Building of Informatics, Technology and Science (BITS)
Volume 6, No 2, September 2024, Page: 1045−1058
ISSN 2684-8910 (media cetak)
ISSN 2685-3310 (media online)
DOI 10.47065/bits.v6i2.5807

Optimasi Rekomendasi Sustainable Development Goals (SDGs) di Indonesia menggunakan Content-Based Filtering dan Algoritma Machine Learning

Alfajri Hulvi*, Kusrini

Fakultas Teknik, Magister Teknik Informatika, Universitas Amikom, Yogyakarta, Indonesia
Email: alfajrihulvi@students.amikom.ac.id, kusrini@amikom.ac.id
Email Penulis Korespondensi: alfajrihulvi@students.amikom.ac.id
Submitted: **17/08/2024**; Accepted: **10/09/2024**; Published**: 12/09/2024**

Abstrak−Lahirnya program tentang Tujuan Pembangunan Berkelanjutan atau Sustainable Development Goals (SDGs) pada tahun 2015 membuat masyarakat di semua negara mulai memandang penting pembangunan berkelanjutan untuk diimplementasikan. Indonesia, sebagai bagian dari komunitas global, juga telah mengadopsi SDGs ini sebagai kerangka kerja dalam upaya mencapai Indonesia Emas 2045. Dengan visi ini, Indonesia bercita-cita menjadi negara maju yang berdaulat, adil, dan makmur tepat pada peringatan 100 tahun kemerdekaannya. Untuk mencapai tujuan secara efektif, penting untuk menerapkan sistem rekomendasi berbasis Artificial Intelligence (AI) yang mempertimbangkan tantangan sosial, ekonomi, dan lingkungan hidup yang dihadapi oleh negara Indonesia di masa mendatang. Content-Based Filtering (CBF) adalah teknik yang populer untuk membangun sistem tersebut. Penelitian ini membahas teknik untuk optimasi CBF menggunakan beberapa algoritma Machine Learning tradisional yaitu SVM, KNN, DT dan algoritma Deep Learning yaitu MLP. Teknik pengambilan sampling dan penyetelan hiperparameter juga diperhatikan dalam penelitian ini. Algoritma Deep Learning MLP dengan teknik sampling menggunakan SMOTE dan Grid Search menghasilkan akurasi tertinggi yaitu 84% dibandingkan dengan algoritma Machine Learning tradisional lainnya.

Kata Kunci: SDGs, Content-Based Filtering, Machine Learning, Deep Learning

Abstract−The birth of the Sustainable Development Goals (SDGs) program in 2015 has made people in all countries begin to view sustainable development as important to implement. Indonesia, as part of the global community, has also adopted these SDGs as a framework in an effort to achieve Indonesia Emas 2045. With this vision, Indonesia aspires to become a developed country that is sovereign, just, and prosperous right on the 100th anniversary of its independence. To achieve the goals effectively, it is important to implement an Artificial Intelligence (AI)-based recommendation system that considers the social, economic, and environmental challenges faced by Indonesian countries in the future. Content-Based Filtering (CBF) is a popular technique for building such a system. This study discusses techniques for CBF optimization using several traditional Machine Learning algorithms, namely SVM, KNN, DT and Deep Learning algorithms, namely MLP. Sampling techniques and hyperparameter tuning are also considered in this study. The MLP Deep Learning algorithm with sampling techniques using SMOTE and Grid Search produces the highest accuracy of 84% compared to other traditional Machine Learning algorithms.

Keywords: SDGs, Content-Based Filtering, Machine Learning, Deep Learning

1. PENDAHULUAN

Pembangunan berkelanjutan sudah diperkenalkan sejak Konferensi Lingkungan di Stockholm tahun 1972. Perserikatan Bangsa-Bangsa (PBB) merilis program yang diberi nama Sustainable Development Goals (SDGs) pada akhir tahun 2015. Program tujuan perencanaan pembangunan disusun dalam dokumen "*Transforming Our World: The 2030 Agenda for Sustainable Development*" dan telah mendapatkan persetujuan dari 190 negara. Dokumen ini memuat 17 tujuan dan 169 target Pembangunan [1]. Tujuan program ini disusun dapat menjawab keterbelakangan pembangunan negara-negara di seluruh dunia, baik negara maju maupun berkembang. Hal ini menjadikan program ini menjadi perhatian semua negara yang terlibat, terutama bagi negara-negara berkembang dimana ketimpangan ekonomi dan kemiskinan merupakan permasalahan global yang sampai saat ini belum terselesaikan.

Indonesia, sebagai bagian dari komunitas global, juga telah mengadopsi SDGs ini sebagai kerangka kerja dalam upaya mencapai Indonesia Emas 2045. Dengan visi ini, Indonesia bercita-cita menjadi negara maju yang berdaulat, adil, dan makmur tepat pada peringatan 100 tahun kemerdekaannya. Untuk mencapai tujuan secara efektif, penting untuk menerapkan sebuah sistem rekomendasi yang mempertimbangkan tantangan sosial, ekonomi, dan lingkungan hidup yang dihadapi oleh negara Indonesia di masa mendatang.

Artificial Intelligence (AI) telah menjadi salah satu teknologi paling transformatif dalam berbagai aspek kehidupan modern, termasuk dalam memberikan rekomendasi yang efektif dan relevan. Salah satu keunggulan utama AI dalam sistem rekomendasi adalah kemampuannya untuk memproses data dengan kecepatan dan akurasi yang jauh melampaui kemampuan manusia. Dengan algoritma machine learning yang terus belajar dan berkembang, AI mampu memberikan rekomendasi yang semakin presisi seiring waktu.

Dalam membangun sebuah sistem rekomendasi berbasis AI penting untuk menggunakan algoritma machine learning yang tepat. Pemilihan algoritma yang tepat sangatlah krusial karena dapat mempengaruhi akurasi dan efektivitas rekomendasi yang dihasilkan. Banyak peneliti melakukan penelitian yang memanfaatkan sistem rekomendasi dengan metode Content-Based Filtering. Penelitian yang dilakukan Rosidah dkk. pada tahun 2024 tentang *Library Book Recommendation System Using Content-Based Filtering* menggunakan metode ini dalam melakukan rekomendasi buku di perpustakaan berdasarkan personalisasi dari pengguna [2]. Metode yang sama juga

Building of Informatics, Technology and Science (BITS)
Volume 6, No 2, September 2024, Page: 1045−1058
ISSN 2684-8910 (media cetak)
ISSN 2685-3310 (media online)
DOI 10.47065/bits.v6i2.5807

dilakukan oleh Leander dkk. dalam melakukan rekomendasi film berdasarkan personalisasi user yang dilakukan optimalisasi dengan algoritma klasifikasi Support Vector Machine (SVM) [3].

Dengan kata lain, metode Content-Based Filtering memungkinkan untuk dilakukan penelitian terhadap sistem rekomendasi pada SDGs. Penelitian ini berfokus untuk menemukan algoritma terbaik di antara SVM, KNN, DT dan MLP yang dapat meningkatkan kinerja dan akurasi sistem rekomendasi dengan mengidentifikasi, menganalisis, dan menguji berbagai teknik dan metode.

Penelitian terkait metode Content-Based Filtering dilakukan oleh Rosidah dkk, penelitian ini terkait sistem rekomendasi buku perpustakaan di SMK Darul Mustofa Bangkalan Madura dengan menggunakan algoritma TF-IDF dan Cosine Similarity untuk mendaptakan mencari kemiripan kata dan kalimat. Masih ada peluang untuk melakukan eksplorasi pembuatan model dengan menggunakan strategi penambahan algoritma ML lain [2].

Sedangkan penelitian oleh Leander dkk. menggunakan metode Content-Based Filtering dan algoritma SVM. Didapatkan hasil penelitian menggunakan algortima SVM akurasi terbaik sebesar 88,5% dengan fungsi kernel RBF, nilai C = 1 dan nilai gamma = 10. Penelitian ini hanya menggunakan 1 teknik algoritma, sehingga masih ada kemungkinan bahwa teknik ML lain dapat mencapai kinerja yang lebih baik [3].

Pada algoritma Deep Learning, metode Multi Layer Perceptron (MLP) pernah dilakukan penelitian oleh Zhang dengan melakukan perbandingan algoritma Deep Learning yaitu CNN, LSTM dan MLP dalam melakukan klasifikasi pada berita. Terbukti algoritma MLP mempunya akurasi yang lebih tinggi dibandingkan dengan algoritma lainnya [4].

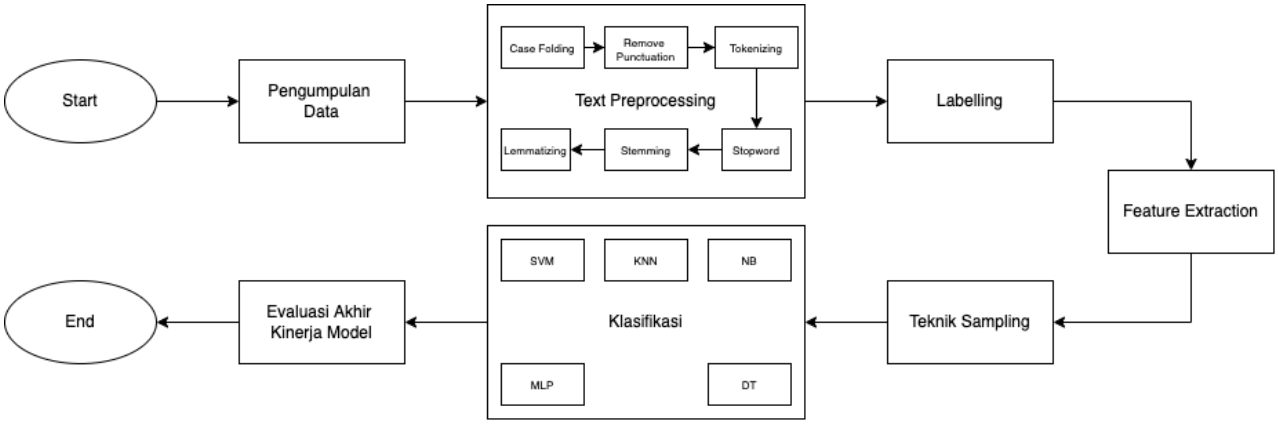
Sebelumnya, penelitian yang dilakukan Rolanda dkk. mengembangkan pencarian oleh pengguna berdasarkan kategori yang sama sehingga pengguna dapat mengurangi keinginan untuk mencari sesuai dengan kategori yang mereka inginkan, pada penelitian ini penulis menggunakan sistem rekomendasi berdasarkan content base filtering. Penelitian ini belum menyajikan metode evaluasi hasil, sehingga masih ada probabilitas untuk mengembangkan sistem yang mempunyai hasil yang lebih maksimal [5].

Pada 2 metode algoritma diatas, ditentukan perbandingan dengan mencari parameter terbaik guna untuk melakukan optimalisasi algoritma. Penelitian yang dilakukan Baroqah Pohan dkk. menggunakan teknik GridSearchCV dan Tuning Hyperparameter untuk menentukan parameter terbaik pada algoritma SVM dan didapat akurasi sebesar 69,95% [6].

2. METODOLOGI PENELITIAN

2.1 Tahapan Penelitian

Tahapan ini dilakukan untuk pembuatan rancangan sistem dan alur pembuatan program yang akan dijadikan sebagai pedoman atau acuan dalam melakukan penelitian. Perancangan dapat berupa alur program, langkah-langkah atau flowchart [7]. Gambar 1 berikut ini menunjukkan gambaran lebih rinci mengenai tahapan metodologi penelitian. Pada Gambar 1 tahapan penelitian dimulai dengan pengumpulan data sampai dengan mendapatkan hasil evaluasi akhir dari kinerja suatu model untuk mendapatkan hasil yang terbaik.



Gambar 1. Tahapan Penelitian

2.2 Pengumpulan Data

Data diperoleh dengan menggunakan *web scraping* pada website perguruan tinggi yang membahas SDGs yang tergabung dalam SDGs Center dari Kementerian Perencanaan Pembangunan Nasional Republik Indonesia (Bappenas RI) dan portal berita membahas terkait SDGs. Data yang diambil pada website SDGs Center dari perguruan tinggi dan portal berita seperti: artikel, berita, kajian dan publikasi. Pada Tabel 1 terdapat daftar Instansi di Indonesia yang dilakukan pengumpulan data.

Building of Informatics, Technology and Science (BITS)
Volume 6, No 2, September 2024, Page: 1045−1058
ISSN 2684-8910 (media cetak)
ISSN 2685-3310 (media online)
DOI 10.47065/bits.v6i2.5807

Tabel 1. Tabel Instansi

No	Nama Instansi	Alamat web
1	Bappenas RI	https://sdgs.bappenas.go.id/berita
2	Universitas Airlangga	https://sdgscenter.unair.ac.id
3	Universitas Hasanuddin	https://sdgscenter.unhas.ac.id
4	Universitas Gajah Mada	https://sustainabledevelopment.ugm.ac.id
5	PBB Indonesia	https://indonesia.un.org
6	Liputan 6	https://liputan6.com
7	BBC News Indonesia	https://bbc.com
8	Antara News	https://antaranews.com

Pada Tabel 1 diatas dinyatakan bahwa 1 instansi dari pemerintahaan, 3 instansi dari kampus yang tergabung kedalam SDGs Center dari Kementerian Perencanaan Pembangunan Nasional Republik Indonesia, 1 dari perwakilan PBB di Indonesia dan 3 media berita online yang membahas terkait SDGs di Indonesia.

2.3 *Text Preprocessing*

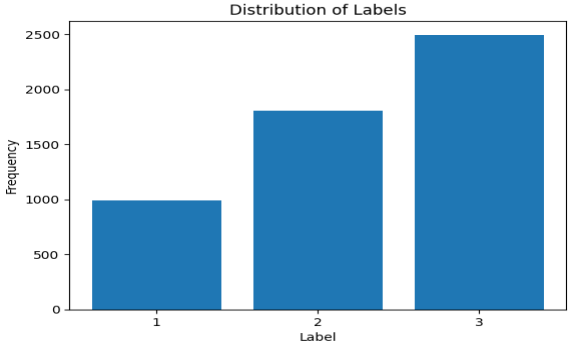
Tahap pra-pemrosesan terdiri dari pembersihan teks, yaitu, case folding, penghapusan angka dan tanda baca, *tokenizing*, *stemming*, *lemmatizing* dan penghapusan kata atau stopword [8]. Normalisasi kata menormalkan kata-kata yang memiliki makna yang sama tetapi karakter yang berbeda. Misalnya, menghapus huruf vokal untuk menduplikasi karakter, menambahkan huruf, dan mengadopsi bahasa yang tidak standar. Normalisasi kata menggunakan corpus dari [9]. Stemming adalah proses untuk menghilangkan imbuhan dan mengubahnya menjadi kata dasar menggunakan Sastrawi Library. Langkah terakhir adalah proses stopword removal dengan mengambil daftar stopword [10].

2.4 Labelling

Memberikan label pada paragraf yang telah dilakukan text preprocessing. Labeling terdiri dari 17 Tujuan SDGs yang dikelompokkan menjadi 3 kategori yaitu sosial, lingkungan dan ekonomi. Tabel 2 merupakan pengelompokkan 3 kategori berdasarkan SDGs, sedangkan pada Gambar 2 merupakan hasil distribusi dari kategori.

Tabel 2. Pengelompokkan Kategori Berdasarkan SDGs

No	Pilar	Goals	Keterangan
1	Sosial	SDGs 1	Tanpa Kemiskinan
		SDGs 4	Pendidikan Berkualitas
		SDGs 5	Kesetaraan Gender
		SDGs 10	Berkurangnya Kesenjangan
		SDGs 16	Perdamaian, Keadilan dan Kelembagaan yang Tangguh
2	Lingkungan	SDGs 6	Air Bersih dan Sanitasi Layak
		SDGs 7	Energi Bersih dan Terjangkau
		SDGs 11	Kota dan Permukiman yang Berkelanjutan
		SDGs 12	Konsumsi dan Produksi yang Bertanggung Jawab
		SDGs 13	Penanganan Perubahan Iklim
		SDGs 14	Ekosistem Lautan
		SDGs 15	Ekosistem Daratan
3	Ekonomi	SDGs 2	Tanpa Kelaparan
		SDGs 3	Kehidupan Sehat dan Sejahtera
		SDGs 8	Pekerjaan Layak dan Pertumbuhan Ekonomi
		SDGs 9	Industri, Inovasi dan Infrastruktur
		SDGs 17	Kemitraan untuk Mencapai Tujuan



Gambar 2. Distribusi Label

Building of Informatics, Technology and Science (BITS)
Volume 6, No 2, September 2024, Page: 1045−1058
ISSN 2684-8910 (media cetak)
ISSN 2685-3310 (media online)
DOI 10.47065/bits.v6i2.5807

Visulisasi pada Gambar 2 menunjukkan bahwa distribusi data pada kategori 1 yaitu kategori sosial memiliki data sebanyak 994 data, sedangkan pada kategori 2 yaitu kategori lingkungan memiliki data sebanyak 1806 data dan pada kategori 3 yaitu kategori ekonomi memiliki data sebanyak 2497 data. Gambar 2 menyatakan bahwa data yang dikumpulkan paling banyak pada kategori ekonomi dan data yang paling sedikit pada kategori sosial.

2.5 *Feature Extraction*

Ekstraksi fitur dilakukan untuk menemukan nilai dari fitur-fitur yang terdapat pada dokumen [11], yang membantu menentukan keberhasilan proses text mining. Jika nilai fitur tidak tepat, maka informasi yang dihasilkan dari text mining tidak dapat memenuhi kriteria yang diinginkan [12]. Pada penelitian ini digunakan ekstraksi fitur unigram TF-IDF (*Term Frequency-Inverse Document Frequency*) cara kerjanya adalah dengan memberikan bobot pada setiap kalimat dalam sebuah dokumen.

$$TF - IDF \ = \ TF \ x \ Log(1 + \frac{N}{df}) \quad (1)$$

TF-IDF adalah *Term Frequency* (TF) yang mengukur frekuensi sebuah kata dalam sebuah dokumen dan Inverse Document Frequency (IDF) yang merupakan penghitung frekuensi untuk sebuah term t dalam dokumen d, sedangkan DF adalah jumlah kemunculan term dalam set dokumen N [13].

2.6 Teknik Sampling

Identifikasi yang akurat dari sampel dapat dipengaruhi oleh distribusi sampel yang tidak seimbang di ketiga kelas. Di sini, metode oversampling diterapkan, yaitu SMOTE [14], yang didasarkan pada pengklasifikasi KNN [15] dengan K=5 dan menciptakan data sintetis [16] pada kelas minoritas (lihat Algorithm 1). Contoh-contoh dalam kelas riwayat penyakit jantung diambil secara berlebihan, sehingga subjek dalam kedua kelas tersebut terdistribusi secara seragam.

Alogritma SMOTE

Input : M (jumlah sampel di kelas minoritas), N (% rasio sampel minoritas sintetik untuk penyeimbangan kelas), K (jumlah tetangga terdekat);

Pilih secara acak subset S dari data kelas minoritas berukuran $S = \frac{N}{100} M$ (sampel sintetik dalam kelas minoritas) sedemikian rupa sehingga label kelas terdistribusi secara merata;

for $s_i \in$ S **do**

1) Temukan K tetangga terdekat;
2) Pilih secara acak salah satu KNN yaitu $\hat{s}_i$;
3) Hitung jarak k = $\hat{s}_i - s_i$ yang dipilih secara acak $\hat{s}_i$ dan instance s_i;
4) Sintentik baru dihasilkan sebagai $s_{syn} = s_i + \delta d_{ik}$, k (dimana δ = $rand\ (0,1)$ adalah angka acak diantara 0 dan 1);

end for

Ulangi langkah nomor 2–4 hingga proporsi kelas atau label minoritas yang diinginkan terpenuhi.

2.7 Klasifikasi

Penelitian ini menggunakan beberapa algoritma yaitu SVM, KNN, NB, DT dan MLP.

a. SVM

SVM adalah sebuah algoritma pembelajaran mesin yang diawasi [17]. SVM bekerja ketika diberikan data pelatihan bersama dengan label terkait. Setelah pelatihan selesai, jika diberikan kumpulan data, model akan memberikan label padanya. Mesin vektor pendukung bekerja dengan baik ketika berhadapan dengan masalah klasifikasi linier. Untuk klasifikasi, mesin ini membuat hyper-plane dengan memilih jarak maksimum antara titik-titik data yang berdekatan. Ini adalah teknik pembelajaran mesin yang fleksibel dan kuat yang digunakan untuk regresi dan klasifikasi. Ketika berhadapan dengan ruang dimensi tinggi, SVM bekerja dengan baik dan menawarkan akurasi yang sangat baik, dan hanya membutuhkan sedikit memori untuk pemrosesan.

Linear Function

$$f(x^1, x^2) \ = \ x^1 . x^2 \quad (2)$$

RBF Function

$$f(x^1, x^2) \ = \ \exp\left(-\gamma ||x^1 - x^2||^2\right) \quad (3)$$

b. KNN

Masalah-masalah dalam domain regresi dan klasifikasi dapat diselesaikan dengan menggunakan algoritma pembelajaran mesin yang terawasi ini [18]. KNN bekerja dengan menghitung jarak antara kueri dan semua contoh (K) yang lebih dekat dengan pertanyaan tersebut dan kemudian memilih label mayoritas yang berulang. Ini dapat dipahami dan diterapkan dengan sangat cepat. Ini memperlambat kinerjanya ketika ada peningkatan dalam sampel data. Algoritma ini menyimpan semua data pelatihan, sehingga algoritma ini menjadi algoritma yang mahal. Ini membutuhkan penyimpanan memori yang tinggi dibandingkan dengan metode lain.

Building of Informatics, Technology and Science (BITS)
Volume 6, No 2, September 2024, Page: 1045−1058
ISSN 2684-8910 (media cetak)
ISSN 2685-3310 (media online)
DOI 10.47065/bits.v6i2.5807

$$f(x,y) = \sqrt{\sum_{i=1}^{k}(x_i - y_i)^2} \quad (4)$$

c. MultinomialNB
Algoritma ini digunakan sebagai pengklasifikasi artikel berita. Ini adalah metode klasifikasi yang lazim digunakan ketika berhadapan dengan klasifikasi multi-kelas. Algoritma ini bekerja berdasarkan teknik probabilistik yang pertama kali diusulkan oleh Lewis [19] dan berakar pada teorema Bayes. Algoritma ini dapat berjalan dengan sangat efisien dengan dataset yang besar. Untuk masalah klasifikasi teks, Naïve Bayes digunakan sebagai standar karena memiliki waktu yang cepat dibandingkan dengan pengklasifikasi lainnya. Naïve Bayes juga digunakan untuk memecahkan masalah seperti deteksi spam. Karena kesederhanaannya, Naïve Bayes mengungguli banyak teknik klasifikasi yang lebih canggih.

$$P(c|x) = \frac{P(x|c).P(c)}{P(x)} \quad (5)$$

d. Decision Tree
Decision tree adalah struktur seperti pohon yang digunakan untuk mengelola kumpulan data yang besar. Struktur ini sering digambarkan sebagai diagram alir, dengan cabang luar yang mewakili hasil dan simpul dalam yang mewakili sifat-sifat set data. Pohon keputusan sangat populer karena efisien, dapat diandalkan, dan mudah dimengerti. Label kelas yang diproyeksikan untuk pohon keputusan berasal dari akar pohon. Langkah-langkah berikutnya dalam pohon diputuskan dengan membandingkan nilai atribut akar dengan informasi dalam catatan. Setelah lompatan pada simpul berikutnya, cabang yang cocok diikuti ke nilai yang ditunjukkan oleh hasil perbandingan. Entropi berubah ketika contoh pelatihan dibagi menjadi kelompok-kelompok yang lebih kecil dengan menggunakan simpul pohon keputusan. Pengukuran perubahan entropi ini adalah perolehan informasi [20].

$$Entropy = \sum_{i=1}^{c} -P_i . log_2(P_i) \quad (6)$$

$$Gain = 1 - \sum_{i=1}^{c}(P_i)^2 \quad (7)$$

e. Multi Layer Perceptron
Multilayer perceptron (MLP) adalah jenis jaringan syaraf tiruan yang terdiri dari beberapa lapisan. Perceptron tunggal hanya dapat menyelesaikan masalah linier, tetapi MLP lebih cocok untuk contoh-contoh nonlinier. MLP digunakan untuk menangani masalah yang kompleks. Jaringan syaraf tiruan feedforward dengan banyak lapisan adalah contoh dari MLP [21].
Fungsi aktivasi lain di luar fungsi langkah biasanya digunakan oleh MLP. Neuron lapisan yang terkubur sering melakukan fungsi sigmoid. Seperti halnya fungsi langkah, transisi yang mulus daripada batas keputusan yang kaku dihasilkan dengan menggunakan fungsi sigmoid [22]. Dalam MLP, pembelajaran juga terdiri dari penyesuaian bobot perceptron untuk mendapatkan kesalahan serendah mungkin. Hal ini dilakukan melalui teknik backpropagation, yang mengurangi MSE.

$$E = \frac{1}{2}\sum_{k}(\gamma_k - y_k)^2 \quad (6)$$

Metode yang digunakan melibatkan pelatihan dan pengujian beberapa pengklasifikasi yaitu SVM, KNN, NB, DT dan MLP. Parameter dari setiap algoritma klasifikasi memanfaatkan Grid Search untuk mencari parameter terbaik, yang ditunjukkan pada Tabel 3.

Tabel 3. Daftar Hasil Parameter

Model	Parameter
SVM	C=100, gamma=1, kernel='linear'
SVM + SMOTE	C=10, gamma=1, kernel='linear'
KNN	n_neighbors=16
KNN + SMOTE	n_neighbors=3
NB	alpha=0.01
NB + SMOTE	alpha=1e-05
DT	criterion='entropy', max_depth=50
DT + SMOTE	criterion='entropy', max_depth=90
MLP	hidden_layer_sizes=10, max_iter=100, activation='relu', solver='sgd',random_state=25, learning_rate='constant', learning_rate_init=0.5
MLP + SMOTE	hidden_layer_sizes=10, max_iter=100, activation='relu', solver='sgd',random_state=25, learning_rate='constant', learning_rate_init=0.5

2.8 Evaluasi Hasil

Hasil performa dari classifier digunakan untuk mengetahui seberapa baik model klasifikasi tersebut. Hasil kinerja dalam klasifikasi umumnya dilakukan dengan menggunakan akurasi. Nilai akurasi klasifikasi mendefinisikan jika

Building of Informatics, Technology and Science (BITS)
Volume 6, No 2, September 2024, Page: 1045−1058
ISSN 2684-8910 (media cetak)
ISSN 2685-3310 (media online)
DOI 10.47065/bits.v6i2.5807

seluruh himpunan label yang diprediksi untuk sampel sesuai dengan himpunan label yang sebenarnya, maka akurasi subset adalah 1, sedangkan jika tidak sama, maka 0 [23].

$$accuracy\ (y, \hat{y}) = \frac{1}{n_{samples}} \sum_{i=0}^{n_{samples}} (\hat{y}_i = y_{i)} \quad (9)$$

Di mana $\hat{y}_i$ adalah nilai prediksi dari sampel i dan $y_{i)}$ adalah nilai aktual.

3. HASIL DAN PEMBAHASAN

3.1 Koleksi Data

Data diperoleh dengan menggunakan teknik web scraping pada website 8 instansi yang membahas seputar SDGs, website 8 instansi tertera pada Tabel 1. Data yang dikumpulkan dari website 8 instansi berjumlah 5139 data. Tabel 4 terdapat 1 data yang bersumber dari instansi Bappenas RI yang belum dilakukan Text Preprocessing.

Tabel 4. Contoh Beberapa Koleksi Data

No	Instansi	Artikel
1	Bappenas RI	Sekretariat Tujuan Pembangunan Berkelanjutan (TPB/SDGs) Kementerian Bappenas bersama-sama dengan Kementerian/Lembaga, Pemerintah Daerah, Akademisi dan Organisasi Masyarakat Sipil (OMS/CSO) melakukan diskusi tentang struktur, sistematika, mekanisme dan pengorganisasian penyusunan Pedoman Rencana Aksi Nasional (RAN) dan Rencana Aksi Daerah (RAD) TPB/SDGs di Jakarta (8/12).

3.2 Text Preprocessing

Data yang telah dikumpulkan memiliki angka yang tidak mempunya makna, tanda baca, karakter khusus, symbol emoticon, penulisan kata yang kurang lengkap, sehinggal perlu dilakukan pemrosesan untuk menghasilkan data yang bersih dan siap diolah. Pada Tabel 5 menunjukkan hasil text prepocessing setiap langkah-langkah.

Tabel 5. Tahap Text Preprocessing

No	Artikel	Case Folding	Remove Punctuation	Tokenizing	Stopword	Stemming	Lemmatizing
1	Sekretariat Tujuan Pembangunan Berkelanjutan (TPB/SDGs) Kementerian Bappenas bersama-sama dengan Kementerian/Lembaga, Pemerintah Daerah, Akademisi dan Organisasi Masyarakat Sipil (OMS/CSO) melakukan diskusi tentang struktur, sistematika, mekanisme dan pengorganisasian penyusunan Pedoman Rencana Aksi	sekretariat tujuan pembangunan berkelanjutan (tpb/sdgs) kementerian bappenas bersama-sama dengan kementerian/ lembaga, pemerintah daerah, akademisi dan organisasi masyarakat sipil (oms/cso) melakukan diskusi tentang struktur, sistematika, mekanisme dan pengorganis	sekretariat tujuan pembangunan berkelanjutan tpbsdgs kementerian bappenas bersama-sama dengan kementerianlembaga pemerintah daerah akademisi dan organisasi masyarakat sipil omscso melakukan diskusi tentang struktur sistematika mekanisme dan pengorganisasian penyusunan pedoman rencana aksi nasional ran	['sekretariat', 'tujuan', 'pembangunan', 'berkelanjutan', 'tpbsdgs', 'kementerian', 'bappenas', 'bersama', 'sama', 'dengan', 'kementerianlembaga', 'pemerintah', 'daerah', 'akademisi', 'dan', 'organisasi', 'masyarakat', 'sipil', 'omscso', 'melakukan', 'diskusi', 'tentang',	['tujuan', 'pembangunan', 'berkelanjutan', 'tpbsdgs', 'bappenas', 'bersama', , 'kementerianlembaga', 'pemerintah', 'daerah', 'akademisi', 'organisasi', 'sipil', 'omscso', 'diskusi', 'struktur', 'sistematika', 'mekanisme', 'organisasi', 'susun', 'pedoman', 'rencana', 'aksi', 'ran', 'rencana', 'aksi',	['tuju', 'bangun', 'lanjut', 'tpbsdgs', 'bappenas', 'sama', 'kementerianlembaga', , 'perintah', 'daerah', 'akademisi', , 'organisasi', , 'sipil', 'omscso', 'diskusi', 'struktur', 'sistematika', 'mekanisme', 'organisasi', , 'susun', 'pedoman', 'rencana', 'aksi', 'ran', 'rencana', 'aksi',	tuju bangun lanjut tpbsdgs bappenas sama kementerian lembaga perintah daerah akademisi organisasi sipil omscso diskusi struktur sistematika mekanisme organisasi susun pedoman rencana aksi ran rencana aksi daerah rad tpbsdgs

Building of Informatics, Technology and Science (BITS)
Volume 6, No 2, September 2024, Page: 1045−1058
ISSN 2684-8910 (media cetak)
ISSN 2685-3310 (media online)
DOI 10.47065/bits.v6i2.5807

	Nasional (RAN) dan Rencana Aksi Daerah (RAD) TPB/SDGs di Jakarta (8/12).	asian penyusunan pedoman rencana aksi nasional (ran) dan rencana aksi daerah (rad) tpb/sdgs di jakarta (8/12).	dan rencana aksi daerah rad tpbsdgs di jakarta	'struktur', 'sistematika', 'mekanisme', 'dan', 'pengorganisasian', 'penyusunan', 'pedoman', 'rencana', 'aksi', 'nasional', 'ran', 'dan', 'rencana', 'aksi', 'daerah', 'rad', 'tpbsdgs', 'di', 'jakarta']	anisasian', 'penyusunan', 'pedoman', 'rencana', 'aksi', 'ran', 'rencana', 'aksi', 'daerah', 'rad', 'tpbsdgs']	'daerah', 'rad', 'tpbsdgs']

3.3 Labelling

Setelah melakukan text preprocessing, tahapan selanjutnya adalah labelling data. Data yang masing-masing memiliki goals pada SDGs dilakukan labelling ke dalam pengelompokkan kategori sosial, lingkungan dan ekonomi. Pada Tabel 6 dibawah ini menunjukkan hasil dari labelling data yang mengacu pada goals SDGs pada setiap artikel atau berita yang terdapat pada instansi tersebut.

Tabel 6. Labelling Data

No	Instansi	Artikel	SDGs	Label
1	Bappenas RI	Sekretariat Tujuan Pembangunan Berkelanjutan (TPB/SDGs) Kementerian Bappenas bersama-sama dengan Kementerian/Lembaga, Pemerintah Daerah, Akademisi dan Organisasi Masyarakat Sipil (OMS/CSO) melakukan diskusi tentang struktur, sistematika, mekanisme dan pengorganisasian penyusunan Pedoman Rencana Aksi Nasional (RAN) dan Rencana Aksi Daerah (RAD) TPB/SDGs di Jakarta (8/12).	17	Ekonomi
2	Innovillage	BANDUNG, Telkom University , Social Project Innovilage harus memiliki unsur kemanfaatan untuk masyarakat, solusi digital, kolaborasi dan sustainable serta memiliki keterkaitan dengan Sustainable Development Goals (SDGs) mendorong terciptanya kemanfaatan sosial dan peningkatan ekonomi yang terukur. Sustainable Development Goals (SDGs) merupakan suatu rencana aksi global yang disepakati oleh para pemimpin dunia, termasuk Indonesia, guna mengakhiri kemiskinan, mengurangi kesenjangan dan melindungi lingkungan. Terdapat total 17 SDGs salah satu kategori Sustainable Development Goals (SDGs) adalah Infrastruktur, Industri dan Inovasi. Dengan membangun infrastruktur yang tangguh dan andal, mendukung industrialisasi yang inklusif dan berkelanjutan, serta membantu perkembangan inovasi. Target dari Tujuan Pembangunan Berkelanjutan yang kesembilan adalah Membangun infrastruktur yang berkualitas, dapat diandalkan, berkelanjutan dan tahan lama, mendorong industrialisasi yang inklusif dan berkelanjutan, meningkatkan akses industri skala kecil dan usaha skala kecil lainnya, khususnya di negara-negara berkembang, meningkatkan mutu infrastruktur dan menambahkan komponen pada industri agar dapat berkelanjutan, menambah penelitian ilmiah, dan meningkatkan kemampuan teknologi dari sektor industri di semua negara, khususnya negara berkembang. Contoh social project Innovilage	9	Ekonomi

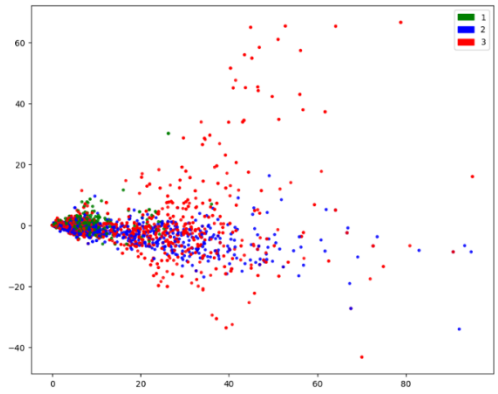
Building of Informatics, Technology and Science (BITS)
Volume 6, No 2, September 2024, Page: 1045−1058
ISSN 2684-8910 (media cetak)
ISSN 2685-3310 (media online)
DOI 10.47065/bits.v6i2.5807

		pada tahun sebelumnya yang termasuk kedalam Sustainable Development Goals (SDGs) Go-Subak social project Innovilage 2021. Sistem Go-Subak hadir untuk memberikan kebermanfaatan bagi para petani di Desa Jatiluwih dalam menciptakan infrastruktur, industri dan inovasi baru yang dapat meringankan serta mengatasi permasalahan yang sedang dialami oleh para petani khususnya dalam mempermudah pengukuran pH dan kelembaban tanah serta proses penjadwalan pemupukan padi yang cepat dan praktis dalam jangka waktu yang panjang.		
3	UGM	Siapa kami, Indonesia dikenal sebagai negara agraris karena perekonomiannya sebagian besar bertumpu pada budidaya tanaman pangan dan lahan pertanian. Pertanian di Indonesia memanfaatkan 30% luas lahan negara, menyerap tenaga kerja 31,74% (38,29 juta orang), dan berkontribusi terhadap 13,7% PDB. Di dunia, negara kita termasuk dalam lima besar negara dengan produksi pertanian tertinggi setelah China, India, Amerika, dan Brazil. , Untuk memaksimalkan potensi pertanian Indonesia, AIC-UGM (Pusat Inovasi Agroteknologi Universitas Gadjah Mada) yang juga dikenal sebagai PIAT (Pusat Inovasi Agroteknologi) berkomitmen untuk menjadi pusat penelitian pertanian terpadu yang diakui secara internasional dengan mendukung pengembangan dan implementasi pertanian terpadu. inovasi teknologi yang dapat langsung dimanfaatkan oleh masyarakat, pemerintah, swasta, dan akademisi. Tema penelitian utama di pusat ini meliputi pengelolaan sumber daya pangan berkelanjutan dan konservasi sumber daya alam berkelanjutan. Penekanan kuat adalah pada inovasi teknologi karena proses pertanian di masa depan memerlukan penerapan teknologi modern. Penelitian kami mempertimbangkan interaksi antara tumbuhan, teknologi, hewan, lingkungan, dan komunitas untuk mendorong pendekatan yang komprehensif. Saat ini, AIC mengelola lahan produktif seluas 35 hektar di Berbah, Sleman dan 151 hektar di Mangunan-Girirejo, Bantul, Yogyakarta., Klik di tautan di bawah untuk mempelajari lebih lanjut tentang sorotan proyek kami., Bank Gen Sayuran, Rumah Pengelolaan Sampah Inovatif	2	Ekonomi
4	Bappenas RI	Hari Surya merupakan salah satu hari peringatan yang jarang diketahui oleh masyarakat. Hari Surya Sedunia diperingati setiap tanggal 3 Mei setiap tahunnya. Peringatan Hari Surya ini adalah sebuah bentuk kampanye atas energi matahari yang merupakan energi ramah lingkungan dan terbarukan.	7	Lingkungan
5	Unair	UNAIR NEWS Kementerian Pemberdayaan Perempuan Badan Eksekutif Mahasiswa (BEM) Universitas Airlangga (UNAIR) mengadakan kajian rutin bertajuk Karin Batch 3 bertajuk Alpha Female: Your Very First Weapon Should be Your Voice pada Sabtu (31/8/2019). Kegiatan di Ruang Kuliah B FKG UNAIR ini merupakan tindak lanjut kajian pemberdayaan perempuan dari Karin Angkatan 1 yang membahas mengenai kekerasan seksual. Lebih lanjut, penelitian yang diikuti lebih dari 200 mahasiswa ini, juga bertujuan untuk memperkenalkan Help Center sebagai wadah konseling yang siap membantu mahasiswa dalam mengatasi permasalahan pribadinya.	10	Sosial

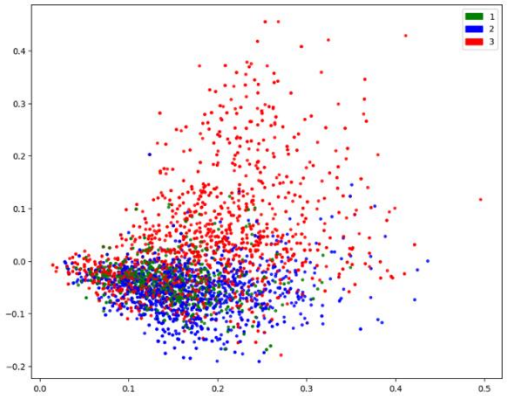
3.4 *Feature Extraction*

Data penelitian sudah mendapatkan labelling yaitu ekonomi, lingkungan dan sosial. Data yang sudah mempunyai label dilakukan *feature extraction* menggunakan TF-IDF untuk diberikan pembobotan teks yang terdapat pada sebuah dokumen. Hasil pembobotan teks tersebut, digunakan suatu metode *Latent Semantic Analysis* (LSA) yang disajikan dalam bentuk grafi scatter plot guna untuk mengekstrasi fitur pada data disetiap kategori. Gambar 3 menunjukkan sebaran hasil vektor pembobotan teks pada metode TF, sedangkan Gambar 4 menunjukkan sebaran hasil vektor pembobotan teks pada metode IDF.

Building of Informatics, Technology and Science (BITS)
Volume 6, No 2, September 2024, Page: 1045−1058
ISSN 2684-8910 (media cetak)
ISSN 2685-3310 (media online)
DOI 10.47065/bits.v6i2.5807



Gambar 3. Grafik pada Term Frequency (TF)



Gambar 4. Grafik Pada Inverse Document Frequency (IDF)

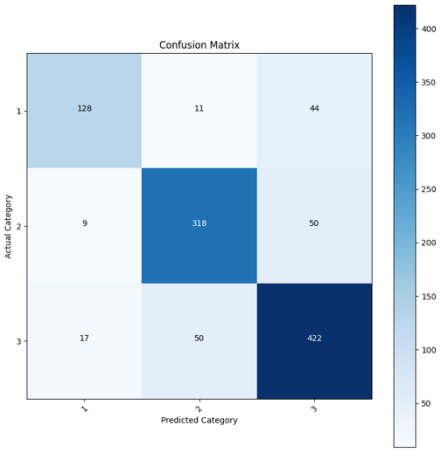
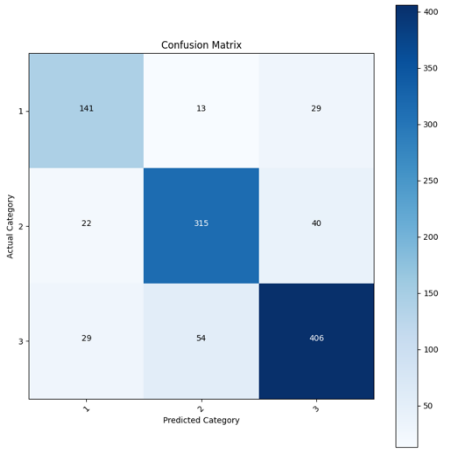
3.5 Teknik Sampling

Pada penelitian ini, data yang diambil melalui web scraping terjadi ketidakseimbangan data atau *imbalanced data.* Data tersebut harus dilakukan teknik sampling dengan tujuan supaya data menjadi seimbang atau balanced. Algoritma yang digunakan pada penelitian ini yaitu algoritma SMOTE dengan metode oversampling. Pada metode oversampling, kategori minoritas yaitu pada kategori 1 dan 2, dilakukan penambahan porsi data supaya jumlahnya sama dengan jumlah data pada kategori 3.

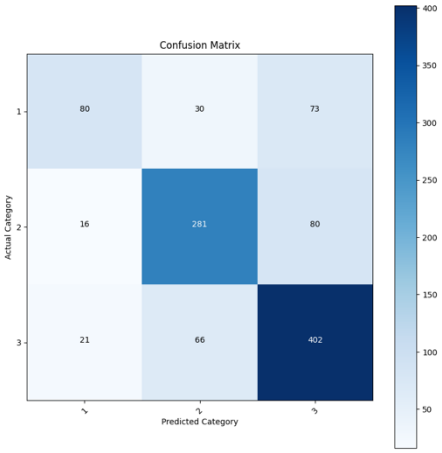
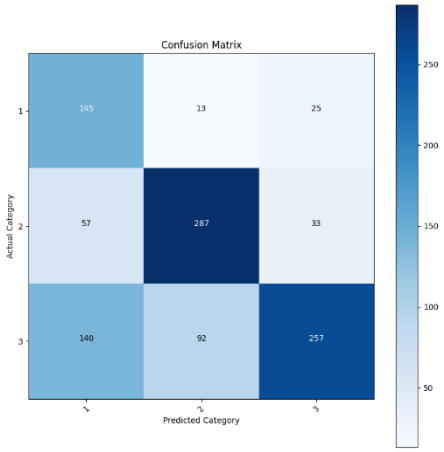
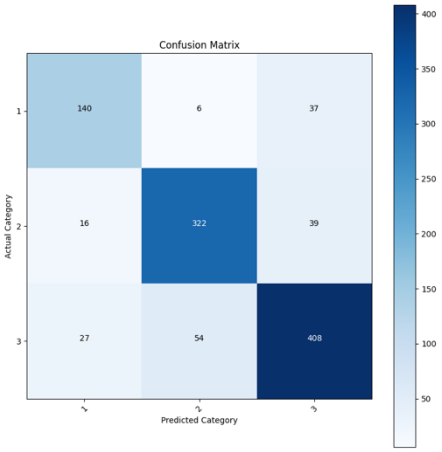
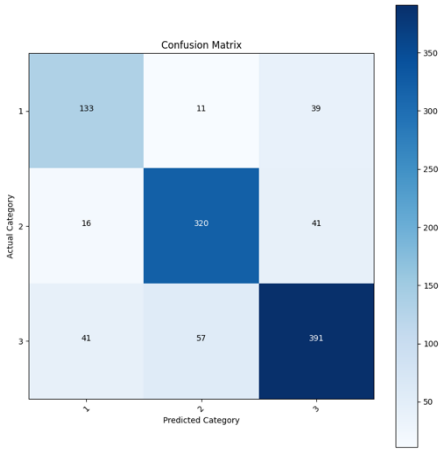
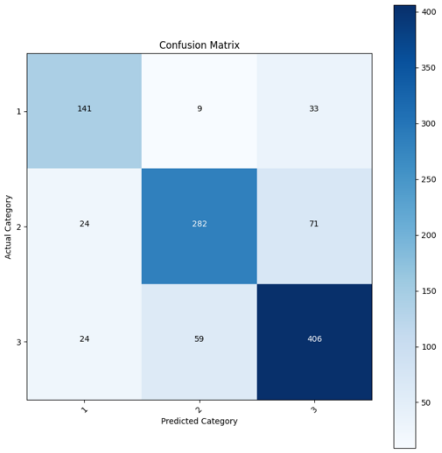
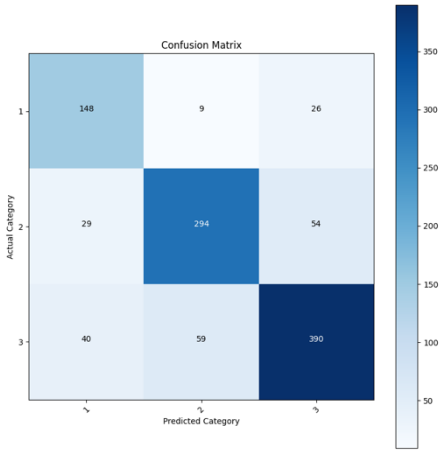
3.6 Klasifikasi

Pada penelitian ini, melakukan pelatihan data yang di implementasikan pada 5 algoritma klasifikasi yaitu: SVM, KNN, MultinomialNB, DT dan MLP. Hasil pengklasifikasikan menghasilkan confusion matrix yang menjadi instrumen kunci untuk melakukan analisis dari prediksi pada 5 algoritma klasifikasi. Pada Tabel 7 tertera hasil confusion matrix dari implementasi 5 algoritma klasifikasi tanpa teknik sampling SMOTE dengan menggunakan teknik sampling SMOTE.

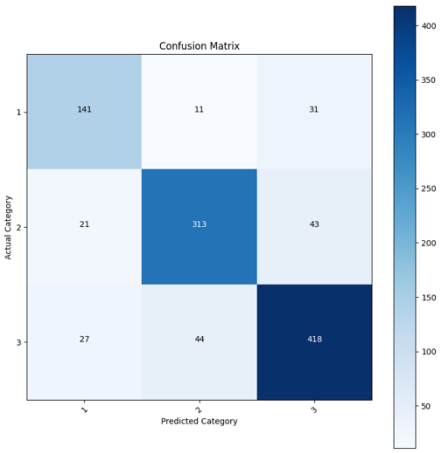
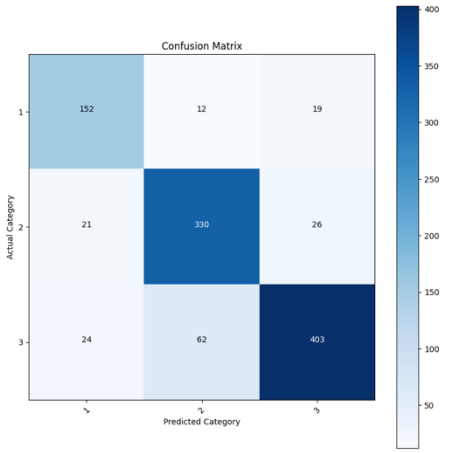
Tabel 7. Hasil Confusion Matrix

No	Algoritma	Confusion Matrix Tanpa SMOTE	Confusion Matrix dengan SMOTE
1	SVM		

Building of Informatics, Technology and Science (BITS)
Volume 6, No 2, September 2024, Page: 1045−1058
ISSN 2684-8910 (media cetak)
ISSN 2685-3310 (media online)
DOI 10.47065/bits.v6i2.5807

		Gambar diatas menunjukkan hasil grafik confusion matrix pada algoritma SVM tanpa teknik sampling SMOTE.	Gambar diatas menunjukkan hasil grafik confusion matrix pada algoritma SVM dengan teknik sampling SMOTE.
2	KNN		
		Gambar diatas menunjukkan hasil grafik confusion matrix pada algoritma KNN tanpa teknik sampling SMOTE.	Gambar diatas menunjukkan hasil grafik confusion matrix pada algoritma KNN dengan teknik sampling SMOTE.
3	MultinomialNB		
		Gambar diatas menunjukkan hasil grafik confusion matrix pada algoritma MultinomialNB tanpa teknik sampling SMOTE.	Gambar diatas menunjukkan hasil grafik confusion matrix pada algoritma MultinomialNB dengan teknik sampling SMOTE.
4	DT		
		Gambar diatas menunjukkan hasil grafik confusion matrix pada algoritma DT tanpa teknik sampling SMOTE.	Gambar diatas menunjukkan hasil grafik confusion matrix pada algoritma DT dengan teknik sampling SMOTE.

Building of Informatics, Technology and Science (BITS)
Volume 6, No 2, September 2024, Page: 1045−1058
ISSN 2684-8910 (media cetak)
ISSN 2685-3310 (media online)
DOI 10.47065/bits.v6i2.5807

5	MLP		
		Gambar diatas menunjukkan hasil grafik confusion matrix pada algoritma MLP tanpa teknik sampling SMOTE.	Gambar diatas menunjukkan hasil grafik confusion matrix pada algoritma MLP dengan teknik sampling SMOTE.

3.7 Evaluasi Akhir

Setelah menyelesaikan pra-pemrosesan, klasifikasi dilakukan dengan beberapa pengklasifikasi pembelajaran dengan teknik sampling menggunakan algoritma SMOTE metode oversampling dan Grid Search. Tabel 8 di bawah ini, menampilkan hasil pengujian 5 algoritma klasifikasi.

Tabel 8. Daftar Hasil Pengujian

Teknik	Akurasi
SVM + Grid Search	83%
SVM + SMOTE + Grid Search	82%
KNN + Grid Search	73%
KNN + SMOTE + Grid Search	66%
NB + Grid Search	83%
NB + SMOTE + Grid Search	80%
DT + Grid Search	79%
DT + SMOTE + Grid Search	79%
MLP + Grid Search	83%
MLP + SMOTE + Grid Search	84%

Berdasarkan hasil pengujian content-based filtering menggunakan beberapa algoritma di atas pada sistem rekomendasi SDGs dapat dilihat bahwa teknik SVM, NB dan MLP dengan teknik Grid Search menghasilkan akurasi yang sama yaitu 83%, namun ketika ditambahkan teknik sampling MLP mengalami kenaikan akurasi 1% menjadi 84%, sedangkan SVM dan NB mengalami penurunan masing masing 1% dan 3% menjadi 82% dan 80%.

4. KESIMPULAN

Berdasarkan hasil penelitian dengan menggunakan 5 teknik algoritma klasifikasi yaitu SVM, KNN, NB, DT dan MLP dengan teknik sampling SMOTE dan pencarian parameter menggunakan Grid Search, didapat nilai akurasi terbaik pada algoritma MLP dengan teknik sampling SMOTE dan pencarian parameter dengan Grid Search. Algoritma MLP menunjukkan dapat melakukan pemecahan masalah pada *Sustainable Development Goals* (SDGs) pada pengelompokkan kategori sosial, lingkungan dan ekonomi berdasarkan data yang diambil pada sumber berita, artikel dan hasil kajian pada instansi pemerintahaan, kampus, maupun portal berita yang membahas SDGs. Algoritma MLP dengan teknik SMOTE memperoleh nilai akurasi terbaik yaitu 84%. Menunjukkan bahwa, implementasi teknik SMOTE dapat melakukan optimalisasi algoritma terbukti algoritma MLP mengalami kenaikan sebesar 1% yang dari nilai akurasi 83% tanpa menggunakan teknik sampling SMOTE. Rekomendasi untuk penelitian selanjutnya: pertama, melakukan eksplorasi metode klasifikasi teks lainnya selain metode klasifikasi diatas. Kedua, melakukan eksplorasi word embedding dengan menggunakan FastText dan GloVe. Ketiga, melakukan eksplorasi terkait seleksi fitur seperti Chi-Square dan PCA. Dan terakhir, melakukan pengkajian teknik sampling selain SMOTE lainnya.

REFERENCES

Building of Informatics, Technology and Science (BITS)
Volume 6, No 2, September 2024, Page: 1045−1058
ISSN 2684-8910 (media cetak)
ISSN 2685-3310 (media online)
DOI 10.47065/bits.v6i2.5807

[1] A. S. Rusydiana and M. H. Khalifah, "Islamic Social Instruments and Sustainable Development Goals (SDGs) Framework," *Management and Sustainability*, vol. 2, no. 2, Jan. 2024, doi: 10.58968/ms.v2i2.382.
[2] L. Rosidah and P. Dellia, "Library Book Recommendation System Using Content-Based Filtering," *Internet of Things and Artificial Intelligence Journal*, vol. 4, no. 1, pp. 42–65, Feb. 2024, doi: 10.31763/iota.v4i1.693.
[3] J. Leander and A. Wicaksana, "Optimizing a Personalized Movie Recommendation System with Support Vector Machine and Content-Based Filtering," *Journal of System and Management Sciences*, vol. 14, no. 1, pp. 490–501, 2024, doi: 10.33168/JSMS.2024.0128.
[4] A. R. Syulistyo, V. M. Agustin, and D. Puspitasari, "Predicting News Article Popularity with Multi Layer Perceptron Algorithm," *Journal of Applied Intelligent System*, vol. 7, no. 2, pp. 193–205, Sep. 2022, doi: 10.33633/jais.v7i2.6826.
[5] V. Rolanda, T. S. Gunawan, and Wanayumini, "Content-Based Filtering Recommendation System Using Categories Search Engine," *International Journal of Research in Vocational Studies (IJRVOCAS)*, vol. 2, no. 4, pp. 120–125, Jan. 2023, doi: 10.53893/ijrvocas.v2i4.177.
[6] Achmad Baroqah Pohan, Irmawati, and A. Kurniasih, "Optimization of Classification Algorithm with GridSearchCV and Hyperparameter Tuning for Sentiment Analysis of the Nusantara Capital City," *Journal of Artificial Intelligence and Engineering Applications (JAIEA)*, vol. 3, no. 3, pp. 808–814, Jun. 2024, doi: 10.59934/jaiea.v3i3.514.
[7] Z. Alhaq, A. Mustopa, S. Mulyatun, and J. D. Santoso, "PENERAPAN METODE SUPPORT VECTOR MACHINE UNTUK ANALISIS SENTIMEN PENGGUNA TWITTER," *Journal of Information System Management (JOISM)*, vol. 3, no. 2, pp. 44–49, Jul. 2021, doi: 10.24076/joism.2021v3i2.558.
[8] A. Habberrih and M. Ali Abuzaraida, "Sentiment Analysis of Libyan Dialect Using Machine Learning with Stemming and Stop-words Removal," in *5TH INTERNATIONAL CONFERENCE ON COMMUNICATION ENGINEERING AND COMPUTER SCIENCE (CIC-COCOS'24)*, Cihan University-Erbil, 2024, pp. 259–264. doi: 10.24086/cocos2024/paper.1171.
[9] M. Ibrohim and I. Budi, "Multi-label Hate Speech and Abusive Language Detection in Indonesian Twitter," Jul. 2019, pp. 46–57. doi: 10.18653/v1/W19-3506.
[10] L. A. Fitriana, A. Mustopa, M. R. Firdaus, and R. Dahlia, "Application of the Finite State Automata (FSA) Method in Indonesian Stemming using the Nazief & Adriani Algorithm," *SISTEMASI*, vol. 13, no. 3, p. 1125, May 2024, doi: 10.32520/stmsi.v13i3.4038.
[11] A. Nurkasanah and M. Hayaty, "Feature Extraction using Lexicon on the Emotion Recognition Dataset of Indonesian Text," *Ultimatics : Jurnal Teknik Informatika*, vol. 14, no. 1, 2022.
[12] A. Maiti, A. Abarda, and M. Hanini, "The impact of feature extraction techniques on the performance of text data classification models," *Indonesian Journal of Electrical Engineering and Computer Science*, vol. 35, no. 2, p. 1041, Aug. 2024, doi: 10.11591/ijeecs.v35.i2.pp1041-1052.
[13] A. F. Abdul Fadlil, I. Riadi, and F. Andrianto, "Improving Sentiment Analysis in Digital Marketplaces through SVM Kernel Fine-Tuning," *International Journal of Computing and Digital Systems*, vol. 15, no. 1, pp. 159–171, Jul. 2024, doi: 10.12785/ijcds/160113.
[14] L. D. Cahya, A. Luthfiarta, J. I. T. Krisna, S. Winarno, and A. Nugraha, "Improving Multi-label Classification Performance on Imbalanced Datasets Through SMOTE Technique and Data Augmentation Using IndoBERT Model," *Jurnal Nasional Teknologi dan Sistem Informasi*, vol. 9, no. 3, pp. 290–298, Jan. 2024, doi: 10.25077/TEKNOSI.v9i3.2023.290-298.
[15] P. Cunningham and S. J. Delany, "k-Nearest Neighbour Classifiers: 2nd Edition (with Python examples)," Apr. 2020, doi: 10.1145/3459665.
[16] E. Dritsas, N. Fazakis, O. Kocsis, K. Moustakas, and N. Fakotakis, "Optimal Team Pairing of Elder Office Employees with Machine Learning on Synthetic Data," Apr. 2021. doi: 10.1109/IISA52424.2021.9555511.
[17] R. Awad Mariette and Khanna, "Support Vector Machines for Classification," in *Efficient Learning Machines: Theories, Concepts, and Applications for Engineers and System Designers*, Berkeley, CA: Apress, 2015, pp. 39–66. doi: 10.1007/978-1-4302-5990-9_3.
[18] A. K. Narayan Vipul and Daniel, "RBCHS: Region-Based Cluster Head Selection Protocol in Wireless Sensor Network," in *Proceedings of Integrated Intelligence Enable Networks and Computing*, V. B. and B. V. and C. R. G. Singh Mer Krishan Kant and Semwal, Ed., Singapore: Springer Singapore, 2021, pp. 863–869.
[19] D. D. Lewis, "Naive (Bayes) at forty: The independence assumption in information retrieval," in *Machine Learning: ECML-98*, C. Nédellec and C. Rouveirol, Eds., Berlin, Heidelberg: Springer Berlin Heidelberg, 1998, pp. 4–15.
[20] C. M. Bhatt, P. Patel, T. Ghetia, and P. L. Mazzeo, "Effective Heart Disease Prediction Using Machine Learning Techniques," *Algorithms*, vol. 16, no. 2, Feb. 2023, doi: 10.3390/a16020088.
[21] M. D. Mohanty and M. N. Mohanty, "Chapter 5 - Verbal sentiment analysis and detection using recurrent neural network," in *Advanced Data Mining Tools and Methods for Social Computing*, S. De, S. Dey, S. Bhattacharyya, and S. Bhatia, Eds., in Hybrid Computational Intelligence for Pattern Analysis. , Academic Press, 2022, pp. 85–106. doi: https://doi.org/10.1016/B978-0-32-385708-6.00012-6.

Building of Informatics, Technology and Science (BITS)
Volume 6, No 2, September 2024, Page: 1045−1058
ISSN 2684-8910 (media cetak)
ISSN 2685-3310 (media online)
DOI 10.47065/bits.v6i2.5807

[22] T. Menzies, E. Kocagüneli, L. Minku, F. Peters, and B. Turhan, "Chapter 24 - Using Goals in Model-Based Reasoning," in *Sharing Data and Models in Software Engineering*, 2015, pp. 321–353. doi: https://doi.org/10.1016/B978-0-12-417295-1.00024-2.

[23] P. Atanasova, "A Diagnostic Study of Explainability Techniques for Text Classification," in *Accountable and Explainable Methods for Complex Reasoning over Text*, Cham: Springer Nature Switzerland, 2020, pp. 155–187. doi: 10.1007/978-3-031-51518-7_7.