi Asal	Alokasi Waktu (jam/minggu)
1	Arif Mudi Priyatno	Ketua	Data Mining	Universitas Pahlawan	3 jam / minggu
2	Fahmi Iqbal Firmananda	Anggota 1	Manajemen	Universitas Pahlawan	3 jam / minggu

3. Objek Penelitian: Perbandingan *Machine learning*, N-Gram, Sentimen Berita

4. Masa Pelaksanaan

Mulai : bulan: September tahun: 2021
Berakhir : bulan: Juli tahun: 2022

5. Usulan Biaya : Rp 10.421.000,-

6. Lokasi Penelitian : Lab Universitas Pahlawan

7. Instansi Mitra (uraikan apa kontribusinya dalam kontrak kerjasama).....

8. Temuan yang ditargetkan : N-Gram untuk perbandingan Machine learning pada sentiment berita

9. Kontribusi mendasar pada suatu bidang ilmu : perbandingan machine learning untuk peramalan

10. Jurnal ilmiah yang menjadi sasaran (tuliskan nama terbitan berkala ilmiah internasional bereputasi, nasional terakreditasi, atau nasional tidak terakreditasi dan tahun rencana publikasi) Journal of Artificial Intelligence and Digital Business (RIGGS).

11. Rencana luaran yang ditargetkan: Jurnal Non-Nasional

DAFTAR ISI

HALAMAN PENGESAHAN.....	III
IDENTITAS DAN URAIAN UMUM.....	IV
DAFTAR ISI	V
DAFTAR LAMPIRAN	IX
ABSTRAK	IX
BAB 1. PENDAHULUAN.....	1
A. LATAR BELAKANG	1
B. TUJUAN PENELITIAN	2
C. URGENSI PENELITIAN.....	2
D. POTENSI LUARAN.....	3
BAB 2. TINJAUAN PUSTAKA.....	4
A. MACHINE LEARNING	4
B. N-GRAM	6
BAB 3. METODE PENELITIAN	8
A. DATA.....	8
B. PRE-PROCESSING	8
C. TERM WEIGHTING	9
D. KLASIFIKASI SENTIMEN	11
E. PENGUKURAN PERFORMA	11
F. DOKUMENTASI	11
BAB 4. BIAYA DAN JADWAL PENELITIAN	12
A. ANGGARAN BIAYA	12
B. JADWAL PENELITIAN.....	12
BAB 5. HASIL DAN PEMBAHASAN	14

BAB 6. KESIMPULAN	17
DAFTAR PUSTAKA	18

DAFTAR LAMPIRAN

Lampiran 1 Justifikasi Anggaran Penelitian	22
Lampiran 2 Susunan organisasi tim peneliti	23
Lampiran 3 Biodata ketua dan anggota tim pengusul.....	24
Lampiran 4 Surat pernyataan ketua peneliti.....	28

ABSTRAK

Sentiment analisis saat ini banyak digunakan pada aplikasi natural language processing ataupun information retrieval. Analisa sentiment analisis bisa memberikan informasi terkait headline berita keuangan yang beredar dan memberikan masukan terhadap perusahaan. Sentimen positif akan memberikan dampak yang baik pula terhadap perkembangan perusahaan, akan tetapi sentiment negative akan membuat reputasi perusahaan berkurang, hal ini akan berpengaruh terhadap perkembangan perusahaan. Penelitian ini melakukan perbandingan metode machine learning pada headline berita keuangan dengan ekstraksi fitur n-gram. Tujuan penelitian ini untuk mendapatkan metode terbaik dalam mengklasifikasikan sentiment headline berita keuangan perusahaan. Metode machine learning yang dibandingkan yaitu Multinomial Naïve Bayes, Logistic Regression, Support Vector Machine, multi-layer perceptron (MLP), Stochastic Gradient Descent, dan Decision Trees.

BAB 1.

PENDAHULUAN

A. Latar Belakang

Teknologi Informasi dan komunikasi pada saat ini berkembang dengan sangat pesat. Berbagai hal telah bisa dilakukan dengan teknologi terutama dengan adanya kecerdasan buatan. Salah satu contoh yang bisa dilakukan yaitu melakukan analisis sentiment terhadap judul berita. Analisa sentiment pada text dapat dilakukan menggunakan text mining. Sentiment analisis adalah proses pengkategorian pendapat / topik pada text, proses ini untuk menentukan apakah pendapat / topik bernilai positif, negative, atau netral [1]. Penelitian sentimen analisis telah banyak dilakukan diberbagai sektor, seperti keuangan [2], politik [3], perusahaan [4], dan lain sebagainya. Analisa sentiment terhadap berita keuangan dan perusahaan akan memberikan efek terhadap kredibilitas suatu perusahaan. Sentimen berpengaruh besar pada perusahaan yang bersifat terbuka atau terdaftar pada bursa saham.

Penelitian ini melakukan perbandingan metode machine learning untuk sentiment analisis dengan fitur n-gram pada kasus berita keuangan dan headline perusahaan. Data digunakan dari penelitian [5] dengan jumlah 4846 judul. Jumlah sentimen netral 2879, jumlah sentimen positif 1363, dan jumlah sentiment negative 604. Data dilakukan preprocessing dengan menghapus hal yang tidak dibutuhkan. Hasil preproceasing diberikan pembobotan dengan metode n-gram. Fitur hasil n-gram dilakukan proses sentimen analyst dengan menggunakan beberapa metode machine learning, diantaranya Multinomial Naïve Bayes, Logistic Regression, Support Vector Machine, multi-layer perceptron (MLP), Stochastic Gradient Descent, dan Decision Trees. Performa sentimen analisis di hitung menggunakan presisi, recall, dan f1-measure.

Penelitian terkait yaitu [6] melakukan analisis sentiment pada artikel berita berdasarkan lexicon. Data yang digunakan yaitu BBC News tahun 2004 hingga 2005 (<http://mlg.ucd.ie/datasets/bbc.html>). Penelitian tersebut mendapatkan kesimpulan bahwa artikel dengan topik bisnis dan olahraga memiliki sentiment positif, sedangkan topik terkait hiburan dan olahraga memiliki sentiment negative.

Penelitian [7] melakukan investigasi terhadap metrik pengukuran sentiment analisis pada artikel berita. Metode pembobotan yang digunakan yaitu *term frequency-inverse document frequency* (TF-IDF). Metode klasifikasi menggunakan Gaussian Naive Bayes classifiers dan *linear support vector machine* (linear SVM). Hasil terbaik didapatkan menggunakan metode

linear SVM sebesar 62%, sedangkan Gaussian naïve Bayes mendapatkan 61%.

Penelitian [8] melakukan analisis sentiment pada berita keuangan VADER (Valence Aware Dictionary and Sentiment Reasoner). Penelitian bertujuan menunjukkan pengaruh sentiment berita keuangan terhadap perubahan harga pasar saham. Penelitian [9] melakukan perbandingan metode analisis pada headline berita di belanda. Perbandingan tersebut yaitu anotasi manual, crowd coding, numerous dictionaries, machine learning, dan deep learning. Kesimpulan didapatkan bahwa hasil terbaik yaitu dilakukan oleh manusia atau crowd coding, numerous dictionaries menunjukkan bahwa tingkat validitas bisa diterima, dan machine learning serta deep learning secara substansi lebih baik dari pada numerous dictionaries tetapi masih jauh dari kinerja manusia.

B. Tujuan Penelitian

Tujuan penelitian Fitur N-Gram Untuk Perbandingan Metode Machine Learning Pada Sentimen Judul Berita Keuangan yaitu:

1. Mengidentifikasi metode machine learning yang paling efektif dalam mengklasifikasikan sentimen pada headline berita keuangan menggunakan fitur n-gram.
2. Mengevaluasi kinerja berbagai metode machine learning termasuk Multinomial Naïve Bayes, Logistic Regression, Support Vector Machine, Multi-Layer Perceptron (MLP), Stochastic Gradient Descent, dan Decision Trees dalam analisis sentimen.
3. Menentukan efektivitas fitur n-gram dalam meningkatkan akurasi, presisi, dan recall dari model machine learning untuk klasifikasi sentimen.
4. Membandingkan hasil evaluasi performa dari metode-metode tersebut dengan menggunakan metrik f1-measure, presisi, dan recall.

C. Urgensi Penelitian

Urgensi penelitian Fitur N-Gram Untuk Perbandingan Metode Machine Learning Pada Sentimen Judul Berita Keuangan yaitu:

1. Peningkatan Keputusan Bisnis: Analisis sentimen yang akurat pada berita keuangan dapat memberikan informasi berharga kepada perusahaan mengenai persepsi publik. Hal ini memungkinkan perusahaan untuk mengambil keputusan strategis yang lebih

baik berdasarkan sentimen pasar dan opini publik.

2. **Optimalisasi Metode Machine Learning:** Ada banyak metode machine learning yang dapat digunakan untuk analisis sentimen, namun belum ada konsensus mengenai metode yang paling efektif. Penelitian ini berusaha untuk mengisi celah tersebut dengan membandingkan beberapa metode dan menentukan metode yang paling optimal.
3. **Pemanfaatan Fitur N-Gram:** Fitur n-gram memiliki potensi untuk meningkatkan kinerja model machine learning dalam analisis sentimen. Penelitian ini menguji efektivitas fitur n-gram dan memberikan wawasan tentang bagaimana fitur ini dapat digunakan secara efektif dalam aplikasi analisis sentimen.
4. **Relevansi pada Industri Keuangan:** Dalam industri keuangan, berita dan sentimen publik dapat mempengaruhi harga saham dan keputusan investasi. Oleh karena itu, kemampuan untuk secara otomatis menganalisis sentimen dari headline berita keuangan memiliki nilai praktis yang tinggi bagi para pelaku industri.

D. Potensi Luaran

Potensi Luaran penelitian Fitur N-Gram Untuk Perbandingan Metode Machine Learning Pada Sentimen Judul Berita Keuangan yaitu:

1. **Publikasi ilmiah:** Kontribusi pada literatur ilmiah di bidang analisis sentimen, natural language processing, dan machine learning.
2. **Model dan Algoritma yang Dioptimalkan:** Menghasilkan model machine learning yang paling efektif untuk analisis sentimen pada headline berita keuangan. Menyediakan algoritma yang dioptimalkan dengan fitur n-gram yang dapat digunakan oleh peneliti dan praktisi dalam aplikasi analisis sentimen lainnya.
3. **Aplikasi Praktis di Industri Keuangan:** Solusi analisis sentimen yang dapat diimplementasikan oleh perusahaan keuangan untuk memonitor dan menganalisis persepsi publik terhadap berita keuangan. Alat bantu untuk perusahaan dalam pengambilan keputusan strategis berdasarkan analisis sentimen dari berita keuangan.

BAB 2.

TINJAUAN PUSTAKA

Bab kajian pustaka membahas tentang referensi terkait pada kegiatan penelitian ini. Kajian pustaka pada bab ini diantaranya machine learning; n-gram.

A. Machine Learning

Machine Learning (ML) adalah cabang dari kecerdasan buatan (Artificial Intelligence, AI) yang berfokus pada pengembangan algoritma dan teknik yang memungkinkan komputer untuk belajar dari data dan membuat keputusan berdasarkan data tersebut. Salah satu definisi klasik dari ML adalah kemampuan komputer untuk belajar dari pengalaman tanpa diprogram secara eksplisit. Teknik ML telah digunakan dalam berbagai bidang, termasuk pengenalan pola, prediksi, dan pengambilan keputusan.

ML secara umum dibagi menjadi tiga kategori utama: pembelajaran terawasi (supervised learning), pembelajaran tak terawasi (unsupervised learning), dan pembelajaran penguatan (reinforcement learning). Pembelajaran terawasi melibatkan pelatihan model pada dataset yang telah diberi label, di mana setiap input memiliki output yang sesuai. Contoh populer dari pembelajaran terawasi termasuk regresi linear, regresi logistik, dan Support Vector Machines (SVM). Dalam pembelajaran tak terawasi, model dilatih pada data yang tidak diberi label, dengan tujuan menemukan struktur atau pola tersembunyi dalam data. Contoh dari pembelajaran tak terawasi adalah clustering dan asosiasi. Pembelajaran penguatan melibatkan pelatihan agen untuk membuat keputusan yang mengoptimalkan hasil tertentu berdasarkan umpan balik dari lingkungannya.

Ada banyak algoritma dan metode yang digunakan dalam ML, masing-masing dengan kelebihan dan kekurangan tersendiri. Beberapa algoritma yang sering digunakan antara lain:

- Regresi Linear dan Regresi Logistik: Algoritma ini digunakan untuk prediksi dan klasifikasi, di mana hubungan antara variabel independen dan dependen diwakili oleh persamaan linear.
- Support Vector Machines (SVM): SVM adalah algoritma yang kuat untuk klasifikasi dan regresi, yang bekerja dengan menemukan hyperplane optimal yang memisahkan data dalam ruang fitur.
- K-Nearest Neighbors (KNN): KNN adalah algoritma yang sederhana namun

efektif untuk klasifikasi dan regresi, yang bekerja dengan menghitung jarak antara titik data baru dan tetangga terdekatnya dalam dataset.

- **Neural Networks dan Deep Learning:** Neural Networks adalah model yang terinspirasi oleh struktur otak manusia, yang terdiri dari lapisan-lapisan neuron yang saling terhubung. Deep Learning, yang merupakan subbidang dari Neural Networks, menggunakan jaringan dengan banyak lapisan untuk menangkap representasi data yang kompleks.
- **Random Forest dan Ensemble Methods:** Random Forest adalah algoritma ensemble yang menggabungkan prediksi dari beberapa pohon keputusan untuk meningkatkan akurasi dan mengurangi overfitting.

ML telah diterapkan dalam berbagai domain untuk memecahkan berbagai masalah. Beberapa contoh aplikasi ML antara lain:

- **Pengenalan Wajah dan Objek:** ML digunakan untuk mengenali wajah dan objek dalam gambar dan video, yang memiliki aplikasi dalam keamanan, media sosial, dan kesehatan.
- **Deteksi Penipuan:** Algoritma ML digunakan oleh lembaga keuangan untuk mendeteksi transaksi penipuan dengan menganalisis pola dalam data transaksi.
- **Prediksi Harga Saham:** ML digunakan untuk menganalisis data pasar saham dan membuat prediksi tentang pergerakan harga saham di masa depan, seperti yang dibahas dalam penelitian ini.
- **Analisis Sentimen:** Algoritma ML digunakan untuk menganalisis teks dari media sosial, ulasan produk, dan berita untuk menentukan sentimen publik terhadap topik tertentu.
- **Diagnosis Medis:** ML digunakan dalam bidang medis untuk mendiagnosis penyakit berdasarkan data pasien, seperti gambar medis dan catatan kesehatan elektronik.

Meskipun ML memiliki potensi besar, ada beberapa tantangan yang perlu diatasi. Salah satunya adalah masalah bias dalam data, yang dapat menyebabkan model ML menghasilkan prediksi yang tidak adil atau diskriminatif. Selain itu, interpretabilitas model ML juga menjadi

isu penting, terutama dalam aplikasi yang membutuhkan penjelasan yang jelas tentang bagaimana keputusan dibuat. Masa depan ML terlihat cerah dengan kemajuan dalam teknologi komputasi, ketersediaan data yang lebih besar, dan pengembangan algoritma yang lebih canggih. Penelitian dalam bidang ML terus berkembang, dengan fokus pada peningkatan akurasi, efisiensi, dan interpretabilitas model.

B. N-Gram

N-gram adalah konsep dasar dalam pemrosesan bahasa alami (Natural Language Processing, NLP) yang merujuk pada serangkaian N kata atau karakter yang berurutan dalam teks. N-gram digunakan untuk mengidentifikasi pola dan struktur dalam data teks, yang kemudian dapat digunakan untuk berbagai aplikasi seperti analisis teks, pengenalan suara, dan mesin terjemahan. Misalnya, dalam kalimat "saya belajar machine learning", bigram (2-gram) akan menghasilkan pasangan kata "saya belajar", "belajar machine", dan "machine learning".

N-gram diklasifikasikan berdasarkan jumlah elemen dalam kelompok: Unigram (1-gram): Mengandung satu kata atau karakter. Contoh: "saya", "belajar", "machine", "learning". Bigram (2-gram): Mengandung dua kata atau karakter berurutan. Contoh: "saya belajar", "belajar machine". Trigram (3-gram): Mengandung tiga kata atau karakter berurutan. Contoh: "saya belajar machine", "belajar machine learning". n-gram (n-gram): Mengandung n kata atau karakter berurutan. Semakin besar nilai N, semakin panjang konteks yang diambil dalam analisis, namun juga membutuhkan lebih banyak data dan komputasi untuk menghasilkan model yang akurat.

N-gram memiliki berbagai aplikasi dalam NLP dan bidang terkait. Beberapa penggunaan utama termasuk: Pengenalan Pola Teks: N-gram digunakan untuk mengenali pola dalam teks, yang dapat membantu dalam analisis sentimen, klasifikasi teks, dan ekstraksi informasi. Prediksi Kata: Dalam aplikasi seperti auto-complete dan koreksi otomatis, N-gram digunakan untuk memprediksi kata berikutnya berdasarkan kata-kata sebelumnya. Mesin Terjemahan: Sistem terjemahan menggunakan N-gram untuk memahami konteks dan menghasilkan terjemahan yang lebih akurat. Analisis Sentimen: N-gram membantu dalam menentukan sentimen dari teks dengan mengidentifikasi frase atau kata-kata yang sering muncul dalam konteks positif atau negatif. Pemodelan Bahasa: N-gram digunakan dalam pemodelan bahasa untuk memperkirakan probabilitas suatu rangkaian kata.

N-gram memiliki beberapa kelebihan yang membuatnya populer dalam analisis teks:

Kesederhanaan dan Efektivitas: N-gram mudah diimplementasikan dan sering kali memberikan hasil yang efektif untuk tugas-tugas dasar dalam NLP. Identifikasi Pola Lokal: N-gram dapat menangkap pola lokal dalam teks, yang berguna untuk banyak aplikasi praktis.

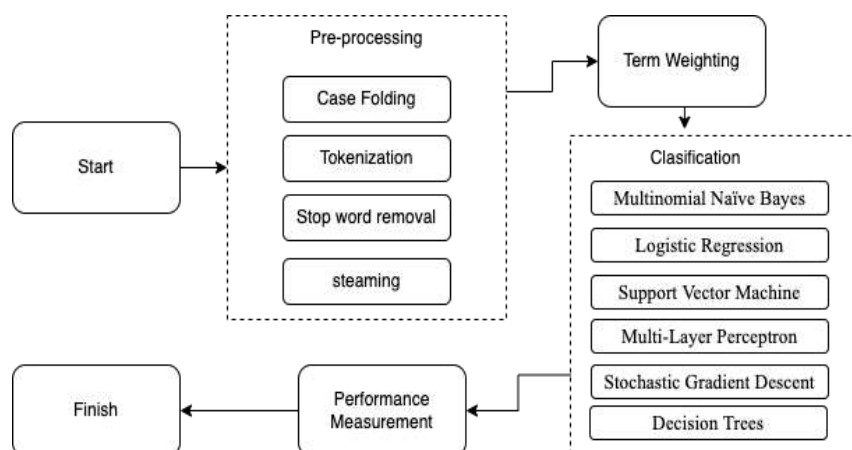
Namun, N-gram juga memiliki keterbatasan: Ketergantungan pada Data Latih: Model N-gram sangat bergantung pada data latih. Semakin besar nilai N, semakin banyak data yang dibutuhkan untuk menghasilkan model yang andal. Masalah Sparsity: Untuk nilai N yang besar, jumlah kombinasi kata meningkat secara eksponensial, menyebabkan masalah sparsity (kelangkaan data) yang dapat mengurangi akurasi model. Tidak Menangkap Konteks Jangka Panjang: N-gram hanya menangkap konteks lokal dan mungkin tidak efektif untuk memahami konteks jangka panjang dalam teks.

Penelitian terbaru dalam penggunaan N-gram telah mencoba untuk mengatasi beberapa keterbatasan tersebut dengan menggabungkan teknik lain seperti pembelajaran mendalam (deep learning) dan representasi kata berbasis vektor (word embeddings). Misalnya, peneliti telah mengembangkan model hibrid yang menggabungkan N-gram dengan model neural networks untuk meningkatkan akurasi prediksi teks dan analisis sentiment. Selain itu, penelitian dalam pengurangan sparsity telah mengarah pada pengembangan teknik-teknik seperti smoothing dan back-off yang meningkatkan kinerja model N-gram.

Dalam analisis sentimen, N-gram digunakan untuk menangkap frase atau kata-kata yang menunjukkan sentimen positif atau negatif. Misalnya, dalam analisis sentimen media sosial, bigram dan trigram dapat membantu mengidentifikasi ungkapan seperti "sangat bagus" atau "sangat buruk", yang memberikan indikasi lebih kuat tentang sentimen daripada unigram. Penggunaan N-gram dalam analisis sentimen juga dapat digabungkan dengan teknik pembelajaran mesin lainnya untuk meningkatkan akurasi klasifikasi sentimen.

BAB 3. METODE PENELITIAN

Secara umum bab ini membahas tentang alur metodologi penelitian. Sub bab berikut ini membahas tentang tahapan-tahapan metodologi penelitian secara lebih rinci.



Gambar 1 Metodologi Penelitian

Proses penelitian ini diantaranya yaitu pengambilan data, pre-processing, pembobotan menggunakan n-gram, klasifikasi sentiment, dan pengukuran performa. Gambar 1 menunjukkan tentang tahapan penelitian yang dilakukan.

A. Data

Data merupakan teks berita keuangan dan siaran pers perusahaan. Penandaan sentiment dilakukan oleh sekelompok 16 annotator dengan latar belakang pendidikan bisnis yang memadai. Data menggunakan penelitian [5] dengan jumlah total 4846 judul berita. Jumlah sentimen netral 2879, jumlah sentimen positif 1363, dan jumlah sentiment negative 604.

B. Pre-processing

Tahapan ini dilakukan untuk mempersiapkan data dapat digunakan pada tahapan-tahapan selanjutnya. Pre-processing membersihkan data text dari kata (token) yang umum dan tidak memiliki makna yang diperlukan, hal ini dilakukan untuk mengurangi noise pada data text. Gambar 1 menunjukkan proses yang terdapat pada pre-processing yaitu case folding, tokenization, stopword removal dan steaming. Case Folding merupakan tahap awal yang dilakukan untuk menyeragamkan huruf dengan cara mengubah semua huruf besar menjadi huruf kecil. Contohnya “According to Gran, the company has no plans to move all production

to Russia, although that is where the company is growing.” menjadi “according to gran, the company has no plans to move all production to russia, although that is where the company is growing”. Tokenization merupakan tahapan memotong kalimat menjadi token (term) yang lebih kecil. Contohnya “according to gran, the company has no plans to move all production to russia, although that is where the company is growing.” menjadi [“according“, ”to“, “gran”, “the”, “company“, “has”, “no”, “plans”, “to”, “move”, “all”, “production”, “to”, “russia”, “although”, “that”, “is”, “where”, “company”, “is”, “growing”]. Stop word removal merupakan tahapan menghapus term yang bersifat umum dan tidak memiliki pengaruh. Contoh term yang dihapus yaitu 'its', 'again', 'for', 'myself', 'his', dan lain sebagainya [10]. Steaming adalah proses pengembalian term menjadi kata dasarnya [11]. Contohnya yaitu “plans” menjadi “plan”, “growing” menjadi “grow”.

C. Term Weighting

N-gram adalah pemotongan kalimat atau string menjadi leboh kecil sesuai dengan N karakter yang ditentukan. Gram didefinisikan sebagai sub urutan dari N karakter yang dikerjakan [12]. Metode n-gram digunakan dengan tujuan memberikan makna dari urutan kata ataupun karakter pada suatu kalimat . Metode n-gram digunakan untuk mengambil potongan karakter huruf sebanyak n dari kalimat yang secara kontinuitas. kontinuitas dimaksud adalah dari awal hingga akhir dokumen.

Tabel 1. Contoh n-gram

N-gram	Result
Unigram	: ‘a’, ‘c’, ‘e’, ‘o’, ‘r’, ‘d’, ‘g’, ‘r’, ‘a’, ‘n’
Bigram	: ‘_a’, ‘ac’, ‘cc’, ‘co’, ‘or’, ‘rd’, ‘dg’, ‘gr’, ‘ra’, ‘an’, ‘n_’
Trigram	: ‘_ac’, ‘acc’, ‘cco’, ‘cor’, ‘ord’, ‘rdg’, ‘dgr’, ‘gra’, ‘ran’, ‘an_’, ‘n__’
Quadgram	: ‘_acc’, ‘acco’, ‘ccor’, ‘cord’, ‘ordg’, ‘rdgr’, ‘dgra’, ‘gran’, ‘ran_’, ‘an__’, ‘n___’

Tabel 2. Contoh n-gram pada penelitian

N-gram	Result
Unigram	: ‘according’, ‘gran’, ‘company’, ‘plan’, ‘move’, ‘production’, ‘russia’, ‘although’, ‘company’, ‘grow’
Bigram	: ‘according gran’, ‘gran company’, ‘company plan’, ‘plan move’, ‘move production’, ‘production russia’, ‘russia although’, ‘although company’, ‘company grow’
Trigram	: ‘according gran company’, ‘gran company plan’, ‘company plan move’, ‘plan move production’, ‘move production russia’, ‘production russia although’, ‘russia although company’, ‘although company grow’
Quadgram	: ‘according gran company plan’, ‘gran company plan move’, ‘company plan move production’, ‘plan move production russia’, ‘move production Russia although’, ‘production russia although company’, ‘russia although company grow’

N-Gram diklasifikasikan berdasarkan n karakter. Secara umum, n-gram dilakukan dengan memberikan tambahan blank di awal dan di akhir [13]. Contoh, kalimat “accord gran” dilakukan proses n-gram, blank dimaksud disimbolkan dengan “_” menghasilkan n-gram pada Table 1.

N-Gram memiliki keuntungan yaitu matching setiap kata dapat digunakan walaupun ada interpetasi kode tertentu seperti kode pos. Pada aplikasi natural language processing, angka dan kode tertentu dianggap sebagai noise. Hal ini dianggap keuntungan karena dekomposisi bagian-bagian kecil yang tetap memiliki makna. Hal ini secara tekstual tidak memberikan pengaruh.

Pada penelitian ini n-gram digunakan tidak dipecah menjadi karakter-karakter kecil, tetapi n-gram memecah kalimat menjadi kata perkata. Contoh kalimat “according gran company plan move production russia although company grow” dilakukan proses n-gram dan hasilnya dapat dilihat pada Table 2. Term weighting pada penelitian ini menggunakan

banyaknya gram yang muncul atau lebih dikenal dengan term frequency (TF).

D. Klasifikasi Sentimen

Penelitian ini membandingkan beberapa metode klasifikasi. Metode yang digunakan yaitu *multinomial naïve bayes* [14][15], *logistic regression* [16][17], *support vector machine* [18][19][20], *multi layer perceptron* [21][22][23][24], *stochastic gradient descent* [25][26], dan *decition tree* [27][28].

E. Pengukuran Performa

Evaluasi performa pada penelitian ini menggunakan confusion matrix. Evaluasi performa berdasarkan confusion matrix diantaranya yaitu presisi, recall, dan f1-measure[11][29].

$$recall = \frac{TP}{TP+FN} \quad (1)$$

$$Precision = \frac{TP}{TP+FP} \quad (2)$$

$$F1 - measure = \frac{2 \times precision \times recall}{precision+recall} \quad (3)$$

F. Dokumentasi

Tahap terakhir dari penelitian adalah membuat dokumentasi dan penyusunan laporan akhir penelitian. Tahap ini bertujuan untuk menyimpan dokumentasi keseluruhan proses dan hasil penelitian yang telah dilakukan.

BAB 4.

BIAYA DAN JADWAL PENELITIAN

Bagian ini membahas tentang anggaran biaya penelitian dan jadwal penelitian yang akan dilaksanakan. Anggaran akan dibahas secara garis besar dan diperjelas pada lampiran. Jadwal penelitian akan membahas waktu dan proses pelaksanaan yang dilakukan.

A. Anggaran Biaya

Total Biaya penelitian tentang *Fitur N-Gram Untuk Perbandingan Metode Machine Learning Pada Sentimen Judul Berita Keuangan* adalah Rp. 10.421.000,- (Sepuluh Juta Empat Ratus dua puluh satu ribu rupiah). Tabel 4.1 merupakan garis besar penggunaan anggaran pada penelitian ini. Lampiran 1 menjelaskan secara detail pengeluaran penelitian yang dilakukan.

Table 4.1 Anggaran Biaya

No	Garis Besar Pengeluaran	Besaran
1	Honorarium	Rp. 3.496.000,-
2	Bahan Penelitian	Rp. 2.495.000,-
3	Pengumpulan Data	Rp. 880.000,-
4	Pelaporan, Luaran Penelitian	Rp. 3.550.000,-
Total		Rp. 10.421.000,-

B. Jadwal Penelitian

Jadwal pelaksanaan penelitian ini dari bulan januari hingga desember tahun 2023. Tabel 4.2 merupakan rincian jadwal kegiatan penelitian ini.

Table 4.2 Jadwal Penelitian

No	Kegiatan	Bulan												
		1	2	3	4	5	6	7	8	9	10	11	12	
1	Studi Literature													
2	Pengumpulan Data													

3	Analisa dan Perancangan												
4	Implementasi Code												
5	Pengujian dan Evaluasi												
6	Penulisan Laporan												
7	Publikasi												

BAB 5.
HASIL DAN PEMBAHASAN

Data digunakan pada penelitian ini berjumlah 4846 judul berita. Jumlah sentimen netral 2879, jumlah sentimen positif 1363, dan jumlah sentiment negative 604. Data dilakukan pembagian data latih dan data uji dengan persentase 70:30, 80:20, dan 90:10. Table 3 menunjukkan hasil dari pembagian data latih dan data uji.

Table 3 Data Latih dan Data Uji

No	Persentase	Data Latih	Data Uji	Total
1	70 %	3392	1454	4868
2	80 %	3876	970	4868
3	90 %	4361	485	4868

Data dilakukan pre-processing untuk membersihkan data text. Proses pre-processing diantaranya case folding, tokenization, stop word removal, dan stemming. Table 4 menunjukkan hasil dari pre-processing.

Table 4 Result Pre-processing

No	Input	Result Pre-processing
1	According to Gran, the company has no plans to move all production to Russia, although that is where the company is growing.	according gran company plan move production russia although company grow
2	Technopolis plans to develop in stages an area of no less than 100,000 square meters in order to host companies working in computer technologies and telecommunications, the statement said.	technopolis plan develop stage area less 100,000 square meter order host company work computer technology telecommunication statement say

3	The international electronic industry company Elcoteq has laid off tens of employees from its Tallinn facility; contrary to earlier layoffs the company contracted the ranks of its office workers, the daily Postimees reported.	international electronic industry company elcoteq laid ten employee tallinn facility contrary earlier layoff company contract rank office worker daily postimees report
4	With the new production plant the company would increase its capacity to meet the expected increase in demand and would improve the use of raw materials and therefore increase the production profitability.	new production plant company would increase capacity meet expect increase demand would improve use raw material therefore increase production profitability

Table 5 Result Percentage 70:30

Method	F-1 Measure	Precision	Recall
Multinomial NaiveBayes	69,79	69,61	70,56
Logistic Regression	73,93	73,27	73,93
Support Vector Machine	66,65	70,02	70,56
Multi-Layer perceptron	70,27	70,04	70,63
Stochastic Gradient Descent	68,89	68,65	69,25
Decision Trees	67,99	68,22	67,81

Table 6 Result Percentage 80:20

Method	F-1 Measure	Precision	Recall
Multinomial NaiveBayes	69,73	69,85	71,03
Logistic Regression	73,94	73,94	74,63
Support Vector Machine	70,07	75,29	73,60
Multi-Layer perceptron	69,32	69,28	69,38
Stochastic Gradient Descent	70,27	70,17	70,41

Decision Trees	67,27	67,03	67,62
----------------	-------	-------	-------

Table 7 Result Percentage 90:10

Method	F-1 Measure	Precision	Recall
Multinomial NaiveBayes	68,65	69,27	69,69
Logistic Regression	73,70	73,81	74,43
Support Vector Machine	68,91	71,71	71,75
Multi-Layer perceptron	72,76	72,79	72,99
Stochastic Gradient Descent	71,16	71,08	71,75
Decision Trees	66,12	65,84	66,60

Hasil pre-processing dilakukan klasifikasi sentiment dengan beberapa metode machine learning. Table 5, 6, dan 7 menunjukkan hasil klasifikasi. Metode multinomial mendapatkan hasil terbaik Ketika pembagian persentase 70:30 yaitu f1-measure 69,79 persen. Metode logistic regression mendapatkan hasil terbaik pada pembagian 80:20 yaitu f1-measure 73,94 persen. Metode support vector machine mendapatkan hasil terbaik pada pembagian 80:20 yaitu 70,07 persen. Metode multi-layer perceptron mendapatkan hasil terbaik pada pembagian 90:10 yaitu 72,76 persen. Hal ini dikarenakan metode multi-layer perceptron membutuhkan data latih yang banyak agar dapat memaksimalkan hidden layer yang dimiliki untuk mengetahui klasifikasinya. Jika jumlah data latih sedikit maka hidden layer multi-layer perceptron tidak berjalan dengan optimal. Hal ini kedepannya jika ingin menggunakan multi-layer perceptron harap memperbanyak data latih, baik dengan menambah manual ataupun dengan metode augmentasi data. Metode stochastic gradient descent mendapatkan hasil terbaik pada pembagian 90:10 yaitu 71,16 persen. Metode decision trees mendapatkan hasil terbaik pada pembagian 70:30 yaitu 67,99 persen. Secara keseluruhan hasil terbaik didapatkan oleh metode logistic regression. Perbedaan hasil yang tidak jauh menunjukkan bahwa proses metode telah sesuai dan tidak adanya underfitting ataupun overfitting. Hasil yang baik ini dapat dioptimalkan lagi dengan menggunakan metode ekstraksi fitur yang lebih baik lagi seperti menggunakan pembobotan fasttext, glove ataupun yang lainnya. Hal ini tentu dengan menambahkan transfer.

BAB 6.

KESIMPULAN

Penelitian ini melakukan perbandingan metode machine learning untuk klasifikasi sentiment dengan ekstraksi fitur n-gram. Data digunakan dengan jumlah total 4868 headline berita. Pembobotan n-gram dan klasifikasi dengan machine learning menunjukkan hasil yang baik. Hasil tertinggi didapatkan yaitu f1-measure, presisi, recall sebesar 73,94; 73,94; dan 74,63. Berdasarkan hasil pengujian yang telah dilakukan, n-gram mampu melakukan ekstraksi fitur dari data yang ada.

DAFTAR PUSTAKA

- [1] S. Kurniawan, W. Gata, D. A. Puspitawati, N. -, M. Tabrani, and K. Novel, “Perbandingan Metode Klasifikasi Analisis Sentimen Tokoh Politik Pada Komentar Media Berita Online,” *J. RESTI (Rekayasa Sist. dan Teknol. Informasi)*, vol. 3, no. 2, pp. 176–183, 2019, doi: 10.29207/resti.v3i2.935.
- [2] K. Mishev, A. Gjorgjevikj, I. Vodenska, L. T. Chitkushev, and D. Trajanov, “Evaluation of Sentiment Analysis in Finance: From Lexicons to Transformers,” *IEEE Access*, vol. 8, pp. 131662–131682, 2020, doi: 10.1109/ACCESS.2020.3009626.
- [3] E. Kušen and M. Strembeck, “Politics, sentiments, and misinformation: An analysis of the Twitter discussion on the 2016 Austrian Presidential Elections,” *Online Soc. Networks Media*, vol. 5, pp. 37–50, 2018, doi: 10.1016/j.osnem.2017.12.002.
- [4] S. M. Shuhidan, S. R. Hamidi, S. Kazemian, S. M. Shuhidan, and M. A. Ismail, “Sentiment Analysis for Financial News Headlines using Machine Learning Algorithm,” in *Advances in Intelligent Systems and Computing*, vol. 739, A. M. Lokman, T. Yamanaka, P. Lévy, K. Chen, and S. Koyama, Eds. Singapore: Springer Singapore, 2018, pp. 64–72. doi: 10.1007/978-981-10-8612-0_8.
- [5] P. Malo, A. Sinha, P. Korhonen, J. Wallenius, and P. Takala, “Good debt or bad debt: Detecting semantic orientations in economic texts,” *J. Assoc. Inf. Sci. Technol.*, vol. 65, no. 4, pp. 782–796, 2014, doi: 10.1002/asi.23062.
- [6] S. Taj, B. B. Shaikh, and A. Fatemah Meghji, “Sentiment Analysis of News Articles: A Lexicon based Approach,” in *2019 2nd International Conference on Computing, Mathematics and Engineering Technologies (iCoMET)*, Jan. 2019, pp. 1–5. doi: 10.1109/ICOMET.2019.8673428.
- [7] P. R. Nagarajan, M. T, S. R. A. M. Hari, K. K, and M. G, “Certain Investigation On Cause Analysis Of Accuracy Metrics In Sentimental Analysis On News Articles,” in *2021 5th International Conference on Electronics, Communication and Aerospace Technology (ICECA)*, Dec. 2021, pp. 1033–1037. doi: 10.1109/ICECA52323.2021.9675846.
- [8] A. Agarwal, “Sentiment Analysis of Financial News,” in *2020 12th International Conference on Computational Intelligence and Communication Networks (CICN)*, Sep.

- 2020, pp. 312–315. doi: 10.1109/CICN49253.2020.9242579.
- [9] W. van Atteveldt, M. A. C. G. van der Velden, and M. Boukes, “The Validity of Sentiment Analysis: Comparing Manual Annotation, Crowd-Coding, Dictionary Approaches, and Machine Learning Algorithms,” *Commun. Methods Meas.*, vol. 15, no. 2, pp. 121–140, Apr. 2021, doi: 10.1080/19312458.2020.1869198.
- [10] J. Nothman, H. Qin, and R. Yurchak, “Stop Word Lists in Free Open-source Software Packages,” in *Proceedings of Workshop for NLP Open Source Software (NLP-OSS)*, 2018, pp. 7–12. doi: 10.18653/v1/W18-2502.
- [11] A. M. Priyatno, M. M. Muttaqi, F. Syuhada, and A. Z. Arifin, “Deteksi bot spammer twitter berbasis time interval entropy dan global vectors for word representations tweet’s hashtag,” *Regist. J. Ilm. Teknol. Sist. Inf.*, vol. 5, no. 1, pp. 37–46, Jan. 2019, doi: 10.26594/register.v5i1.1382.
- [12] R. Aulianita, L. Utami, N. Musyaffa, G. Wijaya, A. Mukhayaroh, and A. Yoraeni, “Sentiment Analysis Review Of Smartphones With Artificial Intelligent Camera Technology Using Naive Bayes and n-gram Character Selection,” *J. Phys. Conf. Ser.*, vol. 1641, no. 1, p. 012076, Nov. 2020, doi: 10.1088/1742-6596/1641/1/012076.
- [13] M. A. P. Subali and C. Fatichah, “Kombinasi Metode Rule-Based dan N-Gram Stemming untuk Mengenali Stemmer Bahasa Bali,” *J. Teknol. Inf. dan Ilmu Komput.*, vol. 6, no. 2, p. 219, 2019, doi: 10.25126/jtiik.2019621105.
- [14] C. D. Manning, P. Raghavan, and H. Schütze, “Text classification and Naive Bayes,” in *Introduction to Information Retrieval*, no. c, Cambridge University Press, 2008, pp. 234–265. doi: 10.1017/CBO9780511809071.014.
- [15] “sklearn.naive_bayes.MultinomialNB — scikit-learn 1.1.1 documentation.” https://scikit-learn.org/stable/modules/generated/sklearn.naive_bayes.MultinomialNB.html (accessed Jul. 30, 2022).
- [16] J. L. Morales and J. Nocedal, “Remark on ‘algorithm 778: L-BFGS-B: Fortran subroutines for large-scale bound constrained optimization,’” *ACM Trans. Math. Softw.*, vol. 38, no. 1, pp. 1–4, Nov. 2011, doi: 10.1145/2049662.2049669.

- [17] “sklearn.linear_model.LogisticRegression — scikit-learn 1.1.1 documentation.” https://scikit-learn.org/stable/modules/generated/sklearn.linear_model.LogisticRegression.html (accessed Jul. 30, 2022).
- [18] C.-C. Chang and C.-J. Lin, “LIBSVM: A library for support vector machines,” *ACM Trans. Intell. Syst. Technol.*, vol. 2, no. 3, pp. 1–27, Apr. 2011, doi: 10.1145/1961189.1961199.
- [19] J. Platt and others, “Probabilistic outputs for support vector machines and comparisons to regularized likelihood methods,” *Adv. large margin Classif.*, vol. 10, no. 3, pp. 61–74, 1999.
- [20] “sklearn.svm.SVC — scikit-learn 1.1.1 documentation.” <https://scikit-learn.org/stable/modules/generated/sklearn.svm.SVC.html> (accessed Jul. 30, 2022).
- [21] X. Glorot and Y. Bengio, “Understanding the difficulty of training deep feedforward neural networks,” *J. Mach. Learn. Res.*, vol. 9, pp. 249–256, 2010.
- [22] K. He, X. Zhang, S. Ren, and J. Sun, “Delving Deep into Rectifiers: Surpassing Human-Level Performance on ImageNet Classification,” in *2015 IEEE International Conference on Computer Vision (ICCV)*, Dec. 2015, vol. 2015 Inter, pp. 1026–1034. doi: 10.1109/ICCV.2015.123.
- [23] D. P. Kingma and J. Ba, “Adam: A Method for Stochastic Optimization,” *3rd Int. Conf. Learn. Represent. ICLR 2015 - Conf. Track Proc.*, pp. 1–15, Dec. 2014, [Online]. Available: <http://arxiv.org/abs/1412.6980>
- [24] “sklearn.neural_network.MLPClassifier — scikit-learn 1.1.1 documentation.” https://scikit-learn.org/stable/modules/generated/sklearn.neural_network.MLPClassifier.html (accessed Jul. 30, 2022).
- [25] B. Zadrozny and C. Elkan, “Transforming classifier scores into accurate multiclass probability estimates,” in *Proceedings of the eighth ACM SIGKDD international conference on Knowledge discovery and data mining - KDD '02*, 2002, p. 694. doi: 10.1145/775047.775151.

- [26] “sklearn.linear_model.SGDClassifier — scikit-learn 1.1.1 documentation.” https://scikit-learn.org/stable/modules/generated/sklearn.linear_model.SGDClassifier.html (accessed Jul. 30, 2022).
- [27] A. Cutler, D. R. Cutler, and J. R. Stevens, “Random forests,” in *Ensemble Machine Learning: Methods and Applications*, Boston, MA: Springer, Boston, MA, 2012, pp. 157–175. doi: 10.1007/9781441993267_5.
- [28] “Random forests - classification description.” https://www.stat.berkeley.edu/~breiman/RandomForests/cc_home.htm (accessed Jul. 30, 2022).
- [29] A. M. Priyatno, F. M. Putra, P. Cholidhazia, and L. Ningsih, “Combination of extraction features based on texture and colour feature for beef and pork classification,” *J. Phys. Conf. Ser.*, vol. 1563, no. 1, p. 012007, Jun. 2020, doi: 10.1088/1742-6596/1563/1/012007.

Lampiran 1 Justifikasi Anggaran Penelitian

No	Uraian	Satuan	Volume	Besaran	Volume x Besaran
1.	Honorarium				
	a. Koordinator Peneliti/Perekayasa	OB	10	Rp. 279.600	2.796.000
	b. Pembantu Peneliti/Perekayasa	OJ	28	Rp. 25.000	700.000
Subtotal Honorarium					3.496.000
2	Bahan Penelitian				
	a. ATK				
	1) Kertas A4	Rim	5	60.000	300.000
	2) Pena	Kotak	1	60.000	60.000
	3) Map	Lusin	1	60.000	60.000
	b. Bahan Penelitian Habis Pakai				
	1) Sewa komputer komputasi	paket	1	3000000	1.575.000
	2) Lisensi Zoom Meeting	paket	1	500.000	500.000
Subtotal Bahan Penelitian					2.495.000
3.	Pengumpulan Data				
	a. Transport	Ok	20	6.500	130.000
	b. Biaya Konsumsi	Ok	20	25.000	750.000
Subtotal biaya pengumpulan data					880.000
4.	Pelaporan, Luaran Penelitian				
	a. Foto Copy Proposal dan Laporan, Kuisisioner dsb	OK	200	Rp. 150	30.000
	b. Jilid Laporan	OK	4	Rp. 5000	20.000
	c. Luaran Penelitian				
	a) Jurnal Nasional Sinta 3	Cond.	paket	3.500.000	3.500.000
	b) Jurnal Nasional Sinta 4				
	c) Jurnal Nasional Sinta 5				
Subtotal biaya Laporan dan Luaran Penelitian					3.550.000
Total					10.421.000

Lampiran 2 Susunan organisasi tim peneliti

NO	Nama	Jabatan	Pembagian Tugas	Instansi Asal	Alokasi Waktu
1	Arif Mudi Priyatno	Ketua	Konseptor Peneliyain	Universitas Pahlawan	6 jam/minggu
2	Fahmi Iqbal Firmananda	Anggota 3	Membantu membangun fitur teknikal indikator	Universitas Pahlawan	6 jam/minggu

Lampiran 3 Biodata ketua dan anggota tim pengusul

A. Biodata Arif Mudi Priyatno Sebagai Ketua Tim Pengusul

A. Identitas Diri

1	Nama Lengkap (dengan gelar)	Arif Mudi Priyatno, S.T., M.Kom
2	Jenis Kelamin	Laki-Laki
3	Jabatan Fungsional	-
4	NIK	14090323059501
5	NIDN	1023059501
6	Tempat dan Tanggal Lahir	Pasir Emas, 23 Mei 1995
7	E-mail	arifmudi@universitaspahlawan.ac.id
8	Nomor Telepon/ HP	082390449323
9	Alamat Kantor	Jl. Tuanku Tambusai No.23 Bangkinang
10	Nomor Telepon/ Faks	(0762)
11	Lulusan yang Telah Dihasilkan	S-1 = - orang, S-2 = - orang, S-3 = - orang
12	Mata Kuliah yang Diampu	1. Pengantar Teknologi Informasi

B. Riwayat Pendidikan

	S-1	S-2	S-3
Nama Perguruan Tinggi	UIN SULTAN SYARIF KASIM RIAU	INSTITUT TEKNOLOGI SEPULUH NOPEMBER	
Bidang Ilmu	Teknik Informastika	Teknik Informatika	
Judul Skripsi/Tesis/Disertasi	Penerapan <i>Wavelet Haar</i> dan <i>Backpropagation</i> Untuk Pengelompokan Diabetik Retinopati Berdasarkan Citra Retina Mata	Deteksi Akun Spammer Berdasarkan Hashtag Dan Aktifitas Komunitas Pada Twitter.	
Nama Pembimbing/Promotor	Suwanto Sanjaya, S.T., M.Kom	Prof. Dr. Agus Zainal Arifin, S.Kom., M.Kom	

**C. Pengalaman Penelitian Dalam 5 Tahun Terakhir
(Bukan Skripsi, Tesis, dan Disertasi)**

No	Tahun	Judul Penelitian	Pendanaan	
			Sumber*	Jml (Juta Rp)
1				

D. Pengalaman Pengabdian Kepada Masyarakat dalam 5 Tahun Terakhir

No	Tahun	Judul Pengabdian Kepada Masyarakat	Pendanaan	
			Sumber*	Jml (Juta Rp)
1				

E. Publikasi Artikel Ilmiah Dalam Jurnal Dalam 5 Tahun Terakhir

No	Tahun	Judul Artikel Ilmiah	Nama Jurnal	Volume/Nomor /Tahun
1	2022	TF-IDF Weighting to Detect Spammer Accounts on Twitter based on Tweets and Retweet Representation of Tweets	SISTEMASI	11/3/2022
2	2022	N-Gram Feature for Comparison of Machine Learning Methods on Sentiment in Financial News Headlines	Journal of Artificial Intelligence and Digital Business (RIGGS)	1/1/2022
3	2022	Dice Similarity and TF-IDF for New Student Admissions Chatbot	Journal of Artificial Intelligence and Digital Business (RIGGS)	1/1/2022
6	2020	Spammer Detection Based on Account, Tweet, and Community Activity on Twitter	Jurnal Ilmu Komputer dan Informasi	13/2/2020

7	2020	Combination of extraction features based on texture and colour feature for beef and pork classification	Journal of Physics: Conference Series	1563/1/2020
8	2019	Automatic image slice marking propagation on segmentation of dental CBCT	TELKOMNIKA (Telecommunication Computing Electronics and Control)	17/6/2019
9	2019	Perbandingan Imputasi dan Parameter Support Vector Regression Untuk Peramalan Cuaca	Simetris: Jurnal Teknik Mesin, Elektro dan Ilmu Komputer	10/2/2019
10	2019	The Application of Haar Wavelet and Backpropagation For Diabetic Retinopathy Classification Based on Eye Retina Image	International Journal Of Science, Engineering, And Information Technology	3/2/2019
11	2019	Deteksi bot spammer twitter berbasis time interval entropy dan global vectors for word representations tweet's hashtag	Register: Jurnal Ilmiah Teknologi Sistem Informasi	5/1/2019

F. Pemakalah Seminar Ilmiah (*Oral Persentation*) dalam 5 Tahun

No	Nama Temu Ilmiah/ Seminar	Judul Artikel	Waktu dan Tempat
1			

G. Karya Buku dalam 5 Tahun Terakhir

No	Judul Buku	Tahun	Jumlah Halaman	Penerbit
1	Deteksi akun spammer pada Twitter menggunakan machine learning	2021		PT. Pena Persada Kerta Utama

H. Perolehan HKI dalam 10 Tahun Terakhir

No	Judul/ Tema HKI	Tahun	Jenis	Nomor P/ID
1				

I. Pengalaman Merumuskan Kebijakan Publik/ Rekayasa Sosial Lainnya dalam 10 Tahun Terakhir

No	Judul/Tema/Jenis Rekayasa Sosial Lainnya	Tahun	Tempat Penerapan	Respon Masyarakat
1				

J. Penghargaan dalam 10 Tahun Terakhir (dari Pemerintah, Asosiasi, atau Institusi Lainnya)

No	Jenis Penghargaan	Institusi Pemberi Penghargaan	Tahun
1			

Semua data yang saya isikan dan tercantum dalam biodata ini adalah benar dan dapat dipertanggungjawabkan secara hukum. Apabila dikemudian hari ternyata dijumpai ketidaksesuaian dengan kenyataan, saya sanggup menerima sanksi.

Demikian biodata ini saya buat dengan sebenarnya untuk memenuhi salah satu persyaratan dalam pengajuan penugasan Penelitian Dosen Pemula

Bangkinang, 01 Februari 2023



Arif Mudi Priyatno, S.T., M.Kom

Lampiran 4 Surat pernyataan ketua peneliti

SURAT PERNYATAAN

Yang bertanda tangan di bawah ini:

Nama : Arif Mudi Priyatno, S.T., M.Kom
NIP : -
NIDN : 1023059501
Pangkat/Golongan : -/-
Alamat : Desa Pasir Emas, Kecamatan Singingi, Kabupaten Kuansing

Dengan ini menyatakan bahwa penelitian saya yang berjudul: “Fitur N-Gram Untuk Perbandingan Metode Machine Learning Pada Sentimen Judul Berita Keuangan” dan diusulkan dalam skim Penelitian Terapan T.A. 2021/2022 **bersifat original dan belum pernah dibiayai oleh lembaga/sumber dana lain.**

Bilamana di kemudian hari ditemukan ketidak sesuaian dengan pernyataan ini, maka saya bersedia dituntut dan diproses sesuai dengan ketentuan yang berlaku dan mengembalikan seluruh biaya penelitian yang sudah diterima ke kas negara.

Demikian pernyataan ini dibuat dengan sesungguhnya dan dengan sebenar-benarnya.

Mengetahui,
Dekan



Dr. Samsurijal Hasan, SP. MM
NIP-TT. 0031126490


Bangkinang, 31 Juli 2022

Ketua Peneliti,



Arif Mudi Priyatno, S.T., M.Kom
NIDN. 1023059501

Menyetujui,
Ketua LPPM Universitas Pahlawan Tuanku Tambusai



Dr. Musnar Indra Daulay, M.Pd
NIP-TT: 096542108