

**KOMPARASI ALGORITMA KLASIFIKASI DAN EKTRAKSI  
FITUR PADA SISTEM AKUNTANSI PINTAR**



Disusun Oleh:

**Bagas Adi Makayasa**

**21206052011**

**PROGRAM STUDI INFORMATIKA ROGRAM MAGISTER**

**FAKULTAS SAINS DAN TEKNOLOGI**

**UIN SUNAN KALIJAGA**

**YOGYAKARTA**

**2023**

## PERNYATAAN KEASLIAN

Yang bertanda tangan di bawah ini:

Nama : Bagas Adi Makayasa  
Nim : 21206052011  
Jenjang : Magister  
Program Studi : Informatika

menyatakan bahwa naskah tesis ini secara keseluruhan adalah hasil penelitian/karya saya sendiri, kecuali pada bagian-bagian yang dirujuk sumbernya.

Yogyakarta, 3 Juli 2023  
Saya yang menyatakan,



METRAI  
TEMPEL  
21218AKX547188436

Bagas Adi Makayasa  
NIM: 21206052011

STATE ISLAMIC UNIVERSITY  
SUNAN KALIJAGA  
YOGYAKARTA

## PERNYATAAN BEBAS PLAGIASI

Yang bertanda tangan di bawah ini:

Nama : Bagas Adi Makayasa  
NIM : 21206052011  
Jenjang : Magister  
Program Studi : Informatika

Menyatakan bahwa naskah tesis ini secara keseluruhan benar – benar bebas dari plagiasi. Jika, dikemudian hari terbukti melakukan plagiasi maka, saya siap ditindak sesuai dengan ketentuan hukum yang berlaku.

Yogyakarta, 3 Juli 2023  
Yang membuat pernyataan,



Nama Lengkap  
NIM 21206052011

STATE ISLAMIC UNIVERSITY  
SUNAN KALIJAGA  
YOGYAKARTA

# LEMBAR PENGESAHAN



KEMENTERIAN AGAMA  
UNIVERSITAS ISLAM NEGERI SUNAN KALIJAGA  
FAKULTAS SAINS DAN TEKNOLOGI  
Jl. Marsda Adisucipto Telp. (0274) 540971 Fax. (0274) 519739 Yogyakarta 55281

## PENGESAHAN TUGAS AKHIR

Nomor : B-2004/Un.02/DST/PP.00.9/08/2023

Tugas Akhir dengan judul : Komparasi Algoritma Klasifikasi dan Ekstraksi Fitur pada Sistem Akuntansi Pintar  
yang dipersiapkan dan disusun oleh:

Nama : BAGAS ADI MAKAYASA, S.Kom  
Nomor Induk Mahasiswa : 21206052011  
Telah diujikan pada : Jumat, 28 Juli 2023  
Nilai ujian Tugas Akhir : A

dinyatakan telah diterima oleh Fakultas Sains dan Teknologi UIN Sunan Kalijaga Yogyakarta

### TIM UJIAN TUGAS AKHIR



Ketua Sidang

Ir. Maria Ulfah Siregar, S.Kom., MIT., Ph.D.  
SIGNED

Valid ID: 64d2fde78e7fb



Penguji I

Dr. Agung Fatwanto, S.Si., M.Kom.  
SIGNED

Valid ID: 64d2451f9519e



Penguji II

Dr. Ir. Bambang Sugiantoro, S.Si., M.T.,  
IPM., ASEAN Eng.  
SIGNED

Valid ID: 64cad326d5624



Yogyakarta, 28 Juli 2023  
UIN Sunan Kalijaga  
Dekan Fakultas Sains dan Teknologi

Prof. Dr. Dra. Hj. Khurul Wardati, M.Si.  
SIGNED

Valid ID: 64d343ad328ad

## NOTA DINAS PEMBIMBING



KEMENTERIAN AGAMA REPUBLIK INDONESIA  
UNIVERSITAS ISLAM NEGERI SUNAN KALIJAGA  
MAGISTER INFORMATIKA  
Jl. Marsda Adisucipto Yogyakarta. Telp (0274) 515856 Yogyakarta 55281

### SURAT PERSETUJUAN TUGAS AKHIR

Hal : Persetujuan Tugas Akhir

Kepada:  
Yth. Dekan Fakultas Sains dan Teknologi  
Uin Sunan Kalijaga  
Yogyakarta Di Yogyakarta

*Assalamu 'alaikum Wr. Wb.*

Setelah melakukan bimbingan, arahan, dan koreksi terhadap penulisan tesis yang berjudul :  
"Komparasi Algoritma Klasifikasi dan Pemodelan Bahasa Pada Sistem Akuntansi Pintar"  
yang di tulis oleh :

Nama : Bagas Adi Makayasa  
NIM : 21206052011  
Jenjang : Magister  
Program Studi : Informatika

Sudah dapat diajukan kepada Program Studi Magister Informatika Fakultas Sains dan  
Teknologi UIN Sunan Kalijaga sebagai salah satu syarat untuk memperoleh gelar Magister  
Informatika.

Dengan ini saya mengharap agar tugas tersebut di atas agar dapat segera dimunaqosahkan.  
Atas perhatiannya saya ucapkan terimakasih.

*Wassalamu'alaikum Wr. Wb.*

Yogyakarta, 3 Juli 2023  
Pembimbing,



Ir. Maria Ulfah Siregar, S.Kom., MIT., Ph.D.  
NIP. 19780106 200212 2 001

## ABSTRAK

Masalah pencatatan keuangan yang tidak sesuai dengan prinsip ilmu akuntansi berpotensi menimbulkan masalah yang tidak perlu. Usaha Mikro, Kecil dan Menengah dengan ciri khasnya masih menghadapi banyak kendala dalam penulisan laporan keuangan. Penelitian ini bertujuan untuk menyelidiki peluang penerapan otomatisasi sistem akuntansi menggunakan pendekatan NLP, menginterpretasikan transaksi keuangan berdasarkan teks yang tertulis pada formulir transaksi ke dalam jurnal akuntansi (debit dan kredit). Eksperimen dilakukan dengan membandingkan kinerja tiga algoritma klasifikasi (yaitu SVM, K-Nearest Neighbor, dan Random Forest) dengan Ekstraksi Fitur (TF-IDF, BOW, dan Word2Vec). Terdapat 200 dataset transaksi keuangan yang terdiri dari sepuluh kelas, data tersebut dibagi menjadi dua bagian, yaitu *balance dataset* dan *imbalance dataset*. Pasangan SVM dan Word2Vec dalam *balanced dataset* memberikan akurasi tertinggi (92,5%), presisi (92,5%), daya ingat/sensitivitas (93,33%), dan skor F1 (92%). Namun jika dibandingkan dengan hasil penelitian semantik terkait (performa rata-rata mencapai 95%), hasil yang diperoleh pada penelitian ini masih lebih rendah. Satu hal yang mungkin berpengaruh signifikan adalah jumlah data dalam korpus yang masih kurang. Peneliti menyarankan untuk menambah jumlah dataset dan mencoba menggunakan kombinasi model bahasa seperti Glove, Bert dll. Penelitian ini juga dapat digunakan sebagai model awal untuk kasus transaksi keuangan yang lebih kompleks pada penelitian selanjutnya.

**Kata Kunci:** *Natural Language Processing*; Akuntansi; Transaksi Keuangan.

## **ABSTRACT**

*The problem of financial records that are not in accordance with the principles of accounting science has the potential to cause unnecessary problems. Micro, Small and Medium Enterprises with their characteristics still encounter many obstacles in writing financial reports. This research aims to explore opportunities for implementing an accounting automation system using the NLP approach, interpreting financial transactions based on the text written on the transaction form into accounting journals (debits and credits). Experiments were carried out by comparing the performance of three classification algorithms (namely SVM, K-Nearest Neighbor, and Random Forest) with Feature Extraction (TF-IDF, BOW, and Word2Vec). There are 200 financial transaction datasets consisting of ten classes, the data is divided into two parts, namely balance dataset and imbalance dataset. The SVM and Word2Vec pair in the balance dataset gave the highest score accuracy (92.5%), precision (92.5%), recall (93.33%), and F1 score (92%). However, when compared with the results of related semantic research (the average performance reaches 95%), the results obtained in this study are still lower. One thing that may have a significant effect is the amount of data in the corpus which is still lacking. Researchers suggest increasing the number of datasets and trying to use a combination of language models such as Glove, Bert etc. This research can also be used as an initial model for more complex financial transaction cases in future studies.*

**Keywords:** *Natural Language Processing; Accountancy; Financial Transactions.*

## KATA PENGANTAR

*Bissmillahirrahmanirrahim,*

*Assalamualaikum warahmatullahi wabarakatuhu*

Puji syukur penulis panjatkan kehadiran Allah Subhanahu wata'ala yang telah melimpahkan segala rahmat dan hidayah-Nya, sehingga penulis dapat menyelesaikan thesis dengan judul “Komparasi Algoritma Klasifikasi dan Ekstraksi Fitur Pada Sistem Akuntansi Pintar” sebagai salah satu persyaratan mencapai gelar Magister Informatika.

Penulis mengucapkan terima kasih kepada semua pihak yang telah memberikan dorongan, semangat, dan ide-ide kreatif sehingga tahap demi tahap penyusunan thesis ini telah selesai. Ucapan terima kasih tersebut secara khusus disampaikan kepada:

1. Bapak Prof. Dr. Phil. Al Makin, S.Ag., M.A., selaku Rektor UIN Sunan Kalijaga Yogyakarta.
2. Ibu Dr. Dra. Hj. Khurul Wardati, M.Si., selaku Dekan Fakultas Sains dan Teknologi UIN Sunan Kalijaga Yogyakarta.
3. Bapak Dr. Bambang Sugiantoro, M.T., selaku Ketua Program Studi Magister Informatika Fakultas Sains dan Teknologi UIN Sunan Kalijaga Yogyakarta.
4. Ibu Ir. Maria Ulfah Siregar, S.Kom., MIT., Ph.D., selaku dosen Pembimbing Tesis yang telah memberikan arahan selama studi sehingga dapat menyelesaikan penyusunan tesis ini.
5. Bapak dan Ibu dosen Program Studi Magister Informatika Fakultas Sains dan Teknologi UIN Sunan Kalijaga Yogyakarta yang sudah membagi ilmu yang bermanfaat.



6. Seluruh Staf Karyawan Fakultas Sains dan Teknologi Universitas Islam Negeri Sunan Kalijaga Yogyakarta yang telah membantu sehingga penyusunan thesis ini berjalan lancar.
7. Orang tua dan istri yang senantiasa memberikan doa selama studi di Magister Informatika Universitas Islam Negeri Sunan Kalijaga Yogyakarta.
8. Teman-teman magister informatika angkatan 2021 Universitas Islam Negeri Sunan Kalijaga atas kerjasama, saran, dan bantuannya.
9. Semua pihak yang tidak bisa penulis sebutkan satu persatu atas bantuannya dalam penyelesaian thesis ini.

Akhirnya penulis hanya dapat bersyukur kepada Allah SWT semoga yang telah dilakukan selamama ini dapat menjadi amal dan bekal di akhirat nanti. Penulis menyadari masih banyak kesalahan dan kekurangan dalam penulisan thesis ini, maka dari itu saran dan kritik sangat diharapkan demi perbaikan. Semoga thesis ini dapat bermanfaat bagi penulis sendiri khususnya dan bagi pembaca pada umumnya, terimakasih.

*Wassalamualaikum warahmatullahi wabarakatuhu*

STATE ISLAMIC UNIVERSITY  
SUNAN KALIJAGA  
YOGYAKARTA

Yogyakarta, 6 Agustus 2023



Bagas Adi Makayasa  
21206052011



STATE ISLAMIC UNIVERSITY  
**SUNAN KALIJAGA**  
YOGYAKARTA

## MOTTO

Bekerjalah 100%, maka hasilnya 100%.  
Berusahalah memberikan 200%, maka hasilnya 100% plus  
kepercayaan. Bekerjalah setengah-setengah maka komitmen mulai  
diragukan.

اللَّهُمَّ صَلِّ عَلَى مُحَمَّدٍ وَعَلَى آلِهِ وَسَلَّمَ



STATE ISLAMIC UNIVERSITY  
SUNAN KALIJAGA  
YOGYAKARTA

## DAFTAR ISI

PERNYATAAN KEASLIAN .....	II
PERNYATAAN BEBAS PLAGIASI .....	III
LEMBAR PENGESAHAN .....	IV
NOTA DINAS PEMBIMBING .....	V
ABSTRAK .....	VI
ABSTRACT .....	VII
KATA PENGANTAR .....	VIII
PERSEMBAHAN .....	X
MOTTO .....	XI
DAFTAR ISI .....	XII
DAFTAR GAMBAR .....	XIV
DAFTAR TABEL .....	XVI
BAB I PENDAHULUAN .....	1
A. LATAR BELAKANG .....	1
B. RUMUSAN MASALAH .....	4
C. BATASAN MASALAH .....	4
D. TUJUAN PENELITIAN .....	4
E. MANFAAT PENELITIAN .....	4
F. KEASLIAN PENELITIAN .....	5
BAB II TINJAUAN PUSTAKA DAN LANDASAN TEORI .....	6
A. TINJAUAN PUSTAKA .....	6
B. LANDASAN TEORI .....	9
1. Transaksi Keuangan dalam Akuntansi .....	9
2. Natural Language Processing .....	12
3. Ekstraksi Fitur .....	21

4. Algoritma Klasifikasi.....	26
<b>BAB III METODE PENELITIAN .....</b>	<b>34</b>
<b>A. METODE PENELITIAN.....</b>	<b>34</b>
1. Input data dan labeling.....	35
2. Preprocessing .....	38
3. Ekstraksi Fitur .....	40
3. Evaluasi.....	42
<b>B. ALAT DAN BAHAN .....</b>	<b>43</b>
1. Alat.....	43
2. Bahan .....	44
<b>BAB IV HASIL DAN PEMBAHASAN .....</b>	<b>45</b>
<b>A. SKENARIO UJI COBA.....</b>	<b>45</b>
1. Input data dan labeling.....	47
2. Preprocessing .....	50
3. Ekstraksi Fitur dan Klasifikasi.....	51
<b>B. HASIL UJI COBA .....</b>	<b>52</b>
<b>C. PEMBAHASAN .....</b>	<b>53</b>
<b>BAB V PENUTUP .....</b>	<b>60</b>
<b>A. KESIMPULAN.....</b>	<b>60</b>
<b>B. SARAN.....</b>	<b>61</b>
<b>DAFTAR PUSTAKA.....</b>	<b>62</b>

## DAFTAR GAMBAR

Gambar 1 Perilaku UMKM dan aktivitas transaksi keuangan .....	3
Gambar 2 Rumpun ilmu NLP .....	14
Gambar 3 CFG ATIS English.....	15
Gambar 4 Contoh parsing tree NLP .....	16
Gambar 5 Alur implementasi NLP .....	18
Gambar 6 Contoh balance dataset dan imbalance dataset .....	19
Gambar 7 Simulasi representasi kedekatan nilai vektor.....	24
Gambar 8 Arsitektur CBOW dan Skip-gram.....	24
Gambar 9 Ilustrasi data linier dan non linier .....	28
Gambar 10 Ilustrasi algoritma SVM linier .....	28
Gambar 11 Ilustrasi SVM dengan kernel data non linier. ....	30
Gambar 12 Ilustrasi algoritma KNN.....	32
Gambar 13 Ilustrasi Algoritma Random Forest.....	33
Gambar 14 Desain Arsitektur Sistem .....	34
Gambar 15 Flowchart metode penelitian.....	35
Gambar 16 Sample input kalimat transaksi .....	42
Gambar 17 Representasi Word2Vec berdasarkan kata kunci.....	42
Gambar 18 Confusion Matrix.....	43
Gambar 19 GUI input transaksi keuangan.....	45
Gambar 20 Simulasi prediksi menggunakan KNN .....	46
Gambar 21 Kode program untuk mengambil data .....	48
Gambar 22 Contoh lima data teratas .....	48
Gambar 23 Label kelas dari dataset .....	49
Gambar 24 Grafik distribusi data .....	50
Gambar 25 Kode program tahap preprocessing .....	51

Gambar 26 Contoh data setelah preprocessing .....	52
Gambar 27 Kode program Ekstraksi Fitur BOW dan TF-IDF .....	52
Gambar 28 Kode program confusion matrix .....	53
Gambar 29 Hasil uji coba .....	54
Gambar 30 Google voice search inputan pendek .....	57
Gambar 31 Google voice search inputan panjang .....	58



## DAFTAR TABEL

Tabel 1 Komparasi penelitian terkait.....	6
Tabel 2 Simulasi pembuatan jurnal akuntansi .....	12
Tabel 3 Contoh penggunaan frasa, klausa, dan kalimat .....	15
Tabel 4 Analogi analisis bahasa .....	17
Tabel 5 Simulasi Word2Vec Skip-Gram .....	25
Tabel 6 Simulasi Word2Vec CBOW .....	26
Tabel 7 Contoh korpus yang sudah dilabeli .....	37
Tabel 8 Contoh Lowercase .....	39
Tabel 9 Contoh Tokenizing .....	39
Tabel 10 Contoh Stopwords Removal .....	39
Tabel 11 Contoh Stemming .....	39
Tabel 12 Ilustrasi BOW .....	40
Tabel 13 Simulasi perhitungan TF-IDF .....	41
Tabel 14 Alat penelitian.....	44
Tabel 15 Bahan Penelitian .....	44

STATE ISLAMIC UNIVERSITY  
SUNAN KALIJAGA  
YOGYAKARTA



# **BAB I**

## **PENDAHULUAN**

### **A. LATAR BELAKANG**

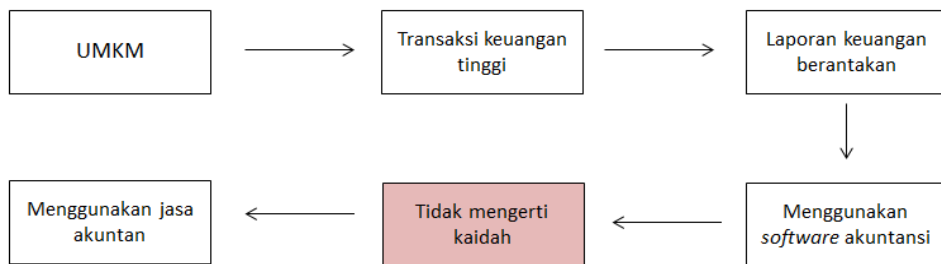
Kementrian Koperasi mencatat jumlah data Usaha Mikro Kecil dan Menengah (UMKM) di Indonesia sebanyak 64,2 juta pada tahun 2018, sedangkan data perusahaan berjumlah 729 pada tahun 2022 (Hisnul, 2022), angka yang menarik untuk menggambarkan potensi transaksi keuangan di dalamnya. Banyaknya UMKM tersebut menggambarkan betapa vitalnya peran usaha kecil sebagai salah satu pilar ekonomi suatu negara. Dalam organisasi perusahaan pasti memiliki laporan keuangan sebagai rangkuman atas perputaran uang yang telah dilakukan, tidak terkecuali UMKM sesederhana apapun laporan tersebut. Contoh seperti laporan laba rugi, laporan perubahan modal, laporan utang piutang dan lain-lain, dalam hal ini akuntansi merupakan cabang keilmuan yang fokus pada pembukuan keuangan.

Dengan mencatat keuangan secara baik dan sesuai standar akuntansi, UMKM akan merasakan manfaat baik secara internal maupun eksternal (pihak luar yang berkepentingan) (Nurul, 2020). Adapun laporan keuangan berguna untuk berbagai macam hal seperti perencanaan bisnis, informasi posisi keuangan setiap bulan, kontrol biaya, pertimbangan pendanaan investasi, pengambilan keputusan bisnis, perhitungan pajak, dan lain sebagainya.

Secara kaidah akuntansi, setiap aktivitas dinyatakan sah apabila telah dicatat setiap kali terjadi transaksi yang berkaitan dengan uang masuk dan uang keluar. Seperti contoh “pemilik UMKM membeli satu buah komputer

senilai dua juta rupiah” maka kegiatan tersebut dicatat sebagai satu transaksi keuangan, “pemilik membeli dua buah printer senilai satu juta rupiah” maka ini juga harus dicatat sebagai satu kegiatan transaksi keuangan. Pencatatan ini harus ditulis satu per satu dan terus menerus agar tidak ada riwayat transaksi yang hilang karena akan berdampak pada laporan keuangan yang dihasilkan nantinya. Bisa dibayangkan jika dalam satu hari saja UMKM harus disibukkan dengan banyaknya transaksi keuangan, sehingga muncul opsi menggunakan software akuntansi untuk mempermudah pekerjaan.

Namun UMKM dan keunikannya memiliki masalah tersendiri dalam pencatatan keuangan, perlu ditekankan bahwa secara umum UMKM harus dilihat dalam kondisi tidak ideal dari sisi manajemen dibandingkan dengan perusahaan. Untuk memperkuat hipotesis masalah, peneliti melakukan wawancara terhadap empat orang perwakilan akuntan dan pelaku UMKM untuk mengetahui perilaku UMKM dan aktivitas transaksi keuangan (Sony et al., 2022). Dari hasil wawancara narasumber menjelaskan bahwa software akuntansi yang ada masih dihadapkan pada dua masalah utama yaitu pemahaman pengguna terhadap kaidah pencatatan yang benar dan kendala dari User Interface (UI) dan User Experience (UX) yang berbeda-beda. Perilaku UMKM terkait transaksi keuangan seperti pada Gambar 1.



Gambar 1. Perilaku UMKM dan aktivitas transaksi keuangan.

Dari permasalahan yang ada, peneliti melihat peluang alternatif untuk melakukan pencatatan transaksi keuangan (jurnal) menggunakan pendekatan Pemrosesan Bahasa Alami atau Natural Language Processing (NLP) untuk pengguna awam atau ‘stupid user’. Adapun NLP merupakan turunan dari bidang Kecerdasan Buatan (Artificial Intelligence) yang berfokus pada pemrosesan teks atau suara, dengan tujuan utama bagaimana komputer mampu memahami informasi layaknya manusia (Jayasurya, 2022).

Skenario penelitian ini dengan melakukan komparasi pada beberapa algoritma klasifikasi yaitu K-Nearest Neighbour, Support Vector Machine, dan Random Forest yang dikombinasikan dengan Ekstraksi Fitur Bag of Words, Term Frequency – Inverse Document Frequency, dan Word2Vec. Adapun pertimbangan memilih ketiga algoritma klasifikasi di atas karena dari beberapa literatur yang didapatkan, algoritma tersebut rata-rata menunjukkan hasil pengujian yang baik. Hasil akhir dari penelitian ini diharapkan mampu melihat perbandingan kinerja dari kombinasi antara algoritma dan Ekstraksi Fitur terhadap studi kasus yang diangkat.

## **B. RUMUSAN MASALAH**

Berdasarkan latar belakang yang telah diuraikan, maka rumusan masalah dalam penelitian ini adalah bagaimana melakukan komparasi kinerja dari kombinasi algoritma klasifikasi dan ekstraksi fitur untuk sistem akuntansi pintar?

## **C. BATASAN MASALAH**

Batasan masalah dalam penelitian ini sebagai berikut:

1. Akuntansi Pintar yang dimaksud ialah dalam konteks pendekatan Pemrosesan Bahasa Alami (NLP) pada studi kasus perusahaan jasa di UMKM.
2. Korpus yang dipakai telah divalidasi dan dilakukan pengelempokan olah akuntan selaku ahli.

## **D. TUJUAN PENELITIAN**

Penelitian ini bertujuan untuk melakukan komparasi algoritma klasifikasi dan ekstraksi fitur pada sistem akuntansi pintar.

## **E. MANFAAT PENELITIAN**

Penelitian ini diharapkan memberi manfaat sebagai berikut:

1. Melakukan komparasi terhadap kombinasi algoritma klasifikasi dan Ekstraksi Fitur.
2. Mengevaluasi performa atas kombinasi terbaik dari algoritma dan Ekstraksi Fitur.
3. Melihat peluang pengembangan sistem akuntansi pintar dengan

pendekatan Pemrosesan Bahasa Alami

#### **F. KEASLIAN PENELITIAN**

Berdasarkan pengamatan dan tinjauan pustaka, penelitian ini belum pernah dilakukan sebelumnya.



## **BAB V**

### **PENUTUP**

#### **A. KESIMPULAN**

Berdasarkan penelitian yang telah dilakukan, diperoleh kesimpulan sebagai berikut:

1. Dari dua belas kombinasi yang diuji, pasangan algoritma SVM dengan Ekstraksi Fitur Word2Vec CBOW memiliki performa terbaik dengan tingkat akurasi 92,5%, presisi 93,5%, recall 93,42%, dan skor F1 92,29% .
2. Dalam mengembangkan model, perlu memperhatikan kondisi dataset dimana dalam penelitian ini balance dataset secara signifikan mempengaruhi hasil kinerja model machine learning.
3. Secara umum Ekstraksi Fitur Word2Vec lebih optimal dibanding TF-IDF dan BOW karena kemampuannya dalam memahami konteks kalimat. Kemudian TF-IDF lebih optimal dibandingkan BoW karena model TF-IDF memanfaatkan informasi tentang frekuensi kemunculan setiap kata dalam kalimat dalam korpus, sedangkan model BoW hanya memanfaatkan informasi tentang keberadaan atau Tidak adanya kata (istilah/istilah) dalam kalimat di dalam korpus.

Berdasarkan hasil yang diperoleh dari penelitian ini, secara statistik performa model yang dibuat masih lebih rendah dibandingkan penelitian Ekstraksi Fitur semantik yang serupa dimana rata-rata menunjukkan performa di atas 95%. Adapun faktor yang mungkin mempengaruhi hasil secara signifikan ini yaitu jumlah data di dalam corpus yang digunakan masih sangat rendah (200 data dengan 10 kelas).

## **B. SARAN**

Berdasarkan hasil uji coba yang telah dilakukan pada penelitian ini, diharapkan agar pada penelitian selanjutnya bisa meningkatkan akurasi pengujian model yang lebih akurat dan tidak overfitting. Oleh karena itu, peneliti mempunyai saran untuk pengembang melakukan penelitian pada masa yang akan datang untuk menambahkan jumlah data dan juga penambahan teknik pada tahap preprocessing yang digunakan, dikarenakan jumlah dataset dan preprocessing juga dapat mempengaruhi hasil performa model. Selain itu, juga dapat menambahkan metode lain pada proses validation menggunakan K-fold cross validation agar mengurangi overfitting pada model dan mengembangkan metode lain pada word embedding seperti GloVe, FastText, dan lain sebagainya.



## DAFTAR PUSTAKA

- Hisnul, Pompong Budi Setiadi, Sri Rahayu, "UMKM di Masa Pandemi COVID 19 Berdampak Pada Teknologi Dan Ddigitalisasi Pada Pusat Oleh-Oleh Rahma di Desa Kendalrejo", *Jurnal Ekonomi dan Bisnis*, Vol. 11 No. 1 Juli 2022.
- Nurul, Lulu, "Pentingnya Penyusunan Laporan Keuangan UMKM Bagi Para Pengusaha Bakery, Cake, dan Pastry (BCP) di Kota Blitar", *Jurnal Graha Pengabdian*, Vol 2 No 2, 2020.
- Wawancara dengan Sony Warsono, Nadya Windi, dan Anggit Firmansyah, mahasiswa dan dosen Magister Akuntansi Universitas Gadjah Mada, pada 04 Oktober 2022.
- Wawancara dengan Anam dan Antariksa, pelaku UMKM, pada 12 Oktober 2022.
- S. Salloum, T. Gaber, S. Vadera and K. Shaalan, "A Systematic Literature Review on Phishing Email Detection Using Natural Language Processing Techniques," in *IEEE Access*, vol. 10, pp. 65703-65727, 2022, doi: 10.1109/ACCESS.2022.3183083.
- G. G. Jayasurya, S. Kumar, B. K. Singh and V. Kumar, "Analysis of Public Sentiment on COVID-19 Vaccination Using Twitter," in *IEEE Transactions on Computational Social Systems*, vol. 9, no. 4, pp. 1101-1111, Aug. 2022, doi: 10.1109/TCSS.2021.3122439.
- Aziz Husain, Fadhila Tangguh Atmojo, and Erma Susanti, "Analisis Perbandingan Performa Metode Klasifikasi pada Dataset Multiclass Citra Busur Panah", *Jurnal Technocom*, Vol. 19, No. 3, Agustus 2020: 286-294
- M. Jayaratne and B. Jayatilleke, "Predicting Personality Using Answers to Open-Ended Interview Questions," in *IEEE Access*, vol. 8, pp. 115345-115355, 2020, doi: 10.1109/ACCESS.2020.3004002.
- Iswandi, Irvan. Iping Supriana Suwardi, dan Nur Ulfa Maulidevi, "Penelitian Awal: Otomatisasi Interpretasi Data Akuntansi Berbasis Natural Language Processing", *Jurnal Sistem Informasi*, Vol 5 No 2,



2013.

- C. Mugisha and I. Paik, "Comparison of Neural Language Modeling Pipelines for Outcome Prediction From Unstructured Medical Text Notes," in *IEEE Access*, vol. 10, pp. 16489-16498, 2022, doi: 10.1109/ACCESS.2022.3148279.
- M. U. Siregar, I. Setiawan, N. Z. Akmal, D. Wardani, Y. Yunitasari and A. Wijayanto, "Optimized Random Forest Classifier Based on Genetic Algorithm for Heart Failure Prediction," 2022 Seventh International Conference on Informatics and Computing (ICIC), Denpasar, Bali, Indonesia, 2022, pp. 01-06, doi: 10.1109/ICIC56845.2022.10006987.
- Syamsul, "Analisis Pencatatan dan Pelaporan Keuangan UMKM di Kota Palu", *Jurnal Keuangan dan Bisnis*, Vol 10 No 1, 2022.
- M. F. Mridha, A. A. Lima, K. Nur, S. C. Das, M. Hasan and M. M. Kabir, "A Survey of Automatic Text Summarization: Progress, Process and Challenges," in *IEEE Access*, vol. 9, pp. 156043-156070, 2021, doi: 10.1109/ACCESS.2021.3129786.
- X. Chen, P. Cong and S. Lv, "A Long-Text Classification Method of Chinese News Based on BERT and CNN," in *IEEE Access*, vol. 10, pp. 34046-34057, 2022, doi: 10.1109/ACCESS.2022.3162614.
- J. Jiang et al., "Enhancements of Attention-Based Bidirectional LSTM for Hybrid Automatic Text Summarization," in *IEEE Access*, vol. 9, pp. 123660-123671, 2021, doi: 10.1109/ACCESS.2021.3110143.
- R. Devika, S. Vairavasundaram, C. S. J. Mahenthara, V. Varadarajan and K. Kotecha, "A Deep Learning Model Based on BERT and Sentence Transformer for Semantic Keyphrase Extraction on Big Social Data," in *IEEE Access*, vol. 9, pp. 165252-165261, 2021, doi: 10.1109/ACCESS.2021.3133651.
- Warsono, Sony, "Dasar-Dasar Akuntansi: Tes Potensi Akuntansi", Yogyakarta, ABPublisher, 2018.
- H. S. Nawaz, Z. Shi, Y. Gan, A. Hirpa, J. Dong and H. Zheng, "Temporal Moment Localization via Natural Language by Utilizing Video

- Question Answers as a Special Variant and Bypassing NLP for Corpora," in IEEE Transactions on Circuits and Systems for Video Technology, vol. 32, no. 9, pp. 6174-6185, Sept. 2022, doi: 10.1109/TCSVT.2022.3162650.
- A. Radhakrishnan, D. Mahapatra and A. James, "Consumer Document Analytical Accelerator Hardware," in IEEE Access, vol. 11, pp. 5161-5167, 2023, doi: 10.1109/ACCESS.2023.3237463.
- H. A. Ahmed, N. Z. Bawany and J. A. Shamsi, "CaPBug-A Framework for Automatic Bug Categorization and Prioritization Using NLP and Machine Learning Algorithms," in IEEE Access, vol. 9, pp. 50496-50512, 2021, doi: 10.1109/ACCESS.2021.3069248.
- R. Sonbol, G. Rebdawi and N. Ghneim, "The Use of NLP-Based Text Representation Techniques to Support Requirement Engineering Tasks: A Systematic Mapping Review," in IEEE Access, vol. 10, pp. 62811-62830, 2022, doi: 10.1109/ACCESS.2022.3182372.
- S. G. Kanakaraddi and S. S. Nandval, "Dynamic Fuzzy Parser to Parse English Sentence Using POS Tagger and Fuzzy Max-Min Technique," 2018 International Conference on Current Trends towards Converging Technologies (ICCTCT), Coimbatore, India, 2018, pp. 1-5, doi: 10.1109/ICCTCT.2018.8551093..
- Ioan, Alexandu Cuza, "The Semantics and Pragmatics of Natural Language", transkrip presentasi University of IASI Faculty of Computer Science, 2018.
- Indri Monika Parapat, Muhammad Tanzil Furqon, Sutrisno, "Penerapan Metode Support Vector Machine (SVM) Pada Klasifikasi Penyimpangan Tumbuh Kembang /Anak", Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer, Vol. 2, No. 10, Oktober 2018, hlm. 3163-3169.
- Markoulidakis, Rallis, Georgoulas, Kopsiaftis, G. Doulamis, A. Doulamis, N.Multiclass Confusion MatrixReduction Method and ItsApplication on Net Promoter ScoreClassification Problem. Technologies2021,9, 81.  
<https://doi.org/10.3390/technologies9040081>.

- S. Sugriyono, and M. U. Siregar, "Prapemrosesan klasifikasi algoritme kNN menggunakan K-means dan matriks jarak untuk dataset hasil studi mahasiswa," *Jurnal Teknologi dan Sistem Komputer*, vol. 8, no. 4, pp. 311-316, Oct. 2020. <https://doi.org/10.14710/jtsiskom.2020.13874>
- A. S. Yazid and A. Fatwanto, "Penentuan Kelas Kata Pada Part of Speech Tagging Kata Ambigu Bahasa Indonesia", *JISKa*, vol. 2, no. 3, pp. 157–166, Mar. 2018.
- M. Khushi et al., "A Comparative Performance Analysis of Data Resampling Methods on Imbalance Medical Data," in *IEEE Access*, vol. 9, pp. 109960-109975, 2021, doi: 10.1109/ACCESS.2021.3102399..

