

Model Klasifikasi Multilabel pada Publikasi Penelitian SDG dengan Pendekatan Multilevel dan Hierarki

Berliana Sugiarti Putri¹, Lya Hulliyatus Suadaa^{1,2}, Efri Diah Utami²

¹ Program Studi Komputasi Statistik, Politeknik Statistika STIS, Jakarta Timur, DKI Jakarta 13330, Indonesia

² Unit Kajian Sains Data, Politeknik Statistika STIS, Jakarta Timur, DKI Jakarta 13330, Indonesia

[Diserahkan: 23 September 2024, Direvisi: 11 Desember 2024, Diterima: 24 Februari 2025]

Penulis Korespondensi: Lya Hulliyatus Suadaa (email: lya@stis.ac.id)

INTISARI — Kemajuan lini penelitian, ditandai dengan meningkatnya jumlah publikasi penelitian, menyebabkan identifikasi implementasi publikasi penelitian makin sulit dilakukan, khususnya terkait SDG. Saat ini, pengategorian publikasi penelitian ke dalam level SDG belum dilakukan. Hal tersebut dibutuhkan Pusat Penelitian dan Pengabdian Masyarakat Politeknik Statistika STIS untuk pemantauan aksi nyata SDG oleh dosen. Penelitian ini bertujuan untuk mengimplementasikan dan mengevaluasi metode transformasi masalah dan algoritma klasifikasi pembelajaran mesin dengan pendekatan multilevel dan hierarki untuk mengategorikan publikasi penelitian ke dalam level SDG. Metode transformasi masalah yang digunakan adalah *binary relevance* (BR), *label powerset* (LP), dan *classifier chains* (CC). Selain itu, algoritma klasifikasi pembelajaran mesin yang digunakan adalah *logistic regression* (LR) dan *support vector machine* (SVM). Adapun masukan yang digunakan adalah judul, abstrak, serta judul dan abstrak. Model filter terbaik yang mengklasifikasikan data ke dalam SDG dan non-SDG adalah model dengan masukan judul dan SVM dengan akurasi 0,8634. Model level terbaik yang mengklasifikasikan data ke level SDG adalah model dengan masukan judul, metode LP, dan SVM dengan pendekatan multilevel. Model level mengklasifikasikan data ke dalam empat pilar, tujuan, target, dan indikator SDG dengan akurasi sebesar 0,8067, 0,7501, 0,6792, dan 0,6194. Hasil penelitian menunjukkan bahwa masukan judul memiliki akurasi terbaik dibandingkan masukan lainnya yang memiliki informasi lebih lengkap. Hal tersebut disebabkan penggunaan dua bahasa secara sekaligus, yaitu bahasa Inggris dan bahasa Indonesia. Dengan demikian, penelitian di masa depan dapat memodifikasi model dengan masukan satu bahasa saja untuk mengoptimalkan proses *term frequency-inverse document frequency* (TF-IDF) agar makna kata dari masing-masing bahasa tidak dianggap kata penting yang berbeda.

KATA KUNCI — Pembelajaran Mesin, Metode Transformasi Masalah, Pendekatan Multilevel, Pendekatan Hierarki.

I. PENDAHULUAN

Dunia, melalui komitmen negara anggota Perserikatan Bangsa-Bangsa (PBB), telah menyepakati *sustainable development goals* (SDG) sebagai tujuan utama dalam meningkatkan kualitas hidup masyarakat. Hal tersebut menunjukkan bahwa kualitas hidup masyarakat adalah hal yang urgen secara universal [1]. SDG mendorong pembangunan berkelanjutan di Indonesia. Berdasarkan data yang dirilis oleh Badan Pusat Statistik (BPS), terdapat penurunan persentase penduduk miskin di Indonesia sebesar 1,86%, dari sebesar 11,22% pada tahun 2015 menjadi 9,36% pada tahun 2023 [2]. SDG juga mendorong peningkatan angka partisipasi kasar (APK) dan angka partisipasi murni (APM) di Indonesia untuk hampir semua jenjang pendidikan di tahun 2015–2022 [3], [4]. Penurunan dan peningkatan nilai tersebut tidak semata-mata menghentikan usaha dan peran Indonesia untuk terus menuntaskan tujuan dan target SDG. Aksi nyata perlu terus dilakukan untuk mendapatkan hasil yang optimal.

Implementasi SDG sebagai bentuk aksi nyata dapat dilakukan oleh perguruan tinggi melalui penelitian, sesuai salah satu kewajibannya dalam tri dharma perguruan tinggi, yaitu menyelenggarakan penelitian. Penelitian adalah kegiatan yang dilakukan dengan kaidah dan metode ilmiah untuk mengumpulkan informasi, data, dan wawasan relevan dengan pengujian disiplin ilmu. Penelitian diharapkan dapat memberi manfaat, baik bagi peningkatan mutu perguruan tinggi maupun pemenuhan pembangunan nasional [5].

Pemerintah juga berperan dalam aksi nyata SDG melalui penelitian mengacu pada Peraturan Menteri Pendayagunaan Aparatur Negara dan Reformasi Birokrasi Nomor 17 tahun

2013 dengan mewajibkan dosen yang ingin memperoleh jabatan akademik untuk melakukan penelitian dan publikasi ilmiah [6]. Hasil penelitian wajib disebarluaskan dengan cara diseminarkan, dipublikasikan, dan/atau dipatenkan oleh perguruan tinggi mengacu pada Undang-Undang Nomor 12 tahun 2012 [5].

Peran pemerintah melalui regulasinya menjadikan penelitian makin berkembang dari segi kuantitas. Berdasarkan data yang dipublikasikan oleh World Class University (WCU) Analysis, jumlah publikasi penelitian cenderung mengalami peningkatan setiap tahunnya untuk semua afiliasi perguruan tinggi [7].

Tren positif terkait kuantitas publikasi penelitian juga terjadi pada afiliasi Politeknik Statistika (Polstat) STIS sesuai data WCU Analysis [7]. Hal tersebut menunjukkan bahwa Polstat STIS turut berperan aktif dalam kemajuan penelitian di Indonesia. Bahkan, Polstat STIS juga berperan aktif dengan menyelenggarakan kegiatan Seminar Nasional Official Statistics setiap tahunnya dan International Conference on Data Science and Official Statistics (ICDSOS) setiap dua tahun. Kegiatan tersebut menjadi wadah bagi para pemakalah untuk mempublikasikan hasil penelitiannya [8].

Perkembangan kuantitas publikasi penelitian menyebabkan proses identifikasi implementasi publikasi penelitian makin sulit dilakukan, termasuk identifikasi implementasi ke dalam level SDG. Hal ini dibutuhkan Pusat Penelitian dan Pengabdian Masyarakat (PPPM) Polstat STIS untuk memantau aksi nyata SDG oleh dosen sebagai salah satu perwujudan tri dharma perguruan tinggi secara konkret.

Pengategorian publikasi penelitian dosen ke dalam level SDG secara otomatis dapat dilakukan dengan mengimplementasikan model klasifikasi multilabel dengan pendekatan multilevel dan hierarki. Klasifikasi multilabel adalah jenis klasifikasi yang memiliki beberapa kelas sekaligus (multilabel) pada *instance*-nya [9]. Pendekatan multilevel menyusun kelas-kelas ke dalam level, tanpa memperhatikan keterkaitan antarlevel. Adapun pendekatan hierarki menyusun kelas-kelas ke dalam level dengan memperhatikan keterkaitan antarlevel secara hierarki [10], [11].

Pada level dasar, publikasi akan difilter ke dalam SDG atau non-SDG. Publikasi SDG diklasifikasikan ke dalam empat level. Pada level pertama, publikasi diklasifikasikan ke dalam satu atau beberapa kategori pada empat pilar SDG secara multilabel, yaitu pilar sosial, ekonomi, lingkungan, serta hukum dan tata kelola. Empat pilar SDG adalah dimensi ringkas SDG yang ditetapkan oleh Badan Perencanaan Pembangunan Nasional (Bappenas) dan berkaitan dengan tujuan SDG. Pilar sosial berkaitan dengan tujuan 1, 2, 3, 4, dan 5. Pilar ekonomi berkaitan dengan tujuan 7, 8, 9, 10, dan 17. Pilar lingkungan berkaitan dengan tujuan 6, 11, 12, 13, 14, dan 15. Pilar hukum dan tata kelola berkaitan dengan tujuan 16 [12]. Pada level kedua, publikasi diklasifikasikan ke dalam satu atau beberapa kategori dari 17 tujuan SDG secara multilabel. Pada level ketiga, publikasi diklasifikasikan ke dalam satu atau beberapa kategori secara multilabel dari 143 target SDG Indonesia. Pada level keempat, publikasi diklasifikasikan ke dalam satu atau beberapa kategori secara multilabel dari 289 indikator SDG Indonesia.

Untuk mengatasi data multilabel, pengategorian publikasi penelitian ke dalam level SDG dilakukan menggunakan metode transformasi masalah. Metode tersebut mengubah masalah multilabel menjadi satu atau beberapa masalah label tunggal, sehingga dapat diselesaikan dengan menggunakan algoritma label tunggal [13]. Algoritma klasifikasi yang digunakan adalah algoritma klasifikasi pembelajaran mesin, seperti *logistic regression* (LR) dan *support vector machine* (SVM), yang terbukti memberikan kinerja terbaik pada model klasifikasi [13]. Selain itu, kedua algoritma tersebut mampu menangani data berdimensi tinggi [14]. Pembelajaran mesin adalah kategori kecerdasan buatan yang memungkinkan komputer belajar sendiri dari data [15]. Penerapannya bergantung pada jenis masalah yang dihadapi. Klasifikasi dikategorikan ke dalam paradigma pembelajaran terbimbing, yang dikenal sebagai pembelajaran melalui data historis.

Model terbaik dipilih untuk mengategorikan publikasi penelitian dosen Polstat STIS ke dalam level SDG. Model terbaik dipilih berdasarkan ukuran evaluasi. Hasil kategori publikasi ke dalam level SDG akan divisualisasikan agar diperoleh wawasan yang berguna bagi Polstat STIS.

Berdasarkan latar belakang tersebut, penelitian ini bertujuan untuk mengimplementasikan dan mengevaluasi metode transformasi masalah dan algoritma klasifikasi pembelajaran mesin dengan pendekatan multilevel dan hierarki untuk mengategorikan publikasi penelitian ke dalam level SDG. Untuk memenuhi tujuan tersebut, penelitian ini juga berkontribusi dalam membangun *dataset* yang digunakan dalam mengimplementasikan dan mengevaluasi model. Selain itu, model terbaik dari hasil implementasi dan evaluasi metode transformasi masalah dan algoritma klasifikasi pembelajaran mesin juga diterapkan untuk mengategorikan publikasi penelitian dosen Polstat STIS ke dalam level SDG sebagai salah satu realisasi nyata dari model yang dibangun.

II. METODOLOGI

Penelitian ini berfokus pada implementasi dan evaluasi model klasifikasi multilabel dengan pendekatan multilevel dan hierarki untuk mengategorikan publikasi penelitian dosen Polstat STIS ke dalam level SDG. Selain itu, tidak semua target dan indikator SDG digunakan karena beberapa target dan indikator SDG masih dikembangkan di Indonesia. Jumlah data untuk setiap label dibatasi minimal lima dengan asumsi nilai sudah mewakili pelabelan. Pelabelan dilakukan berdasarkan judul dan abstrak publikasi, dengan asumsi bahwa judul dan abstrak telah mewakili isi penelitian. Selain itu, tidak semua konten pada publikasi dapat ditemukan.

Dokumen yang dikumpulkan adalah publikasi penelitian, tidak termasuk buku. *Dataset* yang digunakan dimulai dari tahun 2016 karena gagasan SDG diresmikan pada tahun 2015. Model klasifikasi yang dibangun memproses masukan multibahasa, yaitu bahasa Indonesia dan bahasa Inggris, secara bersamaan.

Tahapan penelitian ini mengacu pada tahapan pada Cross-Industry Standard Process for Data Mining (CRISP-DM). CRISP-DM adalah kerangka kerja proyek analitis yang terdiri atas *business understanding*, *data understanding*, *data preparation*, *modelling*, *evaluation*, dan *deployment* [16].

A. BUSINESS UNDERSTANDING

Business understanding bertujuan untuk mencari pemahaman terkait tujuan dan persyaratan, yaitu definisi masalah dan langkah untuk mencapai tujuan [16]. *Business understanding* pada penelitian ini meliputi analisis kebutuhan klien dan studi pustaka. Kedua proses tersebut bertujuan untuk memahami tujuan dan persyaratan dalam mengategorikan publikasi penelitian ke dalam level SDG.

1) ANALISIS KEBUTUHAN KLIEN

Analisis kebutuhan klien dilakukan untuk mendapatkan informasi terkait latar belakang dan spesifikasi kebutuhan. Analisis dilakukan melalui diskusi dengan PPPM. Personel PPPM yang mengikuti diskusi terdiri atas kepala, sekretaris, dan staf PPPM yang bertugas melakukan manajemen data penelitian. Hasilnya adalah dibutuhkan kajian mengenai implementasi dan evaluasi model klasifikasi multilabel dengan pendekatan multilevel dan hierarki untuk mengategorikan publikasi penelitian dosen Polstat STIS ke dalam level SDG.

2) STUDI PUSTAKA

Studi pustaka dilakukan dengan mengumpulkan informasi terkait hasil diskusi dengan klien pada analisis kebutuhan klien. Kegiatan studi pustaka pada penelitian ini terdiri atas mencari informasi latar belakang, tujuan, dan metode untuk mengimplementasikan serta mengevaluasi model.

B. DATA UNDERSTANDING

Fase *data understanding* dimulai dengan mengumpulkan data mentah yang diidentifikasi kualitasnya untuk memperoleh wawasan awal. Wawasan tersebut digunakan guna membentuk hipotesis untuk tahapan selanjutnya [16]. Proses ini terdiri atas pembangunan *dataset* dengan tiga proses, yaitu pengumpulan data, pelabelan data, dan uji reliabilitas antarpemilai.

1) PENGUMPULAN DATA

Data dikumpulkan melalui aplikasi “Publish or Perish” dari basis data Semantic Scholar dengan kata kunci yang ditentukan. Alasan pengambilan informasi publikasi dari aplikasi “Publish or Perish” dan basis data Semantic Scholar adalah karena keduanya mencakup informasi “judul dan abstrak” yang digunakan untuk mewakili dokumen pada penelitian ini.

Kata kuncinya adalah indikator SDG dengan mengambil frasa yang menunjukkan maksud indikator SDG. Dari hasil pencarian menggunakan kata kunci untuk setiap indikator, diambil 15 publikasi. Penggunaan kata kunci tersebut bertujuan untuk memastikan bahwa semua indikator telah terwakili dalam data. Selain itu, diasumsikan bahwa kata kunci sudah mencakup level SDG. Kata kunci tersebut mengacu pada publikasi Metadata Indikator SDG yang diterbitkan oleh Bappenas pada tahun 2020, baik bahasa Indonesia maupun bahasa Inggris. Salah satu contoh data publikasi berdasarkan kata kunci serta pelabelannya tersaji pada Tabel I. Data tersebut termasuk dalam penelitian SDG, dikategorikan ke dalam pilar sosial, tujuan 1: tanpa kemiskinan; serta tujuan 4: pendidikan berkualitas. Data juga dikategorikan ke dalam target 1.2, yang merupakan turunan dari tujuan 1; serta target 4.1 dan target 4.2, yang merupakan turunan dari tujuan 4. Data juga dikategorikan ke dalam indikator 1.2.2, 4.1.2, serta 4.2.2. Indikator 1.2.2 mewakili persentase laki-laki, perempuan, dan anak-anak dari semua usia, yang hidup dalam kemiskinan dalam berbagai dimensi, sesuai dengan definisi nasional; indikator 4.1.2 mewakili tingkat penyelesaian pendidikan jenjang SD/ sederajat, SMP/ sederajat, dan SMA/ sederajat; sedangkan indikator 4.2.2 mewakili tingkat partisipasi dalam pembelajaran yang teroganisasi menurut jenis kelamin.

Publikasi yang telah dikumpulkan selanjutnya difilter untuk mengategorikan publikasi berlabel SDG dan non-SDG menggunakan model API dari *open source* SDG (OSDG). Pada penelitian ini persentase publikasi berlabel SDG dan non-SDG tidak ditentukan secara khusus.

Selain itu, dilakukan penentuan kata kunci yang memiliki kemungkinan terkait non-SDG sebanyak 23 kata kunci, seperti rumus dan matematika murni. Publikasi non-SDG dikumpulkan dari aplikasi "Publish or Perish" dan basis data Semantic Scholar dengan jumlah 500 untuk setiap kata kunci. Terakhir, data publikasi divalidasi dengan model dari OSDG.

Data publikasi penelitian dosen Polstat STIS juga dikumpulkan untuk dikategorikan ke dalam level SDG. Data dikumpulkan dari *website* Science and Technology Index (Sinta) yang tersinkronisasi dengan basis data Scopus, Web of Science (WoS), Garuda, dan Google Scholar. *Website* Sinta adalah sistem informasi penelitian berbasis web untuk mengukur kinerja ilmu pengetahuan dan teknologi [17].

Data yang telah terkumpul diseleksi untuk memperoleh penelitian yang mencakup setiap indikator. Seleksinya terdiri atas melakukan filter tahun terbit; menghapus publikasi selain bahasa Inggris dan bahasa Indonesia; menghapus publikasi tanpa penulis, judul, atau abstrak; mengeliminasi duplikasi data; serta memastikan setiap kata kunci berjumlah minimal lima, dengan asumsi nilai tersebut sudah mewakili label.

2) PELABELAN DATA

Data dilabeli oleh tiga *annotator* dengan label berupa indikator SDG yang diwakili kode indikator. Label akhir ditentukan dengan prinsip *majority voting*. *Annotator* pada tahap ini adalah mahasiswa Polstat STIS yang telah mempelajari mata kuliah *Official Statistics* Lanjutan yang memuat materi SDG, dengan asumsi bahwa *annotator* telah memiliki pengetahuan dasar SDG.

3) UJI RELIABILITAS ANTARPENILAI

Uji reliabilitas antarpenilai dilakukan untuk mengetahui kesepakatan penilai dalam mengategorikan data. Ukuran yang digunakan adalah koefisien *Krippendorff's alpha*. Kesepakatan dikatakan sudah cukup apabila nilai koefisien *Krippendorff's*

TABEL I
CONTOH DATA PUBLIKASI PENELITIAN DENGAN KATA KUNCI "KEMISKINAN
DALAM BERBAGAI DIMENSI" SERTA PELABELANNYA

| Publikasi | | | | |
|--|---------|--------|-------------|-------------------|
| Analisis Faktor-Faktor yang Memengaruhi Kemiskinan di Kota Manado | | | | |
| Kemiskinan merupakan isu penting dalam dimensi pembangunan manusia di Indonesia. Penelitian ini bertujuan untuk mengetahui ... | | | | |
| Label | | | | |
| SDGs | 4 Pilar | Tujuan | Target | Indikator |
| 1 | Sosial | 1;4 | 1.2;4.1;4.2 | 1.2.2;4.1.2;4.2.2 |

alpha minimal pada rentang 0,60–0,80 dengan interpretasi berupa kesepakatan yang sesungguhnya [18].

Koefisien *Krippendorff's alpha* pada penelitian ini dihitung untuk setiap level. Penghitungan koefisien tidak dilakukan untuk setiap kategori pada setiap level karena banyaknya jumlah kategori pada setiap level. Selain itu, koefisien *Krippendorff's alpha* juga sudah mengakomodasi isian multinilai [19].

C. DATA PREPARATION

Fase *data preparation* mencakup kegiatan membangun kumpulan data akhir. Data akhir tersebut digunakan untuk pembangunan model [16]. Pada penelitian ini, *data preparation* meliputi prapemrosesan data, *multilabel binarizer*, dan ekstraksi fitur. Fase ini dilakukan untuk semua *dataset* yang telah melalui proses pembangunan *dataset*.

1) PRAPEMROSESAN DATA

Tahap yang pertama dilakukan adalah *case folding*. Pada tahap ini dilakukan konversi huruf menjadi *lowercase* atau huruf kecil. Penelitian ini menggunakan *library* Pandas untuk melakukan tahap *case folding*.

Tahap berikutnya adalah eliminasi *stopwords* dan karakter. Penelitian ini menggunakan *library* SpaCy untuk eliminasi *stopword* bahasa Inggris atau bahasa Indonesia karena memiliki jumlah koleksi *stopwords* yang banyak [20]. Selain itu, dilakukan eliminasi *stopwords* bahasa Indonesia berdasarkan daftar *stopwords* penelitian [21]. Penelitian ini juga mengeliminasi karakter nonalfabet menggunakan *library* regex.

Selanjutnya, dilakukan proses *stemming*. Tahap ini dilakukan untuk menghilangkan imbuhan kata, sehingga diperoleh kata dasar. *Stemming* untuk data bahasa Inggris menggunakan *library* NLTK, sedangkan data bahasa Indonesia menggunakan *library* Sastrawi.

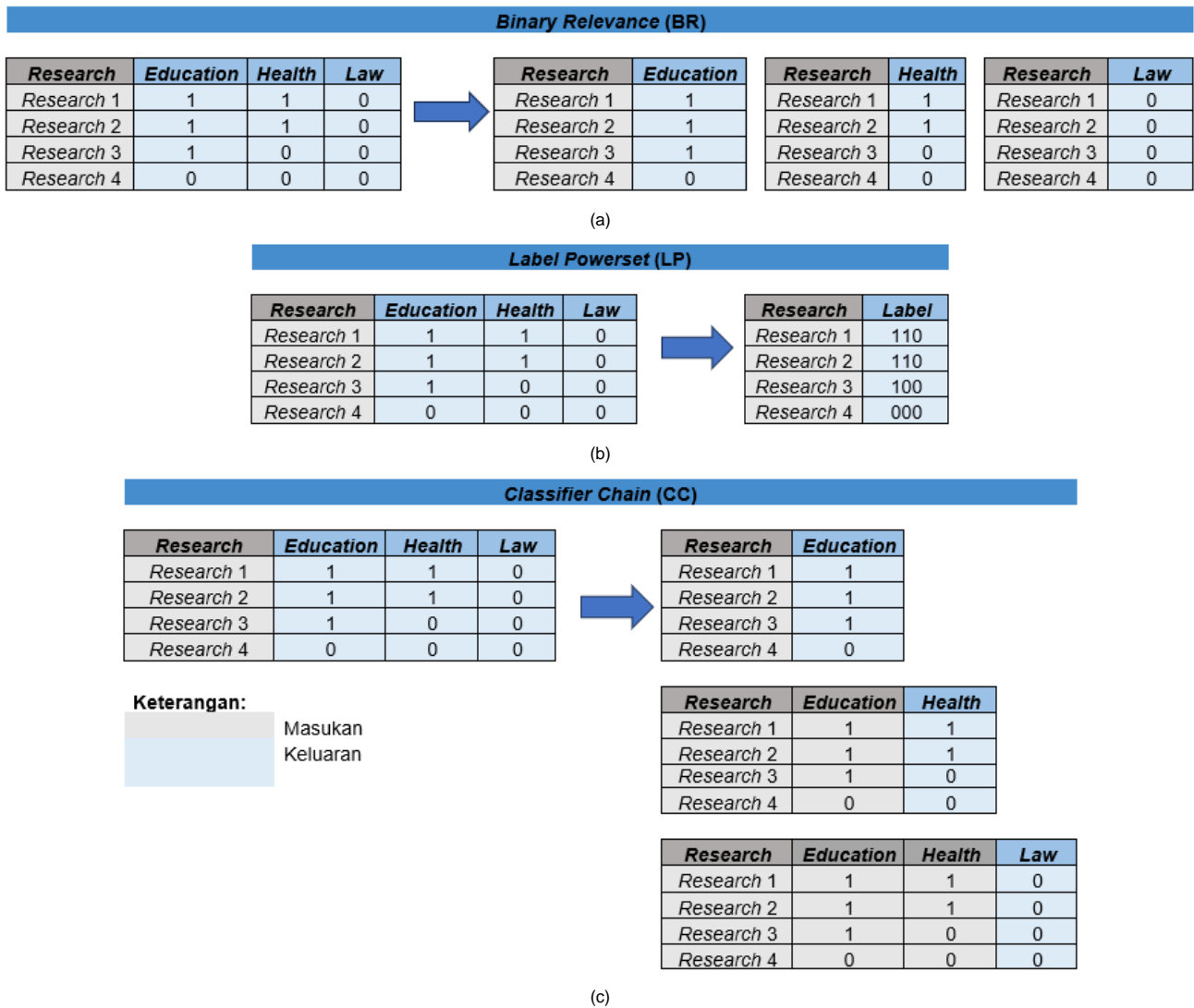
Tahap terakhir dalam prapemrosesan data adalah *tokenization*, yaitu tahap pemisahan data teks menjadi token atau kata yang merupakan unit yang lebih kecil untuk melatih model [22]. Pada tahap ini, digunakan *library* NLTK.

2) MULTILABEL BINARIZER

Multilabel binarizer adalah proses untuk mengubah data target ke format multilabel biner. Kolom mewakili label unik dan baris berhubungan dengan *instance*. Tahap ini menggunakan *class* MultiLabelBinarizer dari *library* scikit-learn.

3) EKSTRAKSI FITUR

Tahap ekstraksi fitur menggunakan teknik *term frequency-inverse document frequency* (TF-IDF). TF-IDF merupakan ekstraksi fitur teks yang memetakan setiap kata dan melakukan perkalian TF dan IDF sebagai pembobotan dokumen [23]. TF-



Gambar 1. Ilustrasi metode transformasi masalah, (a) binary relevance (BR), (b) label powerset (LP), (c) classifier chain (CC).

IDF menggunakan class TfidfVectorizer dari library scikit-learn dengan rumus menggunakan (1) dan (2).

$$TF_{(t,d)} = \frac{f_{t,d}}{\sum_{t \in d} f_{t,d}} \quad (1)$$

$$IDF_{(t)} = \log \frac{1+n}{1+df(t)} + 1 \quad (2)$$

dengan $f_{t,d}$ adalah frekuensi kata t dalam dokumen d , $\sum_{t \in d} f_{t,d}$ adalah total kata t pada setiap dokumen d , n adalah jumlah dokumen pelatihan yang digunakan, dan $df(t)$ adalah jumlah dokumen pelatihan yang memuat $term$.

D. MODELLING

Penelitian ini membangun dua macam model, yaitu model klasifikasi penelitian SDG dan non-SDG sebagai model filter dan model klasifikasi multilabel ke dalam level SDG sebagai model level. Model filter adalah model klasifikasi *multiclass*, sedangkan model level adalah model klasifikasi *multilable*. Oleh karena itu, hanya model level yang menggunakan metode transformasi masalah. Model level dibangun menggunakan pendekatan multilevel dan hierarki. Masukan *dataset* terdiri atas judul, abstrak, serta judul dan abstrak.

Pembangunan model filter dan model level dengan pendekatan multilevel melibatkan *grid search* dalam pencarian

parameter terbaik dan mekanisme *5-fold cross validation*. Pada proses *grid search* juga ditentukan nilai *max_features* menggunakan *pipeline* dari library scikit-learn. Pembangunan model level dengan pendekatan hierarki juga menggunakan parameter dari model level dengan pendekatan multilevel.

1) PEMBANGUNAN MODEL FILTER

Model filter adalah model yang mengategorikan publikasi penelitian ke dalam kategori SDG atau non-SDG. Model ini menggunakan algoritma klasifikasi pembelajaran mesin, yaitu LR dan SVM, yang diperoleh dari library scikit-learn.

2) PEMBANGUNAN MODEL LEVEL

Model level merupakan model klasifikasi multilabel yang mengategorikan publikasi ke dalam level SDG yang terdiri atas empat pilar, tujuan, target, dan indikator SDG. Model level menggunakan metode tranformasi masalah untuk mengatasi permasalahan data multilabel. Selanjutnya, algoritma klasifikasi digunakan untuk mengategorikan data.

Pada pembangunan model level dengan pendekatan multilevel, metode transformasi masalah yang digunakan adalah BR, LP, dan *classifier chains* (CC). BR adalah algoritma yang menguraikan masalah klasifikasi multilabel menjadi masalah klasifikasi biner independen, LP adalah algoritma yang menjadikan setiap kejadian unik dari kumpulan label dalam data pelatihan dianggap sebagai satu kelas untuk

kumpulan data yang baru ditransformasikan, sedangkan CC mengubah masalah pembelajaran multilabel menjadi rantai masalah klasifikasi biner. Setiap pengklasifikasi menggabungkan kelas-kelas yang diprediksi oleh pengklasifikasi sebelumnya sebagai fitur baru [24], [25]. Ilustrasi metode tersebut ditunjukkan pada Gambar 1.

Algoritma klasifikasi yang digunakan, yaitu LR dan SVM, terbukti memberikan kinerja terbaik pada model klasifikasi [13]. Selain itu, kedua algoritma tersebut digunakan karena mampu menangani data berdimensi tinggi [14], sesuai dengan karakteristik data yang digunakan. Metode transformasi masalah diperoleh dari *library* scikit-multilearn, sedangkan algoritma klasifikasi diperoleh dari scikit-learn.

Pada pembangunan model level dengan pendekatan hierarki, metode transformasi masalah yang digunakan adalah salah satu dari BR, LP, atau CC, sesuai hasil *grid search* pendekatan multilevel. Algoritma klasifikasi yang digunakan juga sesuai dengan hasil *grid search* pendekatan multilevel, antara LR atau SVM.

Model level dengan pendekatan hierarki menyusun kelas-kelas ke dalam level dengan memperhatikan keterkaitan antarlevel secara hierarki. Model dibangun secara datar untuk setiap level, dengan prediksi pada level atas menjadi masukan tambahan untuk level berikutnya. Proses tersebut berulang sampai level terbawah (*top-down approach*) [10], [11].

E. EVALUATION

Kinerja model klasifikasi dilihat dari nilai ukuran evaluasi pada data pengujian. Untuk menghasilkan nilai ukuran evaluasi yang lebih konsisten, model dibangun menggunakan *k-fold cross validation* karena *dataset* memiliki peluang sama, yaitu sebesar $1/k$ untuk data pelatihan dan data pengujian. Penelitian ini mengevaluasi model menggunakan *5-fold cross validation*.

Ukuran evaluasi model filter yang digunakan adalah akurasi, *F1-score*, dan waktu eksekusi, sedangkan ukuran evaluasi model level adalah akurasi, *F1-score*, *hamming loss*, dan waktu eksekusi. Akurasi adalah rasio yang dilabeli benar dengan total label, *F1-score* adalah rata-rata harmonik antara *recall* dan presisi, *recall* adalah rasio klasifikasi positif yang benar terhadap jumlah klasifikasi yang seharusnya positif, dan presisi adalah rasio klasifikasi positif yang benar terhadap jumlah data yang diprediksi positif. Sementara itu, *Hamming loss* adalah rasio yang dilabeli salah dengan total label, sedangkan waktu eksekusi adalah waktu algoritma dalam menampilkan hasil pelatihan model [13]. Formula akurasi, *recall*, presisi, dan *F1-score* dinyatakan dalam (3) sampai (6).

$$\text{akurasi} = \frac{TP + TN}{TP + FP + FN + TN} \quad (3)$$

$$\text{recall} = \frac{TP}{TP + FN} \quad (4)$$

$$\text{presisi} = \frac{TP}{TP + FP} \quad (5)$$

$$F1 - \text{score} = \frac{2 \times (\text{recall} \times \text{presisi})}{(\text{recall} + \text{presisi})} \quad (6)$$

TP mewakili model memprediksi kelas positif secara benar, sesuai kondisi aktual; *TN* mewakili model memprediksi kelas negatif secara benar, sesuai kondisi aktual; *FP* mewakili model memprediksi kelas sebagai positif, sedangkan kondisi aktualnya negatif; dan *FN* mewakili model memprediksi kelas sebagai negatif, sedangkan kondisi aktualnya positif.

Model filter terbaik adalah model dengan nilai akurasi dan *F1-score* yang tinggi, serta waktu yang optimal. Model level

terbaik adalah model dengan nilai akurasi dan *F1-score* yang tinggi, *hamming loss* yang rendah, serta waktu yang optimal [13].

F. DEPLOYMENT

Model terbaik diimplementasikan untuk mengategorikan publikasi penelitian dosen Polstat STIS ke dalam level SDG. Hasil kategorinya divisualisasikan sebagai berikut.

1. Distribusi publikasi SDG dan non-SDG yang divisualisasikan dengan diagram lingkaran.
2. Distribusi publikasi SDG berdasarkan bahasa yang divisualisasikan dengan diagram batang.
3. Distribusi publikasi pada kategori level empat pilar dan tujuan SDG yang divisualisasikan dengan diagram batang.

III. HASIL DAN PEMBAHASAN

A. PEMBANGUNAN DATASET

Dataset dilabeli oleh tiga *annotator* dengan label indikator SDG yang mewakili empat pilar, tujuan, dan target SDG. Setelah itu, dilakukan uji reliabilitas pelabelan menggunakan koefisien *Krippendorff's alpha* yang dihitung untuk setiap level. Jumlah data yang dikumpulkan dari kata kunci sebanyak 20.050 data. Setelah itu, data diseleksi untuk memperoleh penelitian yang mencakup setiap indikator. Jumlah data yang terkumpul untuk pembangunan model sebanyak 8.090 data.

1) UJI RELIABILITAS ANTARPENILAI

Nilai *alpha* untuk setiap level tertera pada Tabel II. Secara umum, nilai *alpha* berada di rentang 0,60–0,80, sehingga kualitas data tergolong baik. Nilai *alpha* menurun dari level 1 hingga level 4 karena jumlah kategori bertambah, sehingga kemungkinan *annotator* melabelkan berbeda pun bertambah.

2) KARAKTERISTIK DATASET

Gambar 2 menunjukkan perbedaan distribusi publikasi penelitian SDG dan non-SDG. Persentase nilai masing-masing sebesar 66,5% untuk SDG dan 33,5% untuk non-SDG. Nilai tersebut menunjukkan *dataset* yang tidak seimbang.

Berdasarkan bahasanya, *dataset* terbagi menjadi publikasi bahasa Indonesia dan bahasa Inggris dari bagian judul atau abstrak. Penelitian ini tidak mengatasi ketidakseimbangan data untuk membangun model. Visualisasi distribusi *dataset* berdasarkan bahasa tersaji pada Gambar 3. Bahasa Indonesia mendominasi *dataset* berdasarkan judul, yaitu sebanyak 4.576 publikasi, sedangkan bahasa Inggris mendominasi *dataset* berdasarkan abstrak, sebanyak 4.816 publikasi. Hal tersebut sesuai kata kunci yang digunakan untuk mengumpulkan *dataset*, yaitu kata kunci bahasa Indonesia dan bahasa Inggris yang mengacu pada publikasi metadata indikator SDG. Selain itu, jumlah publikasi penelitian bahasa Indonesia dan bahasa Inggris, baik berdasarkan judul maupun abstrak, tidak jauh berbeda.

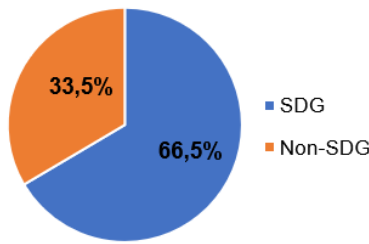
B. MODEL KLASIFIKASI PENDEKATAN MULTILEVEL

1) MODEL FILTER: TERKAIT SDG

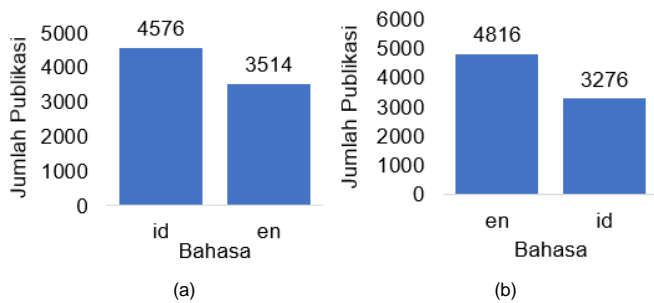
Model filter mengklasifikasikan data ke dalam kategori SDG dan non-SDG. Sebelum klasifikasi, dilakukan *grid search* dengan *5-fold cross validation* untuk mendapatkan parameter terbaik. Berdasarkan hasil evaluasi pada Tabel III, model filter terbaik adalah model dengan masukan judul dan algoritma SVM karena hasil prediksi model mencapai nilai akurasi dan *F1-score* paling tinggi. Selain itu, model filter terbaik tersebut juga memiliki waktu eksekusi yang cukup optimal.

TABEL II
 NILAI KOEFISIEN KRIPPENDORFF'S ALPHA SETIAP LEVEL

| No | Level | Alpha | No | Level | Alpha |
|----|-----------------|-------|----|---------------|-------|
| 1 | Empat Pilar SDG | 0,869 | 3 | Target SDG | 0,760 |
| 2 | Tujuan SDG | 0,815 | 4 | Indikator SDG | 0,734 |



Gambar 2. Distribusi publikasi penelitian SDG dan non-SDG.



Gambar 3. Distribusi publikasi penelitian berdasarkan bahasa dan metadata, (a) judul, (b) abstrak.

Penambahan informasi pada masukan model, seperti abstrak, biasanya meningkatkan ukuran evaluasi karena data pelatihan akan mempelajari lebih mendalam makna sebenarnya dari teks [26], [27]. Namun, ada perbedaan pada penelitian ini, yaitu model dengan masukan judul memiliki nilai ukuran evaluasi yang lebih baik. Hal tersebut sesuai dengan penelitian sebelumnya [28], yaitu bahwa penambahan informasi yang menyebabkan penambahan kata pada vokabuler tidak menjamin dihasilkannya fitur kata penting dengan nilai akurasi yang tinggi.

Sebuah penelitian menggambarkan bahwa penambahan informasi pada masukan model tidak memengaruhi peningkatan nilai ukuran evaluasi [29]. Objek penelitian tersebut berupa publikasi penelitian dari Journal of Universal Computer Science (J.UCS) dan the Association of Computing Machinery (ACM). Pada data J.UCS, model dengan masukan judul saja menghasilkan nilai akurasi dan *F1-score* masing-masing sebesar 0,85 dan 0,91, sedangkan model dengan masukan judul dan kata kunci menghasilkan nilai akurasi dan *F1-score* sebesar 0,65 dan 0,71. Pada data ACM, model dengan masukan judul saja menghasilkan nilai akurasi dan *F1-score* sebesar 0,85 dan 0,92, sedangkan model dengan masukan judul dan kata kunci menghasilkan nilai akurasi dan *F1-score* sebesar 0,79 dan 0,86 [29].

Berdasarkan hal tersebut, terdapat kemungkinan bahwa model dengan masukan judul sudah cukup mewakili kata penting untuk membangun model. Selain itu, model pada penelitian ini juga memproses multibahasa secara sekaligus, yaitu bahasa Indonesia dan bahasa Inggris, sehingga satu makna kata dari dua bahasa dianggap dua kata penting yang berbeda. Contohnya, kata "poverty" dan "miskin" merupakan dua kata penting yang berbeda jika model memproses dua bahasa sekaligus dan menggunakan ekstraksi fitur TF-IDF. Berdasarkan hal tersebut, model dengan masukan abstrak serta judul dan abstrak memiliki nilai kinerja yang lebih rendah

TABEL III
 EVALUASI KINERJA KLASIFIKASI MODEL FILTER

| Ukuran Evaluasi | LR | | SVM | | | |
|-----------------|--------|---------|-----------------|--------|---------|-----------------|
| | Judul | Abstrak | Judul + Abstrak | Judul | Abstrak | Judul + Abstrak |
| Akurasi | 0,8476 | 0,8408 | 0,8435 | 0,8634 | 0,8410 | 0,8516 |
| <i>F1-score</i> | 0,8204 | 0,8102 | 0,8136 | 0,8382 | 0,8065 | 0,8184 |
| Waktu (s) | 0,3123 | 15,390 | 0,8025 | 14,902 | 65,564 | 37,591 |

karena banyaknya kata penting yang beragam yang sebenarnya memiliki kemungkinan makna yang sama. Oleh karena itu, penelitian selanjutnya dapat mengkaji pengembangan model dengan salah satu bahasa saja.

2) MODEL LEVEL 1: EMPAT PILAR SDG

Model level 1 mengklasifikasikan data ke dalam kategori empat pilar SDG. Sebelum klasifikasi, dilakukan *grid search* dengan *5-fold cross validation* untuk mendapatkan parameter terbaik.

Berdasarkan hasil evaluasi pada Tabel IV, model level 1 terbaik adalah model dengan masukan judul, metode LP, dan algoritma SVM. Hal tersebut terjadi karena model memiliki nilai akurasi dan *F1-score* paling tinggi serta *hamming loss* paling rendah, yaitu masing-masing sebesar 0,8067, 0,8353, serta 0,0830. Model level 1 terbaik juga memiliki waktu eksekusi yang cukup optimal.

Metode LP adalah metode terbaik pada penelitian ini karena kemampuannya dalam mempertimbangkan keterkaitan antar kategori yang tidak diatasi metode BR dengan menjadikan label sebagai kombinasi unik terpisah. Metode LP juga tidak mengklasifikasikan label secara berantai dari masukan sebelumnya seperti metode CC. Metode CC mengakumulasi kesalahan *classifier* sebelumnya, sehingga hasil kategori cenderung tidak akurat [27]. SVM adalah algoritma terbaik pada penelitian ini karena kemampuannya dalam memaksimalkan *margin* [30].

3) MODEL LEVEL 2: TUJUAN SDG

Model level 2 mengklasifikasikan data ke dalam kategori tujuan SDG. Sebelum klasifikasi, dilakukan *grid search* dengan *5-fold cross validation* untuk mendapatkan parameter terbaik.

Berdasarkan hasil evaluasi pada Tabel IV, model level 2 terbaik adalah model dengan masukan judul, metode LP, dan algoritma SVM. Hal tersebut terjadi karena model memiliki nilai akurasi dan *F1-score* paling tinggi serta *hamming loss* paling rendah, yaitu masing-masing 0,7501, 0,7922, serta 0,0267. Model level 2 terbaik tersebut juga memiliki waktu eksekusi yang cukup optimal.

Model terbaik yang terpilih pada level ini sejalan dengan penelitian sebelumnya [13]. Model terbaik untuk mengklasifikasikan artikel ke dalam 17 tujuan SDG adalah kombinasi model dengan metode LP dan algoritma SVM [13].

4) MODEL LEVEL 3: TARGET SDG

Model level 3 mengklasifikasikan data ke dalam kategori target SDG. Sebelum klasifikasi, dilakukan *grid search* dengan *5-fold cross validation* untuk mendapatkan parameter terbaik.

Berdasarkan hasil evaluasi pada Tabel IV, model level 3 terbaik adalah model dengan masukan judul, metode LP, dan algoritma SVM karena memiliki nilai akurasi dan *F1-score* paling tinggi, sebesar 0,6792 dan 0,6829. Nilai *hamming loss* berada di urutan keempat terkecil, sebesar 0,0045. Model level 3 terbaik tersebut juga memiliki waktu eksekusi yang cukup optimal.

TABEL IV
EVALUASI KINERJA KLASIFIKASI MODEL LEVEL 1-4 PENDEKATAN MULTILEVEL

| Metode | Algoritma | Akurasi | F1-Score | Hamming Loss | Waktu (s) | Akurasi | F1-Score | Hamming Loss | Waktu (s) | Akurasi | F1-Score | Hamming Loss | Waktu (s) |
|----------------|-----------|---------|----------|--------------|-----------|---------|----------|--------------|-----------|---------|-------------------|--------------|-----------|
| | | | Judul | | | | Abstrak | | | | Judul dan Abstrak | | |
| Level 1 | | | | | | | | | | | | | |
| BR | LR | 0,7291 | 0,8094 | 0,0898 | 0,8517 | 0,6889 | 0,7799 | 0,1011 | 284,61 | 0,7136 | 0,7992 | 0,0923 | 40,388 |
| | SVM | 0,7365 | 0,8217 | 0,0863 | 1000,9 | 0,7153 | 0,8055 | 0,0946 | 1158,0 | 0,7380 | 0,8235 | 0,0863 | 959,44 |
| LP | LR | 0,8058 | 0,8330 | 0,0840 | 0,6421 | 0,7724 | 0,7996 | 0,0983 | 23,201 | 0,7941 | 0,8218 | 0,0877 | 28,715 |
| | SVM | 0,8067 | 0,8353 | 0,0830 | 453,10 | 0,7789 | 0,8083 | 0,0953 | 507,82 | 0,8010 | 0,8296 | 0,0845 | 513,68 |
| CC | LR | 0,7744 | 0,8027 | 0,0978 | 14,548 | 0,7527 | 0,7812 | 0,1073 | 57,685 | 0,7722 | 0,7999 | 0,0977 | 86,564 |
| | SVM | 0,7846 | 0,8120 | 0,0936 | 909,98 | 0,7642 | 0,7941 | 0,1014 | 912,88 | 0,7826 | 0,8122 | 0,0931 | 739,34 |
| Level 2 | | | | | | | | | | | | | |
| BR | LR | 0,6040 | 0,7500 | 0,0297 | 43,674 | 0,5072 | 0,6678 | 0,0372 | 20,208 | 0,5408 | 0,7003 | 0,0342 | 35,509 |
| | SVM | 0,6135 | 0,7482 | 0,0287 | 1335,4 | 0,5395 | 0,6830 | 0,0344 | 2081,8 | 0,5802 | 0,7195 | 0,0312 | 2471,5 |
| LP | LR | 0,7439 | 0,7818 | 0,0285 | 63,880 | 0,6651 | 0,7049 | 0,0375 | 10,959 | 0,7055 | 0,7459 | 0,0324 | 13,276 |
| | SVM | 0,7501 | 0,7922 | 0,0267 | 605,10 | 0,6915 | 0,7307 | 0,0335 | 675,93 | 0,7225 | 0,7640 | 0,0296 | 1125,4 |
| CC | LR | 0,6683 | 0,7446 | 0,0362 | 60,343 | 0,5867 | 0,6685 | 0,0450 | 36,306 | 0,6168 | 0,6990 | 0,0412 | 97,273 |
| | SVM | 0,6845 | 0,7446 | 0,0343 | 1054,2 | 0,6326 | 0,6921 | 0,0399 | 1471,8 | 0,6670 | 0,7273 | 0,0357 | 1921,7 |
| Level 3 | | | | | | | | | | | | | |
| BR | LR | 0,5395 | 0,6395 | 0,0044 | 41,125 | 0,3719 | 0,4706 | 0,0059 | 108,69 | 0,4154 | 0,5157 | 0,0054 | 224,91 |
| | SVM | 0,5300 | 0,6249 | 0,0045 | 393,31 | 0,3938 | 0,4763 | 0,0055 | 2463,8 | 0,4507 | 0,5377 | 0,0050 | 2088,2 |
| LP | LR | 0,6670 | 0,6572 | 0,0050 | 83,534 | 0,5458 | 0,5323 | 0,0067 | 19,284 | 0,6003 | 0,5883 | 0,0058 | 60,395 |
| | SVM | 0,6792 | 0,6829 | 0,0045 | 109,53 | 0,5713 | 0,5551 | 0,0060 | 545,89 | 0,6202 | 0,6099 | 0,0053 | 476,22 |
| CC | LR | 0,5642 | 0,6525 | 0,0043 | 71,177 | 0,3949 | 0,4763 | 0,0059 | 169,90 | 0,4358 | 0,5224 | 0,0054 | 410,90 |
| | SVM | 0,5681 | 0,6469 | 0,0044 | 485,04 | 0,4375 | 0,5031 | 0,0055 | 2810,3 | 0,4902 | 0,5623 | 0,0049 | 2425,4 |
| Level 4 | | | | | | | | | | | | | |
| BR | LR | 0,4841 | 0,5679 | 0,0027 | 63,391 | 0,2964 | 0,3912 | 0,0036 | 267,52 | 0,3288 | 0,4268 | 0,0034 | 466,51 |
| | SVM | 0,4670 | 0,5555 | 0,0027 | 998,18 | 0,3031 | 0,3905 | 0,0034 | 3731,7 | 0,3585 | 0,4465 | 0,0031 | 3022,8 |
| LP | LR | 0,6142 | 0,5846 | 0,0032 | 96,483 | 0,4598 | 0,4504 | 0,0044 | 41,329 | 0,5161 | 0,4990 | 0,0039 | 50,279 |
| | SVM | 0,6194 | 0,6158 | 0,0029 | 137,67 | 0,4953 | 0,4868 | 0,0038 | 533,03 | 0,5462 | 0,5378 | 0,0034 | 382,91 |
| CC | LR | 0,5064 | 0,5778 | 0,0026 | 104,14 | 0,3115 | 0,3973 | 0,0036 | 424,26 | 0,3524 | 0,4357 | 0,0034 | 1105,8 |
| | SVM | 0,5036 | 0,5715 | 0,0026 | 826,44 | 0,3445 | 0,4178 | 0,0034 | 4243,7 | 0,3986 | 0,4763 | 0,0031 | 3916,3 |

5) MODEL LEVEL 4: INDIKATOR SDG

Model level 4 mengklasifikasikan data ke dalam kategori indikator SDG. Pada level ini, dilakukan *grid search* dengan 5-*fold cross validation* untuk mendapatkan parameter terbaik.

Berdasarkan hasil evaluasi pada Tabel IV, model level 4 terbaik adalah model dengan masukan judul, metode LP, dan algoritma SVM karena memiliki nilai akurasi dan *F1-score* paling tinggi, yaitu 0,6194 dan 0,6158. Nilai *hamming loss* berada di urutan ketiga terkecil, yaitu 0,0029. Model level 4 terbaik tersebut juga memiliki waktu eksekusi yang cukup optimal.

C. MODEL KLASIFIKASI PENDEKATAN HIERARKI

Pembangunan model level dengan pendekatan hierarki menggunakan metode LP dan SVM sesuai model terbaik pada pendekatan multilevel. Model level dengan pendekatan hierarki juga menggunakan parameter dari pendekatan multilevel.

Pada Tabel V, model hierarki terbaik adalah model dengan masukan judul karena memiliki akurasi dan *F1-score* tertinggi serta *hamming loss* terkecil di level 3-4. Pada level 1-2, nilai akurasi dan *F1-score* tertinggi serta *hamming loss* terkecil dihasilkan pada model dengan masukan abstrak serta judul dan abstrak. Model dengan masukan judul juga memiliki waktu eksekusi yang paling optimal untuk semua level. Model level terbaik dengan pendekatan ini adalah model dengan masukan judul, metode LP, dan SVM.

Pada level 1 dan 2, model level dengan pendekatan hierarki dan masukan abstrak serta judul dan abstrak memiliki kinerja terbaik. Pada level 3 dan 4, model level dengan pendekatan

hierarki dan masukan judul memiliki kinerja yang terbaik. Hasil tersebut menunjukkan adanya akumulasi kesalahan di level sebelumnya jika masukan lebih beragam. Selain itu, model dengan level dan dimensi tinggi tidak cocok untuk masukan yang lengkap, seperti abstrak atau judul dan abstrak, karena banyak faktor yang harus dipertimbangkan model dalam mengklasifikasikan data. Berbeda dengan masukan judul, faktor yang harus dipertimbangkan model tidak sebanyak masukan abstrak atau judul dan abstrak. Berdasarkan penjelasan tersebut, disarankan agar digunakan masukan judul jika kategori yang digunakan dalam mengklasifikasikan data cukup banyak.

D. MODEL TERPILIH

Model filter terbaik adalah model dengan masukan judul dan SVM. Model level terbaik untuk setiap pendekatan adalah model dengan masukan judul, metode LP, dan algoritma SVM. Ringkasan model terbaik disajikan pada Tabel VI.

Pada Tabel VI, model level dengan pendekatan multilevel memiliki akurasi dan *F1-score* tertinggi serta *hamming loss* terkecil. Namun, waktu eksekusi yang paling optimal adalah model level dengan pendekatan hierarki. Oleh karena itu, model level terpilih adalah model dengan masukan judul, metode LP, dan algoritma SVM dengan pendekatan multilevel.

Model level dengan pendekatan multilevel memiliki kinerja yang lebih baik dibandingkan dengan pendekatan hierarki karena model berdiri sendiri untuk setiap levelnya dalam membangun model. Sementara itu, model level dengan pendekatan hierarki harus memperhatikan hasil level sebelumnya, sehingga kesalahan level sebelumnya diakumulasi untuk level selanjutnya, walaupun model level dengan

TABEL V
EVALUASI KINERJA KLASIFIKASI MODEL DENGAN PENDEKATAN HIERARKI

| Ukuran Evaluasi | Level 1 | Level 2 | Level 3 | Level 4 | Level 1 | Level 2 | Level 3 | Level 4 | Level 1 | Level 2 | Level 3 | Level 4 |
|-----------------|---------|---------|---------|---------|----------|----------|----------|----------|-------------------|----------|----------|----------|
| | Judul | | | | Abstrak | | | | Judul dan Abstrak | | | |
| Akurasi | 0,8041 | 0,7274 | 0,6411 | 0,5874 | 0,8067 | 0,7289 | 0,6294 | 0,5722 | 0,8067 | 0,7289 | 0,6294 | 0,5722 |
| F1-score | 0,8349 | 0,7699 | 0,6483 | 0,5844 | 0,8353 | 0,7725 | 0,6390 | 0,5692 | 0,8353 | 0,7725 | 0,6390 | 0,5692 |
| Hamming loss | 0,0842 | 0,0298 | 0,0051 | 0,0031 | 0,0830 | 0,0296 | 0,0053 | 0,0033 | 0,0830 | 0,0296 | 0,0053 | 0,0033 |
| Waktu (s) | 46,1430 | 61,7350 | 57,8020 | 58,5570 | 211,3600 | 289,6100 | 257,4500 | 179,8800 | 176,8900 | 243,4400 | 216,0400 | 151,5100 |

pendekatan hierarki dapat menjelaskan keterkaitan kategori untuk setiap levelnya.

Pengguna model, yaitu PPPM Polstat STIS, tetap dapat menggunakan model level dengan pendekatan multilevel sesuai kebutuhannya. Jika pengguna membutuhkan kategori tujuan SDG, pengguna dapat menggunakan model level 2 yang memprediksi kategori di level 1 juga. Namun, jika pengguna membutuhkan kategori indikator SDG, pengguna dapat menggunakan model level 4 yang memprediksi kategori di level 1-3 juga. Konsekuensinya, hasil kategori tidak seakurat model level yang lebih tinggi. Namun, model telah mengategorikan publikasi penelitian ke dalam level SDG secara otomatis.

Pada Tabel VI dijelaskan bahwa terdapat hubungan negatif antara jumlah dimensi dan ukuran evaluasi. Peningkatan jumlah dimensi berpengaruh pada penurunan kinerja ukuran evaluasi. Hal tersebut terjadi karena model memiliki tugas yang lebih besar dalam memprediksi data. Selain itu, peningkatan jumlah dimensi memiliki hubungan positif dengan *overfitting* [31].

Waktu eksekusi juga dipengaruhi oleh jumlah vokabuler yang digunakan pada model. Model level 2, baik menggunakan pendekatan multilevel maupun hierarki, memiliki waktu eksekusi yang paling lama dibandingkan dengan model dengan pendekatan hierarki, padahal jumlah dimensinya tidak sebanyak model level 3 dan 4. Hal tersebut sesuai dengan hasil *grid search* pada bagian *max_features* yang berpengaruh pada jumlah vokabuler yang digunakan untuk membangun model. *Max_features* pada level 3 dan 4 sebesar 3117 kata, sedangkan *max_features* pada level 2 sebesar 6235 kata.

E. HASIL KATEGORI

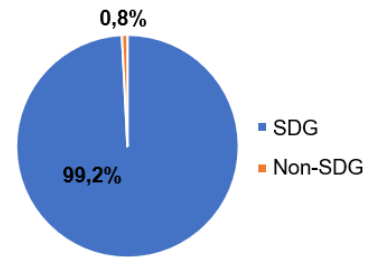
Data dosen Polstat STIS berhasil dikumpulkan sebanyak 1.687 data, dari tahun 1983-2024. Data dikumpulkan melalui *website* PPPM, menu Data Penelitian Sinta pada 28 Mei 2024. Tidak ada proses seleksi untuk data ini. Semua data publikasi penelitian dosen diprediksi menggunakan model terbaik yang sudah dibangun, yaitu model dengan masukan judul dan algoritma SVM untuk model filter serta model dengan masukan judul, metode SVM, dan algoritma SVM dengan pendekatan multilevel.

Pada Gambar 4, tampak bahwa distribusi publikasi penelitian SDG dan non-SDG berbeda. Persentasenya masing-masing sebesar 99,2% dan 0,8% untuk SDG dan non-SDG. Hasil tersebut menunjukkan bahwa mayoritas penelitian dosen sudah berkontribusi dengan kemajuan SDG.

Berdasarkan judulnya, bahasa publikasi terdiri atas bahasa Indonesia, Inggris, dan Jepang. Berdasarkan abstraknya, bahasa publikasi terdiri atas bahasa Indonesia, Inggris, dan Prancis. Selain itu, terdapat publikasi yang tidak disertai abstrak dengan kode "Na" (*not available*). Distribusi data berdasarkan bahasa disajikan pada Gambar 5. Bahasa Indonesia mendominasi data berdasarkan judul, sebanyak 908

TABEL VI
RINGKASAN MODEL TERBAIK

| Model | Dimensi | Model Terbaik | Akurasi | F1-Score | Hamming Loss | Waktu (s) |
|------------------------------|---------|---------------|---------|----------|--------------|-----------|
| Pendekatan Multilevel | | | | | | |
| Filter | 1 | Judul, SVM | 0,8634 | 0,8382 | - | 14,902 |
| Level 1 | 4 | Judul, SVM | 0,8067 | 0,8353 | 0,0830 | 453,10 |
| Level 2 | 17 | LP, SVM | 0,7501 | 0,7922 | 0,0267 | 605,10 |
| Level 3 | 143 | LP, SVM | 0,6792 | 0,6829 | 0,0045 | 109,54 |
| Level 4 | 289 | LP, SVM | 0,6194 | 0,6158 | 0,0029 | 137,67 |
| Pendekatan Hierarki | | | | | | |
| Level 1 | 4 | Judul, SVM | 0,8041 | 0,8349 | 0,0842 | 46,143 |
| Level 2 | 17 | LP, SVM | 0,7274 | 0,7699 | 0,0298 | 61,735 |
| Level 3 | 143 | LP, SVM | 0,6411 | 0,6483 | 0,0051 | 57,802 |
| Level 4 | 289 | LP, SVM | 0,5874 | 0,5844 | 0,0031 | 58,557 |



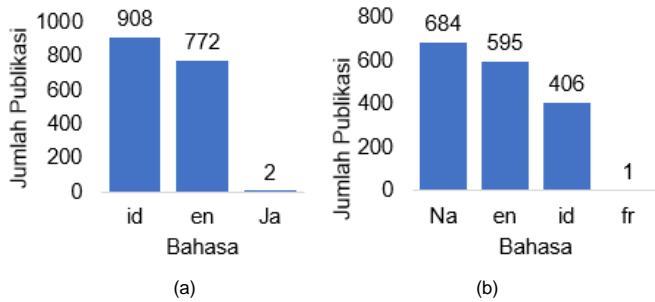
Gambar 4. Distribusi publikasi penelitian SDGs dan non-SDGs.

data, sedangkan bahasa Inggris mendominasi data berdasarkan abstrak, sebanyak 595 data.

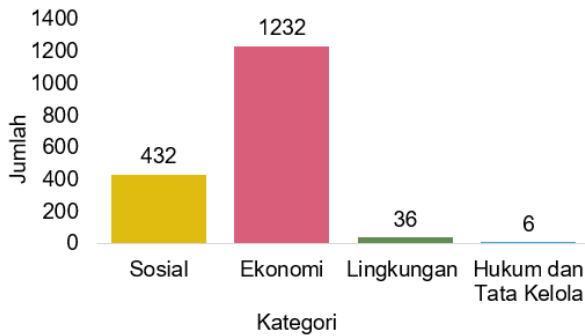
Berdasarkan hasil prediksi, distribusi data publikasi penelitian dosen Polstat STIS pada level empat pilar SDG divisualisasikan dalam Gambar 6. Kategori ekonomi memiliki jumlah publikasi terbanyak, sebesar 1.232 publikasi. Terdapat perbedaan yang cukup jauh antara kategori sosial-ekonomi serta lingkungan-hukum dan tata kelola. Berdasarkan hal tersebut, diketahui bahwa dosen Polstat STIS mayoritas melakukan penelitian terkait sosial dan ekonomi. Topik lingkungan serta hukum dan tata kelola dapat menjadi beberapa topik usulan penelitian bagi dosen.

Berdasarkan hasil prediksi, distribusi data publikasi pada level tujuan SDG divisualisasikan pada Gambar 7. Kategori tujuan 8 memiliki jumlah publikasi paling banyak, sebesar 1495 publikasi, disusul kategori tujuan 17, sebesar 100 publikasi. Tujuan 8 membahas pekerjaan layak dan pertumbuhan ekonomi, sedangkan tujuan 17 membahas kemitraan untuk mencapai tujuan. Adapun kategori dengan jumlah publikasi paling sedikit adalah tujuan 13 dan 14. Tujuan 13 membahas penanganan perubahan iklim, sedangkan tujuan 14 membahas ekosistem lautan. Maka, topik tujuan 13 dan 14 dapat menjadi beberapa topik usulan penelitian bagi dosen.

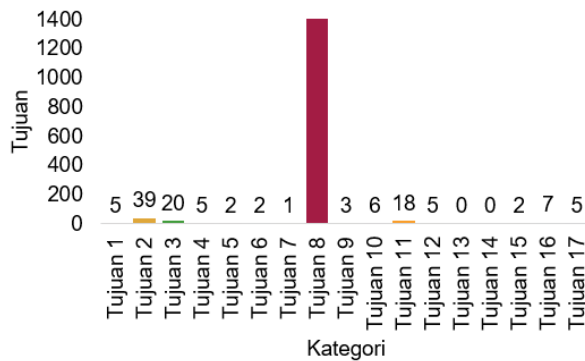
Tabel VII menunjukkan kemampuan model dalam mengklasifikasikan data publikasi penelitian untuk setiap levelnya dengan benar. Tabel VIII menunjukkan kemampuan



Gambar 5. Distribusi publikasi penelitian berdasarkan bahasa dan metadata, (a) judul, (b) abstrak.



Gambar 6. Distribusi publikasi penelitian SDGs berdasarkan empat pilar SDGs.



Gambar 7. Distribusi publikasi penelitian SDGs berdasarkan tujuan SDGs.

model dalam mengklasifikasikan data dengan benar untuk sebagian level.

Model mengklasifikasikan data dengan benar untuk setiap level di Tabel VII karena kata “stunting” tertera jelas pada judulnya. Kata “stunting” tersebut adalah kata kunci indikator 2.2.1 yang digunakan untuk mengumpulkan data. Hal tersebut merupakan salah satu dugaan terkait model dengan mudah mengklasifikasikan data dengan benar untuk setiap levelnya.

Model mengklasifikasikan data dengan benar untuk sebagian level pada Tabel VIII karena adanya kombinasi kata yang lebih beragam pada judulnya, yaitu kata “transportation” dan “growth”. Kata “transportation” dikandung oleh kata kunci indikator 11.2.1.(a) dan 11.2.1.(b), yaitu “convenient access to public transportation” dan “residents are served by public transportation”. Kata “growth” dikandung kata kunci indikator 8.1.1, 8.2.1, 8.9.1, 9.2.1, dan 10.1.1.(e), yaitu “GDP growth rate”, “GDP growth rate per worker”, “growth rate of tourism contribution to GDP”, “manufacturing industry GDP growth rate”, dan “economic growth in underdeveloped areas”. Indikator 11.2.1.(a) dan 11.2.1.(b) adalah tujuan 11 yang membahas pembangunan lingkungan, sedangkan indikator 8.1.1, 8.2.1, 8.9.1, 9.2.1, dan 10.1.1.(e) adalah tujuan 8, 9, dan 11 yang membahas pembangunan ekonomi. Hal tersebut adalah beberapa dugaan terkait model level 1 dan 2

TABEL VII
CONTOH PREDIKSI BENAR UNTUK SEMUA LEVEL

| Publikasi Penelitian | Analisis Faktor-Faktor yang Memengaruhi Stunting di Indonesia Tahun 2021 | | | | |
|---|--|-------------------|--------------|--------------|-----------------|
| | Label Filter | Label Empat Pilar | Label Tujuan | Label Target | Label Indikator |
| Stunting adalah masalah yang timbul akibat kurangnya asupan gizi dan dapat mengganggu pertumbuhan ... | 1 | Sosial | 2 | 2.2 | 2.2.1 |
| Aktual | 1 | Sosial | 2 | 2.2 | 2.2.1 |
| Prediksi | 1 | Sosial | 2 | 2.2 | 2.2.1 |

TABEL VIII
CONTOH PREDIKSI BENAR UNTUK SEBAGIAN LEVEL

| Publikasi Penelitian | Nowcasting the Transportation and Accommodation Sectors Growth using the Google Trends Index | | | | |
|---|--|---------------------|--------------|--------------|-------------------------------|
| | Label Filter | Label Empat Pilar | Label Tujuan | Label Target | Label Indikator |
| This research aims to assess the possibility of the daily and weekly Google Trends Index (GTI) to predict ... | 1 | Ekonomi, Lingkungan | 8, 11 | 8.1, 11.2 | 8.1.1, 11.2.1.(a), 11.2.1.(b) |
| Aktual | 1 | Ekonomi, Lingkungan | 8, 11 | 8.1, 11.2 | 8.1.1, 11.2.1.(a), 11.2.1.(b) |
| Prediksi | 1 | Ekonomi | 8 | 11.3 | 11.3.1.(a) |

mengklasifikasikan data dengan benar. Senada dengan hal tersebut, terdapat beberapa dugaan terkait model level 3 dan 4 mengklasifikasikan data dengan salah karena model kebingungan dalam menentukan kata penting.

Untuk meningkatkan kinerja model, disarankan beberapa hal lain. Pertama, pada model terbaik dapat terus dilakukan pelatihan ulang menggunakan data terbaru. Kedua, model dikombinasikan dengan kata kunci khusus di setiap kategori pada level tujuan SDG dengan TF-IDF untuk meningkatkan representasi fitur. Ketiga, model dimodifikasi dengan teknik ekstraksi fitur yang lebih andal, seperti *pre-trained word embedding*. Keempat, model dimodifikasi dengan *transfer learning* multibahasa, seperti *multilingual bidirectional encoder representations from transformers* (mBERT) yang telah dilakukan pelatihan sebelumnya di korpus yang besar.

IV. KESIMPULAN

Jumlah *dataset* yang berhasil dibangun sebanyak 8.090 data dengan distribusi SDG dan non-SDG sebesar 66,5% dan 33,5%. Terhadap *dataset* tersebut juga telah dilakukan uji reliabilitas untuk mengetahui kualitas data. Kualitas data tergolong baik karena nilai *alpha* berada di rentang $0,60 < \alpha \leq 0,80$.

Berdasarkan ukuran evaluasi, model filter terbaik adalah model dengan masukan judul dan algoritma SVM. Model level terbaik adalah model dengan masukan judul, metode LP, dan algoritma SVM yang dibangun dengan pendekatan multilevel. Model level mengklasifikasikan data ke dalam empat pilar, tujuan, target, dan indikator SDG.

Beberapa topik usulan untuk dosen Polstat STIS yang belum banyak diteliti yaitu lingkungan serta hukum dan tata kelola, khususnya terkait penanganan perubahan iklim dan ekosistem lautan serta keadilan dan kelembagaan yang tangguh. Hasil penelitian berupa model terbaik ini dapat digunakan di lingkungan Politeknik Statistika STIS secara khusus atau perguruan tinggi secara umum untuk pemantauan kontribusi SDG oleh dosen melalui penelitian dan ruang lingkup penelitian yang dapat dilakukan selanjutnya.

Hasil penelitian menunjukkan bahwa masukan judul memiliki akurasi terbaik dibandingkan masukan abstrak atau judul serta abstrak yang memiliki informasi lebih lengkap. Hal tersebut disebabkan oleh penggunaan dua bahasa secara sekaligus, yaitu bahasa Inggris dan bahasa Indonesia. Oleh karena itu, penelitian di masa depan dapat memodifikasi model dengan masukan satu bahasa saja untuk mengoptimalkan proses TF-IDF agar makna kata dari satu bahasa dengan bahasa lain tidak dianggap kata penting yang berbeda.

KONFLIK KEPENTINGAN

Penulis menyatakan bahwa tidak terdapat konflik kepentingan.

KONTRIBUSI PENULIS

Konseptualisasi, Berliana Sugiarti Putri dan Lya Hulliyyatus Suadaa; metodologi, Berliana Sugiarti Putri dan Lya Hulliyyatus Suadaa; penulisan—penyusunan draf asli, Berliana Sugiarti Putri; penulisan—peninjauan dan penyuntingan, Berliana Sugiarti Putri, Lya Hulliyyatus Suadaa, dan Efri Diah Utami; validasi, Lya Hulliyyatus Suadaa dan Efri Diah Utami; supervisi, Lya Hulliyyatus Suadaa.

UCAPAN TERIMA KASIH

Ucapan terima kasih disampaikan kepada rekan-rekan yang telah memberikan dukungan selama penelitian ini berlangsung. Ucapan terima kasih juga penulis berikan kepada Politeknik Statistika STIS, khususnya Pusat Penelitian dan Pengabdian Masyarakat Politeknik Statistika STIS atas fasilitas dan bimbingan yang diberikan.

REFERENSI

- [1] Bappenas, "Sekilas SDGs," Tanggal akses: 1-Okt-2023. [Online]. Tersedia: <https://sdgs.bappenas.go.id/sekilas-sdgs/>
- [2] Badan Pusat Statistik (BPS), "Persentase penduduk miskin (P0) menurut provinsi dan daerah, 2007-2023." Tanggal akses: 1-Okt-2023. [Online]. Tersedia: <https://www.bps.go.id/id/statistics-table/2/MTkyZlZl=/persentase-penduduk-miskin--p0--menurut-provinsi-dan-daerah--persen-.html>
- [3] Badan Pusat Statistik (BPS), "Angka partisipasi kasar (APK) menurut provinsi dan jenjang pendidikan, 2003-2022." Tanggal akses: 1-Okt-2023. [Online]. Tersedia: <https://www.bps.go.id/id/statistics-table/2/MzAzZlZl=/angka-partisipasi-kasar--a-p-k--.html>
- [4] Badan Pusat Statistik (BPS), "Angka partisipasi murni (APM) menurut provinsi dan jenjang pendidikan, 2003-2022." Tanggal akses: 1-Okt-2023. [Online]. Tersedia: <https://www.bps.go.id/id/statistics-table/2/MzA0ZlZl=/angka-partisipasi-murni--a-p-m--.html>
- [5] "Pendidikan Tinggi," Undang-Undang Nomor 12 Tahun 2012.
- [6] "Jabatan Fungsional Dosen dan Angka Kreditnya," Peraturan Menteri Pendayagunaan Aparatur Negara dan Reformasi Birokrasi, No. 17, 2013.
- [7] Kementerian Pendidikan dan Kebudayaan, "WCU Analysis," 2016. Tanggal akses: 1-Okt-2023. [Online]. Tersedia: <https://sinta.kemdikbud.go.id/wcu>
- [8] Politeknik Statistika STIS, "Call for Paper Seminar Nasional Official Statistics 2023." Tanggal akses: 1-Okt-2023. [Online]. Tersedia: <https://semnas.stis.ac.id/call-for-paper>
- [9] C. Vens dkk., "Decision trees for hierarchical multi-label classification," *Mach. Learn.*, vol. 73, no. 2, hal. 185–214, Nov. 2008, doi: 10.1007/s10994-008-5077-3.
- [10] J. Hernández, L.E. Sucar, dan E.F. Morales, "Multidimensional hierarchical classification," *Expert Syst. Appl.*, vol. 41, no. 17, hal. 7671–7677, Des. 2014, doi: 10.1016/j.eswa.2014.05.054.
- [11] H.S. Oh dan Y. Jung, "External methods to address limitations of using global information on the narrow-down approach for hierarchical text classification," *J. Inf. Sci.*, vol. 40, no. 5, hal. 688–708, Okt. 2014, doi: 10.1177/0165551514544626.
- [12] "Koordinasi, Perencanaan, Pemantauan, Evaluasi, dan Pelaporan Pelaksanaan Tujuan Pembangunan Berkelanjutan," Peraturan Menteri Perencanaan Pembangunan Nasional/Kepala Badan Perencanaan Pembangunan Nasional Republik Indonesia, No. 7, 2018.
- [13] R.C. Morales-Hernández, J.G. Juagüey, dan D. Becerra-Alonso, "A comparison of multi-label text classification models in research articles labeled with sustainable development goals," *IEEE Access*, vol. 10, hal. 123534–123548, Nov. 2022, doi: 10.1109/ACCESS.2022.3223094.
- [14] I.H. Sarker, "Machine learning: Algorithms, real-world applications and research directions," *SN Comput. Sci.*, vol. 2, no. 3, hal. 1–21, Mei 2021, doi: 10.1007/s42979-021-00592-x.
- [15] J. Alzubi, A. Nayyar, dan A. Kumar, "Machine learning from theory to algorithms: An overview," *J. Phys. Conf. Ser.*, vol. 1142, hal. 1–15, Des. 2018, doi: 10.1088/1742-6596/1142/1/012012.
- [16] P. Chapman dkk., CRISP-DM 1.0 Step-by-step data mining guide. SPSS Inc, 2000.
- [17] S.L. Octaria, "Analisa Integrasi Data SINTA (Science and Technology Index)," Palembang, Indonesia, Universitas Sriwijaya, 2018. [Online]. Tersedia: http://edocs.ilkom.unsri.ac.id/2906/1/TUGAS%20MTI%20_0903118%201621128_SITI%20LARISTA%20OCTARIA.pdf
- [18] J. Hughes, "Krippendorff's alpha: An R package for measuring agreement using Krippendorff's alpha coefficient," *R J.*, vol. 13, no. 1, hal. 413–425, Jun. 2021, doi: 10.32614/RJ-2021-046.
- [19] K. Krippendorff dan R. Craggs, "The reliability of multi-valued coding of data," *Commun. Methods Meas.*, vol. 10, no. 4, hal. 181–198, Okt. 2016, doi: 10.1080/19312458.2016.1228863.
- [20] M. Sighn, "Stop the stopwords using different Python libraries," Toward AI. Tanggal akses: 1-Okt-2023. [Online]. Tersedia: <https://towardsai.net/p/stop-the-stopwords-using-different-python-libraries>
- [21] F.Z. Tala, "A Study of stemming effects on information retrieval in bahasa Indonesia," Amsterdam, Belanda, Universiteti van Amsterdam, 2003. [Online]. Tersedia: <https://eprints.illc.uva.nl/id/eprint/740/1/MoL-2003-02.text.pdf>
- [22] C. Toraman, E.H. Yilmaz, F. Şahinuç, dan O. Ozcelik, "Impact of tokenization on language models: An analysis for Turkish," *ACM Trans. Asian Low-Resour. Lang. Inf. Proc.*, vol. 22, no. 4, hal. 1–21, Apr. 2023, doi: 10.1145/3578707.
- [23] K. Kowsari dkk., "Text classification algorithms: A survey," *Information*, vol. 10, no. 4, hal. 1–68, Apr. 2019, doi: 10.3390/info10040150.
- [24] Z. Abdallah, A. El Zaart, dan M. Oueidat, "Experimental analysis and comparison of multilabel problem transformation methods for multimedia domain," dalam *2015 Int. Conf. Appl. Res. Comput. Sci. Eng. (ICAR)*, 2015, hal. 1–8. doi: 10.1109/ARCSE.2015.7338147.
- [25] O. Ramadhani, "Klasifikasi multi-label dengan problem transformation menggunakan Python," Skripsi, Institut Teknologi Sepuluh Nopember, Surabaya, Indonesia, 2020.
- [26] G. Mustafa dkk., "Multi-label classification of research articles using Word2Vec and identification of similarity threshold," *Sci. Rep.*, vol. 11, Nov. 2021, Art no. 21900, doi: 10.1038/s41598-021-01460-7.
- [27] M.D. Turner dkk., "Automated annotation of functional imaging experiments via multi-label classification," *Front Neurosci.*, vol. 7, hal. 1–13, Des. 2013, doi: 10.3389/fnins.2013.00240.
- [28] B.J. Hashimoto, "Is frequency enough?: The frequency model in vocabulary size testing," *Lang. Assess. Quart.*, vol. 18, no. 2, hal. 171–187, 2021, doi: 10.1080/15434303.2020.1860058.
- [29] N.A. Sajid dkk., "A novel metadata based multi-label document classification technique," *Comput. Syst. Sci. Eng.*, vol. 46, no. 2, hal. 2195–2214, Feb. 2023, doi: 10.32604/csse.2023.033844.
- [30] C. Cortes dan V. Vapnik, "Support-vector networks," *Mach. Learn.*, vol. 20, no. 3, hal. 273–297, Sep. 1995, doi: 10.1007/BF00994018.
- [31] M.A. Salam, A.T. Azar, M.S. Elgendy, dan K.M. Fouad, "The effect of different dimensionality reduction techniques on machine learning overfitting problem," *Int. J. Adv. Comput. Sci. Appl.*, vol. 12, no. 4, hal. 641–655, Apr. 2021, doi: 10.14569/IJACSA.2021.0120480.